e-IRG Roadmap 2016

Paving the way towards a general purpose European e-Infrastructure

e-IRG Roadmap 2016

December 22, 2016

Final

Version 5.3

http://e-irg.eu

Acknowledgements

The e-IRG delegates

Listed on the e-IRG website
http://e-irg.eu under "Members"

The e-IRG Roadmap Reviewers

Leif Laaksonen

Enzo Valente

Fiorenzo Scaroni

Sergi Girona

Panos Argyrakis

The e-IRG Roadmap Editorial Board

Erik Fledderus – chair of the working group

Marcin Lawenda

Jan Wiebelitz

Fotis Karayannis

Lajos Balint

Rosette Vandenbroucke

Rene Belsø

Norbert Meyer

Arjen van Rijn

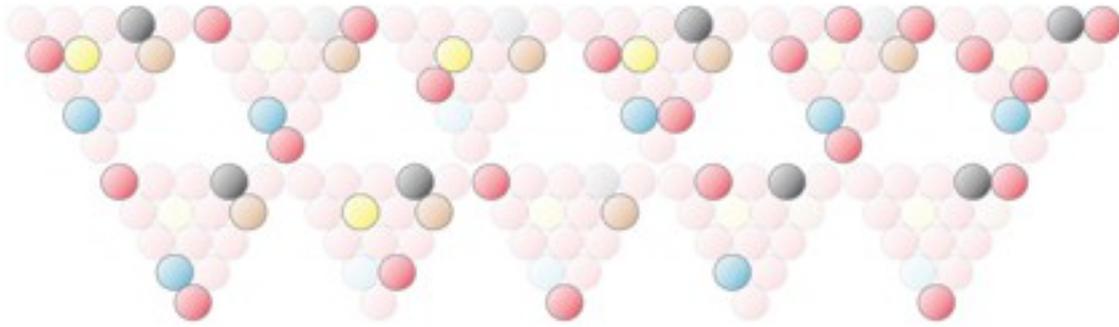Gabriele von Voigt

Françoise Genova

Rossend Llurba

ISBN 978-90-823661-1-2

e-IRG secretariat
P.O. Box 93575,
NL-2509 AN The Hague,
The Netherlands
Phone: +31 (0)70 344 0526
E-mail: secretariat@e-irg .eu

Visiting address

Java Building, Laan van Nieuw Oost-Indië 300
NL-2593 CE The Hague
The Netherlands

Table of Contents

1 Executive Summary

The e-Infrastructure Reflection Group (e-IRG) is a self-regulated and independent body consisting of delegations from the EU Member States and Associated Countries and the European Commission. The e-IRG vision is to facilitate integration in the area of European e-Infrastructures and connected services, within and between member states, at the European level and globally. In its series of Roadmaps, e-IRG builds on this vision to fulfil its mission to *support coherent, innovative and strategic European e-Infrastructure policy making and the development of convergent and sustainable e-Infrastructure services*.

In the previous version of the Roadmap, presented in 2012, e-IRG pointed to the need of *a single e-Infrastructure Commons for knowledge, innovation and science as a living ecosystem that is open and accessible and continuously adapts to the changing requirements of research*. The notion of an e-Infrastructure Commons has now been widely accepted, and significant steps have been taken towards its implementation. However, challenges still remain at the same time as the availability of convergent and sustainable e-Infrastructure services is more and more recognized and accepted as an essential factor for success of the European and national research and innovation systems.

In the Roadmap 2016 e-IRG intends to define a clear route on how to evolve the European e-Infrastructure system further, and turn the vision of the e-Infrastructure Commons into reality for 2020. e-IRG is convinced that the implementation of the e-Infrastructure Commons is a large step towards European leadership in research infrastructures including e-Infrastructures, including the realisation of the European Open Science Cloud and the EU Data Infrastructure, which are part of the Digital Single Market Technologies and the Public Service Modernisation Package.

The key recommendations, further elaborated on in the Roadmap text, are:

- 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.
- e-Infrastructure providers should further increase their efforts to:

- work closely together to fulfil the often complex user needs in a seamless way.

- 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 analyze 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.

- The European Commission should (e.g. in future Work Programmes):

    - provide strong incentives for cross-platform innovations and further support the coordination and consolidation of e-Infrastructure service development and provisioning on the national and the European level.

The Roadmap starts with a brief elaboration on the e-IRG vision and an assessment of the extent to which the current national and international e-Infrastructures already realise this. It then presents a landscape analysis of the current European e-Infrastructure system and identifies the key challenges that hinder e-Infrastructure harmonisation and integration. From the analysis, it is clear that an extended effort on "emphatic co-operation and coordination" among all main stakeholders is required. This involves the providers (the e-Infrastructure developers and operators), the users (research infrastructures including the ESFRI projects, large scientific communities, and users belonging to the "long tail of science"), and the policy makers and funders (the national governments and their agents and the EU). Good coordination can be established through a formal coordination platform among all stakeholders with strong national involvement, inline with the vision of the e-Infrastructure Commons. Also in line with the vision, such a platform can be implemented using a potentially distributed, multi-stakeholder model of governance. Backing for educational process with support of experts providing profiled technical courses is an important aspect for building eScience awareness and capacities among users.

Following the landscape analysis, the Roadmap describes the way forward and provides the recommendations summarized above. The implementation of the e-Infrastructure Commons, loosely integrating the different types of e-Infrastructures, builds on establishing coordinated access to all e-Infrastructure services and tools. The establishment of such a "marketplace" will provide a one stop-shop for the users, providing choices and directing them to a suitable set of services. The marketplace can make use of several technologies and services, such as cloud technologies, a searchable service catalogue and a common identity/authentication/authorisation scheme.

In this way, standardised and single point of access to services will be achieved, without promoting monopolies or implying the need of creating a single integrated provider. Instead, such a solution should be built to be modular and open to new actors, encouraging cooperation, competition and innovation. Also, the national/regional dimension could be made strongly visible: National/regional abstractions and/or instances of the EU marketplace could be available. Such abstractions may provide a sub-set of the European-level services, based on national participation or availability and on national laws and restrictions. On the other hand, additional national/regional services may be available in the different member states or regions. The e-Infrastructure Commons will constitute a coordinated ecosystem among EU and national/regional levels, being automatically synchronised among them. This vision is also consistent with the European Cloud Initiative (ECI) EC Communication, which foresees the development of the European Open Science Cloud (EOSC) as a federated set of services where

Research Infrastructures will be connected and the EU Data Infrastructure (EDI) offering advanced underlying networking, HPC and data services. Some of the above components are already planned to be prototyped or procured, such as the e-Infrastructures Service Catalogue or the European Open Science Cloud.

The Roadmap considers how to achieve the right balance between operation of services and development of innovative ones, and working further on sustainability of the services and their relevance to the user needs is key.

e-IRG believes that the new EC tools of Framework Programme Agreements (FPAs) and the upcoming Operational Grants are in the right direction for increasing users' confidence in the e-Infrastructure long-term sustainability. e-IRG also believes that common spaces with common access and security policies should be gradually implemented as slices of the different e-Infrastructure resources, as it may not be possible to harmonise all the resources of all member states. A common or interoperable identity framework for all e-Infrastructure providers should be developed, supporting eduGAIN and possibly other schemes such as the e-Government IDs (e-IDs).
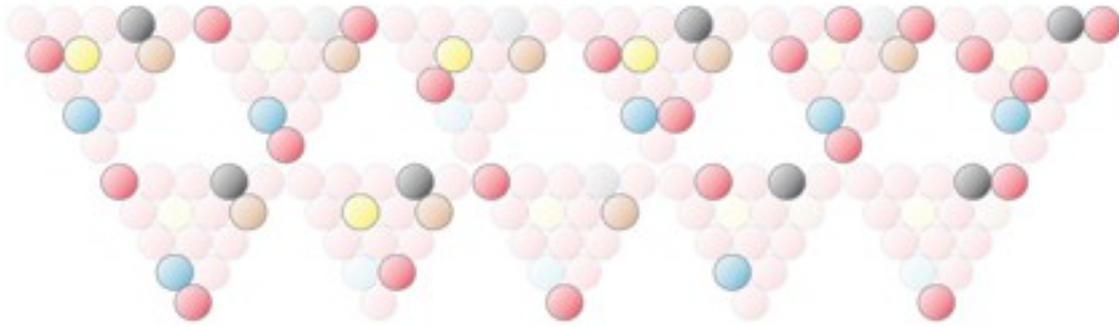
2 Introduction

This document is intended for all related entities relevant to the e-Infrastructure service provision, namely e-Infrastructure service providers and developers, commercial service providers, e-Infrastructure users, policy makers and funding agencies. Emphasis is on policy makers and people close to them, as e-IRG aspires to influence the evolution of the e-Infrastructure service provision towards 2020 and beyond. A brief introduction of the e-Infrastructure history and current outlook is given below. More details on the vision, current landscape, and challenges are provided in the next sections.

e-Infrastructures for research have been around for more than 30 years, a respectable time in the rapidly evolving ICT arena. For networking they started with the creation of several European NRENs (National Research and Education Networks) and soon their Association (RARE), the predecessor of TERENA, which in the early '90s established DANTE, the Operational Unit devoted to building and operating the European research backbone network. DANTE, in cooperation with TERENA, provided the gradually increasing capacity network for research in Europe so that at the end of the '90s the gigabit generations of the backbone (GEANT) had been launched, providing communication and add on services for the entire European research and education community.

The term "e-Infrastructures" came into use with the progress of distributed computing and was rapidly applied to all infrastructures that delivered ICT services. The addition "for research" is meaningful but has almost disappeared while e-Infrastructures are now often simply mentioned as part of the research infrastructures. The emphasis on "for research" intended to stress the fact that, on the one hand, these infrastructures provide production-level services for the research community, but are not intended to directly do research in the information or communication technology (although leading edge technical solutions have also resulted from building the e-Infrastructure). On the other hand, "for research" is also expressing the fact that many national e-Infrastructures are mostly based on public funding and their Acceptable Use Policies (AUPs) instruct (in most cases) avoidance of commercial use so as not to distort market competition. However, nowadays the stress on innovation may lead to an adaptation of policies allowing industrial usage in the pre-competition phase, especially in the case of joint public-private partnership. Nevertheless, a key aspect that describes e-Infrastructures is that the services delivered to the research community have specific and sometimes extremely high demands that in most cases cannot be accommodated by commercial providers.

The creation and evolution of e-Infrastructures have been driven by a combination / dialogue between user needs (*user pull*) and technology progress (*technology push*) both at European and national level. The networking e-Infrastructure GÉANT is a successful example of a federation and as such a combination of national networks with international multi-stakeholder governance to connect them into a pan-European network. High-throughput computing e-Infrastructures have been created by a bottom-up approach as an answer to the computing needs of a research community (leading to the creation of EGI); for high-performance computing local institutes often came together to establish a joint HPC-centre. Similar to GÉANT, these centres entered into a set of agreements (PRACE) where a tier-like structure provided a wide range of large computing facilities and services to researchers.

As a relative new branch on the e-Infrastructure-tree, the data infrastructures and services are still quite fragmented, but with the examples of GÉANT, EGI and PRACE these communities are encouraged to create solutions that fit the whole research world on a global scale. The Research Data Alliance (RDA) in the last few years has been making considerable efforts in reducing barriers to data sharing and exchange, and building the necessary technical and social bridges across technologies, disciplines, and countries. In this domain, the diversity of community cultures and practices has to be taken into account, as well as the requirement that data should be Findable, Accessible, Interoperable and Reusable (FAIR[1]), which requires good data stewardship and significant community involvement.

These efforts are completely in line with the emergence of openness in science and wide accessibility of research data. That's the reason why an integrated approach of exploiting e-Infrastructures together with supporting openness is on the development agenda.

The GÉANT networking infrastructure is practically not challenged by competitors in the research world; it is used by European researchers as a "public utility", and is the basis on which all other e-Infrastructures are built. The other e-Infrastructures cannot claim such recognition at all levels. Until recently most of the computing e-Infrastructures and the upcoming data infrastructures saw each other as competitors, but with EGI, EUDAT and PRACE there is a common understanding on the necessity of co-operating.

e-Infrastructures are in a peculiar situation which is characterised by both a constant intention of collaborating and a position of debating with their users, the research communities and the domain specific (discipline oriented) research infrastructures. It is a huge challenge for an e-Infrastructure to fulfil the diverse and sometimes conflicting requirements of many research environments. This repeatedly results in a dilemma for research infrastructures or communities to choose between picking up services from existing (generic) e-Infrastructures or building their own, more specific, e-Infrastructure. Such dilemmas have been clearly identified with the establishment of the most recent ESFRI roadmapping activities. Answers to this might lie in a joint ESFRI – e-IRG effort to make most widely agreed decisions on e-Infrastructures, but issues like sustainability and different governance structures also need to be mastered. It has to be made clear that the main objective of all e-Infrastructure resources is to support their users and they have to be deployed and managed with that aim.

One way to combine different types of e-Infrastructures is to use the metaphor of a "marketplace" for the provision of a single access to all the services and tools; the marketplace can use among others cloud technologies, a searchable service catalogue and a common authentication / authorisation scheme; in this way users can have access to a one-stop-shop, i.e. a place where all e-Infrastructure services are available, either directly accessible or redirected elsewhere. Commercial ICT services can be also included in a more complete offer to researchers. This requires that public e-Infrastructures profit from commercial providers and broker their interoperation on behalf of the

[1] Open Science http://ec.europa.eu/research/openscience/

user communities (aggregation or needs). Commercial providers should be included from the beginning on the condition that they offer acceptable Service Level Agreements, Terms & Conditions and payment models.

There is also the pressure on e-Infrastructures to contribute to the socio-economic development at European and global levels. Citizen science, open science, open data – answers are needed on how to cope with these new concepts that were non-existent at the set-up of most of the e-Infrastructures.

The next sections will detail the e-IRG view (vision, landscape, recommendations) on how to deal with the challenges and issues in the coming years.
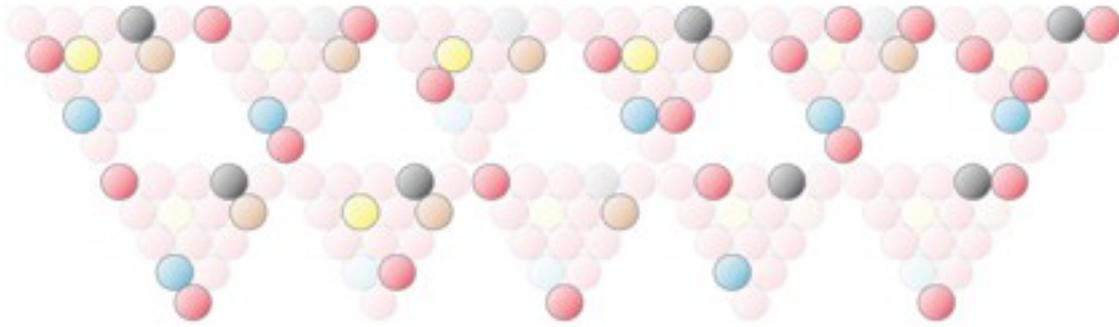
3 Vision

In 2013, the e-Infrastructure Reflection Group identified the need for a more coherent e-Infrastructure landscape in Europe[2] and worded this need in the definition of the e-Infrastructure Commons.

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 and new technological capabilities, 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 is thoroughly justified.

Based on this description, the three distinct functions of the Commons can be summarized as follows:

- 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-Infrastructure services within the Commons[3], promoting a flexible and open approach where user communities are empowered to select the services that fulfil their requirements.

- Implementation of innovation projects aiming for the constant evolution of e-Infrastructures needed to meet the rapidly evolving needs of user communities and take advantage of relevant new technologies when they are mature enough.

The definition mentions various *resources* and *communities* that play a role in defining, using and renewing the Commons. Moreover, in the background the *rules of the game* (*governance*) are required to make it all work in a harmonized way. More general theory on Commons[4] shows that these elements are always present to properly describe its long-term evolution and its role to enable

---

[2]http://e-irg.eu/documents/10920/11274/e-irg-white-paper-2013-final.pdf
[3]E.g. EGI Open Science Commons document https://documents.egi.eu/public/RetrieveFile?
docid=2410&version=1&filename=OpenScienceCommons%20v3.pdf
[4]*Think like a commoner: a short introduction to the life of the Commons*, David Bollier

non-discriminatory access to the resources for the members of a community. These elements will play a further role in the various chapters in this roadmap.

Resources

The research network (Communication Commons[5]), the computational resources, the data capacities and the services constitute the main building blocks of the e-Infrastructure Commons resource, which are shared within the community. The providers[6] of the e-Infrastructure components are implicitly included in the resource. Most of the e-Infrastructure services are universal in the sense that they are independent of the user domains and research disciplines but part of the resources could be dedicated to a domain-specific community. The discussion about resources is continued in Chapter 3.

Community

The community of the e-Infrastructure Commons is primarily the pan-European scholarly community, be it domain-specific or regional, and their use extends to education and applied research communities. Important examples are the scientific communities gathered around the research infrastructures, such as ESFRI. It should also include communities more interested in the 'long tail of data'.

The community is continuously widening with respect to both geographical and application-specific coverage. Extensions towards innovation, as well as towards industrial, health care, or public administration areas represents just specific examples of this widening.

Sometimes this community relation is legally formalised using e.g. an MoU, a cooperative, or an ERIC. This observation bridges to the last topic – governance.

Governance

The Open Access, Open Data, and ultimately Open Science model have introduced the idea of the Commons into the scientific world. While use of (non-open) data and publications is mostly regulated by legal and commercial barriers, other e-Infrastructures provide limited physical resources and consequently have more complex governance rules that may involve long-term commitment and multiple stakeholders. The governance of the e-Infrastructure Commons should be provided by a platform where all stakeholders (e-Infrastructure providers, users, governments, etc.) are represented and have a say in the definition of those governance rules[7].

The governance rules should follow an open approach, which means not necessarily free access, but rather non-discriminatory use of the resources[8] as far as possible, restricted only by legitimate, legal or other regulatory constraints. Moreover, Nobel-prize winner Elinor Ostrom mentions[9] that it has to be ensured that those affected by the rules can participate in modifying the rules. The European Commons Assembly[10] is an example of a civil society forum for discussion on how to apply commons principles.

---

[5]Knowledge without Borders (https://www.terena.org/about/ga/ga36/GA(11)024geg-report.pdf)

[6]This may include providers beyond the public e-Infrastructure providers, e.g. software providers or commercial e-Infrastructure providers.

[7]First implementations of e-Infrastructures in member states are legally formalized using e.g. a cooperative.

[8]"The general values of the commons management strategy are that it maintains openness, does not discriminate among users or uses of the resource, and eliminates the need to obtain approval or a license to use the resource." [Infrastructure - the social value of shared resources, Brett Frischman]

[9]Elinor Ostrom's 8 Principles for Managing A Commmons http://www.onthecommons.org/magazine/elinor-ostroms-8-principles-managing-commmons

[10]European Commons Assembly https://europeancommonsassembly.eu/

Finally, in the e-Infrastructure Commons, where various public and commercial providers can interplay to create the infrastructure layer, standards are a crucial element, which enables cross-domain research, prohibits provider lock-in and facilitates interoperability.

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. It is the mission of e-IRG to support this vision through advocating a coherent, innovative and strategic European e-Infrastructure policymaking and the development of convergent and sustainable e-Infrastructure services. This e-Infrastructure Commons is also a solid basis for building the European Data Infrastructure[11] and the European Open Science Cloud as introduced in the description of the European Data Initiative EC Communication and the Digital Single Market[12], since it contains most of the ingredients needed for an integrated European platform for Open Science. The European Open Science Cloud can be seen as a metaphor of a federation of seamlessly accessible resources and services, well in line with the Commons concept.

---

[11]Building a competitive data and knowledge economy in Europe https://ec.europa.eu/digital-single-market/en/news/communication-european-cloud-initiative-building-competitive-data-and-knowledge-economy-europe
[12]SWD(2015) 100 final accompanying the document "Commons Summary" COM(2015) 192 final, SWD(2015) 100 final
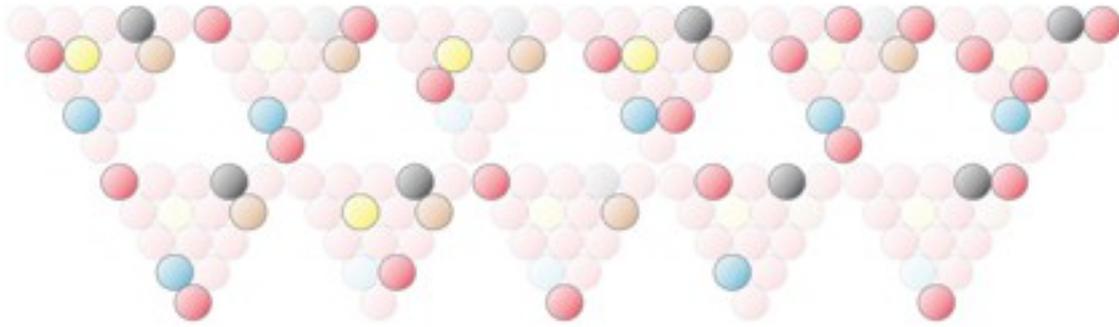
4 Landscape analysis

4.1 Resources – The current e-Infrastructure landscape

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 infrastructure projects. Also, data and cloud infrastructures are developing fast.

It should be noted that the European-level 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 and coherent national e-Infrastructure nodes, and of their cooperation to enable cross-border services for scientific communities.

Below, a brief introduction of the major pan-European horizontal e-Infrastructure initiatives and some examples of services provided are given. GÉANT, the networking infrastructure widely recognized as a 'public utility', is described in more details.

A more complete account of available services can be found in the e-IRG Guidelines Document 2015[13].

Governance is an important aspect of the functioning and sustainability of the various initiatives. In the e-Infrastructure initiatives and their lead organisations it is thus important that all stakeholders are well represented in the governance system.

The e-Infrastructure Commons model suggests a shared governance structure for all research e-Infrastructures with its key elements: a coordination platform where all stakeholders are represented, non-discriminatory use of resources[14], and use of and contribution to standards.

4.1.1 Networking

Connecting research communities across the globe is a prerequisite to stimulate exchange of ideas, data and results. Moreover, it is needed to provide access to unique research facilities and to shared data that are located at specific places, without the need of physical presence; in this way the quest for breakthroughs is sped up, as large communities can simultaneously be part of joint investigations and separate research teams can exploit large infrastructures and data in parallel, by remote access.

---

[13]Best Practices for the use of e-Infrastructures by large-scale research infrastructure http://e-irg.eu/documents/10920/277005/Best+Practices+for+the+use+of+e-Infrastructures+by+large-scale+research+infrastructures.pdf
[14]as far as not restricted by legitimate, legal or other regulatory constraints (see Section Vision)

Already since a few decades the National Research and Education Networks (NRENs) have been connecting universities, research institutes, and sometimes other public institutions in their country. Their governance and sources of income differ from country to country, as well as the way the access to NREN resources is managed. Due to large scientific cross-border collaborations and communities, and to the need of accessing unique research infrastructures installed in various sites around the world, the NREN organisations felt the need to make synergies in order to set up a networking system interconnecting all the NRENs. For that purpose the NREN organisations set up associations (RARE/TERENA/EARN) and operational bodies (DANTE/GÉANT) with the scope of fulfilling the requirements of NRENs and the international user communities. The GÉANT Association has gradually grown into a pan-European organisation, where the associated NRENs link together research communities and provides trans-national access to Research Infrastructures and other research resources. GÉANT provides interconnectivity between NRENs across 43 European countries, by serving an estimated 50 millions of users of practically all research disciplines and thematic domains. The major scientific communities that are connected via GÉANT's collaborative network are amongst others high-energy physics, bio-medical sciences, astronomy/radio astronomy, earth observation and early warning, as well as arts and culture.

In addition to pan-European connectivity, the GÉANT network has international connections to a large set of partner networks (some 60 NRENs) worldwide, in particular through regional agreements – thereby 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 related NRENs and GÉANT for international *point-to-point links* to connect parts of the research infrastructure that are distributed over Europe or beyond. If the project or infrastructure is distributed across national boundaries, GÉANT can *help coordinate with the relevant local NRENs* and *advise on appropriate technical solutions*. GÉANT provides also important experimental services for researchers, such as *innovation test beds* – these test beds will further stretch the available cutting-edge capabilities and are an important source for innovation and renewal.

GÉANT delivers a range of additional services on top of the networking ones at the international level. Most of these services match those offered at national level by the NRENs, among which for instance:

- Research communities need to manage secure and hassle-free access to their services, multiple tools and datasets. *Trust and identity* therefore take up a pivotal position in the e-Infrastructure eco-system. 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. A good example is given by eduGAIN, providing interoperation between national digital identity federations.

- Integration of the local layer is provided by facilitating access to wireless networks in campuses around the world using eduroam.

Practically all ESFRI facilities are connected by GÉANT and the NRENs across the ERA. Widely accepted service-level agreement indicators make the connectivity service reliable and dependable – which is extremely important for the user community.

The different governance and financing principles and practices of the various NRENs are taken into consideration on the GÉANT level by applying a traditional, proven, jointly accepted Cost-Sharing model that indirectly takes into consideration the user aspects within the national user communities represented by the NRENs.

The GÉANT community exploits the opportunities stemming from high volume procurement in building customer relations when interacting with the equipment manufacturers. These relations are built on mutual benefits of the vendors, the NRENs, and also GÉANT.

It is to be mentioned that GÉANT met the 'commons' vision quite early. The service has been available for research traffic to the extremely wide user community without any distinction or restriction practically from the start of the GÉANT service. That is why the EC GÉANT High Level Expert Group has been talking about GÉANT as the "Communication Commons" as early as in 2011.

4.1.2 Computing

The role of computing in science has always been prominent, especially in those areas where models were simulated to predict and explain observed phenomena. Recently computing applied to (big) data has transformed and expanded our way of doing science, and it has further stimulated the need for computing resources. These needs range from powerful PCs via clusters to high-performance computers. Some of these resources are quite generic and are bought by research groups at shops "around the corner", others are highly specific for the domain.

On a national level the varieties of high throughput computing resources are often brought together using a National Grid Infrastructure (NGI). The NGIs are federated in a pan-European high-throughput computing and federated cloud infrastructures (European Grid Infrastructure, or EGI). The HPC national infrastructures are federated in the Partnership for Advanced Computing in Europe (PRACE). EGI focuses on large-scale federated high-throughput and cloud computing solutions, while PRACE offers access to world-class high-performance capability computing facilities and services. EGI and PRACE are respectively managed by the organisations EGI.eu and PRACE AISBL.

Both EGI and PRACE have already established contacts with consortia that operate or prepare European large-scale research infrastructures to understand their needs and find out how these match with available resources and existing policies.

- 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. There are basically two forms of access:

    o preparatory access, intended for short-term access to resources, for code-enabling and porting, required to prepare proposals for Project Access and to demonstrate the scalability of codes; and

    o project access, intended for individual researchers and research groups including multi-national research groups, which can be used for 1-year, as well as for 2-year or 3-year (Multi-Year Access) production runs.

- EGI provides solution frameworks built through a service catalogue that has been evolving over the years. The *EGI Federated Cloud Solution* offers a standards-based and open infrastructure to deploy on-demand IT services that can host 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. Also, the *EGI Federated Operations Solution* and the *EGI Community-Driven Innovation & Support Solution* are provided for management of a heterogeneous infrastructure and helping researchers and research infrastructures accessing and using computational services.

Access to EGI's externally provided resources, and to national HPC resources, is provided through various access modes, such as free grant-based allocations, pay per use, and annual membership fees. For the national HPC infrastructure(s), the access modes are closely connected to the chosen governance.

4.1.3   Data infrastructures and services

Data is a key research infrastructure. This was stressed in the "Riding the Wave" report of the relevant EC High Level Expert Group in 2010 and its follow-up RDA Europe Report "The data Harvest" (2014), as well as in the recent Communication from the European Commission "European Cloud Initiative - Building a competitive data and knowledge economy in Europe". The role of the data in science is strongly increasing, and is fully recognized at the political level, as explained for instance in the strong statement of the G8 Ministers of Research in 2013[15].

To the greatest extent and with the fewest constraints possible publicly funded scientific research data should be open, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests, whilst acknowledging the legitimate concerns of private partners. Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards.

Data has to be open (except for legitimate restrictions such as privacy), FAIR – Findable/Accessible/Interoperable/Reusable – and preserved on the long term. The users of the data infrastructures and services are the data providers and data consumers, who can belong to the scientific community, to the industry, to the public services or can just be a citizen.

Data infrastructures and data services need to be part of the e-Infrastructure Commons. Data infrastructures should be built in an interoperable way and provide all potential users with the capability to store their data and to make this data discoverable and accessible while taking into account European and national data laws (privacy, IPR, …).

The way towards such an interoperable European Data Infrastructure might be long. Initiatives such as EUDAT and OpenAIRE have started to address these issues at the European level and are implementing solutions (with their national and regional strongholds) to store, share, and preserve research data across scientific domains. Also, scientific communities and research infrastructures have been building their framework for data sharing and domain specific standards (formats, metadata ...) and in many cases their own data infrastructures, often at the international level, taking into account their specific needs.   As an example, ELIXIR coordinates, integrates and sustains bioinformatics resources across its member states and enables users in academia and industry to access vital data, tools, standards, compute and training services for their research. National and local authorities also set up data infrastructures. All of them should be incorporated in a European Data Infrastructure, which should be an ecosystem able to include different components.

EUDAT aims to move towards a sustainable research data infrastructure. The agreement signed in September 2016 formalises the roles and responsibilities of the service providers constituting the CDI (EUDAT Collaborative Data Infrastructure). Covering both access and deposit, from informal

---

[15]G8 Science Ministers Statement London UK, 12 June 2013
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/206801/G8_Science_Meeting_Statement_12_June_2013.pdf

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. 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. Disciplinary and generic national and local repositories play a key role in particular for preservation and dissemination of the so-called long tail of research data. If researchers have no access to an institutional or a subject repository, Zenodo, hosted by CERN and which exposes its contents to OpenAIRE, enables them to deposit their articles, research data and software. One important aspect for all these repositories is in particular the creation of persistent identifiers.

4.1.4   Clouds

Cloud technologies have been a hype a few years ago and have well made their entrance in non-research environments where probably the most visible use is "Software as a Service". In the research environment local/national cloud services have gained some success but at the European level the cloud initiatives are scarce and have not lead to a pan-European cloud e-Infrastructure. Initiatives to mention include the Helix Nebula initiative, the EGI federated cloud and the GÉANT cloud services platform. Helix Nebula is providing a channel by which innovative cloud service companies can work with major IT companies and public research organisations. The Helix Nebula Marketplace (HNX) is the first multi-vendor product coming out of the initiative and delivers easy and large-scale access to a range of commercial Cloud Services through the innovative open source broker technology. Also, GÉANT is actively helping NRENs (National Research and Education Networks) to deliver cloud services to their communities. It is also engaging with the existing NREN brokerages to promote an efficient and coordinated pan-European approach, by building on existing experience and supplier relationships. Recent announcements of the European Commission (European Open Science Cloud) and the Cloud Communication have added new pressure for the adoption of cloud technologies and services in the research environment, but this should also be seen, as explained earlier, as a metaphor of a federation of seamlessly accessible resources and services, which should include commercial services in order to provide a more complete offer to researchers.

4.1.5   eScience

eScience enhances and accelerates scientific research through efficient utilisation and re-use of software, e-Infrastructure and data. Many fields of research are becoming increasingly data- and computationally intensive. The increased availability of data offers opportunities for researchers to address new scientific questions and accelerate scientific breakthroughs. Traditionally, the focus of data-driven research was on large volumes of data from sensors and simulations, such as in astronomy, physics and meteorology. Increasingly, a large diversity of data, including the so-called 'long tail', is brought together to answer disciplinary and multidisciplinary questions involving many domains. eScience forms a bridge between the evolving computer and data science, the increasing opportunities offered by the e-Infrastructure (compute, storage and network) and research questions of application domains. eScience is inherently demand-driven from those application domains and is a prerequisite to address questions in modern digitally enhanced research. Trust and security are among the basic requirements.

eScience has a broad scope of activities and involves tailoring general software and technologies to specific user needs (e.g. data management, data analytics, efficient computing). Yet, these tailored methodologies should be kept general enough as to allow re-usability. eScience is conducted by digitally skilled researchers (application experts or eScience research engineers) who work at the

interface of the domain specialists and the e-Infrastructure and computer science. They form the essential interface to make efficient use of data and e-Infrastructure possible in computation intensive science.

It is worth noting here that the "eScience" frontier is a moving one, and that the "digital" skills progressively disseminate in the communities to become part of the common background of researchers and engineers, with some of them at the leading edge to take advantage of the new technological capabilities. It is also essential to keep in mind that proper data stewardship is an essential component of data (re)use, which requires domain expertise, manpower and specific skills to provide truly FAIR data. Trust has technological aspects, but users should also trust the quality of the data that they find in the system, and be assured that the systems they are using have sufficient sustainability. This 'human e-cloud' is the heart of eScience, and the definition and dissemination of knowledge and best practices is mandatory. The RDA has a role to play in particular in the definition of good practices in the different aspects of scientific data sharing.

Within Europe, eScience institutions have been set up in the UK, in the Scandinavian countries and in the Netherlands. In other countries eScience activities are organised at existing HPC and NREN organisations or at universities and research institutes. Recently, PLAN-E (Platform of National eScience Centers in Europe) was set up as a European platform for eScience to share best practices and coordinate activities. Members from over 20 European countries are involved now. Currently, activities in the Centers of Excellence in the H2020 program show closest connections to eScience activities, but there are no dedicated eScience activities within H2020.

The paths of PLAN-E and the e-IRG cross, because conducting eScience involves open borders, open data (and software), sharing and usage of e-Infrastructures at the bleeding edge. So the eScience community has great interest in obstacle-free usage of European e-Infrastructures.

Finally, eScience awareness in research and in education, including fostering specific university courses throughout the curriculum of students and PhDs, are crucial for the successful exploitation of the European e-Infrastructure commons for the benefit of its users.

4.1.6   The current status of the Commons

In 2013 e-IRG has given the following recommendations to the (European) e-Infrastructure organisations and projects:

Join forces and share common challenges towards serving the European user communities, avoiding duplication of efforts in:

- Outreach to and involvement of user communities;
- Services registry, discovery and provisioning;
- Financial, legal, business development and procurement;

As a summary from the previous sections progress with some of these issues is noted. As sketched in the landscape description the European e-Infrastructure is now consistently and reasonably sustainably served with computing, data and networking services through projects and organisations such as EUDAT, OpenAIRE, EGI, GÉANT, PRACE and Helix Nebula. It is recognized that initiatives originating from these projects are clustering their outreach activities towards user communities, e.g. the Joint User Forum (e.g. Digital Infrastructures for Research 2016 conference[16]). Other examples are common positions papers[17] such as the one on the European Open Science Cloud for Research

---

[16]http://www.digitalinfrastructures.eu/

[17]https://www.eudat.eu/news/the-european-open-science-cloud-for-research

by EUDAT, LIBER, OpenAIRE, EGI and GÉANT. Another clear indication of convergence is that most of these major players participate in consortia around the development of crucial, truly generic services such as AAI (e.g. the EC project AARC).

However, apart from the merger of TERENA and DANTE into GÉANT no further meaningful movements towards consolidation and tighter federation of these organisations at the European level have been observed. A frequently heard argument explaining this lack of progress is, that representations in these organisations are not well enough coordinated on the country level, leading to conflicting messages to the management of the e-Infrastructure organisations.

Finally, a word on the pictured match between horizontal (generic) services and vertical (domain-specific) needs. We observe that there is often a discrepancy between the offered generic services and user specific needs. This gap must be addressed by dedicated services in the development of which users must be included, but especially and including a level of support offered by skilled people that are able to bridge this gap. Because of the scarcity of these people, a scalable form of this support is a key challenge.

In conclusion, the need for convergence amongst the European e-Infrastructure organisations is recognized, there have been some steps forward, but progress is too slow. It is expected that the start of the EOSC will be a driver to boost the evolution towards the e-Infrastructure Commons, as discussed in the next section.

4.2   Community – The user-context for e-Infrastructures

Since the introduction of the visions "Communication Commons" in 2011, "e-Infrastructure Commons" in 2013 and the "Open Science Commons" in 2014, various efforts have been made to reduce the "silo" status of existing e-Infrastructures. These efforts include starting a dialogue among the existing e-Infrastructures and between e-Infrastructures and research communities and/or research infrastructures. Good collaboration between all these stakeholders is a pillar for an e-Infrastructure Commons.

e-IRG recommended that the various user communities should organise themselves. Well organised user communities are better capable at formulating their e-Infrastructure requirements, because they are able to identify dedicated teams of experts bridging the gap between scientific instrument(s) and the ICT supporting the data processing. They can also define how to describe and organise their data.

Traditionally the big sciences (astronomy, particle physics) have been pioneering in e-Infrastructure planning. It is now seen that in many other research domains this self-organisation is maturing rapidly. An important driver for this is the *ESFRI Roadmap* procedure, which has incited many (international) Research Infrastructures to take all aspects of designing, building, operating and even decommissioning an RI into account. For the e-Infrastructure aspects it has been particularly helpful that a dedicated section on e-needs was incorporated in the 2016 Roadmap questionnaire. e-IRG has played an important role in the definition of the e-needs questions and in the evaluation. The experiences have been written down in a separate report[18]. This collaboration, which is mutually beneficial for ESFRI and the e-IRG, will be continued. On the other hand, the long tail of data is ubiquitous, including disciplines usually known to deal with big data, and essential to optimally exploit the capacity of Open Science. Its specific requirements have to be fully taken into account. The e-IRG recently produced a document assessing the specificities of the long tail of data and relevant recommendations[19].

---

[18]http://e-irg.eu/documents/10920/239416/Experiences+of+e-IRG+involvement+in+ESFRI+proposal+evaluation.pdf
[19]Long Tail of Data http://e-irg.eu/documents/10920/238968/LongTailOfData2016.pdf

Besides a dialogue to voice the needs of their research community, it is important that the users participate in the innovation of e-Infrastructure services, to make sure that the evolutions are relevant to the needs. This recommendation from the e-IRG white paper in 2013 has kept all of its importance. History shows, that the requirements of advanced research communities have always been a driver for innovation of e-Infrastructure services and it is expected that this will remain true in the future. Also, the different elements of trust (e.g. section 3.1.5) have to be taken into account properly.

In this domain, the creation of the RDA in 2013 also changes the landscape. The RDA has been since the beginning an international forum for discussion, within disciplines for those which do not have their own discussion place already, between disciplines, and between disciplines and groups which work on "generic" building blocks of data sharing, on technological and sociological aspects. The ESFRI Clusters can also play an important role to organise disciplinary-wide discussions and assessment.

Both nationally and on the European (international) level there have been projects where user communities are steering and actively contributing to service developments. e-IRG expects that this contribution to service developments will be continued. It is the ultimate challenge of such projects to attain that these services become as generic as possible, serving a wider remit of research needs. Here e-IRG applauds the initiative of the Commission to have this service development scrutinized by experts as to prevent that services are developed which might be less apt to become sustainable[20].

4.3 Governance – The policy-aspect for e-Infrastructures

4.3.1 The Open Science Cloud

During the November 2015 e-IRG workshop in Luxembourg, a "Marketplace" as a first step towards the Commons and a "lightweight" integration of e-Infrastructure services was presented. Its main features: one-stop shopping, searchable catalogue, and common access for research and industrial services. The national/regional dimension was also presented: national/regional "views" and/or instances of the EU marketplace can be made available, offering only the locally available services or possibly additional local services not available at EU level. This can constitute a coordinated ecosystem of marketplaces across the EU.

With the adoption of the Digital Single Markets strategy in May 2015, the Commission announced the launch of a cloud for research data – the research open science cloud. In conjunction with the DSM, the declaration of a "European Open Science Cloud[21]" followed. The EOSC aims to create a trusted environment for hosting and processing research data to support EU science by offering 1,7 million European researchers and 70 million professionals in science and technology a virtual environment with free access at the point of use, open and seamless services for storage, management, analysis and re-use of the data that are linked to their research activities, across borders and scientific disciplines. The European Council, together with the Member States, as well as the European Parliament did also welcome the initiative.

Moreover, the Parliament did also call on the Commission, in cooperation with all relevant stakeholders, to set up an action plan to lead to the establishment of the European Open Science Cloud by the end of 2016, which should seamlessly integrate existing networks, data and high-performance computing systems and e-Infrastructure services across scientific fields, within a framework of shared policies, standards and investments.

---

[20]see Horizon 2020 Work Programme 2016-2017 INFRASUPP-03-2016 b) and EC Communication „A Reinforced European Research Area Partnership for Excellence and Growth" COM(2012)392 final
[21]European Open Science Cloud http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

The Commission appointed a High Level Expert Group on the European Open Science Cloud to advise on the scientific services to be provided on the cloud and on its governance structure.

In April 2016, the Commission proposed an ambitious plan for the European Open Science Cloud, in the context of the European Cloud Initiative. The European Cloud Initiative [22] will focus on:

1. a European Open Science Cloud: a trusted, open environment for storing, sharing and re-using scientific data and results; and

2. a European Data Infrastructure: a world-class digital infrastructure to securely access, move, share and process data in Europe.

While the e-Infrastructures were considered as an essential building block of the European Research Area, the e-Infrastructure Commons aims to be an essential building block for the European Open Science Cloud.

Many stakeholders presented their views on the EOSC to the EC. An important contribution in the view of the realisation of an e-Infrastructure Commons is the common position paper[17] "European Science Cloud for Research" between EUDAT, LIBER, OpenAIRE, EGI and GÉANT in which they claim that many of the services needed for the EOSC already exist, that technical and policy barriers remain and in which they provide eight elements[23] for success – these elements or principles are well-aligned with the e-Infrastructure Commons vision. Several workshops on this theme have been organised commonly by the mentioned organisations and the process surely will follow in the forthcomings, providing input for the efforts devoted to progressing the e-Infrastructure Commons.

Discussion around open science topics can be conducted in the newly founded forum Open Science Policy Platform (OSPP)[24]. Outcomes of this debate are crucial for the implementation of the e-Infrastructure Commons and must be included in the ongoing discussion.

---

[22]Cloud computing https://ec.europa.eu/digital-single-market/cloud

[23](i) Open in design, participation and use, (ii) publicly funded & governed with the 'commons approach', (iii) research-centric with an agile co-design with researchers and research communities, (iv) comprehensive in terms of universality and inclusiveness of all disciplines, (v) diverse & distributed empowering network effects, (vi) interoperable with common standards for resources and services, (vii) service-oriented as well as protocol-centric, and (viii) social connecting diverse communities.

[24]OSPP http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-policy-platform
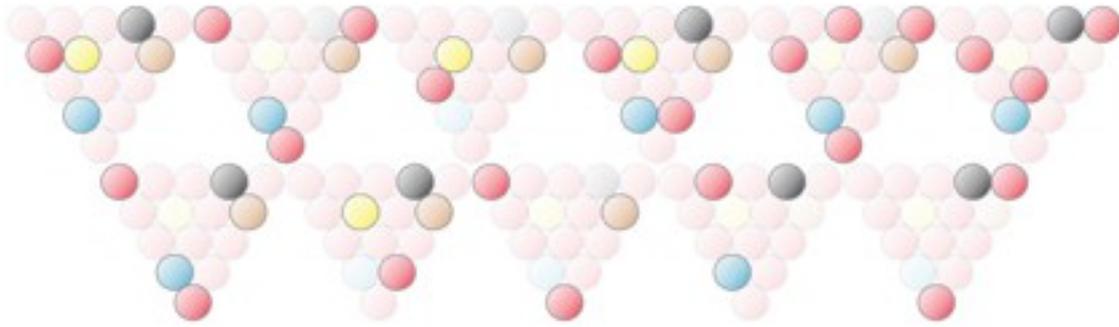
5 Challenges

The realization of a single, unified and open space available for European scientific collaboration as presented in the Vision faces many obstacles. In this chapter an analysis of the key challenges is presented, which hinder implementation of different aspects of the e-Infrastructures Commons but also offer remarkable opportunities to progress towards a desirable outcome. It is very important to identify the most essential ones and take steps to reduce the difficulties as much as possible. The analysis is performed at different levels: policy, organisational, financial, legal and technical.

5.1   Common policies at European level

One of the main obstacles to a straightforward progress is the lack of general guidelines at European level on establishing and maintaining a common e-Infrastructure for scientific collaboration and doing research in general. A clearly defined common outline used by European parties where both major steps and timeline are specified could significantly stimulate the process of e-Infrastructure development as well as e-Infrastructure integration.

A common solution could in principle be a European Research Infrastructure Consortium (ERIC) integrating all those e-Infrastructures. However, arriving there not only involves a huge amount of work, but also hides a number of foreseeable and unforeseeable problems linked in particular to the lack of common policies at the national level. One of them is that a joint e-Infrastructure ERIC in principle calls for a legal entity integrating the national e-Infrastructures owners in each participating member state. Although in some countries this may be relatively easy to do or is already done, in other countries this may require a very long and complex process, demanding to change ministry structures and laws.

The current e-Infrastructures invest a lot in discussions with research infrastructures and research communities in order to understand the user requirements and find solutions to satisfy these requirements. This on-going process provides a clear challenge for the e-Infrastructure providers to commonly address the research infrastructures and research communities and add new services or adapt existing services to fit best the user requirements. The same is true for the Research Infrastructures, who currently have to talk to the different providers separately. All relevant stakeholders could consider to participate in a formal coordination platform between the service providers of the e-Infrastructure Commons and the research infrastructures and research communities.

On European and in most cases also on national level there are different organisations responsible for e-Infrastructure areas devoted to network, computing, data, and security. However, a number of efficiency benefits and advantages can be recognised in those countries where these areas are operated under the same organisational umbrella. These positive and attractive experiences suggest that organisational integration and at least good coordination of the separate e-Infrastructure segments of networking, computing, data etc. is highly desirable both on national and European level in order to increase the national and European capability of jointly taking care of various and complex aspects involved at the different levels. Establishing policies for using and sustaining e-Infrastructures should be a joint effort between users and providers as well as European and national governments.

5.2   Governance structures for integration

Another challenge for the concept of e-Infrastructure Commons is the lack of a governance body responsible for integration aspects on different levels. Indeed, the process of e-Infrastructure integration is rather complex and consists of many issues including organisational, technical, financial and legal.

However, establishing such a body which meets the agreement of all stakeholders is a demanding task. As a consequence, the above challenge should be taken as a signal of the necessity of thinking in a distributed multi-stakeholder model of governance, rather than in a centralised governing body. The role and task of such a multi-stakeholder "assembly" is to agree in common rules and regulations, as well as standardisation decisions and a kind of "modus operandi".

In addition, currently all supercomputing centres have their own policy, which defines access to HPC resources. In many cases it is based on scientific grants evaluated by scientific councils. These policies are different and specific to each HPC centre so unification seems to be difficult if not impossible. A different solution seems to be necessary. A possible option is to designate common spaces for the e-Infrastructure Commons in each regional centre. Admission rules would be subject to a general policy of access defined on the European level. A definition of European grants supervised by scientists representing different domains is one of the potential solutions. All EU projects should have access to such joint e-Infrastructure Commons resources.

5.3   Financing policies

One of the basic problems with e-Infrastructure development and operation is the lack of straightforward business model, which could serve the current operations, guarantee long-term financial sustainability, and moreover be attractive for both the service providers and the users. Since it is very difficult to introduce a top-down controlled development and operation practice, there is also no reality in waiting for a top-level governing and decision-making body. Improving the financial activities and the business model of the e-Infrastructure providers is therefore only possible if joint efforts by the stakeholders leads to a satisfactory solution.

5.4   Legal aspects

The implementation of the e-Infrastructure Commons introduces many legal issues that must be clarified and appropriately treated. One of the most important issue is related to content: the ownership of the knowledge created with the involvement of the common infrastructure is to be seriously considered and dealt with, regarding both availability and accessibility.

Another area of legal issues concerns software licenses. Software licences should allow free use (but not necessarily free of charge) by different scientific communities via the communication (network) infrastructure. Other areas have been dealt with in the past and progress has been made with the

recent laws in the areas of State Aid, Data Protection and Network Regulation developed at European level.

Scientific software offered by national, regional and disciplinary centres is (and should be) under special licenses tailored to the local requirements negotiated by software licence providers.

Software tools provided as a service under the e-Infrastructure Commons must either have a public license or should be offered under special license conditions for the European scientific community.
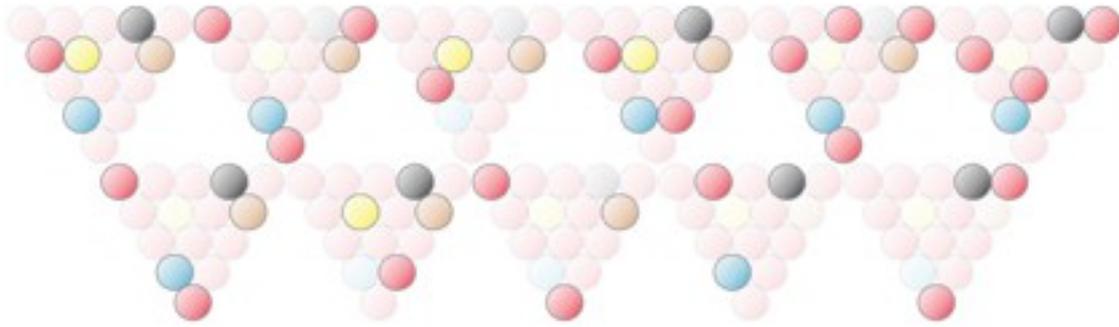
6 Technological aspects

The major technical barriers are stemming from the fact that e-Infrastructures are not only domain-independent (i.e. of horizontal, general character) but in some cases still domain-specific (highly depending on disciplinary areas or thematic fields). Again the border is not at all easy to assign (and also does change over time), but the related tasks are to be well distinguished and shared between the co-operating parties – the providers and the users. Horizontal services should be well covered by the providers, while domain-specific functions should be taken care of by the user groups (the researchers of the various disciplines). Continuous technological development of the e-Infrastructures is a key necessity in keeping the outstanding service quality required by the demanding research and education user community. However there are considerable challenges causing difficulties but also offering stimulating opportunities for the e-Infrastructure developers and operators. Some of them are listed below, without claiming completeness.

6.1.1 The basic challenge: staying at the leading edge

Staying at the leading edge of technological development is obviously the No.1 challenge for the e-Infrastructures. The reason is stemming from the difference between the basic tasks of an e-Infrastructure staff in charge of operations and service provision and an innovator.

Indeed, the major mission of the e-Infrastructure developers and operators is providing stable, reliable and dependable services in a sustainable way, sometimes for a long period of time, by maximum flexibility, outstanding cooperativeness, and high understanding of the users coming from various disciplinary areas, bringing their domain-specific problems and asking for e-Infrastructure support. This is an extremely demanding task, requiring experienced, knowledgeable and motivated experts on the e-Infrastructure behalf. This character is completely different from an "innovator" looking for novel solutions to sometimes rather complex problems, with the freedom of making even extravagant decisions and moving into rather risky ways of solving the problems. A "technological watch" is needed to identify relevant technological progress and their capacity to be sustainable enough to be reliably implemented in an operational environment, and keep the e-Infrastructures and the services at the leading edge. This has to be carefully balanced with the operational needs.

Technological development in the scope of the e-Infrastructure technical areas themselves need to be watched carefully and implemented in the e-Infrastructure facilities and services if that implementation seems to be advantageous and offers contribution to the overall performance of the e-Infrastructure environment.

6.1.2   The last mile challenge

The networks for research have been able to provide a really wide coverage of the user community by supplying the researchers with fast communication possibilities but there is still the challenge to ensure that the performance is available end-to-end in every corner of Europe, the "last or first mile problem". The research networks have to look at all possibilities of increasing the transfer rate to satisfy the demands for low-latency networks. Flexibility is a new requirement related to more automated and responsive networks that needs to be taken up. Mobile networks have been out of the scope of networks for research but this might change in the near future.

6.1.3   HPC technology challenges

Recent key advances in computing deal with massive parallelism, sometimes on a myriad of possible options, and lowering the required energy. Besides that, new ideas arise in the combination of data and computing, as it is creating new opportunities and challenges. In addition, the day the quantum computer becomes available may change the whole of computing for research environment. Furthermore, the impact of parallelism driven by the core count goes up faster than the performance of an individual core. Hence the challenge is how to make all the legacy software capable of exploiting such massive parallelism.

6.1.4   Security challenges

Another example of technological challenge is the lack of common security policy. Security policies and related infrastructure in the member states are heterogeneous at the moment.

Access to the e-Infrastructure Commons should be defined by a distinct policy jointly established by the involved scientific communities and resource providers.

Local authorities should be allowed to influence and, in some well-specified cases, also change a decision taken at a higher level in specific situations (e.g. arrangements violation). While e-IRG acknowledges that Authentication and Authorization Infrastructure (AAI) has done very big steps ahead and works satisfactory in several stand-alone areas (such as the GEANT network), a common or interoperable identity framework[25] for all e-Infrastructure providers should be developed, supporting eduGAIN and possibly other schemes such as the e-Government IDs (e-IDs).

6.1.5   Challenges for data infrastructures

As explained challenges to make data Findable, Accessible, Interoperable and Reusable (FAIR) require the active involvement of the scientific communities. The example of disciplines which are at the leading edge, and how data availability truly revolutionised the way they do science, can be used to convince less advanced communities. Dissemination of skills and best practices is another key element.

Challenges for data infrastructures include more technical aspects and policy aspects, such as harmonizing access to collections and data centres, providing users with a homogeneous interface to heterogeneous resources which would allow them to access the federated nodes without knowing anything about the underlying storage. Other issues include ensuring the integrity of data across the federation, dealing with access policies and monitoring (where communities need to know who has accessed their data to support continued funding); making data available for further analysis, etc. With the increasing amounts of data and the rapid development of new technologies in the ICT industry resulting in rapid obsolescence of media and data formats, it is also becoming more of a challenge to keep data readable and understandable for future use by a new generations of scientists. The primary

---

[25]EUGridPMA https://www.eugridpma.org/

reference over the last few years has been the Open Archival Information System (OAIS) which stresses the importance of preserving the context as well as the content of an information item. In the current Big Data era, the challenge is to scale up these procedures to work efficiently in heterogeneous archiving facilities as a part of a highly distributed, multi-administrative domain storing petabytes of scientific content in complex data formats. At the same time, means for maintaining versions, tracking and recording changes and ensuring data integrity have to be developed, while keeping costs at an affordable level.

6.1.6   Innovation and industrial partnerships challenges

Providing e-Infrastructure services for innovations (technology, product and/or service developments) outside the research and education community is getting more important and is definitely rather challenging – at least in the early phases of establishing Public-Private-Partnerships-based and similar collaborations between the Research sector and their potential industrial partners (e.g. PRACE and its collaborations with industrial partners). On the other hand, integrating industrial commodity services in the e-Infrastructure service portfolio is also getting more important for the users. In order to make PPPs work it is essential to clarify the roles so the parties collaborate rather than compete.

The Internet of Things (IoT) will have a considerable impact in the research world mostly in the form of machine-to-machine (M2M) communication - think about distributed sensors in the fields of environmental protection, biology, agriculture, earthquake observations or remote controllers of big research installations. This will add more data to be transported, stored, computed, preserved and archived, and additional work to deal with the data curation challenges, including the definition of formats and metadata. It could also have an influence on the needed capacity of communication channels.
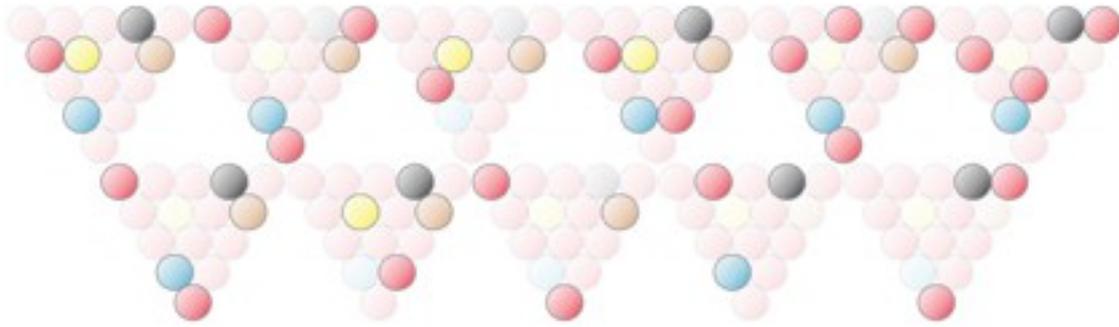
7 Realising the e-Infrastructure Commons – Recommended actions

In this chapter the way forward is proposed taking into account the previous sections, namely the vision, the current landscape and the challenges. Concrete actions expressed as recommendations are also formulated, taking into account the recommendations in the 2013 White Paper.

7.1 The way forward

From the previous analysis it is clear that an emphatic co-operation among all main stakeholders is required: the providers (the e-Infrastructure developers and operators), the users (the scientific communities, both big users including Research Infrastructures and the long tail), and the funders (the EC and the national governments and their agencies). A joint EU e-Infrastructure ERIC still seems to be far away, and thus the only way forward is good coordination through a formal coordination platform among all stakeholders in-line with the Commons, implementing a distributed multi-stakeholder model of governance.

Keeping the right balance between operation of services and development of innovative ones, and working further on sustainability of the services is key. The new EC tools of Framework Programme Agreements (FPAs) and the upcoming Operational Grants are in the right direction for increasing users' confidence in the e-Infrastructure long-term sustainability.

Common spaces with common access and security policies should be gradually implemented as slices of the different e-Infrastructure resources, as it may not be possible to harmonise all the resources of all member states. A common or interoperable identity framework for all e-Infrastructure providers should be developed, compatible with eduGAIN and possibly other schemes such as the e-Government IDs (e-IDs).

A concrete way towards the e-Infrastructure Commons, loosely integrating the different types of e-Infrastructures, is to use a marketplace with a proper governance including a representation of the users as a single point of access to all e-Infrastructure services and tools; the marketplace will act as a one stop-shop for EU researchers, i.e. a place where all e-Infrastructure services are accessible all together, either directly or redirected elsewhere. The marketplace can make use of several technologies and services, such as cloud technologies, a searchable service catalogue and a common authentication/authorisation scheme. In this way, standardised and single point of access to services will be achieved, without promoting monopolies, nor a single integrated provider that has proven to

be very difficult across different e-Infrastructure components. On the contrary, it will be open to new actors, encouraging cooperation, competition and innovation.

The national/regional dimension will be also strongly visible: national / regional perspective / abstractions and/or instances of the EU marketplace will be available, i.e. a sub-set of the EU services, based on national participation or availability of services and also on national laws and restrictions of services at national/regional level, including commercial ones. On the other hand, extra national/regional services may be available in the different member states or regions. This will constitute a coordinated marketplace ecosystem among EU and national/regional levels, being automatically synchronised among them.

Education and training, and dissemination of knowledge and good practices, should play a major role in preparing the actors to participate fully in the endeavour and take full benefit of it.

7.2   Recommended Actions

7.2.1   User communities

The key recommendations for user communities can be summarized as '*organise yourselves*'. Well-organised user communities are better capable at formulating their e-Infrastructure requirements, because they are able to separate out dedicated teams of experts bridging the gap between scientific instrument(s) and the ICT supporting the data processing. The 2014 White Paper contains the following recommendations for user communities:

Drive the long term strategy for their e-Infrastructure needs
While traditionally the big sciences (astronomy, particle physics, biology, etc.) have been pioneering in e-Infrastructure planning, it is now seen that in many other research domains this is maturing rapidly. An important driver for this is the ESFRI Roadmap procedure, which has caused many (international) Research Infrastructures to take all aspects of designing, building, operating and even decommissioning an RI into account. For the e-Infrastructure needs it has been particularly helpful that for the 2016 Roadmap a dedicated section on e-needs has been incorporated and e-IRG has played an important role in the evaluation. The experiences have been written down in a separate report[26]. This mutually beneficial process for ESFRI and e-IRG will be continued.

e-IRG concludes that this recommendation, whilst it is taken up, is still very valid.

Participate in the innovation of e-Infrastructure services
This recommendation has not lost any of its importance. History shows, that the requirements of advanced research communities have always been a driver for innovation of e-Infrastructure services and there is no indication this will not be just as valid in the future.

Both nationally and on the European (international) level there have been, are and will be many projects where user communities are steering and actively contributing to service developments. It is the ultimate challenge of such projects to attain that these services become as generic as possible, serving a wider remit of research needs. Here e-IRG applauds the initiative of the Commission to have this service development scrutinised by experts as to prevent that services are developed which might be less apt to become sustainable.

Community-specific services should be integrated with the European e-Infrastructure Commons. Users have to be able to articulate and communicate their community-specific needs in regards to

---

[26]http://e-irg.eu/documents/10920/239416/Experiences+of+e-IRG+involvement+in+ESFRI+proposal+evaluation.pdf

data and services and translate these needs into clear functional requirements that will drive the development of e.g. VREs (Virtual Research Environments). VREs operators need to link their services with existing underlying e-Infrastructure, building on top of available backbone services.

Contribute to standards and take care of your data!

Compared to the 2013 situation e-IRG gives credit to the establishment of RDA, which has given a worldwide boost to working on standardising data issues. Many of the RDA working groups have roots in research communities.

With regard to data e-IRG sees that the FAIR principles (Findable Accessible, Interoperable and Reusable data) getting universally accepted and implemented as the way to move forward.

Within research communities e-IRG notes an ever-increasing awareness that data quality, data management, data handling, data persistence are essential to get the best scientific results. In fact, e-IRG has noted that for many distributed RIs opting for ESFRI status, guarding and preserving data quality is one of the main drivers to organise themselves!

So users should in all cases take care of their data, involving the definition of the required disciplinary standards as explained above, but also activities in data stewardship, and including data policies and data management in all projects from the start. The top-down requirement from funding agencies, to provide a Data Management Plan for all project proposals, is also moving the lines.

All in all, e-IRG notes a considerable progress with regard to the original recommendations for user communities.

7.2.2   (European and national) e-Infrastructure providers

The key recommendations for e-Infrastructure providers can be summarized as '*work closely together*'. Based on the landscape analysis, e-IRG concludes that convergence amongst the European e-Infrastructure organisations is taking place, but progress is too slow.

One of the step forward is assurance of a good coordination through a formal coordination platform among all stakeholders inline with the Commons, implementing a distributed multi-stakeholder model of governance. It may allow a staged approach towards a common ERIC.

e-IRG concludes that a coordination platform among all stakeholders inline with the Commons, along with a distributed multi-stakeholder model of governance is needed. One of the proposed solutions and step forward could be the introduction of interoperable service of catalogues. Only then users may be able to enjoy a single point of access and as widely as possible common access and security policies, as well as long-term sustainable services. Preferably this coordination platform is built up from strong national building blocks (see next paragraph). It should also include the current "gems", as underlined in the report of the EOSC High Level Expert Group.

7.2.3   National governments/funding agencies

The 2013 White Paper contains the following recommendations for this actor group:

1.   Provide a basic funding level for the national e-Infrastructure, in particular devoted to its continuous innovation;

2.   Empower and fund national user communities for the use of e-Infrastructure services, enabling them to influence the development of the national e-Infrastructure;

3.   Remove existing national regulatory or political constraints for accessing publicly funded e-Infrastructures for private research and public-private research ventures;

4. Provide input for the strategy setting and coordination bodies for their national e-Infrastructures;

5. Encourage the actors in the national e-Infrastructures to collaborate and join forces with their counterparts in other countries and at EU level.

With regard to these recommendations e-IRG sees many indications of progress. In an increasing number of European countries, e-Infrastructures are appearing as a separate item on national Roadmap for (large) research infrastructures, recognizing the importance of these horizontal infrastructures. The national research funders also increasingly require data management paragraphs in research proposals.

Furthermore e-IRG sees an increase in countries, where the various cornerstones of e-Infrastructure development and provisioning (computing, data, networking), that used to act quite independently and uncoordinated, start to team up. This is driven by the sheer necessity to address the institutions and their users in a more coherent fashion, which requires a coordinated planning and funding of e-Infrastructure service development and innovation. In some cases this coordination results in consolidation into a single organisation, with governance models varying between countries (such as the SURF cooperative in NL or the CSC - IT Center for Science (limited company) in Finland).

National e-Infrastructure organisations have a pivotal role. Usually they have strong formal links (such as through memberships) with research institutions and universities, where the 'long tail of science' resides. On the other hand they form the link to European e-Infrastructure projects and organisations. It is nearly impossible to reach this long tail on the European level: you need the intermediate national and institutional connections!

Currently national e-Infrastructure services are quite often not open by default to user groups from other countries and this makes coordinated provisioning at European level difficult.

Specific attention should be devoted to the challenge to serve the long tail of science. It is clear that infrastructure provision is indeed very different depending on to whom one wants to provide infrastructure for. The variation can be described as two continuums:

1. From very high level of technical e-Infrastructure competence, via lower levels to close to zero understanding.

2. From internationally very well organised research communities, via moderately organised to unorganised individual researchers in need of infrastructure.

These various preconditions related to the dichotomy may call for different infrastructure provisions strategies, organisational set-ups, policies and funding schemes. Again, the role of national and local level is very important in this domain, as well as in some cases the one of disciplinary level, and the necessity to integrate infrastructures provided at different levels should be recognised.

The e-IRG recommends strongly, that e-Infrastructure coordination and consolidation on the national level is embraced in full force in every European country. A strong European e-Infrastructure is dependent on strong national building blocks.

However, given always tight budgets for research, governments and funding agencies keep struggling with implementing mechanisms to fund a (generic) national e-Infrastructure and meanwhile keeping a balance with (other) research infrastructure priorities. Taken to the extremes, there are two mechanisms:

1. 'market approach': fund researchers and research communities and let them develop the services or buy them from whatever provider they want;

2. "public provider approach": fund a (generic) national e-Infrastructure and offer these to researchers and research communities for free at the point of use;

Clearly both approaches have their pros and cons. Intermediate and mixed models (e-Infrastructure service development co-funded by users and national e-Infrastructure funds, including the commitment to a long lasting relationship between both) are also frequently advocated. A distinct variety is funding the development of a discipline specific e-Infrastructure on top of a generic (national) e-Infrastructure.

e-IRG observes that in various countries funding agencies are using different models but clearly there is no systematic approach as to assessing their effects and there is no forum to exchange best practices. This lack of clarity on who funds what and assigns what to whom also hinders the resource provisioning on the European level (i.e. for the benefit of European organised user communities).

In this context e-IRG refers to one of the conclusions of the Competiveness Council May 28-29, 2015, of the European Council conclusions on *open, data-intensive and networked research: "ESFRI is invited to explore mechanisms for better coordination of Member States' investment strategies in e-Infrastructures, covering also HPC, distributed computing, scientific data and networks."* For this purpose ESFRI has established a working group with e-IRG representation. The main goal of the group is to analyse the investment strategies of the Member States in e-Infrastructures for research and innovation and formulate recommendations for how mechanisms for coordinating these strategies could be implemented. Important issues to be discussed in connection to national investment strategies include coordinated governance of service delivery, life cycle management for services, and coordinated funding streams for investments and operations.

e-IRG believes, that the funding system must facilitate the right incentive structure to reach the ideal situation, believed to be consisting at least of a balanced mix of:

1. base funding for the innovation of the (national) e-Infrastructure;

2. funding by users derived from service delivery by the providers;

3. top up funding based on (national) priorities for (demanding and well organised) research communities;

e-IRG supports strongly that national e-Infrastructure funding and governance mechanisms are analysed, so that best practices can be identified, which can contribute to more specific recommendations to national governments and funding bodies.

7.2.4 European Commission

The 2013 White Paper contains the following recommendations for the EC:

1. Establish a European harmonised framework for the funding of e-Infrastructure innovation;
2. Empower and fund European user communities, such as the ESFRI projects, to influence the development and use of transnational access to the e-Infrastructure;
3. Enable and promote the use of Structural Funds for e-Infrastructure development in less favoured areas;
4. Provide input for the European strategy setting and coordination bodies and their umbrella forum;
5. Strive towards harmonisation to avoid regulatory conflicts with existing regulations for (among others) state aid or competition rules;
6. Provide clear guidelines for 'regulation proof' participation of private research in the use of e-Infrastructure services.

The key recommendations for EC can be summarized as '*develop the necessary harmonised scope, framework and instruments for improving Research Infrastructures, including e-Infrastructures in terms of operations, innovation and sustainability*'.

It has already been mentioned, that the ESFRI process has been an enormous driver for strategic planning of RIs on the national and European level, with e-Infrastructure strategy setting and planning in its slipstream. It is further observed that in its Horizon2020 work programme (Research Infrastructures including e-Infrastructures) the European Commission is strongly supporting many of these directions. With regard to e-infra this is evident from the defined themes in WP16/17:

1. integration and consolidation of e-Infrastructure platforms supporting European policies and research and education communities;

2. prototyping innovative e-Infrastructure platforms and services for research and education communities, industry and the citizens at large; both platform driven and user driven.

e-IRG also applauds that innovation and service development is now linked to a process in which the service life cycle management is taken into account, involving area experts appointed after consulting e-IRG.

However, European funding for infrastructure innovation and development is scarce and usually distributed to the well-established actors in the e-Infrastructure landscape. e-IRG also observes that in some respects there is still a separation of funding opportunities for the various European e-Infrastructure organisations, which somewhat reduces the incentive for cross-platform coordination and development.

In general the domain of universities and public e-Infrastructure development agencies (national providers) is typically far from resembling to the business incubators which often have not only put new life in the Internet economy, but have also developed and now operate more and more of the e-Infrastructures favoured by researchers. There is, therefore, an overall need to focus on introducing and testing new e-Infrastructure types and approaches, not just regarding funding, but also services, advice, connections and even physical space and sandbox hardware.

To initially address this challenge there is a supplementary funding need, which aptly can be termed venture capital not only to allow innovation and development of new infrastructure, but also to allow for more disruptive technology and disruptive actors – complementing, challenging or, if prove to be more efficient, more helpful and more dependable, even substituting traditional e-Infrastructure providers and their services.

> *e-IRG recommends that in future Work Programmes the EC provides strong incentives for cross platform innovations, thereby further supporting the need for coordination and consolidation of e-Infrastructure service development and provisioning on the national and the European level.*

A final remark is the following: the European e-Infrastructure landscape has recently experienced a policy shift regarding European funding schemes – basically into 1) operations; and 2) innovation and development. It is suggested that the validity and the correctness of the above policy shift is to be further investigated in order to either confirm or invalidate the pertinence of that policy shift. If the usefulness of the policy shift is not proven then the EC should be convinced to discontinue shifting their related policy, otherwise the reasons for and preconditions determining this split are to be further elaborated and broadly analysed in several funding scenarios, namely. The problem of funding the development of key 'sociological' elements in particular for enabling data sharing in a framework based on technological indicators should also be addressed.
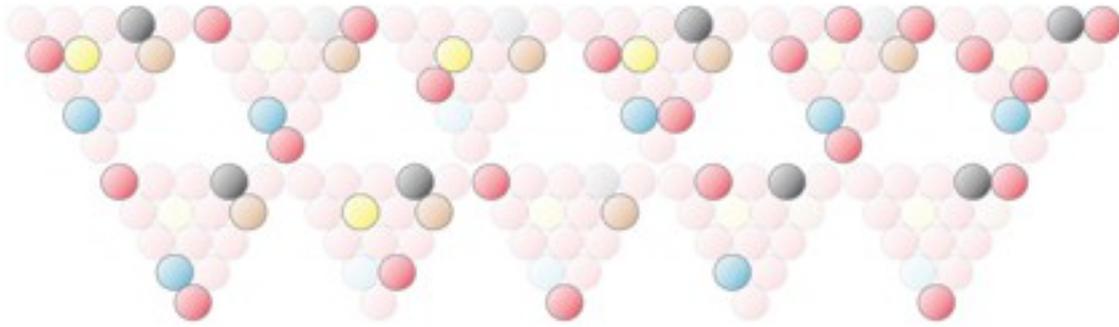
8 Glossary

Term	Explanation
AAI	Authentication and Authorization Infrastructure
AARC	Authentication and Authorization for Research and Collaboration (EU-funded project)
AUP	Acceptable Use Policy
CERN	European Organization for Nuclear Research
DANTE	Delivery of Advanced Network Technology to Europe
EARN	European Academic and Research Network
EC	European Commission
ECI	European Cloud Initiative
EDI	European Data Infrastructure
EGI	European Grid Initiative, a federation of resource centres and coordinated by EGI.eu
EOSC	European Open Science Cloud
ERA	European Research Area
ERIC	European Research Infrastructure Consortium
ESFRI	European Strategy Forum for Research Infrastructures
EU	European Union
EUDAT	European Data Infrastructure (EU-funded project)
e-IRG	e-Infrastructure Reflection Group
e-ID	e-Government Identification Data
FAIR	Findable, Accessible, Interoperable and Reusable
FPA	Framework Programme Agreement
G8	Group of 8 (now G7 due to Russia'a suspension) is a governmental political forum of the leading industry nations and the EU

GÉANT	Pan-European network that connects the NRENs in Europe and beyond and name of the organisation that operate the network
HNX	Helix Nebula Marketplace
H2020	Horizon 2020 (EU Research and Innovation programme 2014
ICT	Information and Communications Technologies
ID	Identification Data
IoT	Internet of Things
LIBER	Ligue des Bibliothéques Européennes de Recherche (Association of European Research Libraries)
MoU	Memorandum of Understanding
NGI	National Grid Initiative
NREN	National Research and Education Network
OpenAIRE	Open Access Infrastructure for Research in Europe
PLAN-E	Platform of National eScience Centers
PRACE	Partnership for Advanced Computing in Europe
RDA	Research Data Alliance
RARE	Réseaux Associés pour la Recherche Européenne (Name of the association that became TERENA)
RI	Research Infrastructure
TERENA	Trans-European Research and Education Networking Association
VRE	Virtual research Environment