RISCAPE

Mapping the Research Infrastructure Landscape

Landscape analysis report in the e-infrastructure sector

Date	18 November 19
Activity	WP10
Lead Partner	EGI.EU
Document Status	VERSION FOR PROJECT LANDSCAPE

Abstract

This RISCAPE deliverable is the landscape analysis about e-infrastructures worldwide. Digital infrastructures, commonly known as e-infrastructures or cyberinfrastructures, address the needs of researchers and innovators for transnational communication networks, high performance and high throughput computing, multidisciplinary data management and collaborative scientific software. E-infrastructures empower researchers with easy and secure online access to facilities and resources and enable them to deliver reusable and reproducible research and innovation outputs. The document reviews the global e-infrastructure landscape in four areas: Networks; High Throughput Computing (HPC); Grids & clouds; Data. A global picture of infrastructures and infrastructure-development projects is provided for each area.

This version will be integrated into the overall Research Infrastructure landscape prepared by RISCAPE. The landscape will be published in December 2019.

COPYRIGHT NOTICE

DELIVERY SLIP

	Name	*Partner/Activity*	*Date*
From:	Gergely Sipos (author)	EGI.EU / WP10	24/07/2019
Moderated by:			
Reviewed by	Mark Dietrich Ari Asmi	External advisory board University of Helsinki / WP1	28/06/2019 04/07/2019
Approved by:			

DOCUMENT LOG

Issue	*Date*	*Comment*	*Author/Partner*
v.0.1	12/06/2019	First version for review by peer infrastructures and advisory board.	Gergely Sipos/EGI.EU
v.0.5	03/07/2019	Version addressing comments from reviewer 1	Gergely Sipos/EGI.EU
v.1	23/07/2019	Version addressing comments from reviewer 2	Gergely Sipos/EGI.EU
v.2	13/11/2019	Version for integration into RISCAPE final landscape	Gergely Sipos/EGI.EU

Contents

Executive summary

This RISCAPE deliverable is the landscape analysis about e-infrastructures worldwide. Digital infrastructures, commonly known as e-infrastructures or cyberinfrastructures, address the needs of researchers and innovators for transnational communication networks, high performance and high throughput computing, multidisciplinary data management and collaborative scientific software. E-infrastructures empower researchers with easy and secure online access to facilities and resources and enable them to deliver reusable and reproducible research and innovation outputs.

The document starts with a short introduction of the European e-infrastructure landscape, covering four types of e-infrastructures: Network (Connectivity); High Performance Computing (supercomputers); Grids & Clouds (clusters, grids and IaaS-PaaS-SaaS compute services); Data (storage and data management infrastructures as well as the data they host). GÉANT and its NRENs, PRACE and its HPC facilities, EGI and its NGIs, EOSC, RDA, EUDAT, OpenAIRE, CoreTrustSeal are introduced within these four areas. PLAN-E, RDA, HPC Centre of Excellences and EOSC-hub Competence Centres are also described to complement the European e-infrastructure landscape with related initiatives.

The main part of the document focuses on the landscape outside Europe. It is approached also from the same four areas: Network; High Performance Computing; Grids & Clouds; Data. Related e-infrastructures and infrastructure development projects are introduced within each area and, where possible, compared/positioned with the European peers.

The document closes with 14 observations and recommendations. Six of these relate to the methodology we used to pull together data for this report and make specific observations and recommendations to Europe concerning e-infrastructures in other regions of the world. The other 8 recommendations relate to certain types of e-infrastructures and point out common trends across the world that Europe should monitor and possibly follow.

1 Introduction

Today the European Union (EU) has 1.7 million researchers and 70 million science and technology professionals engaged in the creation of new knowledge, products, services and processes 5. Underpinning their efforts, digital infrastructures, commonly known as "e-infrastructures," foster innovation and scientific progress across disciplines and between the private and public sectors. Digital infrastructures are the virtual backbone of research and a vital driver for innovation. Without them there would be no sharing of data, no exchange of know-how and no collaboration.

EU-funded digital infrastructures address the needs of European researchers and innovators for transnational communication networks, high performance and high throughput computing, multidisciplinary data management and collaborative scientific software. They empower researchers with easy and secure online access to facilities and resources and enable them to deliver reusable and reproducible research and innovation outputs. But e-infrastructures are set to play a much larger role in society. The overarching aim of the e-infrastructure activities in Horizon 2020 is to achieve by 2020 a single and open European space for on-line research where researchers enjoy leading-edge, ubiquitous and reliable services for networking and computing, and seamless and open access to e-Science[1] environments and global data resources [2].

The major changes in Horizon 2020 with regard to the previous framework programme (FP7) include the emphasis on innovation and the development of human resources, addressing industry as an e-infrastructure supplier and user, more resolve towards service orientation and service integration, and more emphasis on data infrastructure development. In Horizon 2020 the e-infrastructure activities are part of the "European Research Infrastructures, including e-infrastructures" programme. The indicative budget for e-infrastructures from 2014 to 2020 is 890 million euros.

Successful research and innovation needs access to first class research infrastructures, and development of top-class research infrastructures is one of the key areas in the implementation of the European Research Area. In the overall context of research infrastructures, e-infrastructures play a more and more important role. Today, almost all large-scale research activities include or are supported by several e-infrastructure components. Major scientific breakthroughs are increasingly achieved by an international and cross-disciplinary team transferring, storing and analysing vast data collections, and performing advanced simulations using different types of computing facilities. Understanding the global e-infrastructure landscape is therefore of upmost importance for Europe – for both researchers and for those funding, designing, implementing and operating e-infrastructures. Collaboration with e-infrastructures outside Europe can enable new modes of science where resources, in particular the other research infrastructures, are shared and used remotely to overcome fragmentation and to cope with increasing costs and complexity. Areas of expertise can be shared between Europe and the rest of the world, helping communities in adopting good practices.

[1] e-Science is computationally intensive science that is often carried out with the use of e-infrastructures.

This document provides a review of the global e-infrastructure landscape, starting with a short introduction to the European e-infrastructures (Section 2), then continuing with the landscape outside Europe (Section 3). After the landscape review the document closes with a section on lessons learnt and recommendations for the European e-infrastructure stakeholders, including the European Commission (Section 4). Data for this report was gathered by the EGI Foundation using online and face-to-face interviews with both international and European e-infrastructure representatives, harvested from presentations given at regional e-infrastructure events and from public websites and publicly available documents. Although the prime focus of the report is on infrastructures located outside of Europe, it was inevitable to include in our review also those projects that advance such infrastructures with respect to the offered services, and/or with the interconnection and alignment to the European ones. In some regions and for certain types of e-infrastructures the line between 'being an infrastructure or being a project' is blurry, as many e-infrastructures are funded through short term project cycles (2-3 year long projects) without the long-term business and sustainability plans that RISCAPE has identified as an important characteristic of an "infrastructure".

At the end of the report, in Appendix I. we provide a table listing the *e-infrastructures* covered by the report and in Appendix II. a similar table about the EC-funded or EC-related e-infrastructure *projects*.

2 The European e-infrastructure landscape

Before reviewing the global landscape, we start with a short presentation of the European landscape – we do so because then we apply the classification we use for e-infrastructures in Europe, to the rest of the world:

- Network (Connectivity)
- High Performance Computing (supercomputers)
- Grids and clouds (clusters, grids and IaaS-PaaS-SaaS compute services)[2]
- Data (storage and data management infrastructures as well as the data they host)

Pan-European e-infrastructures are dependent on the existence of strong and coherent national e-infrastructure nodes and their cooperation and complementarity to enable cross-border services for scientific communities. The main strategic body facilitating the overall integration of European e-Infrastructures is e-IRG, the 'e-Infrastructure Reflection Group'.

The pan-European e-infrastructures for networking, high-performance computing (HPC) and high-throughput computing (HTC)[3] (clusters built from more commodity-type hardware) are already well established and provide services used by international research and research infrastructures. Also, data and cloud infrastructures are developing fast. It should be noted that the European-level e-infrastructure services are often provided by national e-infrastructures in a collaborative setting and that the pan-European initiatives are dependent on the existence of strong and coherent national e-infrastructure nodes and their cooperation and complementarity to enable cross-border services for scientific communities. Available e-infrastructures and their related services might not always fulfil the user's needs. But collaboration and coordination between e-infrastructures and between e-infrastructures and research infrastructures is ongoing. The main strategic body facilitating the overall integration of e-Infrastructures and connected services within and among European member states is e-IRG, the 'e-Infrastructure Reflection Group' [89]. The following subsections present a brief introduction of the major pan-European e-infrastructure initiatives mostly reusing content from [3].

2.1 Networks

Connecting research communities across the globe is a prerequisite to stimulate exchange of ideas, data and results. 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,

[2] In this document we use the 'grids and clouds' category to refer to a number of different types of infrastructures: high throughput compute infrastructures (institutional clusters and grids), as well as different types of utility cloud services (Infrastructure as a Service, Platform as a Service, Software as a Service).

[3] HTC generally involves running multiple independent instances of a computational task on multiple processors at the same time. HPC generally involves running a single instance of complex parallel software over many processors.
Serial systems, such as large number of desktops or servers are suitable for HTC, while HPC requires supercomputers where cores are connected with higher bandwidth interconnects.

as well as the way the access to NREN resources is managed. The GÉANT Association [61] has gradually grown into a pan-European organisation, where the associated NRENs link together research communities and provide trans-national access to research infrastructures and research resources. GÉANT provides interconnectivity between NRENs across 43 European countries, serving an estimated 50 millions of users of practically all research disciplines and thematic domains. In addition to pan-European connectivity, the GÉANT network has international connections to a large set of partner networks (some 70 NRENs) worldwide, in particular through regional agreements – thereby enabling international collaboration for 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.

2.2 High Performance Computing

The High Performance Computing (HPC) national infrastructures are federated at the European level in the Partnership for Advanced Computing in Europe (PRACE) [62]. PRACE offers access to world-class high-performance capability computing facilities and services. PRACE is managed by the PRACE AISBL organisation. 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 " Excellence of science" and supported by scientific peer-review. There are basically two forms of access:

- preparatory access, intended for short-term access to resources, for code-enabling and porting, required to prepare proposals in the category "project access" and to demonstrate the scalability of codes;
- project access, intended for individual researchers and research groups including multi-national research groups, which can be used for 1-year, 2-year or 3-year (Multi-Year Access) production runs.

For the national HPC infrastructure(s), the access modes are closely connected to the ruling national governance.

A new legal and funding entity, the European High-Performance Computing Joint Undertaking (EuroHPC JU) will pool European and national resources to develop top-of-the-range exascale[4] supercomputers for processing big data, based on competitive European technology [63]. EuroHPC JU develops a pan-European supercomputing infrastructure and supports research and innovation activities during the development and later in the exploitation of the HPC infrastructure.

[4] Exascale computing refers to computing systems capable of at least one exaFLOPS, or a billion billion (i.e. a quintillion) calculations per second. Such capacity represents a thousandfold increase over the first petascale computer that came into operation in 2008.

2.3 Grids and clouds

In this document we use the 'grids and clouds' category to refer to a number of different types of compute infrastructures: high throughput compute infrastructures implemented in the form of institutional clusters and compute grids, cloud compute infrastructures implemented as Infrastructure as a Service, Platform as a Service or Software as a Service.

A rich variety of such infrastructures exist within the academic sectors of European countries. On the national level these are often brought together using national infrastructures (NGIs[5]). The NGIs are federated into the EGI[6] pan-European computing infrastructure [64]. EGI is managed by the EGI Foundation. The EGI Foundation and its NGI members provide solutions built through a service catalogue [65] that has been evolving during many years. The EGI Federated Cloud Solution offers a standards-based and open infrastructure to deploy on-demand IT services that can host, process and serve datasets of public or commercial relevance and can be flexibly expanded in capacity and capability by integrating new providers. This is complemented by the EGI High Throughput Computing (HTC) Solution that provides a global high-throughput data analysis infrastructure, linking large number of compute clusters from independent organisations and delivering computing resources with high scalability.

Access to EGI's resources (cloud or HTC) is provided through various access modes, such as free grant-based allocations, pay per use, and annual membership fees. Moreover, EGI has agreements with resource/service providers from outside Europe to increase the interoperability and availability of compute and data resources from those regions. At the time of writing agreements are in place with Africa-Arabia, Asia-Pacific, China, Canada, India, Latin America, Ukraine and the US [95].

The European Commission adopted on 19 April 2016 the communication "European Cloud Initiative – Building a competitive data and knowledge economy in Europe" [66] that sets out a strategy to strengthen Europe's position in data driven innovation, improve its cohesion and help create a Digital Single Market in Europe. The European Cloud Initiative – which started implementation in 2017 under the name 'European Open Science Cloud' (EOSC) – will provide European science, industry and public authorities with world-class data infrastructures, high-speed connectivity and increasingly powerful computers and networks of computers.

Since January 2018 the EGI Foundation coordinates the 33m euro EOSC-hub project [67] that puts in place the core elements of EOSC. EOSC-hub builds on the EGI Federated Cloud and HTC services, as well as services brought in by partner projects and other providers, most notably EUDAT [70] (see in next section), and various Research Infrastructures. EOSC-hub encompasses a broad and growing spectrum of e-infrastructure and research infrastructure services that 'can be accessed as utilities' (hence the word 'cloud in the name of EOSC) via the EOSC Portal [104].

[5] "NGI" used to mean "National Grid Infrastructure" but EGI broadened the meaning to include cloud and data e-infrastructure services.

[6] EGI used to stand for 'European Grid Infrastructure' but this long name was abandoned to indicate the broader types of e-infrastructure services that EGI started to provide over the years.

The EOSC landscape significantly broadened in 2019 with the start of over 10 new H2020 EOSC-contributor projects [68], and with the establishment of the EOSC Executive Board and EOSC Governance Board (gathering representatives from the Member States and the Commission) [69].

2.4 Data

While Research Infrastructures primarily operate at the beginning of the data production pipeline with generating, filtering raw data and creating data products that are 'consumable' by the broader scientific community, e-infrastructures are mostly concerned with developing and offering tools and services that help RIs implement such pipelines and enable the sharing of the data products for scalable access and data analysis.

Data is a key component of research infrastructures, a fundamental scientific product offered for scientific and commercial exploitation. This was stressed in the "Riding the Wave" report [4] of the relevant EC High Level Expert Group in 2010 and its follow-up RDA Europe Report "The data Harvest" (2014) [5], and recently in the ESFRI 2018 Roadmap [6] where the ESFRIs "generally understood EOSC as a European Open Science (Data) Commons". The storage, curation, archival and sharing of scientific data for download and for online analytics is a shared challenge of e-infrastructures and Research Infrastructures. While Research Infrastructures primarily operate at the beginning of the data production pipeline with generating, filtering raw data and creating data products that are 'consumable' by the broader scientific community, e-infrastructures are mostly concerned with developing and offering tools and services that help RIs implement such pipelines and enable the sharing of the data products for scalable access and data analysis. This symbiosis between RIs and e-infrastructures is in continuous evolution, and EOSC is expected to give a boost to the interactions[7].

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 industry, to the public sector or can be citizens.

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, intellectual property rights (IPR), …). Creating such an interoperable European Data Infrastructure may take many years. Initiatives at the European level have been started to offer various services (e.g. storage, permanent identification, access, anonymisation, discovery, monitoring, semantic linking, validation, data management planning) for research data in general (EUDAT services [70]) for publications and a growing range of other research outputs (OpenAIRE [71] and its services such as the Zenodo repository [72]), and for the caching and staging of research data to/from compute resources (EGI data services [65]). Scientific communities and research infrastructures have been building frameworks for data sharing

[7] For example, with the recently started INFRAEOSC ESFRI Cluster projects where the ESFRIs bring data and services into EOSC.

(formats, metadata …) and in many cases building their own data infrastructures, often at the international level, taking into account their specific needs. National, regional and local authorities have also set up data infrastructures. All of them should be interconnected in a 'European Data Infrastructure', which should be an ecosystem able to include different components. Other initiatives contribute to this with for example: EOSC defining interoperability guidelines, CoreTrustSeal [73] defining certification requirements that reflect the core characteristics of 'trustworthiness' for data repositories (and recently adopted in the FAIRsFAIR project [93]), and the re3data.org project [74] providing a global registry of research data repositories.

2.5 Other initiatives

Besides the infrastructure provisioning initiatives, there are other types of pan-European initiative that can be considered on the e-infrastructure landscape. These engage with and support e-infrastructure users, and provide them with guidance, consultancy, hands-on support and policy development so they can choose the most appropriate e-infrastructure services and can apply those in an optimal way to their own use cases. PLAN-E, the platform of National eScience Centers in Europe; RDA, the Research Data Alliance (RDA); the OpenAIRE National Open Access Desks (NOADs); WISE, a global community that enhances best practice in information security for IT infrastructures for research; the cluster of HPC Centres of Excellence H2020 projects and the EOSC-hub Competence Centres can be mentioned here.

The principal goal of PLAN-E [7] is to act in support of *Enhancing Science.* The objective of PLAN-e is to bring together leading influential e-Science centres across Europe to help coordinate ongoing innovation in scientific methods and exploitation of infrastructure. The goals of PLAN-E cover all the topics that help promoting the eScience approach and strengthening the groups and centers conducting eScience.

The Research Data Alliance (RDA) [8] was launched as a community-driven initiative in 2013 by the European Commission, the United States Government's National Science Foundation (NSF) and National Institute of Standards and Technology (NIST), and the Australian Government's Department of Innovation. The goal of RDA is to build the social and technical infrastructures that enable open sharing and re-use of data. With more than 8,100 members from 137 countries (March 2019), RDA provides a neutral space where its members can come together through focused global Working and Interest Groups to develop and adopt infrastructure that promotes data-sharing and data-driven research, and accelerate the growth of a cohesive data community that integrates contributors across domain, research, national, geographical and generational boundaries. Working groups (WGs) about FAIR-ness are especially active recently, for example the FAIR Data Maturity Model WG [94] that develops as an RDA Recommendation a common set of core assessment criteria for FAIRness and a generic and expandable self-assessment model for measuring the maturity level of a dataset.

Open Science is a critical and collective endeavour to open up research and scientific workflows. All players are involved in making research outputs open, free, available, reusable and FAIR (Findable-

Accessible-Interoperable-Reusable). OpenAIRE supports these efforts via its services and its network of 34 National Open Access Desks (NOADs) [103], which comprises experts working on transferring and translating EU policies to a local level. NOADs and their organizations are the de-facto national nodes for Open Science in the majority of their countries, already highlighting the connection to EU policies, and ready to step up and become an integral part of national EOSC structures. An Open Science Helpdesk bringing coherence to the EOSC training and support landscape, and trains stakeholders to build local support networks for researchers and data practitioners.

WISE [96] is a global community of security experts who advance the security of IT infrastructures for research, and builds trust between such systems, i.e. all the various types of distributed computing, data, and network infrastructures in use today. Through membership of working groups and attendance at workshops these experts participate in the joint development of policy frameworks, guidelines, and templates. WISE so far produced a 'Security for Collaborating Infrastructures Trust Framework' and a 'Risk assessment template'.

HPC Centres of Excellence (CoEs) are funded by the H2020 programme to support the uptake of HPC systems in various scientific areas [75]. Nine CoEs are now running, supporting Biomolecular Research, Global Systems Science, Computational Biomedicine, simulation and modelling, Energy sciences, Weather and Climate simulation, Materials design, Novel materials discovery, Performance optimization. These CoEs strengthen Europe's existing leadership in HPC applications, develop and support tools that improve HPC applications' performance.

The EOSC-hub project includes a Work Package that is composed of 8 Competence Centres (CCs) [76]. Each CC fosters the use of the advanced digital capabilities and resources of EOSC-hub within early adopter research communities, supporting them in data- and computing-intensive science. CCs are driven by well-established and mature research infrastructures from ESFRI or by international scientific collaborations that require integrated data and computing services from multiple European e-infrastructure providers, typically from EGI, EUDAT and GÉANT. CCs operate independently from each other but share the interest and needs in using common solutions from the EOSC-hub service catalogue for the setup of community-specific services for EOSC.

3 The global e-infrastructure landscape

The global landscape of e-infrastructures was reviewed by EGI in collaboration with some of the key pan-European e-infrastructures that best know their peers worldwide (PRACE, GÉANT, OpenAIRE). The landscape was assembled by interviewing the European e-infrastructures about their understanding of the rest of the world, by collecting information about non-European e-infrastructures with online surveys, with teleconference interviews and through face-to-face meetings attended in various regions of the world. We complemented this data with harvested data from public websites and deliverables that are referenced in the report. We feel that, compared to other Research Infrastructure domains, the e-infrastructure landscape is rather well known by, and well connected to, the main players of the European landscape. Moreover the e-infrastructures have informative and fairly up-to-date public websites both in Europe and worldwide.

In the next subsection (3.1) Networks are presented in two separate subsections: one about infrastructures and the other about infrastructure development projects. This structure was chosen because in this field the infrastructures and the projects can be rather clearly separated.

In subsections 3.2, 3.3 and 3.4 HPC, 'Grids and Clouds' and Data infrastructures and projects are presented in another way: according to their geographic distribution. This approach was chosen because the distinction between infrastructures and projects is not so clear in these areas and the geographical location is felt to be a stronger distinguishing feature.

3.1 Networks

3.1.1 Infrastructures worldwide

Network provisioning to research and educational institutes happens similarly in both Europe and the rest of the world: by National Research and Education Network (NREN) that exist as specialised internet service providers dedicated to supporting the needs of the research and education communities within a country. NRENs are usually distinguished from commercial internet service providers by support for a high- speed backbone network, often offering dedicated channels for individual research projects, as well as by specialized 'value added' services on top of network connectivity, specifically for the research and educational sector.

Network provisioning to research and educational institutes happens similarly in both Europe and the rest of the world: by National Research and Education Networks (NRENs). Each NREN organisation reflects the specific environment in which it grew, with country-specific peculiarities such as the political situation, the history of the organisation and its relations with user groups, funding agencies, and the status of research and education in that country.

NRENs are a large and diverse family. At the time of writing 122 NRENs exist around the globe: 14 in East and Southern Africa[8], 5 in South Africa[9], 8 in West and Central Africa, 7 in the Indian subcontinent, 19 in Asia Pacific[10], 16 in the US[11], 11 in Canada, 17 in Latin America[12], 4 in the Caribbean, 12 in Middle Asia[13], 9 in Central Asia and 46 in Europe (including the Nordics). The NREN community keeps a list of NRENs up to date in a public Wikipedia page [13].

Each NREN organisation reflects the specific environment in which it grew, with country-specific peculiarities such as the political situation, the history of the organisation and its relations with user groups, funding agencies, and the status of research and education in that country all woven into its fabric. Another important aspect is the difference between the leading communities that formed the NRENs – each NREN was set up in a form that suited a country's needs and background. The development and support of an NREN infrastructure is often determined by the vision, resource and funding levels in a given country – and this differs between national authorities. Whether an NREN is or is not connecting a specific institution to the network is also dependent upon an acceptable use policy, which varies by NREN: some can connect primary and secondary schools, while others' mandates may extend also to private R&D firms.

An example infrastructure difference is outlined here: some NRENs have a hierarchical architecture that includes a national backbone, which interconnects a number of separately managed regional networks, which, in turn, connect end users (For example Internet2 in the US [83]). Such an NREN will have an indirect view of the organisations connected to its network. By comparison, there are networks without regional branches, which connect institutions directly and have a more direct view of an NREN's connected institutions (For example KENET in Kenya [84]). Such diversity, although puzzling at times, is an inherent characteristic that the NREN community strives to preserve, because the current shape in which NRENs have grown comes from their unique environments, which have fostered their scientific, educational and cultural communities.

The GÉANT Compendium [14] provides an authoritative reference source of information about the status and development of research and education networking in Europe and beyond. Published since 2001, the Compendium provides information on key areas such as users, services, traffic, budget and staffing. The online database can be updated by NRENs on a voluntary basis, can be viewed online, and each year a report is also generated summarising the results. The last Compendium that includes responses from outside Europe is the 2016 edition [15], providing details about Australia, Bosnia Herzegovina, New Zealand, Korea and Tajikistan. As stated in the compendium, the findings about these non-European NRENs do not bring surprise to European colleagues:

---

[8] These 14 NRENs also participate in the 'UbuntuNet Alliance'.
[9] These 5 NRENs also participate in the WACREN umbrella NREN.
[10] These 19 NRENs also participate in the APAN non-for-profit association.
[11] Although advocated since the 1980s, the U.S. does not have one single NREN.
[12] These 17 NRENs also participate in the RedCLARA international organization.
[13] Some of these NRENs, alongside North-African Arabic NRENs, also participate in the ASREN Arab region NREN umbrella.

- The size of each NREN varies greatly.
- NRENs connect universities, institutes of further education, research institutes, and libraries. A large majority are also authorized to connect primary and secondary schools.
- For-profit organisations, are, in general, not covered by the acceptable use policies of these countries, (with the exception of New Zealand).
- 1 Gbps is the typical connection speed in all the NRENs who responded.
- The spectrum of services the NRENs offer (besides connectivity) is growing.

The additional interviews we performed in RISCAPE with the UbuntuNet Alliance and with Internet2[14] did not add new findings to this picture.

GÉANT exists as a membership organisation for European NRENs to support them and through and with them other NRENs in the world. Similar membership organisations exist in other regions too:

- TEIN*CC in Seoul, supporting NRENs in Asia (currently 21)
- RedClara in Montevideo, supporting NRENs in Latin America (currently 13)
- UbuntuNet Alliance, supporting NRENs in East and South Africa (currently 16)
- WACREN, supporting NRENs in West and Central Africa (currently 13)
- ASREN, supporting NRENs in the Arabic countries of North Afrika and the Middle East (currently 22)

GÉANT helped all the formation of all these organisations and supported them in their early years of operation.

3.1.2 Infrastructure projects

GÉANT has been a trusted partner of the European Commission for many years, as the coordinator of network projects co-funded by the European Union and NREN organisations in Europe, and by those in other world regions. Projects with other regions are funded by the DG DEVCO (Directorate General International Cooperation and Development) and DG NEAR (Directorate General Neighbourhood and Enlargements Negotiations) and currently include:

- AfricaConnect2 [17]: AfricaConnect2 supported the development of high-capacity internet networks for research and education across Africa. It builds on existing networks in Eastern and Southern Africa and North Africa and will extend connectivity into West and Central Africa. With links to the GÉANT network AfricaConnect2 established an African gateway for global collaborations.
- CAREN [18]: Launched in 2010, CAREN interconnects researchers, academics and students at over 500 institutions in Kyrgyzstan, Tajikistan and Turkmenistan with Kazakhstan and Uzbekistan also candidate countries.
- EaPConnect [19]: Eastern Partnership Connect (EaPConnect) project sets out to create a regional high-speed internet network dedicated to research and education across Armenia,

[14] UbuntuNet Alliance (regional association of NRENs in Africa); Internet2 (largest NREN in the USA)

Azerbaijan, Belarus, Georgia, Moldova and Ukraine. The network will interconnect the NRENs in these six countries and integrate them with the pan-European GÉANT network.

- EUMEDCONNECT3 [20]: EUMEDCONNECT3 provides a high-capacity dedicated Internet network for the research and education communities across the southern and eastern Mediterranean region.
- TransAfrican Network Development [105]: TANDEM supported dialogue between the EU and African research and education networks, with special attention to the Western and Central Africa region. It is a successor to the WACREN project.
- MAGIC [106]: Middleware for collaborative Applications and Global virtual Communities aims to establish a set of agreements for Europe, Latin America and other participating world regions (North Africa and the Middle East, West and Central Africa, Eastern and Southern Africa, Central Asia and Asia-Pacific) to create a marketplace of services and real-time applications for international and inter-continental research communities. GÉANT is one of 21 project partners from across the globe in this successor to the ELCIRA project, led by RedCLARA.

Besides these focused projects, several European NRENs meet on a regular basis with Chief Executives of other NRENs in the 'International NREN CEO Forum'. This forum sponsors a series of global initiatives (such as the Global Network Infrastructure [16]), all aimed at developing an interoperable global infrastructure and services tailored for research and education.

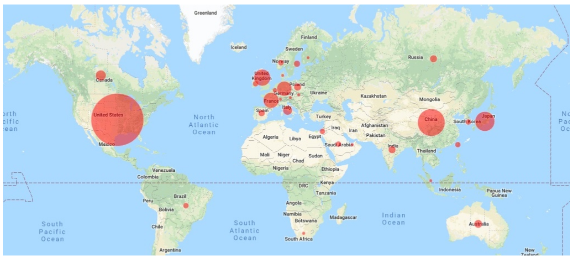
3.2 High Performance Computing

3.2.1 Global overview

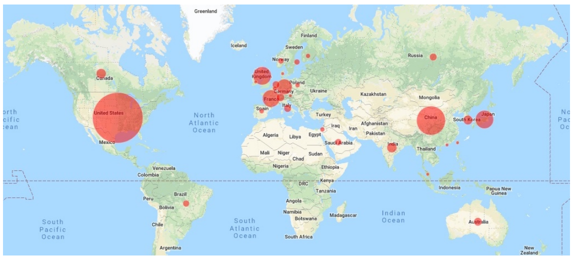
The 'Top500 list' [9] is a globally used measure to assess the status and to compare HPC systems and countries hosting those. This section first reviews the global HPC landscape using the Top500 list and the analysis that PRACE made [80] at the end of 2018 based on the Nov 2018 list of Top500 machines. (The June 2019 version of the Top500 list was announced at ISC 19 but is not analysed here – for the purposes of this report there have not been any significant changes since the November 2018 list, although the rankings of individual systems may have fallen slightly.)

Throughout the years China's dominance in HPC significantly increased, while the USA's dominance decreased, and Japan's and Europe's have remained constant. Despite China's dominance in number of Top500 machines, it is the USA which dominates in terms of total computational capacity. Europe requires unified effort, such as the EuroHPC initiative, to be a considerable player in the race towards the exascale.

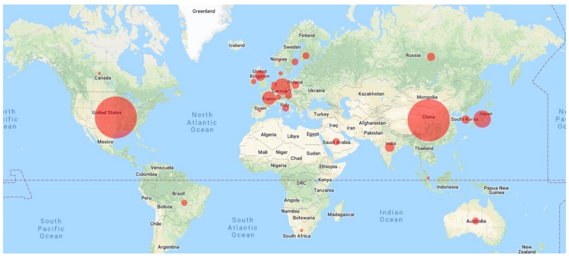
The following figures present the evolution of the HPC global landscape between 2012 and 2018, presenting for each year the number of Top500 systems in each global region.



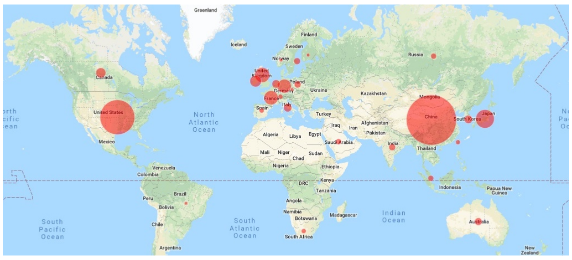
2012 Global HPC Landscape



2014 Global HPC Landscape



2016 Global HPC Landscape



2018 Global HPC Landscape

Figure 1: Mapping the evolution of the global HPC Landscape, from 2012 to 2018, based on Top 500 list. The size of each circle is proportional to the number of Top 500 HPC systems in a country.

From the above figures it is clear to see that the major HPC players are China, Europe, Japan and the USA. Throughout the years China's dominance significantly increased, while the USA's dominance decreased, and Japan's and Europe's have remained constant. Despite China's dominance in numbers, it is the USA which dominates in terms of total computational capacity – as seen in the following map which identifies the location of the top 25 HPC systems (based on the November 2018 Top 500 list).

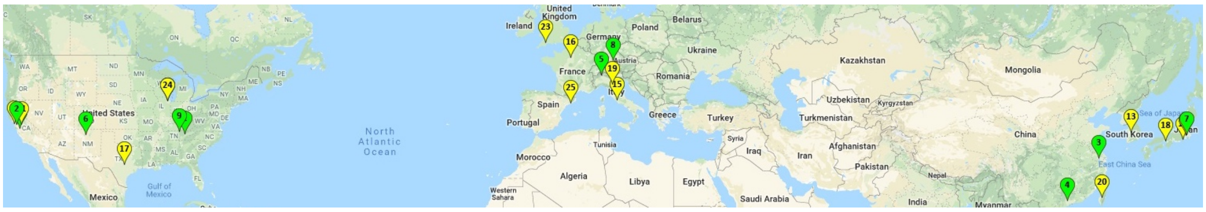


Figure 2: Top 25 HPC System, based on Top 500 list of November 2018.

The growing demand for High Performance Computing (HPC) resources in scientific computing has triggered a number of initiatives globally. The National Strategic Computing Initiative (NSCI) [10] and the Department of Energy's related Exascale Computing Project [11] in the USA, the Japanese FLAGSHIP 2020 project [12], the Tianhe-3 and Sunway exascale projects and quantum computing initiatives in China, the PRACE, ETP4HPC, and EuroHPC projects in Europe are examples of such large-scale efforts.

The highlight of the latest Top500 list (November 2018) is the presence of five US Department of Energy machines in the top 10, including the top two, Summit and Sierra, respectively (see the table below). China had already lost the number one spot in the June 2018 listing, and now losing the second slot too, is occupying number three and four with its Sunway TaihuLight and Tianhe-2A machines. The rest of the top 10 machines consists of three US based, one Japanese, and two European systems, including the #5 Piz Daint (Switzerland) and the upgraded #8 SuperMUC-NG (Germany), the latter re-entering the top 10 after falling from it in June 2014. Both Piz Daint and SuperMUC-NG are part of the PRACE Tier-0 ecosystem.

From the numbers above, it is clear that China and the US have achieved a significant lead in the amount of HPC resources available for scientific research, and this is not an issue that can be tackled by other countries individually in the short run. This suggests that a unified effort by the entire European community, such as the EuroHPC initiative, is necessary if it is to be a considerable player in the race towards the exascale.

For decades, the notion of "performance" has been synonymous with "speed" (as measured in FLOPS, short for floating-point operations per second). This particular focus has led to the emergence of supercomputers that consume egregious amounts of electrical power and produce so much heat that extravagant cooling facilities must be constructed to ensure proper operation. In order to raise awareness to other performance metrics of interest (e.g., performance per watt and energy efficiency for improved reliability), the Green500 list was established in 2006 [101]. In 2016 the Green500 project got integrated with TOP500 and now, while the two lists are separate, they share a joint submission process and publish under the TOP500.org domain. Since Nov 2016 the Green500 system indicates also the ranking of their systems on the Top500 list – allowing an easy comparison of energy efficiency and computational power. By looking at the top 10 systems on the Green500 systems we can see that the greenest machines are getting a higher score also on the Top500 list[15], indicating that the most recent systems (and therefore higher placed in the Top500 list) are on average becoming more and more energy efficient.

3.2.2 United States

The National Science Foundation (NSF) provides millions of hours of computing time to researchers in the US through the NSF-funded XSEDE program [21]. In addition to having extensive outreach programs towards scientists at all levels, the XSEDE program also allocates substantial amount of free of charge computational resources to researchers. The effective review model of XSEDE evaluates the scientific merit of the projects only if the researchers do not already have any grant from independent funding agencies. NSF has also funded a new system (Frontera at TACC) which will succeed Blue Waters as the NSF’s largest HPC system. In general NSF manages funding for specific HPC systems on a competitive project basis, awarding operators with funding based on the quality of new proposals and their historical performance record.

US have been also investing in the next generation of supercomputers, which are basically quantum computers. In this area, USA might stay behind China for a while, however the recent investment of

[15] The average place of the top 10 systems of the Green500 on the Top500 list: 213 in June 2018; 170 in November 2019; 136 in June 2019.

about 1.2 billion dollars on the national quantum initiative might reduce the gap in quantum technology between these two countries.

3.2.3 China

China has a wide variety of HPC investment programs since 2002. Early supercomputers within the network of the China National Grid (CNGird) have been replaced by the world's most powerful supercomputers since 2010. CNGrid is supported by 17 national HPC centers each of which has a system within the Top500. China is supporting more than 20 R&D projects towards exascale computing and is considered to be the world leader in quantum computing research with its 10 billion dollars investment plan in the field.

3.2.4 Latin America

While there is no machine on the Top500 list from Latin America, the High-Performance Computing Latin America Community (HPCLatAm) is a growing platform that brings together HPC actors such as researchers, developers, HPC users to discuss new ideas, experiences, and problems. Their previous conference, CARLA (Latin America High Performance Computing Conference) was in Costa Rica in September 2019 [23].

3.2.5 Middle East, Africa, Asia Pacific, Australia

Saudi Arabia represents the Middle East with one machine around #85. Two machines represent Africa, both hosted in South Africa and around #350-400 on the list. Having 5 supercomputers for HPC research and industry, Australia provides computational power to its researchers through the National Computational Infrastructure (NCI) and Pawsey Supercomputing Centre. Singapore's National Super Computing Center (NSCC) was established in 2015 and provides HPC resources for academic and industrial needs in the field of science and engineering. It supports a 1 petaflop system, a 10 Petabyte data service coupled with dark fibre network to support the Singapore Advanced Research and Education Network.

The European Commission Directorate General for Research and Innovation (DG RTD) recently established the EU-ASEAN[16] High Performing Computing (HPC) Coordination Group with the aim to support the establishment of the ASEAN-EU Research and Innovation Policy Exchange Platform. The Coordination Group brings together EU Member States' HPC policy experts and ASEAN HPC officials from policy and technical levels, enabling them to exchange HPC strategies and plans on regional aspects.

[16] The Association of Southeast Asian Nations (ASEAN) was established on 8 August 1967. The Member States of the Association are Brunei Darussalam, Cambodia, Indonesia, Lao PDR, Malaysia, Myanmar, Philippines, Singapore, Thailand and Viet Nam. On 31 December 2015, the ASEAN Community was formally established. The ASEAN Secretariat is based in Jakarta.

3.2.6 Russia

Even though several Russian systems are listed in the Top500 list each year, their standing in the list has been falling since 2011. Like PRACE, The Supercomputing Consortium of Russian Universities [24] provides free of charge CPU cycles to researchers in the field of HPC. PRACE has initiated an action to setup a link to this only Russian organization on supercomputer research and technologies.

3.2.7 Japan

Japan has recently started to develop the most powerful supercomputer in the field of computational astronomy. According to the National Astronomical Observatory of Japan's (NAOJ) announcement, the supercomputer with the nickname "NS-05 ATERUI II", contains 2,010 Intel Xeon 6148 20-core (2.4 GHz) processors and provides 3.087 peak petaflops for astronomy related research. Japan is targeting to commission its first exascale machine, Post-K, by early 2022. The Japanese government has so far contributed 110 billion yen (~91 billion USD) toward the Post-K computer, including research, development and acquisition, and application development.

PRACE collaborates with XSEDE, Compute Canada and Japan (with RIKEN Center for Computational Science) on experience sharing and on the organisation of International HPC summer schools, with the previous event held in Kobe, Japan in July 2019 [22].

3.2.8 Canada

Compute Canada is a partnership of over 30 universities as well as the regional organizations ACENET, Calcul Québec, Compute Ontario, and WestGrid, providing advanced research computing systems, storage and software solutions to all Canadian researchers. Through Compute Canada, researchers can access four large HPC systems (Niagara, Beluga, Graham and Cedar) installed across the country. The Canadian government has committed Cdn $572 million to digital research infrastructure, of which an estimated Cdn $360 million will be managed by a new advanced research computing organization being established in 2019, with the intent of ensuring continued computing access for Canadian researchers – either through additional investment at existing systems or the launch of new systems in the future. SciNet, Canada's largest supercomputer centre will host the next International HPC Summer School in Toronto in July 2020.

Until around 2012 compute grids dominated the distributed computing landscape. In the past few years cloud facilities started to dominate over grid clusters and new scientific projects plan their compute infrastructures pre-dominantly on clouds unless their experiment requires grand HPC systems or a highly

3.3 Grids and clouds

distributed infrastructure for batch HTC computing.

In this document we use the 'grids and clouds' category to refer to a number of different types of compute infrastructures: high throughput compute infrastructures implemented in the form of institutional clusters and compute grids, cloud compute infrastructures from both academia and industry and covering the full stack from Infrastructure as a Service, Platform as a Service, Software as a Service.

The 'Grids and clouds' landscape is changing more rapidly than the previously discussed network and HPC areas, and most of the infrastructures in this area still do not have sustainable business models for operations, they often fund operation from projects. Therefore, we do not separate the discussion within this section into subsections about infrastructures and about projects, we rather structure the content based on geographical regions.

Until around 2012 compute grids dominated the distributed computing landscape. Grid is a paradigm of accessing remote facilities, most often compute clusters, owned by multiple, independent institutes. The resources are federated into wide area pools and can be accessed by scientific groups in a controlled fashion through a 'grid middleware'. While the grid paradigm can be used for many types of resources, it is typically adopted for compute clusters that are pooled together with the grid middleware for wide area 'high throughput computing' (HTC). In HTC very large number of computational tasks are run in non-interactive mode across many computers at multiple institutes.

In the past few years cloud sites started to dominate over grid clusters and new scientific projects plan their compute infrastructures on clouds (except those that require grand HPC systems or a highly distributed infrastructure for batch HTC computing). Clouds put more emphasis on the 'utility-type' accesses to computers and on software allowing metered, interactive and fluctuating use of a system by diverse and independent groups of users. Clouds can be built from open source software produced and maintained by professional software developers (while grid middleware are/were maintained by academic institutes and staff). Besides also being suitable for HTC, clouds are suitable for hosting scalable data analytics platforms, portals, scientific gateways and virtual research environments. Moreover, grid compute clusters are nowadays often hosted inside Infrastructure-as-a-Service clouds, and/or are often complemented by cloud compute services to absorb peak CPU demands from the grid sites. This trend is expected to continue, and even the largest experiments plan to diversify their compute setup for the next phases of their operation, giving more visible role to clouds and HPC [25].

Existing, international experiments that invested in into grid-based operation model the past – such as the Large Hadron Collider, the VIRGO-LIGO interferometers, WeNMR structural biology, or the Cerenkov Telescope Array – continue to depend on grid compute infrastructures for their production data analysis. New experiments will mostly plan and build their compute setup on clouds.

A global alliance, called Open Research Cloud Alliance (ORCA) [91] was initiated in 2017 to form a global community what would establish and promote research technology standards that foster interoperability between and among scientific research clouds. ORCA organizes 2-3 meeting each year, and started drafting a Declaration [92] which outlines the issues research communities currently have to face in the cloud computing and data management landscape and the tasks that should be accomplished to eliminate these barriers through policies and technical services, tools and protocols. ORCA was initiated from the US but by now attracted members from all over the world, including EGI, GÉANT and other stakeholders from Europe.

3.3.1 United States

In the US the Open Science Grid (OSG) [26] is the largest compute grid e-infrastructure. The Open Science Grid consists of computing and storage elements at over 100 individual sites spanning the United States. These sites, primarily at universities and national labs, range in size from a few hundred to tens of thousands of CPU cores. The setup is very similar to the High Throughput Compute service of EGI in Europe, and OSG and EGI actually use common operational tools and practices and serve common user communities (primarily in High Energy Physics).

The US National Science Foundation (NSF) runs various types of projects to make cloud compute services more prominent on the US cyberinfrastructure landscape. We attempt to highlight the initiatives with the highest relevance to Europe:

- Jetstream [27], led by the Indiana University Pervasive Technology Institute (PTI), expands the XSEDE cyberinfrastructure with a Level-1 site using OpenStack cloud technology. The computing environment consists of two homogenous clusters at Indiana University and TACC, with a test environment at the University of Arizona. The system provides over a half petaflops of computational capacity and 2 petabytes of block and object storage. Access management is integrated with the rest of XSEDE.
- "Exploring Clouds for Acceleration of Science (E-CAS)" is a cooperative agreement between NSF and the Internet2 NREN to build partnerships with commercial cloud computing providers (Amazon Web Services and Google Cloud) and support 6+2 science applications [28] in new and more effective uses of cloud computing capabilities. E-CAS will explore how scientific workflows can leverage advancements in real-time analytics, artificial intelligence, machine learning, accelerated processing hardware, automation in deployment and scaling, and management of serverless applications for a wider range of science. Six applications have been recently selected, the two final projects will be selected in July 2020 for another year of support. Each phase of the project will be followed by a community-led workshop to assess lessons learned and to define leading practices. The approach is similar to the one the EOSCpilot took in Europe H2020, but while in Europe the cloud services came from academic providers, the E-CAS project mobilises services from commercial providers.

- CyVerse [79] provides life scientists with powerful computational infrastructure to handle huge datasets and complex analyses, thus enabling data-driven discovery. It provides data storage, bioinformatics tools, image analyses, cloud services, APIs, and more.

The Internet2 NREN of the US runs the NET+ Cloud Services Program [29] to support the adoption of cloud services within the academic and scientific sectors. The program is similar to the Cloud Framework of GÉANT in Europe and helps NREN members and other qualified institutions access cloud services through a variety of ways, including:

- Leverages the capacity of the Internet2 Network connection for enhanced performance and delivery of cloud services.
- Expands adoption of InCommon Federation identity and access management among cloud services.
- Facilitates deployment and ease of integration of commercial cloud services uniquely configured for research and education needs, legal terms negotiated via the power of a large community, and frequently consistent and predictable pricing.
- Peer evaluation of key service components includes: functional; technical; security; accessibility; business and legal; and deployment.
- Facilitates peer-to-peer and community knowledge sharing of best practices, tools and capabilities.
- Influences industry to develop services meeting the particular needs of the research and education (R&E) community.
- Encourages a strategic relationship between the community and service providers, and provides a basis for long-term collaboration on R&D and product roadmaps

3.3.2 Asia

A regional grid infrastructure was established in Asia Pacific in the early 2000s to serve WLCG, the Worldwide Large Hadron Collider Computing Grid. The infrastructure is still in use today and consists of a WLCG Tier-1 centre in Taipei (at Academia Sinica) and several, smaller Tier-2 sites in the region. During the late 2000s and early 2010s the infrastructure gradually opened up to other science disciplines and was complemented with additional sites, thanks to various EC co-funded projects, such 'Enabling Grid for E-sciencE' (EGEE) [30] and EUAsiaGrid in the 7th Framework Programme. This work resulted in various grid-based 'Virtual Organisations'[17] (VOs) to offer batch compute and file storage infrastructures for projects and researchers in the region. While some of these VOs are still operational, their scope narrowed, and their relevance declined in the past years.

There seems to be a more nationalistic approach to compute infrastructures since the EU co-funded projects have ended. Countries upgrade, replace or shut down their sites with pace, scope and scale depending on national e-infrastructure funding and policies. Similarly to the rest of the world, cloud

[17] For example the twgrid, euasia.euasiagrid.org and vo.france-asia.org Virtual Organisations

computing is becoming the dominant or desired type of infrastructure besides HPC in Asia Pacific [35], with the following cloud infrastructures already in place:

- China: The Chinese grid infrastructure was deployed in alignment with, and based on European grid infrastructures in the mid-2000s, and resulted in the CNGrid infrastructure. Since the EUChinaGrid FP7 project finished in 2007 there was basically no exchange of e-infrastructure technology between Europe and China. While WLCG Tier-2 sites remained in operation, CNGrid evolved into an HPC infrastructure. We could not find information about institutional clouds, we suspect that some institutes may operate clouds for research and education. The AlibabaCloud became one of the most significant global players, but its role in research use is not significant (yet) [36]. A recent cloud federation, CSTCloud [102] is expected to bring an academia-led cloud infrastructure for the Chinese education, research, scientific and technical communities, government departments and hi-tech enterprises. The project is under the leadership of the Chinese Academy of Science and they recently started to exchange information with EGI on practices and tools for federating cloud sites.
- India: After the FP7-funded EU-IndiaGrid project [40] has finished in 2008, the Indian grid infrastructure evolved independently from European grid efforts. The Indian Grid, Garuda, is today India's national grid infrastructure of HPC systems, connecting 70 academic and research institutions across 17 cities. Garuda is coordinated by C-DAC, the Centre for Development of Advanced Computing, provider of other e-infrastructures [41]. C-DAC has developed a complete open source based cloud software stack named 'MEGHDOOT' for setting up a private cloud to offer basic cloud services such as Infrastructure, Platform, and Software services. We found no information about the scale at which MEGHDOOT is deployed and used.
- Indonesia: The National Institute of Science (LIPI=Lembaga Ilmu Pengetahuan Indonesia) operates a grid cluster in WLCG, but also service other applications from weather forecast and molecule modeling. Insitut Teknologi Bandung runs two GPU clusters to support firewatch and other environmental science applications [33].
- Japan: National Institute of Informatics (NII) hosts a Center for Cloud Research and Development (CCRD) as well as a Gakunin Cloud Adoption Support Service. The service provides a cloud testbed based on AWS and Google to run Proof of Concepts, and helps projects and institutes select and procure commercial cloud services for production runs [32].
- Pakistan: the National Centre for Physics (NCP) supports Physics and related applied disciplines in the country. One of the support activities is the operation of e-infrastructures for researchers in the field, and in this context the institute operates an OpenStack based private cloud. The cloud is hosting the national WLCG Tier-2 site and an HPC cluster for other communities.
- Philippines: the Advanced Science and Technology Institute provides NREN, HPC, science cloud infrastructures, as well as operates satellite ground stations to serve national research. The science cloud delivers cloud-based services to researchers and students, and

enables private sharing of data among specific groups. It provisions virtual machines and support projects and researchers in the areas of Disaster Management, Rice Genomics, Health R&D, etc. [34]

- Taiwan: Academia Sinica operates a distributed cloud environment for national researchers, based on OpenStackOpen. The system offers IaaS as well as higher level capabilities through a home-grown front-end, called DiCOS [31].

3.3.3 Australia and New Zealand

The Australian research cloud, called Nectar [37], provides flexible scalable computing power to all Australian researchers, with computing infrastructure, software and services that allow the research community to store, access and run data, remotely, rapidly and autonomously. Nectar was established in 2009 and it is implemented as a federation of OpenStack sites. The architecture is similar to the EGI Cloud in Europe, the two infrastructures evolved in parallel and there are regular meetings between the two teams. Nectar also hosts online virtual research environments, called 'Nectar Virtual Labs' for various disciplines. In 2017 Nectar, the Australian National Data Service (ANDS) and the Research Data Services joined forces under the name 'Australian Research Data Commons' (ARDC) to offer "a more integrated, coherent and reliable system to deal with the various needs of data-intensive, cross-disciplinary and global collaborative research" [38]. Australia (and South Africa) will host one of the Square Kilometre Array astrophysics observatories and this instrument is expected to boost the national e-infrastructure landscape.

In New Zealand the New Zealand eScience Infrastructure (NeSI) [107]enables researchers across a wide range of communities and disciplines to tackle large or highly complicated problems and to investigate scientific challenges that were previously impossible. While a few years ago NeSI operated one site in Nectar, these days its service portfolio consists of (1) HPC and analytics; (2) Consultancy; (3) Data transfer and share and (4) Training services.

3.3.4 Africa

During the mid-2000s' FP7 projects, EUMedGrid and EUMedGrid-Support facilitated the deployment and operation of grid infrastructures and science gateways in the North African region. Since the end of these projects the infrastructures decayed, and we could not locate national/regional grid or cloud e-infrastructures on the African continent. South Africa seems to be the most organised country in the e-Science support perspective, with its National Integrated Cyber Infrastructure System (NICIS) [39]. NICIS promotes scientific and industrial development through the provision of high-performance computing capability, high-speed network capacity and a national research data infrastructure integrated hierarchically into globally connected systems and into local systems, providing seamless access for the research and education communities of South Africa. It is a national initiative of the Department of Science and Technology and implemented by the CSIR. South Africa (and Australia) will host one of the Square Kilometre Array astrophysics observatories and this instrument is expected to boost the national e-infrastructure landscape.

3.3.5 Rest of the world

Brazil

Cooperation between Brazil and European Union started in early 90s and after few partnerships during the following years, coordinated calls started between Brazil and EU in information and communications technology (ICT). So far there have been 4 coordinated calls (in 2010, 2012, 2015, and 2017) [59]. 5, 4, 4, 6 IT R&D projects have been funded in these calls respectively. The last round of projects are finishing in 2019 [60], with 2 of them relating to cloud application development (but not cloud infrastructure development):

- ATMOSPHERE: ATMOSPHERE aims to design and implement a framework and platform relying on lightweight virtualization, hybrid resources and Europe and Brazil federated infrastructures to develop, build, deploy, measure and evolve trustworthy, cloud-enabled applications.
- NECOS: The NECOS project addresses the limitations of current cloud computing infrastructures to respond to the demand of new services, as presented in two use-cases, that will drive the whole execution of the project.

Russia

Besides providing a Tier-1 and Tier-2 sites in WLCG, we could not find information about other grid or cloud e-infrastructures in Russia.

3.4 Data

While some of the Research Infrastructures built their own services for the management of digital data, others rely on generic e-infrastructure services and expand those with science discipline-specific tools, data and support to achieve a customised setup for data depositing, short/mid-term storage, archival, processing, analysis and data dissemination. This work is supported by national, regional and continental e-infrastructures, and RDA and CODATA as global community platforms.

Data infrastructures could be categorized as supportive e-infrastructures or stand-alone research infrastructures (RIs). While shared data management capabilities are offered by e-infrastructures such as EGI, EUDAT, OpenAIRE in Europe, most RIs have discipline-specific data production, curation, preservation and sharing as part of their core institutional mission. Some RIs built their own services for such purposes, others rely on generic e-infrastructure services and expand those with science discipline-specific tools, data and support to achieve a complete setup for data depositing, short/mid-term storage, archival, processing, analysis and data dissemination. In this section we review the landscape of generic data management services that are offered by e-infrastructures and related projects. These services are independent from a given scientific discipline and offer capabilities for any fields of science. Similarly to the previous section we do not cover infrastructures and projects in separate subsections, because data services are still often sustained from project funding. We present them by geographical regions instead. But before going into regional infrastructures and project, let us present two global initiatives.

The Research Data Alliance (RDA) was already introduced in Section 2.5. RDA is not a technical infrastructure. It serves as a global platform for scientific communities and e-infrastructure communities to capture and share good practices and standards for data management, sharing and analysis, and to facilitate the uptake of those good practices within different disciplinary areas. The work of RDA is funded by [42] the European Commission, as well as the Australian Government Department of Education and Training (AU), the National Institute of Standards and Technology (US), the National Science Foundation (US). Other funders are (typically through hosting RDA plenaries) the Japan Science and Technology Agency (JP), Research Data Canada (CA), University of Montreal (CA), Alfred P. Sloan Foundation (US), JISC (UK), MacArthur Foundation (US) and the Wellcome Trust (UK). RDA is a truly international platform – both in respect to its members, and its funders.

A similar global effort is CODATA [57], the Committee on Data of the International Science Council (ISC). CODATA exists to promote global collaboration to advance Open Science and to improve the availability and usability of data for all areas of research. CODATA works also to advance the interoperability and the usability of such data: research data should be FAIR (Findable, Accessible, Interoperable and Reusable). By promoting the policy, technological and cultural changes that are essential to promote Open Science, CODATA helps advance ISC's vision and mission of advancing science as a global public good. Similarly to RDA, CODATA also runs Task Groups and Working Groups but also supports the Data Science Journal and collaborates on major data conferences like SciDataCon and International Data Week.

3.4.1 United states

The US nurtures a diverse and growing 'data services' landscape. Let us begin the review with data repositories. Systems, such as Dryad [43], figshare [44] Harvard Dataverse [45], Open Science Framework [46], Mendeley Data [47] are based in the US and offer data repository for researchers from any discipline to store and to share data. Although based in the US, these services attract an international user base, including researchers from Europe, thanks to their recognition by the largest publishers [48][49] who require authors to deposit data in disciplinary repositories, or in one of these generic ones. These repositories basically work with one of the following two business models, or a combination of those:

- The repository is free for the users, but upload is limited (e.g. Open Science Framework or Zenodo in Europe). In this case the cost of operation is covered by funders (e.g. NSF, H2020) or from sponsors and donations.
- The repository charges the users (e.g. Dryad). In this case the costs are recovered from the usage fees.

Some of the repositories combine the two models and offer free services for users up to certain capacity limits, then introduce usage fees.

Another, more complex generic data service in the US is the National Data Service [50] (NDS). NDS is an emerging vision for how scientists and researchers across all disciplines can find, reuse, and publish data. It builds on the data archiving and sharing efforts already underway within specific communities and links them together with a common set of tools designed around the following capabilities:

- Search: The NDS will allow users to easily search for data across disciplinary boundaries. As users home in on data of interest, they can easily switch to discipline-specific tools.
- Publish: The NDS will connect users to tools for building and sharing collections of data. It will help users find and deliver data to the best repository for data-publishing.
- Link: The NDS will create robust connections between data and published articles. When researchers reference an article, they have ready access to the underlying data.
- Reuse: The NDS will not only provide access to data for download, it will provide tools for transferring data to processing platforms or allow analysis to be attached to the data.

NDS shows remarkable similarities with the EOSC, with more focus on data and less on services and tools for data analysis. With this scope it is actually close to the ESFRI's understanding of EOSC (See that described in Section 2.4).

Another group of services we reviewed could be called as 'research publishing services'. This group includes diverse and innovative services to facilitate Open Science by enabling the sharing and discovery of open access research content (primarily papers and data).

ScienceOpen [77] is a discovery platform with interactive features for scholars to enhance their research in the open, make an impact, and receive credit for it. ScienceOpen offers services for three distinct user groups:

- Publishers: ScienceOpen offers content hosting, context building and marketing services.
- Institutions: ScienceOpen offers solutions and services to promote and share work, to build up your own branding for Open Access publications, to develop new open access publishing paradigms, to create an independent Open Access publishing environment.
- Researchers: ScienceOpen offers search and discover of relevant research in over 56 million Open Access articles, sharing of expertise and receiving credits by publicly reviewing any article, promoting research and tracking readership with article- and author-level metrics, creation of topical collections

Unpaywall [78] is an open database of more than 23 million free scholarly articles that the site harvested from over 50,000 publishers and repositories. The service comes with a Chrome browser extension that indicates during browsing that the user is reading an article record for which Unpaywall has the full text available.

'Dataset Search' [51], the service launched by Google in September 2018. It is, as the name clearly explains, a search tool to find datasets. Similar to how Google Scholar works, Dataset Search lets users find datasets wherever they are hosted – a publisher's site, a digital library, or an author's personal web page. Website owners shall enrich their site with metadata based on schema.org [52] for Dataset Search. These metadata provide salient information about datasets: who created the dataset, when it was published, how the data was collected, what the terms are for using the data, etc. So opposite to the 'push model' that data repositories typically use for gathering metadata from researchers' about datasets, Google Dataset Search applies a pull model reusing Google's web crawler infrastructure.

3.4.2 Canada

Portage, Compute Canada (CC) and the Canadian Association of Research Libraries (CARL) are collaborating to provide a scalable federated platform for digital research data management (RDM) and discovery. They are pleased to announce that the Federated Research Data Repository (FRDR) service has finished Beta and is now in Limited Production [82].

3.4.3 Latin America

In Latin America we could find only the LA Referencia repository service [53]. La Rederencia gives visibility to the scientific production of higher education and research institutions in Latin America, promoting open and free access to full text, with special emphasis on publicly financed results. LA Referencia is a federated access layer to the open access repositories of 10 Latin American countries. LA Referencia has a strong partnership with the OpenAIRE repository services of Europe and facilitates the use of the Zenodo repository for Latin American researchers and institutions [97]. LA Referencia stores scientific papers, articles, reports, doctoral and master theses, over 1.6m items in total. LA Referencia was born as a project coordinated by the Latin American Cooperation of Advanced Networks (RedCLARA) and financed by the Public Goods Fund of the Inter-American Development Bank (IADB) between 2010 and 2013. At the end of the project, in 2013, the network is articulated as a service between the Science & Technology bodies of the region with the support of RedCLARA (local organization for the NRENs, see Section 3.1.1), with the objective of consolidating an open access strategy for access to scientific publications, based on national repositories policies and the generation of public goods.

3.4.4 Australia

The key data initiative in Australia is the Australian National Data Service (ANDS) [54], now merged with RDS [90] and Nectar [37] in the Australian Research Data Commons (ARDC). The ANDS is a partnership led by Monash University, working in collaboration with the Australian National University and the Commonwealth Scientific and Industrial Research Organisation. ANDS' core purpose is to make Australia's research data assets more valuable for researchers, research institutions and the nation. Since being formally established in 2008, ANDS has supported numerous research data projects around Australia. It has also been playing an important role in the international research data community. ANDS' flagship service is the Research Data Australia discovery portal [55] where one can find, access and reuse data for research from Australian research organisations, government agencies and cultural institutions. ANDS does not store the data itself here but displays descriptions of, and links to, the data held by their data publishing partners or contributors. In a simplistic sense ANDS is the Australian version of Google Dataset Search, using the RIF-CS schema [56].

NCI Australia is the nation's most highly integrated high-performance research computing environment. NCI operates the National Research Data Collection, Australia's largest collection of research data, encompassing more than 10 PB of nationally and internationally significant datasets [81].

3.4.5 Africa

In Africa the most relevant generic data project we could find is the African Open Science Platform initiative (AOSP) [58]. AOSP is funded by the South African Department of Science and Technology (DST) through the National Research Foundation, and implemented and managed by the Academy of Science of South Africa. AOSP is a pan-African project for Africa by Africa, with direction provided by CODATA (See Section 3.4 above). The 3-year project was launched in December 2016 and ended in October 2019, possibly with a second phase starting in 2020. Until now AOSP facilitated the exchange of good practices, tools, approaches for Open Science by organising schools and other events, but sharing information through the Web and other channels.

3.4.6 China

China mandated data availability in national data centers in 2018 when the Chinese government has decreed that all scientific data generated in China must be submitted to government-sanctioned data centers before appearing in publications [98]. At the same time, the regulations, posted in March 2018, call for open access and data sharing. The Chinese government supports the implementation of both policies by developing 20 national data centres, covering all types of research data [99]. These national data centres are planned to feed into an overarching cloud infrastructure called CSTCloud, similar to the European Commission's vision for the European Open Science Cloud (EOSC). Whilst development of the CSTCloud and the 20 national data centres is ongoing, there are many Chinese repositories with a more focused scope making good progress in key aspects of open data. A notable example is the Fudan University Social Science Data Repository [100].

3.4.7 Rest of the world

We could not find other generic data e-infrastructure or project in other regions of the world.

4 Lessons learnt and recommendations

4.1 Generic

1. We were pleased to see that the e-infrastructure landscape is quite well connected between the EU and the rest of the world. GÉANT, PRACE, EGI, OpenAIRE have active collaborations worldwide.
2. E-infrastructure facilities/capabilities are well developed and organized in the US, Canada, Australia and Japan, but how they relate to one another is often in flux:
    a. US: XSEDE has been renewed but has a more limited role as "glue" to integrate the separate HPC systems. Leadership HPC investments are now with the Department of Energy.
    b. Canada: quite a bit of change expected in the organizational structure with a new advanced research computing organization being established in 2019.
    c. Australia: NCI, Pawsey and ARDC are reformulating the e-infrastructure space and have a big potential to deliver overarching services that cover the whole spectrum of scientific computing.
    d. Japan: continued strong investment with a diverse ecosystem but in more isolation from Europe than the other developed regions.
    e. Other ASEAN: recognition of the importance of e-infrastructure in most other ASEAN countries, continued investment and development of both networks and facilities, although less so in data-infrastructures.
3. From the high-middle income countries[18] China and Russia were found hard to assess and we do not believe we could fully review their landscape. The main reasons are the language, the size of those countries and that the EU did not have joint e-infrastructure activities for some years with them. Joint EU-regional initiatives (such as calls for projects like in Brazil) could facilitate regional activities and/or more intense international exchange of information.
4. Africa, India and the Middle-East seem to be lagging behind in e-infrastructure availability, compared to the rest of the world.
5. The e-infrastructure websites are quite content rich, and we managed to obtain or double-check most of the data for this report from there. Attending regional e-infrastructure

[18] Using the World Bank classification here: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

conferences were also a big help to identify the key players of the field and to engage with them face-to-face.

6. When we used interviews we did not receive significantly more data and knowledge than with the above-mentioned methods. Responses were often pointers to specific webpage sections or public documents.

4.2 Specific to Networks

7. The NREN concept is adopted worldwide, and collaborations between GÉANT and NRENs of other regions are in place.
8. The EC runs initiatives to develop NRENs in regions where NRENs do not exist yet. GÉANT with its expertise is recognized and used in those projects.

4.3 Specific to HPC

9. The topic is competitive in nature. Europe recognized that it can remain competitive with the US and China only if national fragmentation is eliminated from HPC development. The EuroHPC initiative therefore started, and it is on its way to help Europe enter the 'Exascale club'.
10. PRACE participates in joint initiatives with US, Japan and the ASEAN countries. The EU-ASEAN HPC task force could be a model for the grids & clouds and for the data e-infrastructure landscape too, as an instrument to facilitate collaboration and harmonisation between EU and Asia.

4.4 Specific to Grids & clouds and Data

11. The grid as a technology is disappearing, and cloud fills/filled the gap.
12. Although Europe lacks big commercial cloud providers, and therefore lags behind the US and China in terms of available cloud capacity, the EOSC vision is ahead of the rest of the world for facilitating Open Science. EOSC is a concept and soon a 'product' that the European academic world could export to the rest of the world. Europe should think about the best strategy and approach for this.
13. The grids, clouds and data areas are more diverse and dynamic than networking and HPC. Therefore, these areas need stronger 'human support' for user uptake, i.e. consultancy, information sharing and training. (about topics such as Open Science; Data science; specific technologies) Europe should strengthen support in these domains, especially because human support is one of the most important distinguishing feature of academic solutions over commercial offerings (e.g. over commercial clouds).

14. There are developments outside Europe (primarily in the US) that are complementing the ongoing projects and development directions in Europe. Stakeholders in Europe should monitor these, possibly adopt and benefit from their outcomes:
    a. Full-stack cloud-based setups for disciplinary areas, covering the full research data lifecycle in a single offering (e.g. CyVerse discussed in Section 3.3.1)
    b. New business models with the involvement of pay-for-use models, free and paid tiers, donations (e.g. Dryad and figshare discussed in 3.4.1)
    c. The different ways of involving commercial providers in the support of scientific applications. (e.g. the E-CAS project in the US; Google and Microsoft cloud grants)

5 Acknowledgement

The author would like to thank Debora Cresti (CINECA/PRACE) for her valuable input and review of the HPC landscape, Annabel Grant (GEANT) for her review of the Network landscape, Paolo Manghi, Najla Rettberg and Iryna Kuchma for their input to the Data landscape, Mark Dietrich and Tiziana Ferrari for their input and constructive feedback on the overall report.

6 References

[1] E-infrastructures: making Europe the best place for research and innovation. Booklet by the European Commission, 2016, https://ec.europa.eu/digital-single-market/en/news/e-infrastructures-making-europe-best-place-research-and-innovation,

[2] E-infrastructures in the Horizon 2020 programme: https://ec.europa.eu/programmes/horizon2020/en/h2020-section/e-infrastructures

[3] e-infrastructure Reflection Group (e-IRG): Guide to e-infrastructure requirements for European Research Infrastructures, http://e-irg.eu/catalogue/eirg-1004

[4] Riding the wave – How Europe can gain from the rising tide of scientific data, Oct 2010: http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=707

[5] The Data Harvest Report – sharing data for knowledge, jobs and growth. An RDA Europe report, Dec 2014: https://rd-alliance.org/data-harvest-report-sharing-data-knowledge-jobs-and-growth.html

[6] ESFRI Roadmap 2018: http://roadmap2018.esfri.eu/

[7] Platform of National eScience Centres in Europe (PLAN-E): https://plan-europe.eu/

[8] Research Data Alliance: https://rd-alliance.org

[9] HPC Top500 list: https://www.top500.org/lists/

[10] National Strategic Computing Initiative (USA): https://www.nitrd.gov/nsci/

[11] Exascale Computing Project (USA): https://www.exascaleproject.org/

[12] Flagship 2020 Project (Japan): https://www.r-ccs.riken.jp/en/overview/exascalepj

[13] List of National research and education networks (NRENs): https://en.wikipedia.org/wiki/National_research_and_education_network

[14] GÉANT Compendium: https://compendiumdatabase.geant.org/

[15] GÉANT 2016 Compendium: https://www.geant.org/Resources/Documents/Compendium2016.pdf

[16] Global Network Architecture (GNA): https://gna-re.net/

[17] AfricaConnect2 project: https://www.africaconnect2.net

[18] CAREN project: http://caren.geant.org

[19] EaPConnect project: https://www.eapconnect.eu/

[20] EUMEDCONNECT3 project: http://www.eumedconnect3.net/Pages/Home.aspx

[21] XSEDE programme: https://www.xsede.org/

[22] International HPC Summer School, Kobe, Japan, July 2019: https://ss19.ihpcss.org/

[23] Latin America HPC Conference (CARLA), Turrialba, Costa Rica, September 2019, http://carla2019.ccarla.org/index.php/archive/

[24] Supercomputing Consortium of Russian Universities: http://hpc.msu.ru/node/136

[25] Ian Bird: Evolving distributed computing for the LHC, EGI Conference, May 2019, Amsterdam: https://indico.egi.eu/indico/event/4431/session/1/contribution/43

[26] Open Science Grid: https://opensciencegrid.org/

[27] JetStream Cloud (US): https://jetstream-cloud.org
[28] Scientific applications supported by the E-CAS project: https://www.internet2.edu/vision-initiatives/initiatives/exploring-clouds-acceleration-science/#phase1
[29] NET+ Cloud Services Program: https://www.internet2.edu/vision-initiatives/initiatives/internet2-netplus/
[30] Enabling Grids for E-sciencE projects (2004-2010): http://eu-egee-org.web.cern.ch/eu-egee-org/index.html
[31] DiCOS - operating system for the distributed Cloud environment: https://dicos.grid.sinica.edu.tw/
[32] Kento Aida, Report about e-Science activities in Japan, International Symposium on Grids & Clouds, Taipei, March 2019: http://indico4.twgrid.org/indico/event/8/session/17/contribution/78
[33] Basuki Suhardiman, Report about e-Science activities in Indonesia, International Symposium on Grids & Clouds, Taipei, March 2019: http://indico4.twgrid.org/indico/event/8/session/17/contribution/76
[34] Jelina Tetango, Report about e-Science activities in the Philippines, International Symposium on Grids & Clouds, Taipei, March 2019: http://indico4.twgrid.org/indico/event/8/session/17/contribution/73
[35] Future Requirement Survey for ASEAN Countries by ASEAN Task Force, Presented by Suhaimi Napis, International Symposium on Grids & Clouds, Taipei, March 2019: http://indico4.twgrid.org/indico/event/8/session/17/contribution/77
[36] Gang Chen, Report about e-Science activities in China, International Symposium on Grids & Clouds, Taipei, March 2019: http://indico4.twgrid.org/indico/event/8/session/17/contribution/79
[37] Nectar cloud: https://nectar.org.au/research-cloud/
[38] Australian Research Data Commons: https://www.ands-nectar-rds.org.au/
[39] South African National Integrated Cyber Infrastructure System (NICIS): https://www.csir.co.za/national-integrated-cyber-infrastructure-system
[40] EU-IndiaGrid project: https://cordis.europa.eu/project/rcn/80159/factsheet/en
[41] Indian Centre for Development of Advanced Computing: https://www.cdac.in/index.aspx?id=hgc
[42] RDA funders: https://rd-alliance.org/about-rda/our-funders.html
[43] Dryad digital repository: http://datadryad.org/
[44] figshare: http://figshare.com/
[45] Harvard Dataverse: http://dataverse.harvard.edu/
[46] Open Science Framework: http://osf.io/
[47] Mendeley Data: https://data.mendeley.com/
[48] Springer Nature – Recommended repositories to submit data: https://www.springernature.com/fr/authors/research-data-policy/repositories/12327124

[49] Nature – Recommended repositories to submit data: https://www.nature.com/sdata/policies/repositories
[50] National Data Service (US): http://www.nationaldataservice.org/
[51] Google Dataset Search: https://toolbox.google.com/datasetsearch
[52] Schema.org: https://schema.org/
[53] LA Referencia: http://www.lareferencia.info/en/
[54] Australian National Data Service: https://www.ands.org.au
[55] Research Data Australia discovery portal: https://www.ands.org.au/online-services/research-data-australia
[56] RIF-CS schema: https://www.ands.org.au/online-services/rif-cs-schema
[57] CODATA: http://www.codata.org
[58] African Open Science Platform initiative: http://africanopenscience.org.za/
[59] EU-Brazil Cloud Calls: https://www.eubrasilcloudforum.eu/en/eu-brazil-cloud-research
[60] Projects funded in the 4th EU-BR Coordinated Call (2017): https://www.eubrasilcloudforum.eu/en/4th-eu-br-coordinated-call-2017
[61] GÉANT: https://www.geant.org/
[62] PRACE: http://www.prace-ri.eu/
[63] EuroHPC Joint Undertaking: https://eurohpc-ju.europa.eu/
[64] EGI federation: https://www.egi.eu
[65] EGI Service catalogue: https://www.egi.eu/services/
[66] The European Cloud Initiative: https://ec.europa.eu/digital-single-market/en/%20european-cloud-initiative
[67] European Open Science Cloud hub project (EOSC-hub): https://www.eosc-hub.eu/
[68] European Open Science Cloud projects: https://eosc-portal.eu/about/eosc-projects
[69] European Open Science boards: https://eosc-portal.eu/governance/eosc-board
[70] EUDAT: https://eudat.eu
[71] OpenAIRE: http://openaire.eu
[72] Zenodo data repository: https://zenodo.org/
[73] CoreTrustSeal data repository certification: https://www.coretrustseal.org/
[74] Re3data – Registry of Research Data Repositories: https://www.re3data.org/
[75] HPC Centres of Excellence projects: https://ec.europa.eu/programmes/horizon2020/en/news/overview-eu-funded-centres-excellence-computing-applications
[76] EOSC-hub Competence Centres: https://wiki.eosc-hub.eu/display/EOSC/WP8+Competence+Centres
[77] ScienceOpen: https://www.scienceopen.com/
[78] Unpaywall: https://unpaywall.org
[79] CyVerse: http://www.cyverse.org/about
[80] PRACE-5IP project, D2.2 – Report on stakeholder engagement, to be available at http://www.prace-ri.eu/prace-5ip-public-deliverables/

[81] NCI National Research Data Collection (Australia): https://nci.org.au/services/nci-national-research-data-collection/
[82] Federated Research Data Repository (Canada): https://www.frdr.ca
[83] Internet2 (US): https://www.internet2.edu
[84] KENET (Kenya): https://www.kenet.or.ke
[85] TEIN*CC: http://www.tein.asia
[86] RedClara: https://www.redclara.net
[87] UbuntuNet Alliance: https://ubuntunet.net/
[88] WACREN: https://www.wacren.net/
[89] e-Infrastructure Reflection Group (e-IRG): http://e-irg.eu
[90] Australian Research Data Services (RDS): https://www.rds.edu.au/
[91] Open Research Cloud Alliance (ORCA): http://www.openresearchcloud.org
[92] ORCA Declaration: http://www.openresearchcloud.org/declaration/
[93] Fostering FAIR data practices in Europe H2020 project (FAIRsFAIR): https://www.fairsfair.eu/
[94] Research Data Alliance FAIR Data Maturity Model Working Group: https://www.rd-alliance.org/groups/fair-data-maturity-model-wg
[95] EGI Agreements with peer infrastructures around the world: https://www.egi.eu/about/collaborations/
[96] WISE 'information security for IT infrastructures for research' community: https://wise-community.org/
[97] LA Referencia and CERN enter into agreement to promote and expand open science in Latin America: http://www.lareferencia.info/es/blog-masonry/noticias/item/218-la-referencia-and-cern-enter-into-agreement-to-promote-and-expand-open-science-in-latin-america
[98] China asserts firm grip on research data: https://www.sciencemag.org/news/2018/04/china-asserts-firm-grip-research-data
[99] Is China ready for Open Data?: https://researchdata.springernature.com/users/69696-varsha-khodiyar/posts/54209-is-china-ready-for-open-data
[100] Fudan University Social Science Data Repository: https://dataverse.org/blog/featured-dataverse-repository-fudan-university
[101] Green500 HPC list: https://www.top500.org/green500
[102] CSTCloud federation (China): http://cstcloud.net/
[103] OpenAIRE National Open Access Desks: https://www.openaire.eu/contact-noads
[104] European Open Science Cloud portal: https://eosc-portal.eu/
[105] TANDEM project – "TransAfrican Network Development": https://www.tandem-wacren.eu/
[106] MAGIC project – "Middleware for collaborative Applications and Global virtual Communities": https://cordis.europa.eu/project/rcn/196897/factsheet/en
[107] New Zealand eScience Infrastructure (NeSI): https://www.nesi.org.nz/