Impact report

The mission of EGI is to create and deliver open solutions for science and research infrastructures by federating digital capabilities, resources and expertise between communities and across national boundaries. The scientists relying on EGI services work in large international organisations, in research infrastructures, projects, university labs, or as individual researchers. Today, EGI provides both technical and human services, from integrated and secure distributed high-throughput computing and cloud computing, storage and data resources to consultancy, support and co-development.

The exploitation of the project Key Exploitable Results (described at the end of this chapter) ensured broad impact in different areas like the advancement of scientific knowledge, the adoption of digital innovation in science, and the implementation of European policies and priorities for the European Research Area.

# Impact on Science

EGI-Engage had a mission to expand the capabilities of a European backbone of federated computing services to serve research at all scales. The impact of EGI services is felt across many scientific disciplines and at all scales of the research landscape, from large research communities & Research Infrastructures to the individual researchers and groups described as the long-tail of science.

This chapter includes a few examples of scientific results and innovation that would not have been possible without EGI.

### Individual scientists and university groups (the long-tail of science)

**

**Genetics of pathogens**

*Salmonella* infections end up with many unpleasant symptoms and a likely trip to the hospital. Konrad Förstner, a bioinformatician working at the University of Würzburg (Germany), investigated the invisible genetic battle going on inside the cells.

Förstner and his team took a groundbreaking approach and analysed the combined genetic material from the *Salmonella* bacteria and the host. They analysed the genetic sequences with READemption, a pipeline designed to process the computational tasks using EGI Cloud Compute.

They found that a piece of *Salmonella* RNA has a strong influence on what happens inside the host cells. This result is key for a better understanding of the infection processes and immune responses and was published in *Nature* and *Bioinformatics*.

KU Förstner et al. 2014. Bioinformatics. doi: 10.1093/bioinformatics/btu533

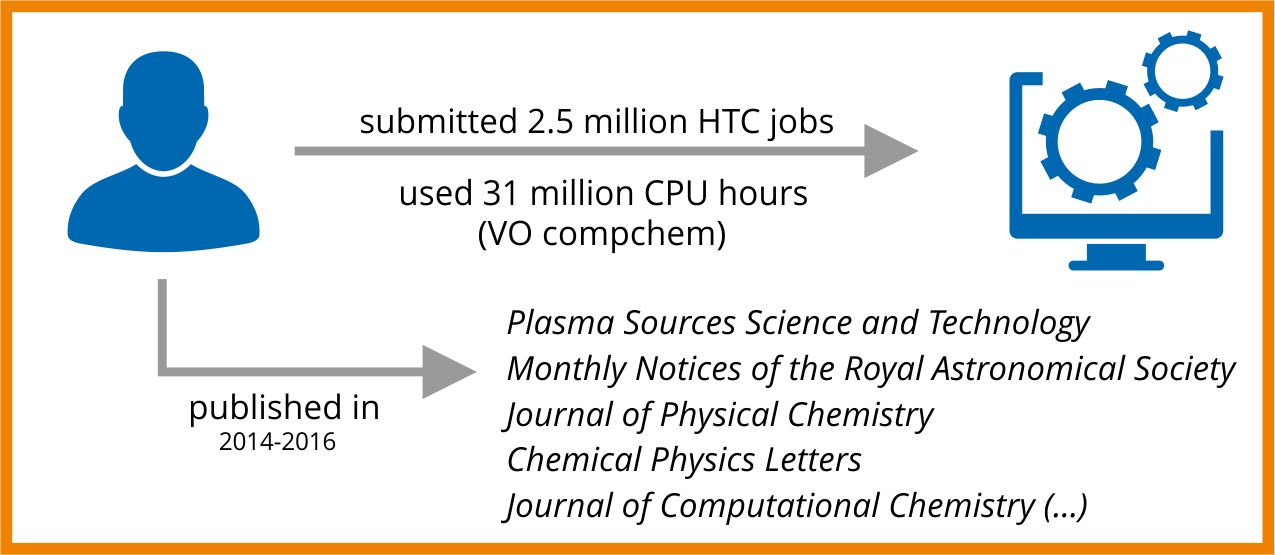
AJ Westermann et al. 2016. Nature. doi:10.1038/nature16547

<https://www.egi.eu/use-cases/research-stories/the-genetics-of-salmonella-infections/>

**Theoretical chemistry**

Chemical reactions are at the core of everything that happens in the Universe. From the thermonuclear fusion that powers the Sun, to how antibiotics help to fight pneumonia, everything depends on what happens when molecules collide and interact to form new compounds.

Ernesto García, based at the University of the Basque Country in Spain creates computational models to describe chemical reactions. In the last two years, García worked in projects ranging from astronomy, to applied chemistry and atmospheric science. He has submitted about 2.5 million High-Throughput Compute jobs for a total of 31 million of CPU hours and published eight papers in peer-reviewed journals, including high-impact *MNRAS* and *Chemical Physics Letters*.



<https://www.egi.eu/use-cases/research-stories/what-happens-when-molecules-collide/>

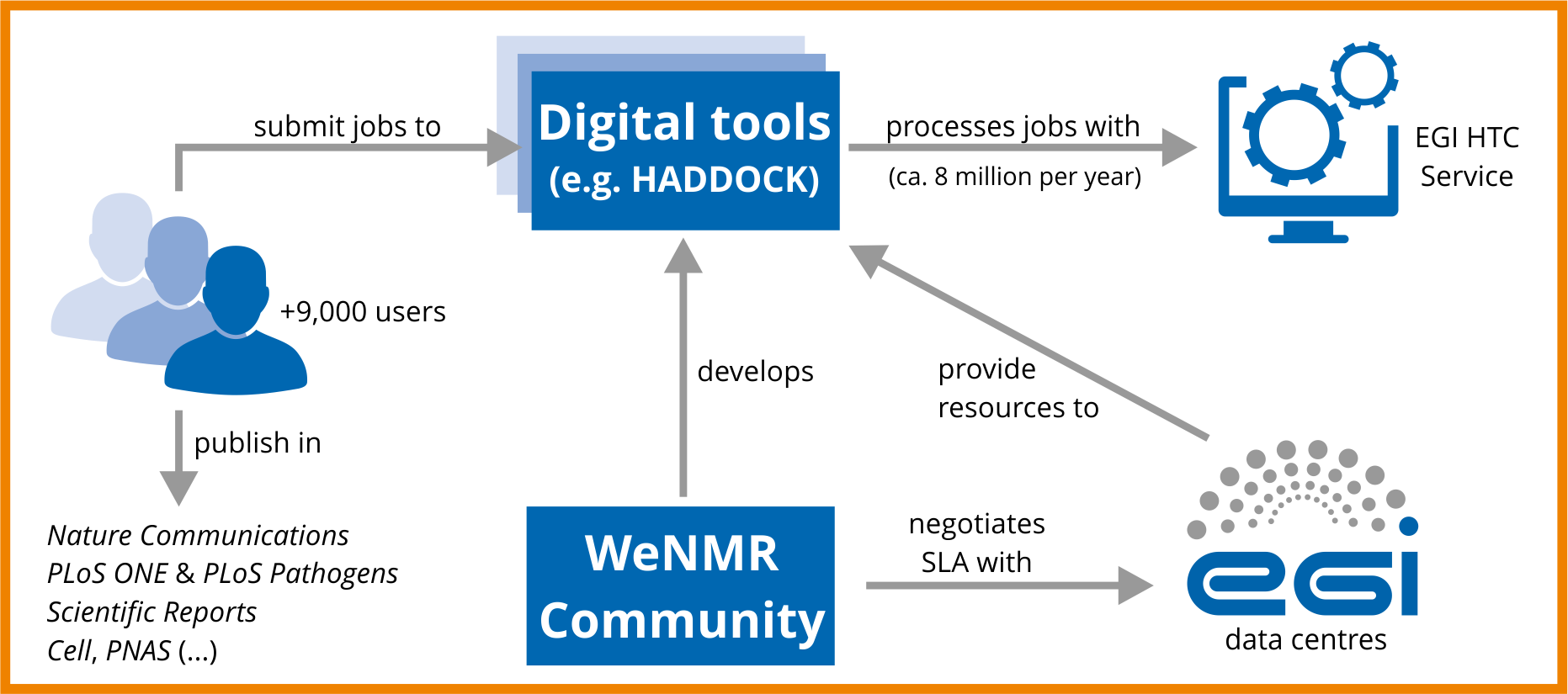
### Community-led innovation

**Digital tools for structural biology**

Structural biology provides an understanding of inner workings of proteins, aminoacids and all the biomolecules that are essential to life. It's a field at the forefront of innovation in medicine and health sciences for the insights it gives to the development of new pharmaceutical drugs, for example. Structural biology research relies on accessible and reliable computational resources for basically everything: from storing huge databases of molecular structures to the processing power required to run complex simulations.

WeNMR is a research community that has developed over the years a rich portfolio of digital tools for structural biology. An example is HADDOCK, a web portal that offers tools for to model the structure of complexes of proteins via a user-friendly interface. HADDOCK consistently ranks at the top of protein prediction experiments and is one of the best programs available for molecular docking. The WeNMR web portals hide the computational complexity from the scientists and process the simulation with the EGI High-Throughput Compute service.

The WeNMR community is supported by the EGI federation through an Service Level Agreement, negotiated within the EGI-Engage project, with the national e-Infrastructures of Belgium, France, Germany, Italy, the Netherlands, Poland, Portugal, Spain and the UK. So far the resources have been used to process over 130,000 job submissions from about 9,000 registered users, which translates into about 8 million jobs per year on the EGI infrastructure.



### Large research collaborations

**Deciding the location of the Cherenkov Telescope Array (CTA)**

The CTA collaboration brings together 1350 scientists and engineers from 32 countries with the goal of building the world’s largest and most sensitive very high-energy gamma-ray observatory. The CTA facility will be used to understand the role of high-energy particles in the most violent phenomena of the Universe and to search for annihilating dark matter particles.

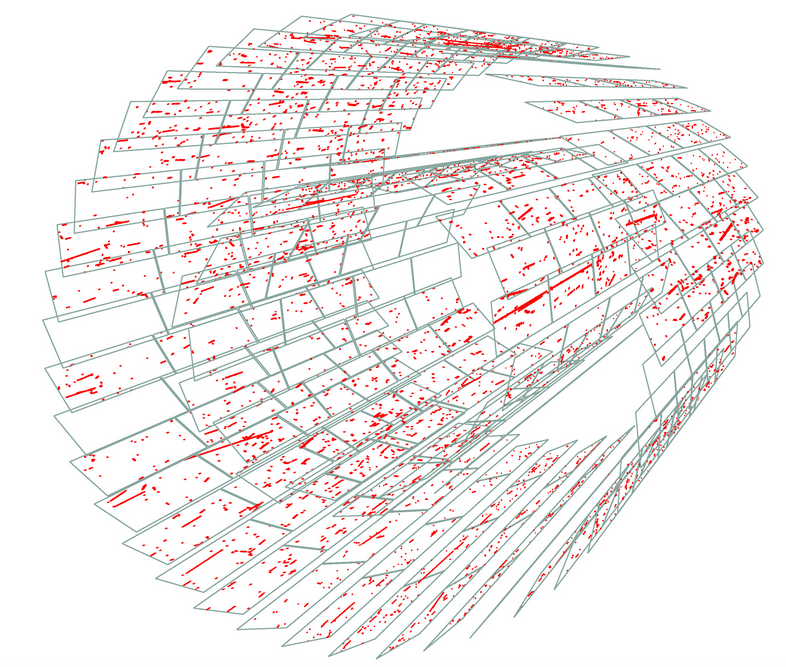
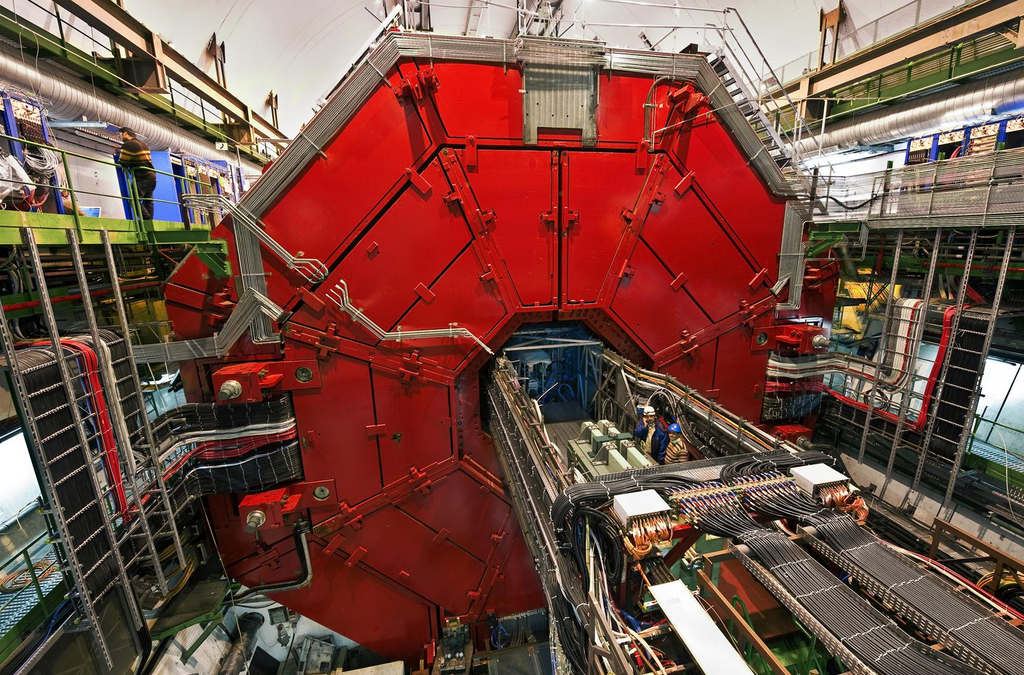
During the EGI-Engage period, the team consumed +360 million core hours in High-Throughput Compute jobs made available by six EGI data centres. With these resources, the collaboration preformed Monte Carlo simulations to decide the best locations and the best geometrical configuration for the telescope arrays. This led to the selection of La Palma in Spain and Paranal in Chile. Construction is expected to start soon.

**Strange and charming particles**

On April 24 2017 in a paper published in Nature Physics (doi: doi:10.1038/nphys4111), the ALICE collaboration reported the observation of so-called strange hadrons in collisions where a large number of particles are created. Strange hadrons are well-known particles with names such as Kaon, Lambda, Xi and Omega and contain at least one so-called strange quark. The enhanced production of strange particles that was observed is a feature of quark-gluon plasma - a very hot and dense state of matter that existed for just a few millionth of a second after the Big Bang. Now, for the first time ever the phenomenon was unambiguously observed and this result is likely to challenge existing theoretical models that do not predict an increase of strange particles in these events.

A few months later, on July 6 2017, the LHCb collaboration announced the first observation of a doubly charmed particle. This particle, called the Ξcc++, is a baryon (particle composed of three quarks) containing two charm quarks and one up quark, resulting in an overall doubly positive charge. The existence of doubly charmed baryons was suspected since the 1970s. The confirmation arrives now, thanks to the high production rate of heavy quarks at the LHC and to the unique capabilities of the LHCb detector that can spot decay products with excellent efficiency and purity.

ALICE and LHCb are two of the four experiments hosted by the Large Hadron Collider (LHC). The data produced by the LHC detectors is immense, at the scale of many Petabytes per year. This data is collected, stored, shared and analysed by hundreds of data centres in the EGI Federation.

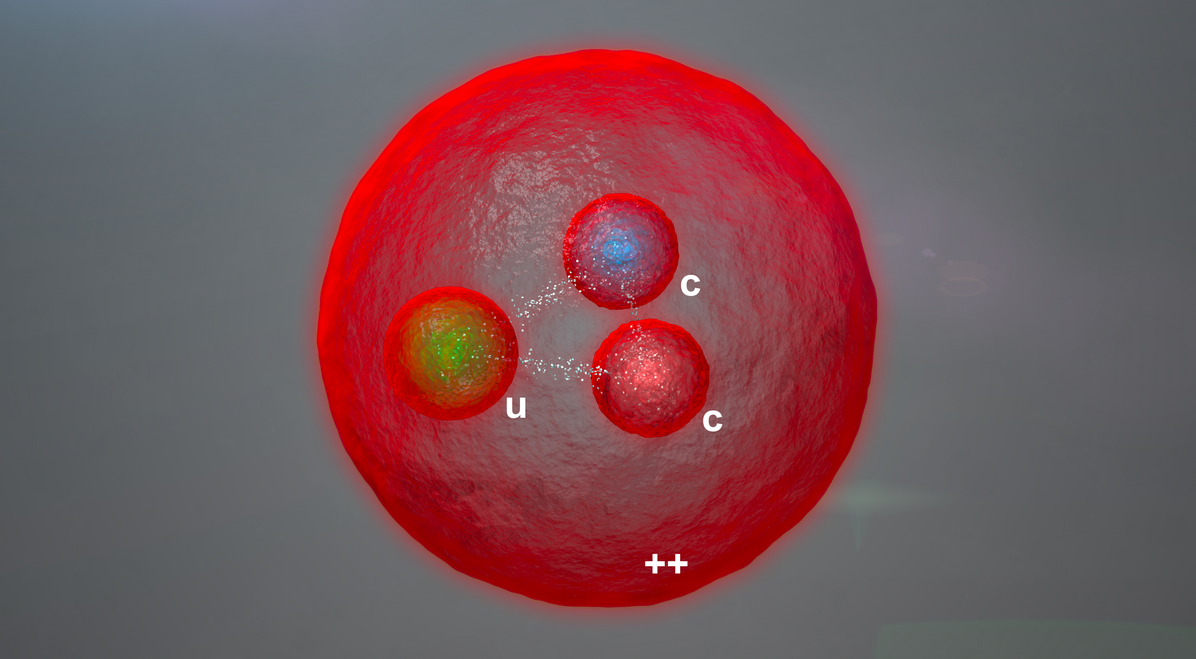


Figure 1. First muon tracks at LHC (top left) recorded by the ALICE Silicon Pixel detector (top right), and an artist representation of the doubly charmed baryon. Credits: CERN, the ALICE Collaboration and the LHCb Collaboration.

**The discovery of gravitational waves with LIGO and Virgo**

For centuries, mankind has observed the sky, first with the naked eye then with increasingly bigger and more powerful telescopes. Today scientists study the universe mainly by observing forms of electromagnetic radiation (e.g. radio waves, gamma rays) and high-energy particles such as cosmic rays or neutrinos. The data collected by large detectors, requires a large amount of computational resources in order to be processed.

On February 11 2016, almost 100 years after Einstein’s prediction, the discovery of gravitational waves was announced. Gravitational waves are thought to be emitted by spinning stars in binary systems, black holes or massive stellar explosions. The discovery, accepted for publication in the journal *Physical Review Letters*, was made by the LIGO Scientific Collaboration and the Virgo Collaboration using data from the two LIGO detectors. A second detection of a gravitational wave event, thought to be originated by the coalescence of two stellar-mass black holes, was announced in January 2017.

The observations earned the Nobel Prize for Physics in 2017, attributed to scientists Rainer Wiss, Barry C. Barish and Kip S. Thorne of the LIGO/Virgo collaboration on 3 October 2017.

The LIGO Scientific Collaboration is a group of more than 1000 scientists from universities around the United States and in other 14 other countries. The Virgo Collaboration consists of more than 250 physicists and engineers affiliated with European institutions such as CNRS, INFN, Nikhef, Wigner RCP, the POLGRAW group and EGO. Their computational model's complexity is proportional to the size of the collaboration: data is taken by three detectors in Europe and USA. Online analysis happens on the measurement sites and data is stored in a circular cache for distribution to different computing centres: the LIGO site in the US, IN2P3 in Lyon and INFN-CNAF in Bologna. The two Virgo centres for offline analysis (IN2P3 and INFN-CNAF) are complemented by HTC compute services provided by INFN grid sites; all of these data centres are federated in EGI. In the USA LIGO uses Condor Clusters federated in Open Science Grid – one of the major e-Infrastructures collaborating with EGI - and HPC resources from XSEDE.

In EGI, the Virgo Virtual Organistion (VO) was responsible for managing thousands of HTC computations submitted to resource centres of the EGI federation. During the EGI-Engage lifetime, the computational needs of Virgo have increased by +623%. The Virgo VO consumed collectively 40 Million CPU hours in 2015 and 2016.

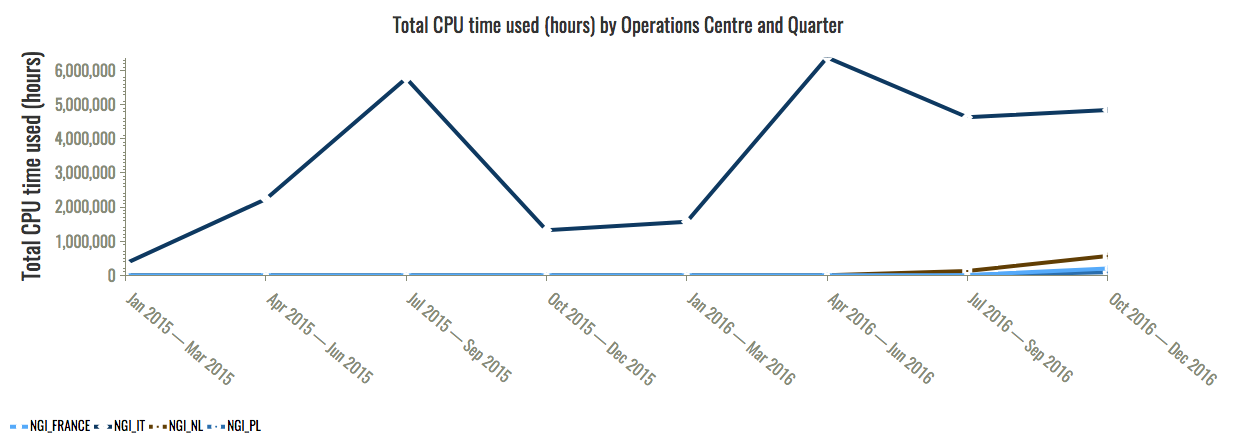
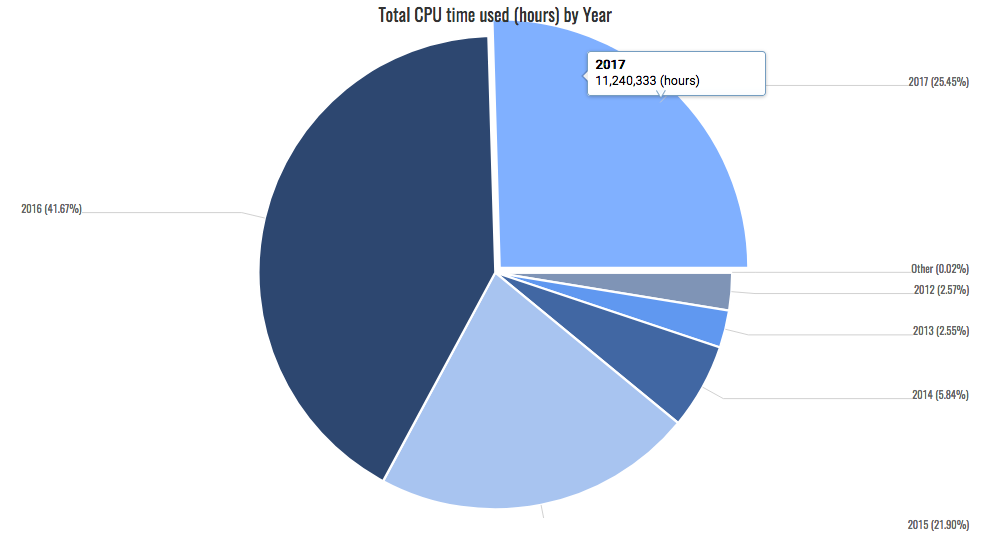


Figure 2. 40 million CPU hours were consumed in total by the VIRGO collaboration during 2015 and 2016. Source: EGI Accounting Portal.



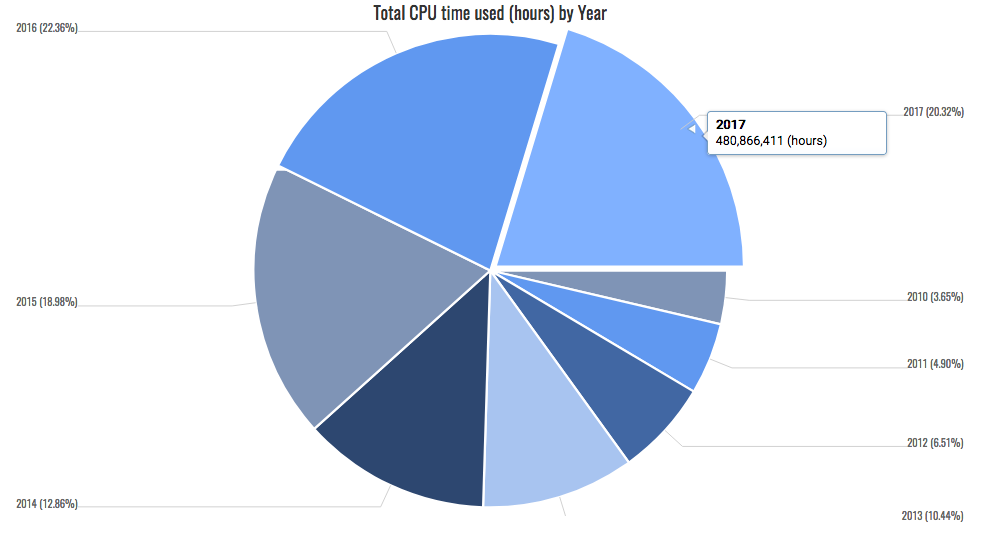


Figure 3. The diagrams plot the increase in compute utilization from 2010 to date of Virgo (top) and ALICE (bottom). Both research collaborations have been increasing exponentially the need of computing resources following the trend of the increasing data production rate and under the push of scientific boundaries thanks to new recent groundbreaking discoveries.

**Tilling up the papers**

The scientific production enabled by EGI services and EGI-Engage key results, amounts to more than 2,000 peer-reviewed papers in 2015 and 3,000 ones in 2016. In total, 22,300 publications can be attributed to projects and research collaborations supported by the EGI Federation since 2008.

The full list can be browsed and consulted in the OpenAIRE repository under the EGI tag (<https://www.openaire.eu/egi-stats>).

Scientific results have a long-term influence on society and sustainable development. The full impact of the research supported by advanced computing will be appreciated in the long term. Thanks to the variety of disciplines supported by EGI, we believe many sectors of our society will be positively influenced by EGI-Engage in the future.

# Digital innovation for science

## Adoption of storage and computing infrastructures

During EGI-Engage the number of registered users increased from 23,520 to 30,508, an overall relative increase of 30%.

A similar trend was observed for users accessing EGI services via thematic services and community-specific portals, which offer data, data products, software and collaborative tools on top of generic EGI services. The total number of users increases to approximately 61,000 users (+61% during EGI-Engage).

The project supported the engagement with 40 new research communities. Among these, 19 are Research Infrastructures and 13 are projects and platform developers/providers supported by national and H2020 funding. The target number of communities enaged was exceeded by +106%.

The increase of the user base was driven by the Physics branch of the Natural Science discipline thanks to the start of pre-production computing activities in various large experiments and to new research frontiers opened by recent discoveries and the increase of sensitivity of scientific observatories.

The increase of the user base was also driven by the expansion of a few national long tail of science groups, and the integration of an increasing number of platforms and applications that have been ported and running on the federated cloud, e.g. Peachnote (Art and Humanities, 1.6 Million users), the National Bioinformatics Infrastructure Sweden tools (10,000 users), the Geohazards ESA Thematic Exploitation platforms (more than 500 users) and the BlueBridge Virtual Research Environments on EGI (more than 250 users). Finally, through the Competence Centres, technical support, training and engagement towards new research groups, EGI-Engage succeeded in bringing new scientific communities to become testers and early adopters of advanced computing services.

For what concerns the business sector, more than 65 contacts were established of which two thirds involve SMEs. The business engagement programme[[1]](#footnote-1) achieved the full support of seven concrete business cases. 25 additional ones are currently in progress.

In particular, through the Competence Centres:

* The DARIAH data repository was fully integrated with EGI Compute services.
* EISCAT\_3D tested and co-designed a DIRAC-based solution for acquisition, curation, access to and processing of research data.
* ELIXIR prototyped a data analysis infrastructure to serve some high-impact life science scientific user cases.
* EPOS demonstrated the use of IaaS, PaaS and SaaS services for data exploitation.
* WeNMR brought to production novel tools to run molecular dynamics simulations and other suited image processing tasks on accelerated computing nodes.
* LifeWatch became an active user of distributed cloud computing for the analysis of distributed heterogeneous data sets.

38 Additional Virtual Organizations (VOs) where enabled during the project. Following the PY1 project recommendations, the technical support activities were primarily focused on large international user groups from Research Infrastructures to maximize impact, and two new unfunded Competence Centres – EMSO and ICOS – where activated.

User communities that completed the engagement process, where further supported in full production by ‘connecting’ them to service providers interested in long-term support via Service Level Agreements (SLAs). 11 SLAs were defined during the project, and 34 additional corporate agreements were defined with users of the Applications on Demand, currently in Beta.

A new service level management process was defined and adopted in EGI-Engage as component of the EGI Integrated Management System. The process aims at maintaining a service catalogue, and at defining, agreeing and monitoring service levels with customers by establishing SLAs, supportive operation level agreements (OLAs) and underpinning agreements (UA). The process is part of the Integrated Service Management framework set up to comply with ISO 9000 and 20000 standards, which got formally audited and certified at the end of 2016.

## Increased availability of scientific data

To date EGI federates 650 PB of stored data. Large research infrastructures were supported for their data processing and analysis needs. With EGI-Engage the number of international long-term research collaborations and infrastructures supported by the EGI Federation increased by 48%.

To date 31 large-scale international collaborations are supported by compute services of the EGI Federation; the number includes eight ESFRI landmarks and three ESFRI projects.

A few examples provide an indication of the scale at which data is being produced and analyzed in the EGI Federation. The Belle II experiment consumed over 1.6 billion CPU hours. In Belle II the computing system has to handle an amount of beam data eventually corresponding to several tens of Petabytes per year.

The Cherenkov Telescope Array preparatory activities conducted between 2013 and 2016 involved 11 PB of data.

The LHC experiments have been producing a sustained rate of about 15 PBs of raw data each year that must be stored, processed, and analyzed: on 29 June 2017, the CERN data centre passed the milestone of 200 PB of data permanently archived in its tape libraries. The data centre coped with 49 PB of data received from the LHC data[[2]](#footnote-2) just in 2016. During 2016 alone, on average 80 Petabytes of data were transferred per month, with peaks at 96 Petabytes during summer. This corresponded to more than 1 billion files per month transferred to thousands of particle physicists working across the world.

LOFAR, the Low-Frequency Array, exceeded 11 Petabytes of archived data.

To date, EGI compute services are being tested and/or used for production-level data analysis by eight ESFRI Landmarks (BBMRI, DARIAH, ELI, ELIXIR, EMSO, HL-LHC, INSTRUCT and LifeWatch) and three Projects (CTA, EPOS and KM3NeT 2.0). EGI accounting data shows that among the 31 large-scale research initiatives currently supported, those which reached full adoption in 2015, have been largely increasing usage in 2016 and 2017 (see **Error! Reference source not found.**).

An additional group of research collaborations and infrastructures is currently involved in service co-design and testing. The group involves: EISCAT-3D, BBMRI, ELIXIR, EPOS, ICOS and SKA.

During EGI-Engage 63 organizations from research and industry have been engaged as part of the project result exploitation plan. Of these 25 organizations are currently actively involved in service piloting.

One of the general challenges affecting today’s research enterprises, is the lack of data experts who can bring frontline ICT developments to science. This problem has been highlighted in the first report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud[[3]](#footnote-3) and is a challenge for effective research data exploitation. The document stressed how “*the lack of core intermediary expertise has created a chasm between e-infrastructure providers and scientific domain specialists*”. EGI-Engage addressed this challenge with a network of Competence Centres that brought together experts from research communities, technology providers and service providers, with support actions towards additional external user groups and with a business programme dedicated to industry and SMEs.

Accessibility and reuse of research data were made possible by the following project activities:

* The technical support to data-driven research infrastructures and communities.
* The prototyping of an Open Data platform to bring relevant public research data close to computation (e.g. LifeWatch diversity conversation core data resources and Copernicus sentinel data). DataHub, offers the possibility to access, store, process and publish data using global data storage backed by computing centers and storage providers worldwide with transparent access to distributed data sets, without unnecessary staging and migration.
* The increased accessibility data thanks to the rolling to production of EGI Check-in for use of federated identifies in the EGI authentication and authorization infrastructure.
* The porting of virtual research environments and various thematic services with EGI baseline Compute and Data Management solutions through the Competence Centres, and the provisioning of a portfolio of scientific applications as Software as a Service through the Applications on Demand beta service, allowing individual researchers and small research teams to perform compute and data-intensive simulations on large, distributed networks of computers, and currently offering 13 different applications[[4]](#footnote-4).

## More efficient use of IT equipment for research

In April 2016 the European Cloud Initiative - Building a competitive data and knowledge economy in Europe – was launched by the European Commission[[5]](#footnote-5). The letter emphasized that fragmentation can hamper data-driven science: *data infrastructures are today split by scientific and economic domains, by countries and by governance models. Access policies for networking, data storage and computing differ. Disconnected and slow data and computing infrastructures hinder scientific discovery, create silos and slow down the circulation of knowledge*.

A European-scale sharable facility for computational, storage and data analysis was recognized to be missing. Integrated provisioning of data, open data analysis tools and connected computing facilities was advocated to better serve the great majority of researchers in Europe.

EGI-Engage addressed this challenge through the coordinated offer of scientific applications, federated computing and storage. The EGI Federation comprises to date about 730,000 cores, 300 PB of online storage and 346 PB of nearline storage.

The compute capacity increased by +23% in the first reporting period, and by +12.3% in the latter one. Online and nearline storage experienced a similar trend: +12% and 42% respectively in the first reporting period, and +5% and +23.3% in the last one. Today more than 200 research collaborations are benefitting by the resulting baseline technical infrastructure.

The project issued a study for cross-border procurement of e-Infrastructure services, which provides guidelines for service procurement in the future European Open Science Cloud aiming at optimizing the cost of providing and operating data-intensive compute infrastructures in Europe. The document defines barriers that inhibit procurement and examines a set of case studies to identify best practices that can overcome such procurement barriers. Analysis of the barriers, case studies and best practices led to a series of opportunities being proposed for the future planning of EGI and the European Open Science Cloud initiative of the EC.

## No lock-in to particular hardware and software platforms

The EGI Federation is designed to allow data and software portability across multiple heterogeneous facilities. Via reference standards and interoperability best practices, the federated model allows the combination of resources (data, software, applications, publications and other digital artifacts) and services from multiple suppliers.

The problem of portability of data and applications is particularly critical when adopting cloud services due to lack of adoption of open standards by many commercial providers, and the proliferation of different Middleware Cloud Stacks in research clouds. During EGI-Engage 14 international research collaborations and their research platforms became active tester and adopters of federated cloud services. A broad range of scientific disciplines (natural sciences, health and medicine, agricultural sciences and humanities) benefited from this.

These research communities involved in cloud computing include 5 Landmarks (ELI, DARIAH, ELIXIR, EMSO and LifeWatch) and one Project (EPOS) in the ESFRI Roadmap 2016. Adoption was possible thanks to the piloting activities conducted by the Competence Centres, by technical support activities towards external projects and Research Infrastructures.

During the project the expected target number of engaged communities was exceeded by 106%.

The EGI-Engage exploitation plan and related communications activities aimed at promoting the usage of the enabled platforms in the context of community events like ESFRI Cluster Project meetings (e.g. CORBEL and ENVRI+). The following section provides information about the increase in yearly usage rates for each supported community.

## Contribution to Digital Innovation Megatrends

‘Megatrends are large-scale social, economic, political, environmental or technological changes that are slow to form but which once they have taken root, exercise a profound lasting influence on many if not most human activities, processes and perceptions’ [[6]](#footnote-6). The project contributed to the further development of various major digital innovation areas mentioned in the Organisation for Economic Co-operation and Development (OECD) ‘Science, Technology and Innovation Outlook 2016 report’ (Figure 4). In particular, EGI-Engage was active to bring innovation to researchers through Cloud Computing, Grid Computing, Modeling Simulation and Big Data Analytics.

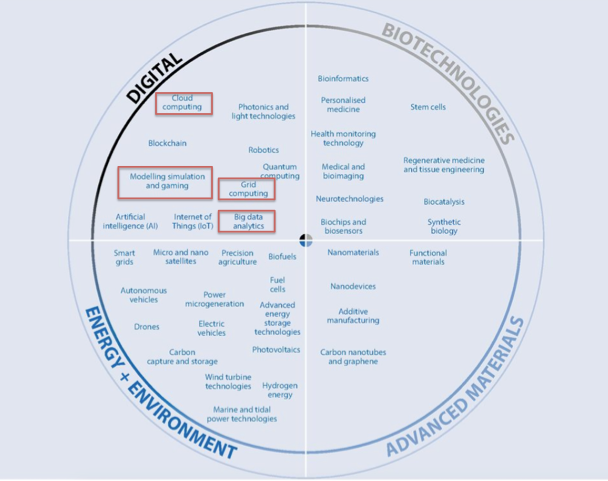


Figure 4. The EGI-Engage contribution (red rectangles) to the development and adoption of emerging future technologies according to the OECD Science, Technology and Innovation report 2016.

### Cloud computing

In the area of Cloud Computing EGI-Engage established a blueprint consisting of best practices and reference standards to achieve interoperability across multiple publicly-funded and commercial cloud providers. To date, the EGI Federated Cloud is the only existing publicly-funded distributed research cloud in Europe, offering on average provides an 7 Million CPU hours per year to researchers from all disciplines.

Natural Sciences, Health and Medicine and Art and Humanities are the disciplines that mostly benefited from EGI-Engage support actions as shown in Figure 5.

The EGI Federated Cloud is a IaaS-type cloud, made of academic private clouds and virtualised resources and built around open standards. Its development is driven by requirements of the scientific community. The Federated Cloud architecture is based on the concept of an abstract Cloud Management Framework that supports a set of cloud interfaces to communities. Each Federated Cloud provider operates an instance of this framework according to its own technology preferences and integrates it with the federation by interacting with EGI core components (e.g. AAI, service registry, accounting, monitoring) as shown in Figure 6.

In the EGI Federated Cloud PaaS and SaaS services are offered by various partners across an heterogeneous set of cloud providers using a single authentication and authorization framework that allows the portability of workloads across multiple providers and enable bringing computing to data.

Integration in EGI-Engage was performed by using public interfaces of the supported Cloud Management Frameworks, thus minimizing the impact on site operations. Providers are organised into the Open Standards and OpenStack realms, each realm exposing a homogeneous interface.

EGI-Engage developed key software components, services and policies to enable federated access to multiple cloud providers via federated identity provisioning, authentication and authorization, and to enable portability of applications and data across a hybrid cloud federation. To date, the EGI Federated Cloud comprises 21 publicly funded clouds and 1 commercial cloud.

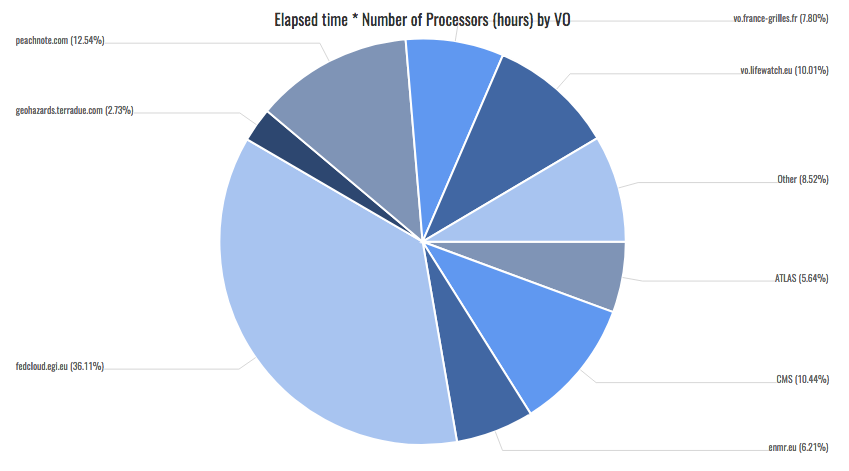


Figure 5. Utilization of Cloud Compute across the main EGI Federated Cloud research communities including the ‘fedcloud’ catch all group (unit: Elapsed time \* Number of Processors (hours) by VO and Year, Jan 2015-August 2017). Source: EGI accounting portal.



Figure 6. EGI Federated Cloud infrastructure and the contributing European cloud providers.

### Grid computing

The EGI Federation is largest HTC distributed computing infrastructure in the world, which federates data centres from Europe but also the Asia Pacific region, the Africa Arabia region, Canada and Latin America and interoperates with the Open Science Grid e-Infrastructure in the USA.

The EGI Grid Federation comprises European national e-Infrastructures (NGIs), data centres from European research organizations, and other major cyber infrastructures from other regions of the world.

Through EGI-Engage EGI continuously worked with e-Infrastructures in the world to ensure an integrated service catalogue for research collaborations.

Thanks to this, during EGI-Engage the overall consumption of HTC scientific computing in EGI saw an enormous increase equal to +40% each year.

The largest HTC infrastructure in the USA is the NSF-funded Open Science Grid (OSG)[[7]](#footnote-7), which federates about 100 distributed centres. EGI and OSG are partners in supporting international research collaborations like the LHC experiments, LIGO and Virgo, structural biologists (WeNMR and SBGrid[[8]](#footnote-8)) and the Large Synoptic Survey Telescope project (LSST). Altogether EGI and OSG respectively provide large scale scientific computing of 7.9 million and 2 million core CPU hours per day.

EGI has established service exchange programs in various regions of the world to ensure data processing and advanced computing are accessible regardless of the country of origin.

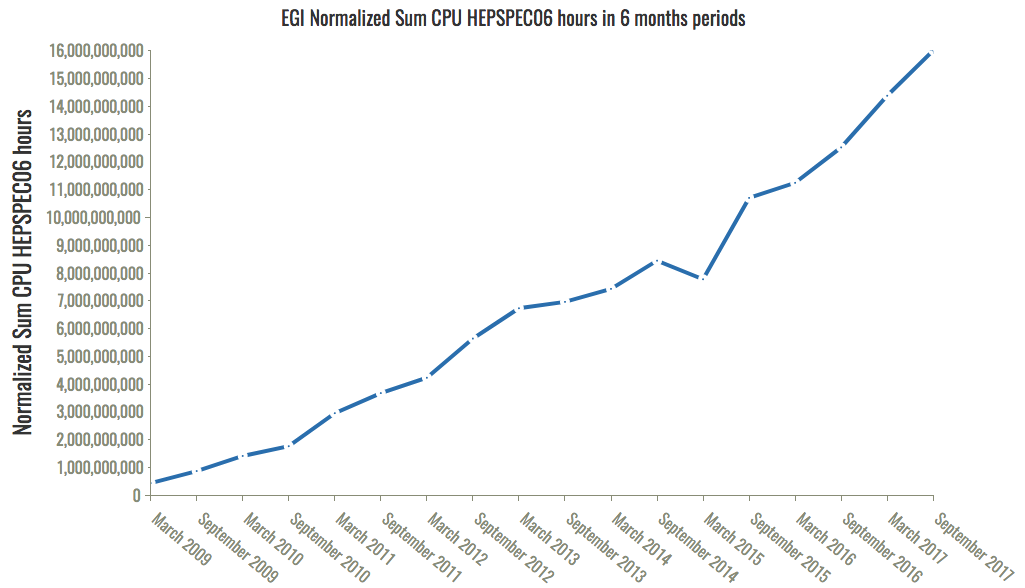


Figure 7. Utilization of EGI HTC compute service from 2009 to date from all user communities. The diagram shows a dramatic change of trend from 2015, when demand started to grow exponentially as a response to the increasing production rate of research data to be processed. Source: EGI Accounting portal.

### Modeling Simulation and Big Data Analytics

Scientific applications providing simulation, data processing and analysis functionalities, data visualization etc. are key enablers for a broad adoption of EGI Compute services. These capabilities are often delivered in the form of scientific gateways and Virtual Research Environments for a web-based access of different digital artifacts like data, software, and processing resources that are managed by diverse systems in separate administration domains.

EGI-Engage supported the technical integration and provisioning of 19 different thematic services[[9]](#footnote-9).

In addition the Applications on Demand (AoD) service – one of the key results of EGI-Engage – was launched as beta service at the end of the project. AoD offers many application and data analysis frameworks, and is expected to evolve over time by adding new features according to the user demand: Jupyter Notebook, Docker, Apache Tomcat, Hadoop, Marathon, and Chronos (generic applications); Galaxy, ClustalW2, Chipster, NAMD and AutoDock Vina (Life Sciences); GnuPlot, Octave and the Statistical R for Computing (for data analysis) and the parallel Semantic Search Engine (Humanities).

AoD provides user-friendly access to compute, storage and applications services in order to carry out data/compute intensive science and innovation. It allows the reuse of existing technology building blocks, requiring minimal new developments; it also allows the hosting and sharing of custom applications.

# Implementation of the ERA

EGI-Engage contributed to the advancement of the implementation of the European Research Area by:

* Strengthening the national systems – the national e-Infrastructures, NGIs, which were better ‘connected’ to an increasing number of international research collaborations.
* Optimizing transnational cooperation by enabling access to national resources and services via an integrated service management system and an internal catalogue of services for providers of the EGI Federation.
* Allowing knowledge circulation and sharing by connecting experts from different research infrastructures, service providers and technology providers in a network of 10 Competence Centres.
* Enabling international cooperation through various new cooperation agreements involving Compute Canada, GEANT and PRACE. These enriched the existing portfolio of collaborations agreements involving major digital infrastructures from all over the world[[10]](#footnote-10).
* Addressing grand societal challenges and United Sustainable Development Goals[[11]](#footnote-11) by: promoting research data exploitation in the agrifood and marine health sectors in collaboration with the AGINFRA initiative and the BlueBridge H2020 project, and providing specific support to the Health sector via the ELIXIR and BBMRI competence centres. In addition, EGI-Engage worked closely with EMSO - a Research Infrastructures in the ESFRI roadmap dedicated to a better understanding of marine ecosystems. Through the establishment of an SLA with EGI cloud providers, EMSO was able to develop and deploy the prototype of their Data Management Platform, a key tool for their mission of creating new knowledge from the data collected by a network of thousands of ocean float observatories.

# European Open Science Cloud (EOSC) and FAIR services

## EOSC Principles

According to the EGI vision researchers from all disciplines should have easy, integrated and open access to the advanced digital capabilities, resources and expertise needed to collaborate and to carry out data/compute intensive science and innovation. In order to achieve this vision, the mission of EGI is to create and deliver open solutions for advanced computing for science and research infrastructures by federating digital capabilities, resources (compute, storage, data) and expertise between communities and across national boundaries.

During EGI-Engage EGI endorsed the principles of the EOSC and advocates the European Open Science Cloud to be the initiative addressing the needs of open access, sharing within and across research communities, ensuring sustained funding to digital research infrastructures.

EGI commits to leverage and exploit the results of the project in the context of EOSC through the following actions that altogether contribute to the implementation of the EOSC roadmap. The tables below map specific actions to the supporting EGI-Engage Key Exploitable Results (KERs).

|  |
| --- |
| ***Governance and funding*** |
| EGI support the definition, implementation and operation of the governing structure benefiting from its long-standing experience with hundreds of facilities worldwide. |
| Contribute its best practices and experience to the definition of the EOSC “rules of engagement”, i.e. the policies for the EOSC actors (e.g. end-users, customers and providers) leveraging for the lightweight federation of services capabilities ensuring interoperability among multiple suppliers at worldwide scale. |
| ***Data culture and FAIR data*** |
| Provide and improve implementation guidelines for FAIR services in the area of advanced compute, federated identity provisioning, authentication and authorization, and contribute to the definition of the EOSC minimum set of reference standards and interoperation agreements and service accreditation models for EOSC. |
| Develop the skills and certification schemes necessary to become users or operators of digital research infrastructures and EOSC through a network of community-lead competence centres involving multiple research communities, technology experts and service providers. |
| ***Research data services and architecture*** |
| Operate the EOSC Hub, the Service Integration and Management system (SIAM) accountable for ensuring that all EOSC service providers perform to provide a seamless service that is compliant to contractual obligations made with customer organizations. Activities include: communicate/manage alignment of services to defined policies and standards, define/maintain/review service level agreements, conduct service audits and quality assurance reviews, manage end-to-end service level management performance retaining overall accountability, provide the tools for aggregating demand and supply, make services and research artifacts discoverable and accessible at European level etc. |
| Operate an open hybrid e-infrastructure offering data-driven advanced compute (Cloud and HTC) and data services from publicly funded and commercial organisations and maintain/evolve the related implementation guidelines to make the related services FAIR. |
| Operate a federated identity provisioning, authentication and authorization services for the EOSC users and service providers. |

## FAIR Services

EGI-Engage contributed to the definition and maintenance of policies, best practices and tools, to make the services of the federation compatible with the FAIR principles. The EGI-Engage contribution to the establishment of FAIR principles for services is outlined below.

* *Findable*: during the project a marketplace (<https://marketplace.egi.eu/>) was designed, implemented and rolled to production. In addition, the EGI internal service catalogue[[12]](#footnote-12) and external service catalogue[[13]](#footnote-13) were defined.
* *Accessible*: accessibility was improved via federated identity provisioning (eduGAIN), and federated authentication and authorization via the Check-in service, and by defining access policies applicable to EGI services[[14]](#footnote-14).
* *Interoperable*: EGI-Engage defined guidelines for Compute (HTC and Cloud) and Data Management interoperability across heterogeneous facilities and multiple suppliers. These resulted in a community-defined standards roadmap[[15]](#footnote-15). In addition, FAIR principles were concretely supported by distributing validated software products adopting the defined standards roadmap. The EGI Federation was integrated with relevant services from the EUDAT CDI (B2SAFE, B2SHARE, and B2ACCESS).
* *Reusable*: the project produced an updated body of security policies for users and providers including the general e-Infrastructure security policy, and the Acceptable Use Policy and Conditions of Use and others[[16]](#footnote-16). In addition, the project enforced adherence to the reference standards roadmap via a lightweight service provider certification procedure[[17]](#footnote-17).

# Key Exploitable Results

This section provides an overview of the project Key Exploitable Results (KERs) that altogether contributed to the project impact described in the previous sections. These belong to two categories: (1) Policy, Processes and Procedures; and (2) Software and services.

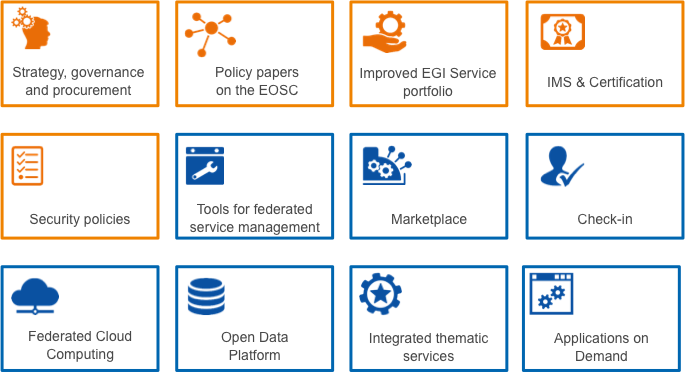


Figure 8. The EGI-Engage Key Exploitable Results grouped in (1) Policy, Processes and Procedures (in orange) and (2) Software and services (in blue).

## Services and software

### Tools for federated service management

Operational tools solve common problems with federated operations. Every distributed infrastructure can benefit from some or all the tools operated by EGI. Accounting/Monitoring can be offered as a service or used as an added value to market the EGI federation to new members. The tools can support the creation of new service federations, or the extension of the existing ones.

### Marketplace

The Marketplace has the ambition of becoming the platform where the services offered by EGI providers and partners can be promoted, discovered, shared and accessed. This includes EGI Services as well as discipline- and community-specific tools and services enabled by EGI and/or provided by third parties under defined agreements.

When operational, the Marketplace will become the unique place where a new customer could discover a service (or group of services), get information about it by browsing the service catalogue, and submit an order, specifying quantity, quality and duration. It will expose all the live EGI services, following the same structure of the service catalogue, exposing service options to allow customers to properly define their orders.

The Marketplace implementation is grounded on an initial analysis and development of a legal, policy and business framework for a marketplace capability that would allow the request, provision, accounting, billing of e-Infrastructure services.

### Federated authentication and Authorization (Check-in)

EGI Check-in is a complete AAI solution that operates as a central hub to connect federated Identity Providers (IdPs) with EGI service providers. Perun was also integrated in this service, in order to support users management, user enrolment and users synchronization from existing VOMSs servers or users' identity management systems.

The EGI Check-in service enables research communities to access the EGI services in a user-friendly way, while preserving security and user privacy. Researchers from home organizations that participate in one of the eduGAIN federations are able to access the EGI services using the same credentials they are using at their home organization. Furthermore, the EGI Check-in service supports user authentication with social media identities, enabling even those users who do not have a federated account at a home organization (such as many users that belong to the “Long Tail of Science”), to be able to access the EGI services in a seamless way without compromising the security of the EGI platform.

The Check-in service can connect to existing community based AAIs and it can be offered as ”Identity Access Management as a Service” to those communities, which do not have or do not want to operate their own AAIs.

### Federated Cloud Computing

The EGI Federated Cloud platform has been expanded with new IaaS capabilities, provided by both native APIs (e.g. OpenStack) and standard based interfaces (e.g. OCCI interface). In particular, the following improvements were produced in EGI-Engage:

* User tools to simplify the management of IaaS resources in a distributed infrastructure, providing orchestration and infrastructure as code deployment of resources.
* Simpler and less-invasive components that facilitate new providers joining the federation and allow the creation of custom federations for specific communities.
* The VMOps Dashboard, a GUI for IaaS operations that hides the heterogeneity of the federation and is completely integrated with the AppDB Cloud MarketPlace for automatic discovery of the capabilities of each provider.

The evolution of the service increased usability and stability of the cloud resources, making them more suitable to a higher number of communities.

### Open Data Platform

The EGI Open Data Platform, built on OneData technology, is being developed to provide capabilities to publish, use and reuse openly accessible data (including, but not limited to, scientific data sets released into the public domain, publicly funded research papers and project deliverables, and software artifacts and demonstrators coming out of public research projects). Other functionality to be provided by the EGI Open Data Platform will include: policy-based publication, sharing and linking of open data sets; integration of open data access with community portals; data access across federations and support for data provenance.

The Open Data Platform can be deployed at multiple EGI data centers (or a private computing cloud site), connecting to various storage systems including Lustre, Amazon S3, Ceph, and NFS and other infrastructures.

### Thematic services are integrated

**Description**: <Gergely to add KER description>

**Category**: Services and software

Innovation level and motivation: VERY HIGH. Increase Federation service offer with specialized services which support the specific needs of scientific communities.

**Innovation Capacity and Motivation**: HIGH. Creation of thematic Data Competence Centers, providing integration of data analysis and visualization services, and real-time data services

**Exploitability level:** VERY HIGH

### Applications on Demand

The Application on Demand service (AoDs) is the EGI’s response to the requirements of researchers, scattered across Europe, without dedicated access to computational and storage resources, as well as other facilities needed to run applications. The service offers:

* Applications that are offered "as a service" through online graphical environments.
* Science Gateways and application-hosting frameworks where custom applications can be executed on EGI Cloud Compute and High-Throughput Compute services and offered as scalable, online services to researchers worldwide.
* Cloud and high-throughput compute resources suited for both compute/data intensive applications and for the hosting of scientific services.
* A network of consultants and supporters who can provide guidance on the use of the service.

The AoDs operates as an open and extensible ‘hub’ where any provider and e-Infrastructure user support teams can integrate and share applications and compute/data components.

### Improved EGI service portfolio

The EGI service portfolio has been improved with service definitions, and the creation of two portfolios and publication of service catalogues. The two portfolios are a direct reflection of what the EGI Foundation offers the participant organisations to enable the Federation itself, and what EGI offers collectively as a federation to the individual researchers and research communities**.**

## Policies, Processes and Procedures

### Update of the Strategy, Governance Evolution and Procurement

*Analysis of barriers and opportunities for cross-border procurement*. Report analysing opportunities and barriers for cross-border procurement of e-Infrastructure services, identifying best practices that could enable RIs or large research collaborations to acquire services to support their research agenda collectively.

*Governance Evolution*. Document assessing the suitability of the EGI governance model in relationship to the evolution of the strategy and the business models.

The document includes the definition of a set of recommendations for supporting the evolution of the governance model.

*Strategy Update*. Documents assessing the update of the EGI strategy and relative implementation plan. The Strategy documents sets out a concrete plan to help research communities develop and enhance their digital capabilities so that they can exploit today’s advanced IT infrastructures to undertake new and better research.

### Policy papers on the EOSC

EGI, together with other leading European initiatives EUDAT, LIBER, OpenAIRE and GÉANT, have shared their joint vision for the European Open Science Cloud for Research with eight elements of success for a concrete contribution to the Digital Single Market.

The joint policy publication “European Open Science Cloud for Research” sets out the partners’ strategic vision for the European Open Science Cloud’s organisation, sustainability and governance as a contribution towards the practical realisation of the EC’s vision.

The paper details eight elements for the success of the EOSC: open, publicly funded and governed, research-centric, comprehensive, diverse & distributed, interoperable, service-oriented and social.

### Integrated Management System and Certification

EGI has defined a system to plan, implement, monitor and continually improve all business processes under the responsibility of EGI Foundation.

This resulted in:

* The implementation of an Integrated Management System (IMS) which integrates all of the organization's systems and processes into one complete framework, enabling an organization to work as a single unit with unified objectives.
* The ISO 9001:2015 and the ISO/IEC 20000-1:2011 certifications. Those certifications show that the EGI management systems put in place to plan, implement, monitor and continually improve all processes regarding the EGI Service Catalogue follow the requirements of the ISO 9001:2015 and ISO/IEC 20000-1:2011 standards. The ISO 9001:2015 certificate covers all business processes including administration and finance, human resources, quality management, risk management, business relationships and continuous improvement.
* ISO/IEC 20000-1:2011 certification is a specialization of the previous designed to cover all IT-related services including compute, storage and data as well as internal services enabling Federation.
* Certification of Auditors from the EGI Foundation to support the adoption of ISO standards by the service providers.

### Security Policies

During the EGI-Engage project, it was necessary to revise the EGI security policies to address issues related to the evolution of EGI services and technology and to mitigate risks identified in recent security risk analyses. The outputs of this work are:

* [NEW] Security Policy for the Endorsement and Operation of Virtual Machine Images, and Policy on the Processing of Personal Data.
* [UPDATED] Acceptable Authentication Assurance.
* [UPDATED] maybe add another example here
* EGI Security Policy

To summarise, the result includes:

* The evolution of EGI security policies, procedures and best practices to mitigate the security risks arising from new trust models, new technology and new services deployed in EGI.
* IGTF trust developments enabling the EGI AAI platform that allows the integration of service providers with identity federations.
* **The d**efinition/update of a security policy framework to deal with the evolution of the EGI services and also to make them more general and reusable by other initiatives.

The full list of currently adopted security policies is always available on the EGI policies and procedures wiki**[[18]](#footnote-18)**.

1. <https://www.egi.eu/business/> [↑](#footnote-ref-1)
2. <https://home.cern/about/updates/2017/07/cern-data-centre-passes-200-petabyte-milestone> [↑](#footnote-ref-2)
3. <https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf#view=fit&pagemode=none> [↑](#footnote-ref-3)
4. <https://marketplace.egi.eu/42-applications-on-demand> [↑](#footnote-ref-4)
5. <http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:52016DC0178> [↑](#footnote-ref-5)
6. <http://www.oecd.org/sti/oecd-science-technology-and-innovation-outlook-25186167.htm> [↑](#footnote-ref-6)
7. <https://swc-osg-workshop.github.io/2015-03-03-iupui/novice/DHTC/01-IntroGrid.html> [↑](#footnote-ref-7)
8. <https://www.opensciencegrid.org/sbgrid-uses-the-osg-to-accelerate-structural-biology-research/> [↑](#footnote-ref-8)
9. <https://www.egi.eu/use-cases/scientific-applications-tools/> [↑](#footnote-ref-9)
10. <https://www.egi.eu/about/collaborations/> [↑](#footnote-ref-10)
11. <http://www.un.org/sustainabledevelopment/sustainable-development-goals/> [↑](#footnote-ref-11)
12. <https://www.egi.eu/internal-services/> [↑](#footnote-ref-12)
13. <https://www.egi.eu/services/> [↑](#footnote-ref-13)
14. <https://www.egi.eu/access-policy/> [↑](#footnote-ref-14)
15. <https://wiki.egi.eu/wiki/Standards> [↑](#footnote-ref-15)
16. <https://wiki.egi.eu/wiki/SPG:Documents> [↑](#footnote-ref-16)
17. <https://wiki.egi.eu/wiki/PROC09> [↑](#footnote-ref-17)
18. <https://wiki.egi.eu/wiki/Policies_and_Procedures> [↑](#footnote-ref-18)