

EGI-Engage: Impact & Results



Advanced Computing for Research
www.egi.eu

Purpose

The **EGI-Engage project** (full name: Engaging the Research Community towards an Open Science Commons) ran from March 2015 to August 2017, coordinated by EGI and co-funded by the European Union (EU) Horizon 2020 program under grant number 654142.

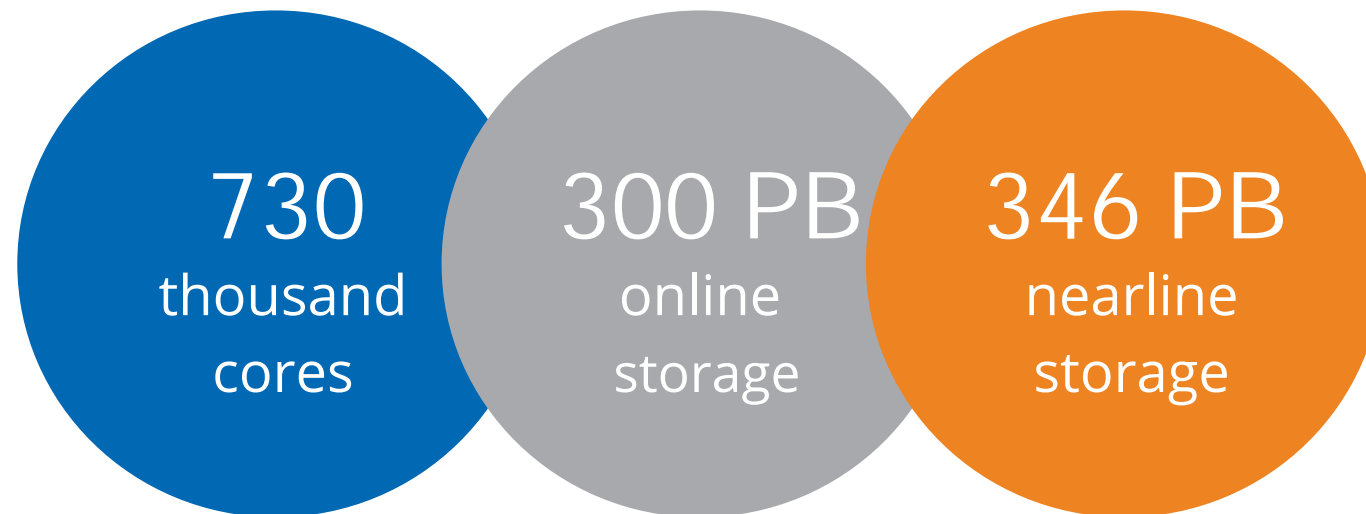
EGI-Engage had a mission to expand the capabilities of a backbone of federated services for compute, storage, data, communication, knowledge and expertise, complementing community-specific capabilities.

This report publication showcases the results of the project and their impact on science and society.

EGI-Engage in numbers:



EGI-Engage supported science at all scales



Building on the EGI service catalogue, EGI-Engage supported a wide range of scientific disciplines at all scales, from large research communities & Research Infrastructures to small research groups and individual researchers.

“ *You can see this increasing demand for distributed computing at every scale, from the theoretical chemist using **5 million core hours** a year, through to major collaborations like WeNMR in structural biology or the Large Hadron Collider, which bring together thousands of scientists and routinely transfer something like **50 petabytes of data** per month.”*

Tiziana Ferrari, EGI-Engage Project Coordinator.

Figures correct as of November 2017

From individual researchers...

Studying chemical reactions

Chemical reactions are at the core of everything in the Universe.

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 has submitted about 2.5 million **High-Throughput Compute** jobs for a total of 31 million CPU hours and published papers in **MNRAS** and **Chemical Physics Letters**.



31
million
core
hours



2.5
million
compute
jobs

Providers: compchem VO, supported by 17 data centres in FR, GR, IT, PL & ES.

Predicting the onset of epilepsy

Epilepsy affects about 2.4 million of people per year.

Massimo Rizzi and his colleagues at the Mario Negri Institute for Pharmacological Research researched the markers that predict the start of epilepsy before symptoms emerge.

By using **High-Throughput Compute** to perform their calculations, they saved years of research time. The results of the study are published in **Scientific Reports**.



200
thousand
compute
jobs

Providers: the biomed VO, supported by 60 data centres & the Italian Grid Infrastructure.

Detecting social media trends

Social networks nowadays are big data production engines.

Athena Vakali and her colleagues at the Aristotle University of Thessaloniki in Greece worked on a new model of detecting social media trends.

Vakali used **Cloud Compute** resources to help them run the experiments. They used about 48 CPU cores and 46 GB of available memory. The results are published in **Advances in Big Data**.



48
CPU
cores



46
GB
memory

Providers: GRNET & Okeanos, part of the EGI Federated Cloud.

to research infrastructures...

MoBrain: from molecule to brain

The **MoBrain Competence Center (CC)** has developed online portals for life scientists worldwide. In 2016, the Mobrain CC partnered with seven EGI data centres to secure **High-Throughput Compute** and **Online Storage** resources for their activities.

In total, the data centres offered around 75 million hours of computing time & over 50 TB storage capacity.

The MoBrain portals powered by the EGI are: HADDOCK, DisVis, AMBER, CS-Rosetta, FANTEN & PowerFit.

EMSO: large-scale, marine RI

The **EMSODEV project** developed a Data Management Platform (DMP) to set up a flexible and scalable data management service for a long-term and (near)-real-time monitoring.

EMSODEV developed the DMP on top of the **EGI Federated Cloud** with the EGI support on virtualisation, storage, networking and security.

The experimental prototype of the DMP is now deployed in the RECAS-cloud and fully integrated with EGI AAI services.

ESA: European Space Agency

Terradue is a SME tasked by ESA to lead the development of a cloud infrastructure that supports the Geohazards and Hydrology thematic exploitation platforms.

Terradue needed cloud resources to make this possible and to be able to handle massive data streams.

Seven **EGI cloud providers** from Italy, UK, Greece, Germany, Poland, Belgium and Spain committed the cloud resources necessary to make the project happen.

71
million
core
hours

50
TB
storage

340
CPU
cores

9
TB
storage

360
CPU
cores

800
GB
memory

More use cases are available on the EGI website: <https://www.egi.eu/use-cases/>

and large research collaborations.

CTA: Cherenkov Telescope Array

The **Cherenkov Telescope Array** (CTA) will be the world's leading gamma-ray public observatory.

CTA is using EGI **High-Throughput Compute** & **Online Storage** services to handle the computational demands during the project's first phase.

Between 2013-2016, CTA consumed:

- **360 million** CPU hours
- **11 Petabytes** of data transferred
- **2 Petabytes** currently in storage
- **11 million** computation jobs.

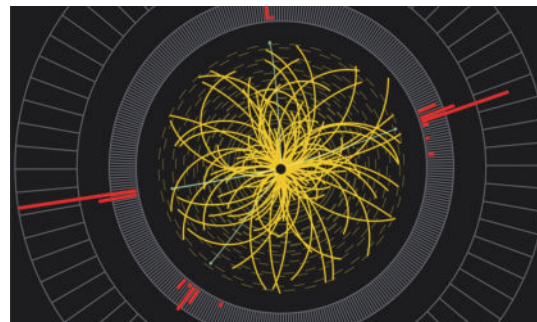


Worldwide LHC Computing Grid

WLCG is a global collaboration of more than 170 computing centres in 42 countries, linking up national and international grid infrastructures.

The collaboration between WLCG and what is now EGI spans over ten years old: WLCG has been involved in every step of the development of EGI and is the biggest consumer of **EGI resources**.

The four largest EGI Virtual Organisations are all LHC experiments: **ATLAS, ALICE, CMS and LHCb**.



LIGO/Virgo collaboration

The work on detecting gravitational waves by the **Virgo & LIGO Scientific Collaborations** won them the **Physics Nobel Prize** in 2017.

Virgo data is collected at the European Gravitational Observatory site, but its final repositories are the CCIN2P3 computing centre in Lyon & the INFN-CNAF centres in Bologna.

Wave analyses are run through EGI via the **Virgo Virtual Organisation** (VO), which consumed collectively **40 million CPU hours** in 2015 & 2016.



Adoption of computing & storage resources

The adoption of EGI services increased during EGI-Engage. Below are the trends observed during the project:

Research communities

+ **30% registered users**, increase driven by:

- Physics, Research Infrastructures
- SaaS operated on top of the EGI Federated Cloud



+ **40 collaborations**, including:

- 19 Research Infrastructures
- 31 RIs and e-Infrastructures integrated



+ **38 virtual organisations**,
in response to outreach activities
& the Competence Centre programme



+ **11 service level agreements** (SLAs):

- Peachnote (music platform)
- EMSOdev (ESFRI)



Business programme

EGI partnered with industry and SMEs to co-develop support solutions for their computing needs.

Use cases:

- 150 recorded
- 20 in progress
- 11 completed
- 4 memorandums of understanding (MoUs)

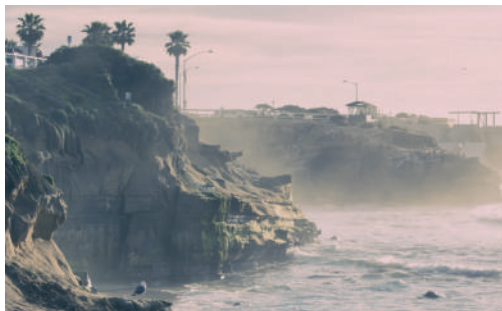


Selected business partners:

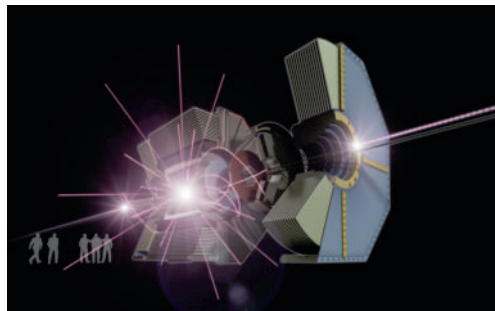


Increased availability of scientific data & efficient use of IT for research

During the [EGI-Engage project](#), the number of international research collaborations and infrastructures supported by the EGI Federation increased by **48%**. EGI supported data & analysis needs at all scales:



From:
OpenCoasts portal
2 TB per year

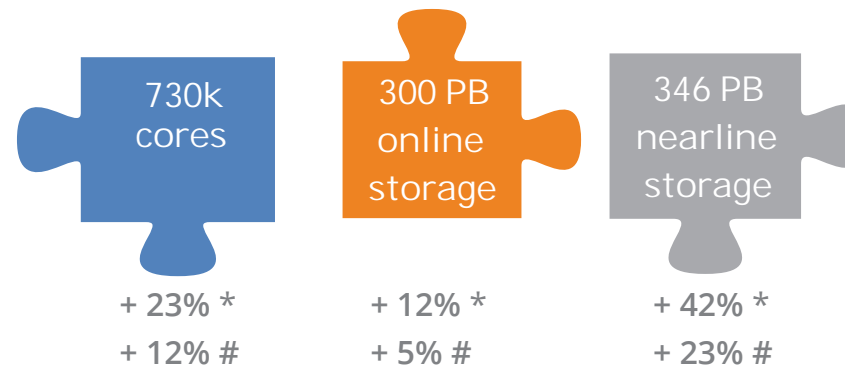


And:
Belle experiment
10s PB per year



To:
LHC
+ 200 PB archived data

Increase of compute capacity:



The compute capacity increased substantially during EGI-Engage. Today, more than **200 research collaborations** benefit from the EGI technical infrastructure.

*Project year 1
#Project year 2

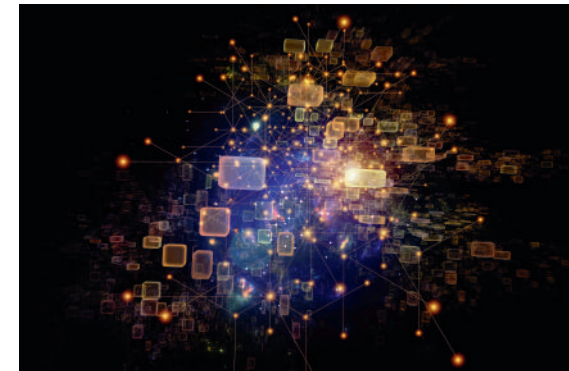
Contribution to cloud megatrends

In the area of cloud computing, the **EGI-Engage project** established a blueprint consisting of best practices to achieve interoperability across multiple cloud providers.

To date, the **EGI Federated Cloud** is the only existing publicly-funded distributed research cloud in Europe, offering on average 7,000,000 CPU hours per year to researchers from all disciplines. It is now made of 21 publicly funded clouds & one commercial cloud.

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

EGI-Engage also 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.



Adoption of FAIR principles

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**:

- **Findable**: the EGI Marketplace was designed & implemented during the project. The EGI internal service catalogue and external service catalogue were also defined.
- **Accessible**: accessibility was improved via federated identity provisioning (eduGAIN), and federated authentication and authorization via the Check-in service.
- **Interoperable**: EGI-Engage defined guidelines for compute and data management interoperability across multiple facilities and suppliers. This resulted in a community-defined standards roadmap.
- **Reusable**: the project produced security policies both for users & providers, including the general e-Infrastructure security policy.



Key Exploitable Results (1)

EGI Marketplace

The **Marketplace** service (in beta) is designed as an electronic market: it's a platform where services can be advertised and where customers can easily order & access them.

The Marketplace will also enhance visibility for resource and service providers, raising awareness of what they can provide as well as helping to promote cross-disciplinary research.



EGI Service Portfolio

The **EGI service portfolio** has been improved with service definitions & the creation of external and internal service catalogues.

The two catalogues are a reflection of what EGI is offering to the participant organisations to enable the federation itself, and what EGI is offering collectively as a federation to researchers and research communities.



Federated service management tools

Operational tools solve common problems with federating operations. The tools can support the creation of new service federations, or the extension of existing ones.

For example, the **Accounting** and **Monitoring** systems can be offered as services or used as an added value to market the EGI federation to new members.



Key Exploitable Results (2)

EGI Check-in service

The **EGI Check-in** service provides a reliable and interoperable AAI solution that can be used as a service for third parties. Check-in enables single sign-on to services through eduGAIN and other identity providers.

Users without institutional accounts can access services through social media or other external accounts, including Google, LinkedIn or ORCID.

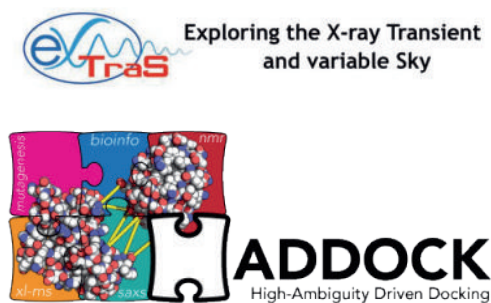


Thematic services

Thematic services are scientific applications, tools, environments integrated with EGI's e-Infrastructure services & typically exposed via web portals. They are not part of the EGI Service Catalogue but they rely on EGI services to run.

Thematic services are co-designed and co-developed with structured scientific communities.

Examples:



EGI Applications on Demand

The **Applications on Demand** service (in beta) gives access to online applications and application-hosting frameworks for compute-intensive data analysis.

This service targets individual researchers, research groups & early-stage research infrastructures, especially those with limited access to dedicated computing/storage resources.



Key Exploitable Results (3)

EGI Open Data Platform

The **EGI Open Data Platform** is a solution designed to make data discoverable and available in an easy way across all EGI resources.

The Open Data platform is founded on the OneData technology and can offer scalable data access and compute capabilities around scientific datasets for scientific groups at a large scale.



EGI Federated Cloud Computing

The **EGI Federated Cloud** is a IaaS-type cloud, made of academic private clouds & virtualised resources & is built on open standards.

During EGI-Engage, the Federated Cloud was expanded with new capabilities, now integrating commercial & public IaaS Cloud deployments & e-Infrastructures with the current production infrastructures.



IMS & Certification

EGI has defined a system to plan, implement & improve the business processes under the responsibility of the EGI Foundation, resulting in:

- the implementation of an **Integrated Management System** (IMS) which unifies all organisational processes into one framework.
- two **ISO certifications**: ISO 9001:2015 & ISO/IEC 20000-1:2011.



Key Exploitable Results (4)

Policy papers on the EOSC

EGI and other leading European initiatives have shared their joint vision for the European Open Science Cloud for Research with several elements of success that contribute to the Digital Single Market.

The publication "**European Open Science Cloud for Research**" sets out the partners' vision for the EOSC's governance & organisation.



EGI Security Policies

EGI has a full set of security policies that accommodate the operation of distributed infrastructures supporting international collaborations.

During the EGI-Engage project, these **security policies** were revised to address issues related to the evolution of EGI services and technology and to mitigate risks identified in recent security analyses.



Strategy, governance & procurement

The following progresses were made during EGI-Engage:

- Analysis of barriers and opportunities for cross-border procurement.
- Governance evolution: assessing the suitability of the EGI governance model in relationship to the evolution of the strategy & organisational business models.
- A **strategy update** for 2015-2020.



Looking to the future: from EGI-Engage to EOSC-hub

During the **EGI-Engage project**, EGI endorsed the principles of the EOSC and advocated 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.

Building on the achievements of EGI-Engage, the **EOSC-hub project** kicked-off in January 2018 with a mission to build the Hub: a central point for all European researchers and innovators to discover, access and use a broad spectrum of resources for advanced data-driven research.

The project is coordinated by the EGI Foundation and brings together 100 beneficiaries and linked third parties including research infrastructures, e-Infrastructure providers, SMEs and academic institutions.



Colophon

This publication was prepared by the EGI Foundation Communications Team.

We would like to thank all the participants of EGI-Engage for making the project happen.

For more technical information about the EGI-Engage project, please see: go.egi.eu/egi-engage

The content of this publication is correct as of November 2017.

Copyright: EGI-Engage Consortium, Creative Commons Attribution 4.0 International License.

The EGI-Engage project was co-funded by the European Union (EU) Horizon 2020 program under grant number 654142.



EGI-Engage's impact is observable at all scales of the European Research Area: from individual research, to large collaborations and businesses, promoting digital innovation and the implementation of the EOSC vision and FAIR principles.



Advanced Computing for Research
www.egi.eu