

# EGI support for Research Infrastructures

EGI: advanced computing for research

[www.egi.eu](http://www.egi.eu)



# Foreword

Dear Reader,

EGI traces its origin back to the early days of distributed computing and was initially designed to cater for the computational needs of the Large Hadron Collider.

Today, EGI operates the largest, federated e-infrastructure of the world and is committed to support all fields of science, from physics to astronomy, environmental sciences to humanities, life sciences, chemistry and many other disciplines.

On behalf of the EGI community, I am happy to present you this brochure: a selection of collaborations through which we support the setup and operation of European Research Infrastructures.

EGI now supports 31 Research Infrastructure and communities. This brochure features a small, but representative subset, completed during the EGI-Engage project. The use cases display the diversity of challenges that research facilities raise toward IT providers, and the richness of responses that the EGI community offers to tackle these challenges.

I hope the booklet will also serve as a catalogue of blueprints for emerging Research Infrastructures that are going to explore the power of e-infrastructures in the coming years.

Gergely Sipos  
User Community Support Manager  
EGI Foundation

*The collaborative work described in this publication was co-funded through the EGI-Engage project, supported by the European Union Horizon 2020 program, under grant number 654142.*

## Environmental sciences



page 4



page 5



page 6



page 7



page 8

## Humanities



page 9

## Physics & Astronomy



page 10



page 11

## Life & Biomedical sciences



page 12



page 13



page 14



page 15



# EGI: Advanced computing for research

EGI is a federation of almost 300 data and compute centres and 21 cloud providers united by a mission to support research activities, business and innovation with advanced computing services. EGI provides technical and human services, from distributed high-throughput computing and cloud computing, storage and data resources to consultancy, support and co-development.

The federation is governed by the EGI Council and coordinated by the EGI Foundation, with headquarters in Amsterdam, the Netherlands. Since its establishment in 2010, the EGI e-infrastructure has been delivering unprecedented data analysis capabilities to tens of thousands of researchers from over a hundred virtual research communities covering many scientific disciplines.

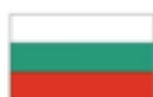
In March 2017, EGI became the first European-wide publicly-funded e-infrastructure to be certified to ISO standards, a sign of our dedication to continuously improve our service offering.



## Countries represented in the EGI Council



Belgium



Bulgaria



Croatia



Czech Republic



Estonia



Finland



France



Germany



Greece



Italy



Rep of Macedonia



Netherlands



Poland



Portugal



Romania



Slovakia



Slovenia



Spain



Sweden



Switzerland



Turkey



United Kingdom



CERN participates in the EGI Council as an international research organisation

# How EGI supports Research Infrastructures

EGI reaches out to Research Infrastructures (RIs) to discuss usage scenarios and assess requirements and preferences. The discussion generates an initial joint service configuration, development and delivery plan.

The document is then used to build a **Competence Centre** (CC): a distributed working group where national e-infrastructure initiatives, scientific institutes, technology and service providers join forces. The CC collectively refines and implements the collaboration workplan.

CC membership is based on joint interest and matching priorities of the RIs and members of the EGI Community.

The CCs collect and analyse requirements, integrate community-specific applications with EGI services, foster interoperability across infrastructures and evolve the services through a user-centric development model.

CCs also take ownership of the resulting services, configuring and providing them through Service Level Agreements to scientific user communities brought in by the RIs, and jointly promoting through events and channels.



## EGI Services featured in this publication



Cloud Compute



High-Throughput Compute



Cloud Container Compute



Online Storage



Data Transfer



Validated Software & Repository



Check-in



Workload Manager



Accounting



Attribute Management



Operational Tools



Configuration Database



Security Coordination



Helpdesk

**Full list of  
EGI services:**  
[go.egi.eu/ext](https://go.egi.eu/ext)  
[go.egi.eu/int](https://go.egi.eu/int)

# EISCAT\_3D: The next generation incoherent scatter radar system

## *A webportal for scientific data discovery and analysis*

### About EISCAT\_3D

EISCAT\_3D will be the world's leading facility for research into the upper atmosphere and near-Earth space or the geospace environment. Construction kicked off in September 2017, with the first stage of the radar system expected to become operational in 2021.

The EISCAT\_3D facility will be distributed across three sites in Scandinavia, each with a 70 metre circular area and 10,000 antennas.

### Challenge

The EISCAT\_3D radar opens up new opportunities for physicists to explore a variety of research fields, but it comes with significant challenges in handling large-scale experimental data. The incoming data rates of 99Tb/s need to be reduced to a few PB/year to be within funding limits.

### EGI & EISCAT\_3D Collaboration

EISCAT and EGI set up a Competence Centre (CC) in the context of the EGI-Engage project to provide researchers with data analysis tools to improve their scientific discovery opportunities. The team developed a web portal for researchers to discover, access and analyse the data generated by EISCAT\_3D.

The CC opted for the EGI Workload Manager tool, based on DIRAC interware. The service provides a web-based graphical interface and a command line interface to interact with data search and job management.

The system also facilitates the development of data models and modelling tools within the EISCAT\_3D community, and is a pilot of a central portal service for scientists to interact and compute with EISCAT data.

### Services used by EISCAT\_3D



Cloud  
Compute



High-Throughput  
Compute



Workload  
Manager

### Partners involved

EISCAT Scientific Association

EGI: CNRS, CSC, EGI Foundation, SNIC & UAB

**Website:** <http://eiscat3d.se>



# ICOS: Integrated Carbon Observation System

## Piloting applications on the EGI Federated Cloud

### About ICOS

ICOS is a pan-European research infrastructure for quantifying and understanding Europe's greenhouse gas (GHG) balance. Its mission is to collect high-quality observational data and to promote its use, e.g. to model GHG fluxes or to support verification of emission data. ICOS brings together 120+ measurement stations across the atmosphere, ecosystem and marine domains.

### Challenge

ICOS needed cloud resources to support the development of the Footprint Tool, a new web application for calculating the GHG footprint of a given location over time, based on meteorological analyses and emissions data. In addition, ICOS wants to deploy cloud resources for near real time processing of Ecosystem GHG Flux data.

ICOS was at the 1st 'Design your e-infrastructure' Workshop in 2016 and is also a test case for interoperability of EGI and EUDAT services.

### EGI & ICOS Collaboration

In the Footprint Tool case, data is stored on B2SAFE and taken to/from the ICOS Carbon Portal with B2STAGE. Users interact only with the ICOS Carbon Portal, which instantiates virtual machines (VMs) in the EGI Federated Cloud. Within EGI, data is saved in storage linked to the EGI Data Hub and accessible by all the ICOS VMs. These include a VM hosting a web interface handling model run requests and output visualisation, a set of worker node VMs, and one VM that manages data sharing across the worker nodes.

For the Ecosystem GHG Flux case, ICOS would build on a current pilot under development in the framework of ENVRIplus. This is currently using the gCube Data Analytics platform, combining resources from D4Science and EGI to orchestrate parallelised HTC computational processes.

### Services used by ICOS:



Cloud Compute



Cloud Container  
Compute



Online Storage

### Partners involved

The ICOS use case is still in a testing phase and is supported through the incubator VO, [fedcloud.egi.eu](https://fedcloud.egi.eu) and uses resources from the CESNET-MetaCloud cloud provider.

**Website:** <https://www.icos-ri.eu/>



# EMSO: European Multidisciplinary Seafloor and water-column Observatory

## *Access to cloud resources from EGI Federation providers*

### About EMSO

EMSO is a large-scale research infrastructure of seafloor & water-column observatories, set up to monitor long-term environmental processes and their interactions. The observatories measure many chemical and physical parameters, such as water temperature and acidity, or the ground/water interactions during earthquakes and tsunamis. The EMSODEV H2020 project is working on the data management systems.

### Challenge

Collecting and analysing ocean data on a day-to-day basis is crucial to monitoring environmental processes. To do this, EMSODEV needed to develop a Data Management Platform (DMP) to set up a flexible and scalable data management service for a long-term, high-resolution and (near)-real-time monitoring.

### EGI & EMSO Collaboration

EMSODEV expressed interest in evaluating the use of EGI Federated Cloud resources for hosting the DMP. EGI secured a pool of cloud resources provided by data centres in Italy, Portugal and Spain through a Service Level Agreement (SLA). In total, 4 cloud providers committed 9 Terabytes of storage capacity and about 340 virtual CPU cores.

EMSODEV developed the DMP platform 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-BARI cloud and fully integrated with the EGI Federated Cloud and the EGI Authentication and Authorisation services.

#### Services used by EMSO:



Cloud Compute



Online Storage

#### Partners involved

EMSO: INGV

EGI: CESGA, EGI Foundation, INFN-Padova, NCG-INGRID-PT, RECAS-BARI

**Website:** <http://www.emsodev.eu>



*Cloud resources to enable two Thematic Exploitation Platforms*

**About ESA**

The European Space Agency (ESA) is Europe’s gateway to space. Its mission is to shape the development of Europe’s space capability and ensure that investment continues to deliver benefits to the citizens of Europe and the world. ESA is an international organisation with 22 member states.

**Challenge**

Terradue is a company specialised in delivering cloud services for earth sciences and was tasked by ESA to lead the development of a cloud infrastructure to support the Geohazards and Hydrology thematic exploitation platforms. Terradue needed cloud resources to make this possible and to be able to handle massive and complex data streams.

**EGI & ESA Collaboration**

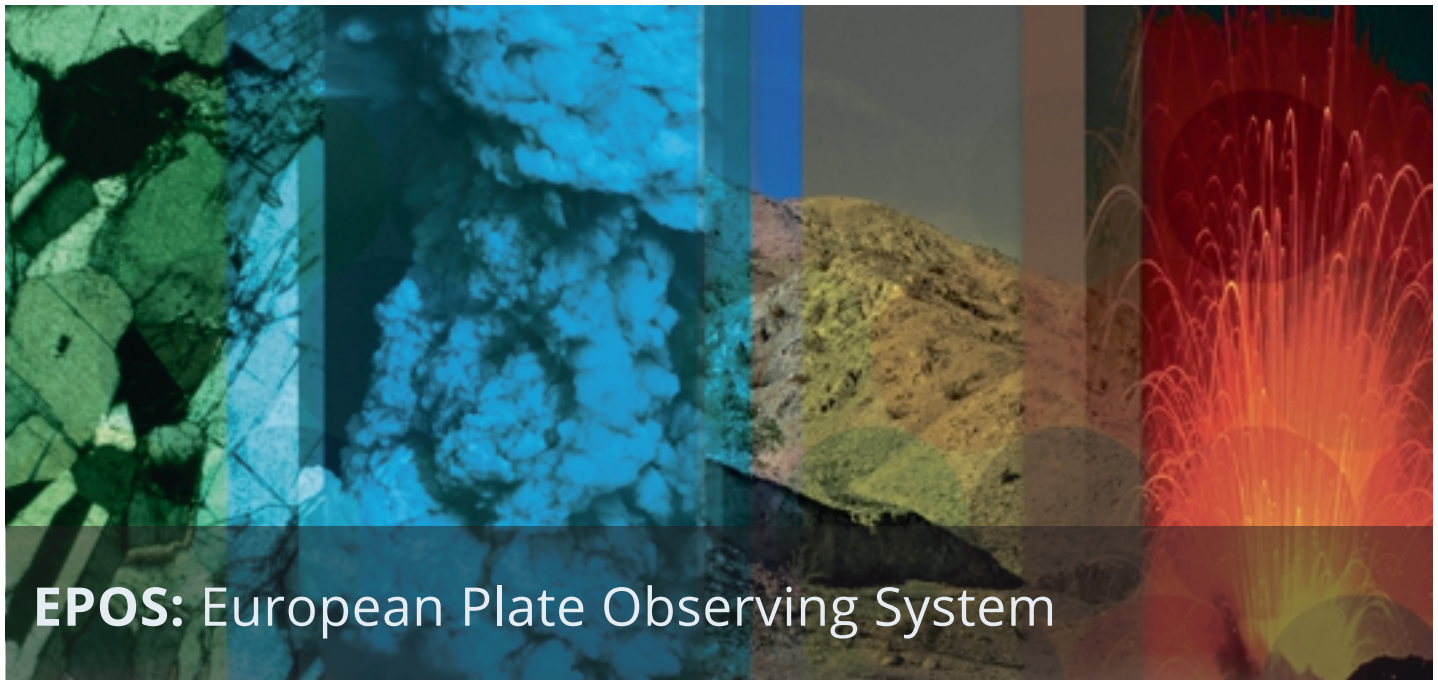
The two thematic exploitation platforms were connected to the EGI Federated Cloud to guarantee enough computational power for their use cases.

This task was successfully completed by developing an interface between Terradue Cloud Platform, one of the available interfaces of the EGI Federated Cloud.

Seven EGI cloud providers from Italy, UK, Greece, Germany, Poland, Belgium and Spain committed the cloud resources necessary for the project through Service and Operational Level Agreements.

<p><b>Services used by ESA / Terradue</b></p> <div><p>Cloud Compute</p><p>Online Storage</p></div>	<p><b>Partners involved</b></p> <p>ESA: Terradue</p> <p>EGI: 100%IT, BEgrid-BELNET, CESGA, CYFRONET-Cloud, EGI Foundation, GoeGrid, HG-09-Okeanos-Cloud, RECAS-BARI</p> <p><i>Website: <a href="https://www.terradue.com/">https://www.terradue.com/</a></i></p>
--	--





# EPOS: European Plate Observing System

## Online applications, big data and user identification

### About EPOS

EPOS is the ESFRI initiative for the solid Earth sciences. EPOS wants to establish a European infrastructure where datasets, metadata and resources can be searched and accessed across sites. This will create new e-science opportunities to monitor and better understand the dynamics of earthquakes, volcanic eruptions and other geological processes.

### Challenge

The understanding of solid Earth dynamics and tectonic processes relies on the analysis of seismological data, ground deformation from terrestrial and satellite observations, geological and petro-chemical studies and lab experiments. Setting up a RI to integrate and enable secure access 100s of Terabytes of data and use them for 3D simulation including data staging and visualisation is an incredible challenge.

### EGI & EPOS Collaboration

During the EGI-Engage project, EGI established a Competence Centre for EPOS to collect, analyse and compare Earth Science community needs with EGI technical offerings. The work of the EPOS Competence Centre covered three areas:

- (1) Establish a prototype service for the authentication and authorization of EPOS users to access federated e-infrastructure services;
- (2) Setup a cloud-based virtual research environment to conduct MISFIT earthquake simulations;
- (3) Enable the computation of Sentinel-1 data on a scalable cloud platform, in collaboration with the European Space Agency (see also page 7).

### Services used by EPOS:



Cloud Compute



Online Storage

### Partners involved

EPOS: KNMI, INGV

EGI: CNRS, CYFRONET, EGI Foundation, Fraunhofer SCAI, GRNET

**Website:** <https://www.epos-ip.org/>



# DARIAH: Digital Research Infrastructure for Arts and Humanities

## *A science gateway to host applications and access to cloud resources*

### About DARIAH

DARIAH EU is a pan-European infrastructure for arts and humanities scholars working with computational tools. Research in arts and humanities often involves huge data sets that can benefit from digital research methods. The goal is to provide tools and services to enhance innovation and generate new knowledge.

### Challenge

DARIAH EU members required a way to provide transparent access to cloud storage resources in order to organise, store and process their datasets and to improve the resilience of the data.

They also wanted to investigate ways to make digital applications and databases available to arts and humanities scholars.

### EGI & DARIAH collaboration

DARIAH EU and EGI set up a joint team as a Competence Centre (CC) in the context of the EGI-Engage project. The CC started by identifying pilot applications, for example: a Simple Semantic Search Engine (SSE) to allow users to search in the Knowledge Base, or the DBO@Cloud, a cloud-based repository with a centuries-old collection of Bavarian dialects.

The CC developed the DARIAH Science Gateway to host the applications and provide access to file transfer and cloud compute services (available at: [go.egi.eu/dariahSG](http://go.egi.eu/dariahSG)). The DARIAH Science Gateway is now in production and relies on the resources committed by two EGI Federated Cloud data centres: INFN-Bari and INFN-Catania.

### Services used by DARIAH EU:



Cloud Compute



Online Storage

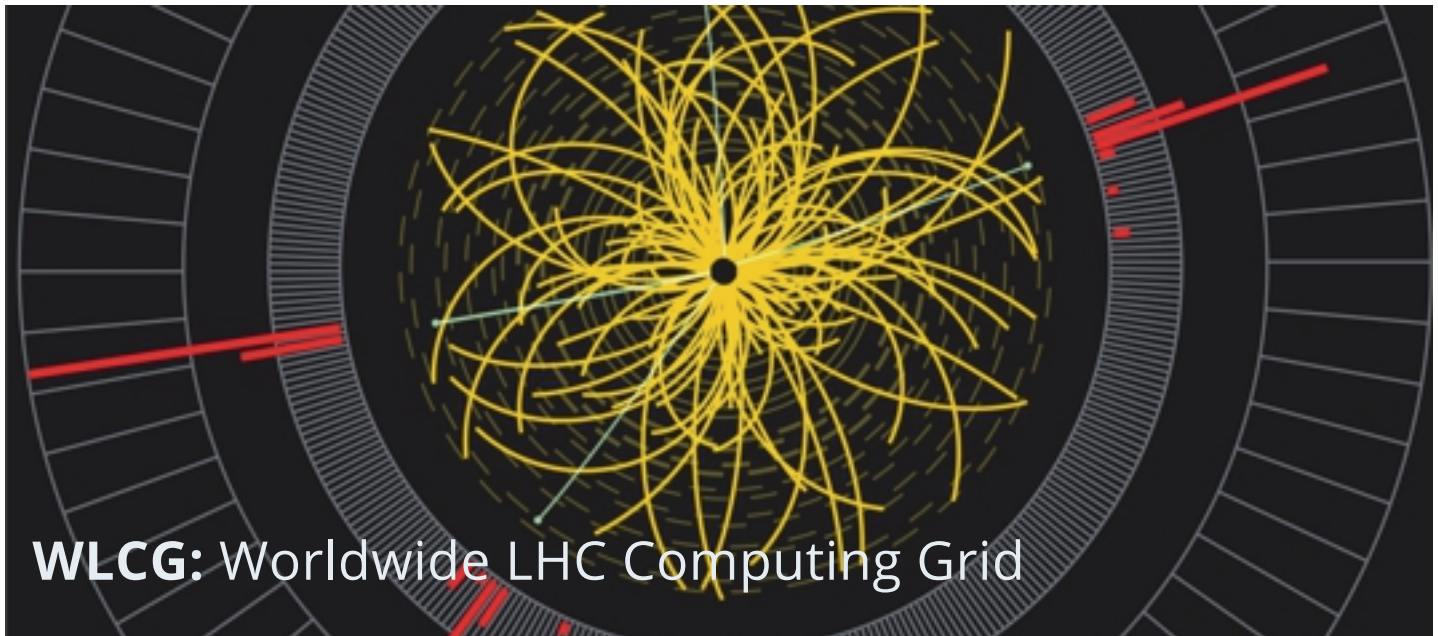
### Partners involved

DARIAH: AAS, DANS, RBI

EGI: EGI Foundation, GWDG, INFN-Bari, INFN-Catania, MTA-SZTAKI

**Website:** <http://www.dariah.eu/>





*Storing, sharing and analysing an unprecedented scale of LHC data*

About WLCG

WLCG is a global collaboration of more than 170 computing centres in 42 countries, linking up national and international grid infrastructures. The mission is to provide global computing resources to store, distribute and analyse the data generated by the High Energy Physics experiments hosted by the Large Hadron Collider (LHC) at CERN.

Challenge

The observations collected by the LHC detectors generate an enormous amount of data that needs to be stored, distributed across hundreds of scientific institutions worldwide and then analysed by thousands of scientists. This regularly involves data transfers of +80 Petabytes / month and ~700 million computing jobs / year, which consume about 5 billion CPU hours.

EGI & WLCG

The services provided by the EGI Federation are essential to run a global exascale, distributed computing infrastructure such as WLCG.

The collaboration between WLCG and what is now EGI is over 10 years old: WLCG has been involved in every step of the development of EGI and is the biggest consumer of EGI compute resources. The four largest EGI Virtual Organisations are all LHC experiments: ATLAS, ALICE, CMS and LHCb.

The computing is in itself an achievement, but the scientific results also endorse the approach. WLCG managed the computational challenges that led, for example, to the discovery of the Higgs Boson in 2012 by ATLAS and CMS, or the LHCb discovery of ‘charmonium particles’ earlier in 2017.

Services used by WLCG:

 HTC	 Data Transfer	 Software Repository	 Operational Tools	 Configuration Database
 Online Storage	 Accounting	 Attribute Management	 Helpdesk	 Security Coordination

Resource providers: EGI Federated Data Centres | Website: <http://wlcg.web.cern.ch/>



# CTA: Cherenkov Telescope Array

## HTC and storage solutions for CTA's computational challenges

### About CTA

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

### Challenge

CTA is a team of 1350 scientists and engineers from 32 countries and will be a distributed array of more than 100 telescopes built in Spain and the European Space Observatory site in Chile.

This collaborative model comes with its challenges: transferring data from telescopes to scientists worldwide, archiving and finding the processing power for data reduction and large-scale Monte Carlo simulations.

### EGI & CTA Collaboration

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

The compute and storage services are provided via the CTA Virtual Organisation, one of the most active user groups of EGI. Since 2012 the consortium has used EGI services to guide the choice of the best sites to host CTA telescopes in the North and in the South hemispheres.

CTA EGI usage (2013-2016)

- > 360 million HS06 CPU hours
- > 11 Petabytes of data transferred
- > 2 Petabytes currently in storage
- > 11 million computation jobs

### Services used by CTA:



High-Throughput  
Compute



Online Storage

### Resource providers

CC-IN2P3, CYFRONET-LCG2, DESY-ZN, GRIF, IN2P3-LAPP, INFN-T1

With additional resources provided by the National e-Infrastructures of the Czech Republic, France, Germany, Italy, Poland and Spain.

**Website:** <https://www.cta-observatory.org/>



## LifeWatch: European Infrastructure for Biodiversity and Ecosystem Research

### *Computing resources, access to data, tools and virtual labs*

#### About LifeWatch

LifeWatch is a Research Infrastructure set up to support the fields of ecosystems research and biodiversity by equipping scientists with access to data, analytical tools and state-of-the-art virtual laboratories.

#### Challenge

Developing the LifeWatch objectives requires a close collaboration between the scientific groups leading the development of digital tools and IT experts that can advise on the best technologies to carry out the vision.

These tools will allow for new avenues of research, but they also come with a high demand for computational resources.

#### Collaboration

LifeWatch partnered with EGI through an EGI-Engage Competence Centre (CC) to:

- > deploy the basic tools required to support data management, data processing and modelling for ecological observatories,
- > evaluate the services required to support workflows for the deployment of virtual labs
- > and support the participation of citizens in LifeWatch's observation records, e.g. in uploading and processing of images.

The work led to a total of seven LifeWatch services for data management and modelling currently in production using the computational resources of the EGI Federated Cloud.

#### Services used by LifeWatch:



Cloud Compute

#### Partners involved

LifeWatch: CESGA, ICETA, INRA, UNIZAR, UPVLC, VLIZ

EGI: CSIC, EGI Foundation, INFN, LIP

**Website:** <http://www.lifewatch.eu/>



## ELIXIR: bringing together life science resources from across Europe

### *A federated cloud for scientific applications*

#### About ELIXIR

ELIXIR unites Europe's leading life science organisations in managing and safeguarding the increasing volume of data being generated by publicly funded research. It coordinates, integrates and sustains bioinformatics resources – such as databases, compute services, applications and training – across its member states and enables users in academia and industry to access what is vital for their research.

#### Challenge

The ELIXIR community initiated the development of the reference architecture for biomedical applications, called the 'ELIXIR Compute Platform'. The platform is envisaged as a federation of compute and storage resources, operated according to community-agreed principles and with the use of centrally provided user access management services.

#### EGI & ELIXIR Collaboration

ELIXIR was one of the Competence Centres (CCs) in the EGI-Engage project. The CC had two main activities:

- > Established a cloud infrastructure: the CC brought together four providers (EMBL-EBI, CNRS, GRNET, CESNET) into a federated cloud resource pool accessible for biomedical researchers with ELIXIR user IDs. The team simplified the way to federate clouds, performed a compatibility study with a US-based cloud system (JetStream) and integrated the Terraform orchestrator technology to simplify porting of scientific applications.
- > Building on this cloud, the CC set up a set of scientific demonstrators to pilot its use. They were: META-Pipe (a metagenomics pipeline), cBioPortal, Marine Metagenomics, PhenoMeNaI and Insyght Comparative Genomics.

#### Services used by ELIXIR:



Cloud Compute



Operational  
Tools



Check-in

#### Partners involved

The ELIXIR Competence Centre:

CESNET, CSC, CNRS, EGI Foundation, EMBL-EBI, GRNET, SURFsara, University of Indiana

**Website:** <https://www.elixir-europe.org/>





## MoBrain: Translational Research from Molecule to Brain

### *Compute resources to power online portals for biomedical researchers*

#### About MoBrain

The MoBrain Competence Center (CC) has developed online portals for life scientists worldwide, by integrating structural biology and medical imaging services and data.

The MoBrain collaboration relies on the existing expertise available within the WeNMR and N4U projects and technology providers.

#### Challenge

MoBrain provides biomedical scientists with a portfolio of portals for the study of molecular forces and biomolecular interactions to improve drug design, treatments and diagnostics.

To do this, the MoBrain community needs advanced computational power to sustain the work of their portals.

#### EGI & MoBrain Collaboration

In 2016, Mobrain signed an SLA (service level agreement) with seven EGI data centres allowing the CC to use High-Throughput Computing and Online Storage services needed to develop portals for life and brain scientists worldwide. In total, the providers pledged around 75 million hours of computing time and 50 Terabytes of storage capacity.

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

To give an example, HADDOCK has so far processed more than 178,000 submissions from over 9,900 scientists, which translates into about 9 million High-Throughput Compute jobs per year on the EGI infrastructure.

#### Services used by MoBrain:



High-Throughput  
Compute



Online Storage

#### Partners involved

MoBrain: University of Utrecht, CIRMMP-Florence, INFN-Padova and the consortium partners

EGI: CESNET-MetaCloud, EGI Foundation, INFN-Padova, NCG-INGRID-PT, NIKHEF, RAL-LCG2, SURFsara, TW-NCHC

**Website:** <http://go.egi.eu/MoBrain>



## ERIC: European Research Initiative on chronic lymphocytic leukemia

### *A dedicated VRE to facilitate open access to health related information*

#### About ERIC

ERIC is a European organisation dedicated to improving the outcome of patients with chronic lymphocytic leukemia (CLL) and related diseases. As of November 2017, the initiative counts 907 members from 64 countries.

#### Challenge

The majority of published research is either hidden behind subscription walls or lost in the fog of nondescript information.

The ERIC initiative wants to address this issue in the field of chronic lymphocytic leukemia by implementing a hub of accurate CLL-specific information within a federated environment and incorporating both data and computational services.

#### EGI & ERIC Collaboration

ERIC was selected as a use case of the second 'Design your e-infrastructure' Workshop in 2016. During the event, participants identified the need of a Virtual Research Environment (VRE) for data repository (patient and clinical trial data) and the computational services to support it.

EGI is supporting this initiative and facilitates the allocation of cloud resources to integrate the CLL VRE and its applications into the EGI Cloud services. The VRE is made of a portal, a database of CLL datasets, analysis applications and a Galaxy workflow system. EGI and ERIC have completed the migration of the portal onto the EGI Federated Cloud. The portal has now been extended to support federated authentication mechanisms using the OIDC protocol.

#### Services used by ERIC:



Cloud Compute



Online Storage

#### Partners involved

ERIC

EGI: BEgrid-BELNET, EGI Foundation

*Website: <http://www.ericll.org/>*

## November 2017

This publication was written and prepared by the Communications and User Community Support Teams of the EGI Foundation. We would like to thank:

- > all the researchers and teams of the Research Infrastructures and projects for their support.
- > all the EGI service & resource providers for making the collaborations possible.

The content of this brochure is correct to the best of our knowledge as of November 2017. More information about EGI's support for Research Infrastructures is available on our website:

<https://www.egi.eu/use-cases/research-infrastructures/>

Design: EGI Communications Team

Cover design concept: Tom Jansen

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

The EGI-Engage project was co-funded for 30 months by the European Union (EU) Horizon 2020 program under grant number 654142, as a collaborative effort involving more than 70 institutions in over 30 countries.

The butterfly image in the cover: © Derek Ramsey / derekramsey.com (used with permission, via Wikimedia Commons). The other Wikimedia Commons images used in this publication are credited in the caption. A special thanks is in order to all Wikicommons photographers.







EGI Foundation | Science Park 140, 1098 XG Amsterdam | The Netherlands

+31 (0)20 89 32 007 | [www.egi.eu](http://www.egi.eu)