Inspired

ISSUE 29
DECEMBER 2017

news from the EGI community

TOP STORIES

MORE

Advanced Computing for Research

www.egi.eu

Welcome to issue 29!

We hope you enjoy reading this last edition of 2017 and we want to wish everyone Happy Holidays and a wonderful 2018 ahead!

Your feedback and suggestions are always welcome!

Send an email to Sara & Iulia at:

press@egi.eu

So long, and thanks for all the fish

Eight years with EGI: Peter Solagna recounts the highlights

The EGI infrastructure has been established in February 2010, as a long term sustainable evolution of the EGEE projects.

At the same time, the EGI Foundation was created as a governing body to support long term sustainability and to coordinate the infrastructure's activities. Today, EGI federates more than 20 cloud providers and hundreds of data centres, spread across Europe and worldwide.

I had the pleasure to work in the operations team of the EGI Foundation from almost the beginning, where I had a privileged observation point to appreciate the evolution of the production infrastructure of EGI and its operations activities.

From the beginning, EGI has been among the largest infrastructures supporting research in Europe and worldwide.

Some numbers

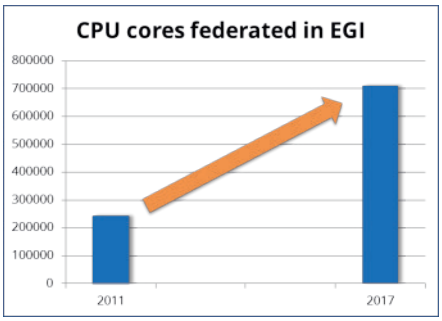
The usage of the EGI infrastructure has traditionally been CPU-intensive and the number of CPU cores available in the EGI production infrastructure is a good metric to track its size.

One of the first times I looked into the number of cores in EGI was back in in 2011: the total was about 240,000 cores.

A few months ago I calculated about 710,000 cores, which means that the CPU capacity of EGI has almost tripled in only six years of existence.

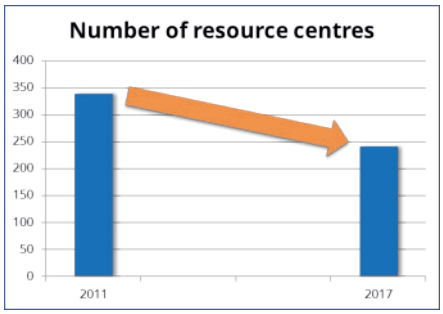
This upturn is responding to the increase of CPU demand of new and existing research communities led by the LHC experiments. The CPU tasks accounted by the EGI infrastructure, are currently using ~80% of the available capacity.

Another interesting metric is the number of data centres, or sites



as the old-fashioned operations people (like me) call them. The total number of data centres has decreased in the past years, going from ~330 to about ~240 today.

The backbone of the production infrastructure is solid and it is here to stay, but in the past years some changes of plans in the national infrastructures or in the local institutions caused the decommissioning of some of the smaller data centres. On the other hand, the coordi-nation work of the national operations centres improved and the total capacity has never stopped to increase.



Services

EGI's approach to communities and to our members has greatly evolved during the past years, thanks to the work of all my colleagues. The EGI infrastructure is now more service and user-oriented than ever before. The new service catalogue is much richer, better-structured and fulfills the requirements of new research communities.

In parallel, the services that are offered to the members of the federation have been consolidated with a plan for a long-term, fee-based support. These services include for example service monitoring, accounting, security coordination and helpdesk.

The operational tools in this portfolio have evolved as well to be more reliable and flexible. For example, monitoring is now centrally provided by ARGO to better support new monitoring probes and be able to react to requirements quickly; the Operations Portal now allows browsing the information system in a GUI, and the messaging network has a new HTTP API that can be easily re-used by new use cases. Looking at the current service portfolio, it can be noted how the burst of new services involved to some extent the federated cloud. And it has been a great achievement of EGI to design, build and bring to production a new set of services, based on new technologies that brought new capabilities.

A lot of hard work has been spent from 2014 onward in tuning the new services to the operational framework of EGI, and the other way around, in adapting the operations to the new requirements. Security coordination evolved the security operations to deal to the "new dimension" added by the cloud, and so did the operational tools and support activities.

Looking to the future

What are the future challenges for the EGI Operations?

The project EOSC-hub will start in January 2018, and it is the first EOSC-related project that focuses on the provisioning of production services.

EGI already has a solid experience in providing production services, but the project brings together both EGI providers and external ones.

The challenge will be to integrate the various best practices implemented by the service providers and find a lowest common denominator to build what will be the EOSC operational framework.

The biggest contribution of EGI Operations to the future European Open Science Cloud is our combined experience (EGI Foundation and 38 operations centres) in building production operations.

It has been a pleasure to work together and to see how EGI evolved during these years, and also an honour to be able to give my small contribution to this process.

While leaving my position at the EGI Foundation, I would really like to thank everybody who contributed at every level to the EGI Operations during these years, and wish all the best for the exciting times ahead for EGI!

More information

Peter Solagna was the EGI Foundation Senior Operations Manager until November 2017, when he departed for multicoloured pastures.

The EGI Foundation team wishes him all the best!

LIGO/Virgo: the search for gravitational waves

Sara Coelho on the computing challenges of the collaboration and how EGI is supporting their groundbreaking work

Gravitational waves are tiny ripples of energy that propagate through the Universe. Researchers believe that gravitational waves are emitted by spinning stars in binary systems, black holes or massive stellar explosions. Albert Einstein predicted their existence in his general theory of relativity.

Almost 100 years after, the Virgo and LIGO Scientific Collaborations announced the first confirmed observation of gravitational waves. 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, followed by two more in August.

The observations paved the way to the Physics Nobel Prize awarded in 2017 to Rainer Weiss, Kip Thorne and Barry Barish for their role in the detection of gravitational waves.

But the detection of gravitational waves is not a three man job.

The LIGO Scientific Collaboration is a group of more than 1000 scientists from universities in the US and in other countries. The two LIGO detectors are located in Hanford and Livingston in the US.

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 the European Gravitational Observatory near Pisa.

The two collaborations share scientific data and computing resources, technologies, computing teams and publications.

Looking for the ripple effect

Gravitational wave detectors are based on the concept of a Michelson interferometer. The system takes in the infrared light of a laser and splits it in two beams that are injected in two long arms placed at 90 degrees of one another. The laser beams are recombined at the end of the arms and reflected toward a photo-detector. If a gravitational wave passes through, it generates a small interference that can be spotted by the detector.

And "small" in this context means really-really small: the amplitude of the interference is in the order of 10-20m, the size of what is left of a proton if you divide it into ten million parts. Just the fact that it is possible to detect such a difference is a massive technical achievement.

What happens to the data?

The two LIGO detectors send their raw and reduced data to a central data repository at Caltech, and store a local copy for redundancy. From Caltech, the reduced data is distributed across the LIGO Data Grid, a network of large dedicated computing clusters run by LIGO, and also the Condor Clusters federated in Open Science Grid- one of the major e-Infrastructures collaborating with EGI - and HPC resources from XSEDE.

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

The lion share of the data analysis is performed by dedicated LIGO Data Grid clusters. Parts of the analyses are submitted as computational grid jobs both in the US (to Open Science Grid) and in Europe (to EGI). Continuous wave analyses are run thorough EGI via the Virgo Virtual Organisation (VO) mainly at CNAF.

The Virgo VO consumed collectively 40 Million CPU hours in 2015 and 2016.

With thanks to:

Michele Punturo, Virgo Data Analysis Software Computing Coordinator, and to **Peter Couvares**, LIGO Data Analysis Computing Manager, for their help with writing this article.

How EGI supports Research Infrastructures

Gergely Sipos summarises our accomplishments in working with RIs

Research Infrastructures (RIs) and research collaborations are vital partners of the EGI community.

They are the main adopters of our services and they operate community cloud/storage/HTC and data systems. They also contribute with thematic services such as scientific portals and tools, made scalable and accessible through the use of EGI services.

EGI is currently working with over 30 research infrastructures from all fields of science: from physics to astronomy, environmental sciences to humanities, life sciences, chemistry and many other disciplines.

Here is an overview of our work and achievements with research infrastructures during the EGI-Engage project:

Environmental sciences

EMSO is a research infrastructure of seafloor & water-column observatories, set up to monitor environmental processes and their interactions. EMSO uses EGI Cloud Compute resources to help them scale-up their data management systems.

The European Space Agency (ESA) is Europe's gateway to space. ESA wanted to develop a cloud infrastructure to support the Geohazards and Hydrology thematic exploitation platforms (TEPs). The two TEPs were integrated within the EGI Federated Cloud to give them enough computational power for their work.

EISCAT_3D will be the world's leading facility for research of the upper atmosphere and the geospace environment. EGI & CNRS partnered with EISCAT to develop a web portal for researchers to discover and analyse the data generated by the EISCAT_3D radar.

EPOS is the ESFRI initiative for the solid Earth sciences. During the EGI-Engage project, EGI supported EPOS in collecting, analysing and comparing Earth Science community needs with EGI technical offerings.

ICOS is a pan-European research infrastructure for quantifying and understanding Europe's greenhouse gas (GHG) balance. EGI is currently offering cloud resources to support the ICOS Carbon Portal and the development of a GHG-footprint calculation tool.

Life & Biomedical Sciences

ELIXIR unites Europe's leading life science organisations in managing the volume of data generated by publicly-funded research. EGI supported ELIXIR in establishing a cloud infrastructure to support life science use cases and to setup scientific demonstrators to pilot its use.

The MoBrain Competence Center (CC) has developed online portals for life scientists worldwide, by integrating structural biology and medical imaging services and data. Mobrain collaborated with seven EGI data centres allowing the CC to use High-Throughput Computing and Online Storage services needed to develop their portals.

ERIC is a European organisation dedicated to improving the outcome of patients with chronic lymphocytic leukemia (CLL) and related diseases. EGI supports ERIC with cloud and storage resources in the development of a VRE for data repository (patient and clinical trial data).

LifeWatch is European infrastructure for biodiversity and ecosystem research. LifeWatch's work with EGI led to a total of seven LifeWatch services for data management and modelling that use the computational resources of the EGI Federated Cloud.

Psysics & Astronomy

The Cherenkov Telescope Array (CTA) will be the world's leading gamma-ray public observatory. CTA is using EGI's High-Throughput Compute and Online Storage services to handle computational demands during the project's preparatory phase.

The CTA Virtual Organisation is one of the most active users of EGI's HTC services.

WLCG is a global collaboration of over 170 computing centres with a 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. WLCG is the biggest consumer of EGI compute resources. The four largest EGI Virtual Organisations are all LHC experiments: ATLAS, ALICE, CMS and LHCb.

Arts & Humanities

DARIAH EU is a pan-European infrastructure for arts and humanities scholars working with computational tools. DARIAH partnered with EGI to make databases and digital applications available to scholars and provide transparent access to cloud resources via the DARIAH Science Gateway.

EGI is dedicated to supporting both established and newly emerging RIs. The next phase of work will start with the EOSC-hub project in January 2018.

More information

Gergely Sipos is the EGI Foundation Customer and Technical Outreach Manager

The OpenAIRE datathon

The OpenAIRE datathon started on 30 November 2017 and will take place until the end of February 2018.

The purpose of the datathon is to encourage developers and data scientists to analyse the OpenAIRE Information Space and improve its user base and third-party services.

The OpenAIRE information space consists of a scholarly communication graph inter-linking publications, datasets, software, research organi-sations, funders, and projects. The graph is the result of harvesting and harmonising metadata from about 3000 data providers and counts around 60 million objects.

The graph is accessible via APIs and a web portal and is used to offer research impacts statistics, access trends, and discovery of interlinked scholarly products.

The datathon encourages teams of computer scientists, data scientists and experts from other fields to join the challenge of studying and analysing the OpenAIRE graph to enhance its discovery and statistical capabilities.

Four main OpenAIRE datasets will be made available as Linked Open Data, Scholix exchange format (JSON scholarly graph representation), XML collections, and full-text collections.

Topics:

- Enabling multi-disciplinary or discipline-specific discovery and stats functionality.
- Novel techniques to enable measurement of scientific impact.
- Innovative techniques to measure scientific impact.
- Enabling reproducibility.
- De-duplication of the information space.

Prizes: an Apple iPad Pro and a featuring on the OpenAIRE web site!

More information

The datathon will last for 3 months, from 30 November to 28 February 2018. Registration is open until 15 January 2018!

Find all the details and how to register: https://datathon.openaire.eu

AEGIS group endorses AARC guidelines on exchange of membership information

Licia Florio writes about the group's first achievement

The AARC Engagement Group for Infrastructures (AEGIS) brings together representatives from research and e-infrastructures, operators of AAI services and the AARC team to bridge communication gaps and make the most of common synergies.

The group was set up during the summer and has recently endorsed the AARC "guidelines on expressing group membership and role information". By agreeing to these guidelines, AEGIS has taken its first step in providing practical support towards the wider uptake of AARC's interoperable federated access solutions.

The current members of AEGIS who endorsed the guidelines are from five e-infrastructures (EGI, EUDAT, GÉANT, PRACE and XSEDE) and two domain-specific research infrastructures (ELIXIR and DARIAH).

Why do we need guidelines for group membership?

Information about group membership is commonly used by Service Providers (SP) to authorise user access to protected resources. Apart from the group information that is managed by the user's home Identity Provider (IdP), research communities usually operate their own group managment services. Such services often act as Attribute Authorities (AA) , maintaining additional information about the users, including VO membership, group membership within VOs, as well as user roles. It is therefore necessary that all involved SPs and IdPs/AAs can interpret this information in a uniform way.

Specifically, the following challenges need to be addressed:

Standardising the way group membership information is expressed:

- Syntactically: uniform formatting; for example, representing group membership as URNs within a specific namespace and a set of rules for the NSS portion

- Semantically: common representation of equivalent concepts; for instance, "admin" and "manager" should be communicated to end SPs as "manager"

- Indicating the entity that is authoritative for each piece of group membership information

- Expressing VO membership and role information

- Supporting group hierarchies in group membership information

The guidelines document endorsed by AEGIS agrees on a common way to exchange group information across different infrastructures, ensuring that everyone uses the same concepts and formats.

The recommendations were defined based on experiences from multiple parties in the AARC project and have subsequently been discussed and tested through the Service Activity 1 Pilots (SA1) attribute management pilot [AARC-SA1-AMP]. Furthermore, it should be noted that a group membership representation scheme following these recommendations has already been adopted to enable cross-infrastructure exchange of group information between the EGI and the ELIXIR AAI.

Bi-directional channel

AEGIS provides a bi-directional channel so that AARC and the research- and e-infrastructures communities can advise each other on developments and implementation aspects of the project. AEGIS in practice provides a mechanism to ensure that AARC results are fit for purpose and are known to those that will need to deploy them. AEGIS helps participating infrastructures to understand the importance of adopting AARC frameworks, and helps to uncover issues that may otherwise emerge during the deployment phase.

More information

Licia Florio is the coordinator of the AARC2 project.

AARC website: https://aarc-project.eu/

The EGI Federation was recognised as a European Innovation Space

Roberta Piscitelli on the recent i-Space award from the Big Data Value Association

The European Innovation Spaces (or i-Spaces) are trusted data incubators with a mission to accelerate the uptake of data-driven innovation in commercial and non-profit sectors. i-Spaces enable stakeholders to develop new businesses, facilitated by advanced Big Data Value (BDV) technologies, applications, and business models.

The foundation of i-Spaces is an existing infrastructure that can be based on a geographic, sectoral or company-ground or a combination of them.

What are i-Spaces?

- Bring technology and application developments together and catering for the development of skills, competence, and best practices.

- Ensure that data is at the centre of Big Data Value activities. i-Spaces make accessible data assets based on industrial, private and open data sources.

- Serve as incubators for testing and benchmarking of technologies, applications, and business models.

- Develop skills and sharing of best practices linking with other existing initiatives at European and national-levels.

- Foster data-driven communities and accelerate value creation.

I-Spaces contribute to community building and act as incubators of data-driven innovation.

BDVA i-Spaces ensure that innovations and research on BDV technologies and novel BDV applications can be quickly tested, piloted and exploited.

The main goal is to develop skills and share best practices, to expand new business models and evaluate societal impact. In particular, at societal level, i-Spaces support the adoption of digitalisation in SMEs, they bring services for local governments and they provide digital solutions for policy development.

What do i-Spaces mean to EGI and to the Big Data Value?

The Big Data Value Association advertised and granted a label for European Innovation Spaces to help drive Big Data adoption across all domains of European industry. Every year, existing hubs can obtain a label as a "BDVA i-Space".

The Big Data Value Association (BDVA) has granted eight i-Space labels this year and one of them was awarded to the EGI Federation.

The EGI Federation is now accredited as one of the key data-driven innovation centres in both commercial and non-profit domains.

The announcement was made during the European Big Data Value Forum in Versailles. Labelled i-Spaces exchange good practices to promote data innovation all over Europe.

For the EGI Federation this means:

- Access to European funding instruments to enable innovation and federation of existing i-Spaces

- An increased visibility in EU Big-Data community

- Possibilities to connect and benchmark to other iSpace practices

More information

Roberta Piscitelli is a Strategy and Policy Officer at the EGI Foundation.

How AGINFRA+ serves the Agriculture & Food domain

Panagiotis Zervas and Nikos Manouselis give an overview of the project

AGINFRA+ is an EINFRA H2020-funded project, which started on 1 January 2017 and will last for 36 months. AGINFRA+ aims to exploit core existing e-infrastructures such as EGI, OpenAIRE and D4Science, to serve European scientists from user communities around Agriculture and Food.

More specifically, AGINFRA+ project aims to provide European scientists with Virtual Research Environments (VREs) facilitating them to collaborate and execute resource-intensive experiments (by liaising with D4Science), offering them access to a vast amount of research publications and datasets (by liaising with OpenAIRE) and cloud and grid resources for intensive computational applications (by liaising with EGI).

AGINFRA+ targets three main user communities, which address problems on:

- Predicting short-term and long-term agricultural yields in the face of both operational, within season decision making, and longer-term decision and policy making with respect to climate smart agriculture and climate adaptation.

- Identifying emerging food safety issues at an early stage and handling of large-scale datasets, to visualize complex data, mathematical models as well as simulation results and to deploy generated data processing workflows as web-based services.

- Food security, namely the need to efficiently analyse data produced by plant phenotyping

and its correlation with crop yield, resource usage and local climates. AGINFRA+ project partners have already initiated a strategic discussion with the key players trying to shape the thematic cloud on agri-food as part of the European Open Science Cloud (EOSC) and have participated in various workshops and braistorming meetings.

Thus, the vision of AGINFRA+ project is to develop a common technical infrastructure that could initially serve the three aforementioned user communities and it could be evolved to an AGINFRA food cloud demonstrator that will be positioned as the EOSC agri-food thematic cloud.

This is expected to be achieved by:

- Implementing the e-infrastructure roadmap for open science in agriculture and food, which is currently being developed by the eROSA project

- Running pilots in larger scale aiming to address all stages across the food value chain

To this end, the AGINFRA+ project will collect and promote appropriate and mature use cases through the AGINFRA+ dissemination channels.

The project will also organise hands-on workshops jointly with other H2020-funded projects where use case stakeholders can test the AGINFRA+ project services.

Finally, AGINFRA+ will develop a pipeline of mature use cases that could demonstrate the potential of a large scale piloting in an AGINFRA food cloud demonstrator.

More information

Nikos Manouselis and **Panagiotis Zervas** coordinate the AGINFRA+ project.

Website:
http://www.plus.aginfra.eu/