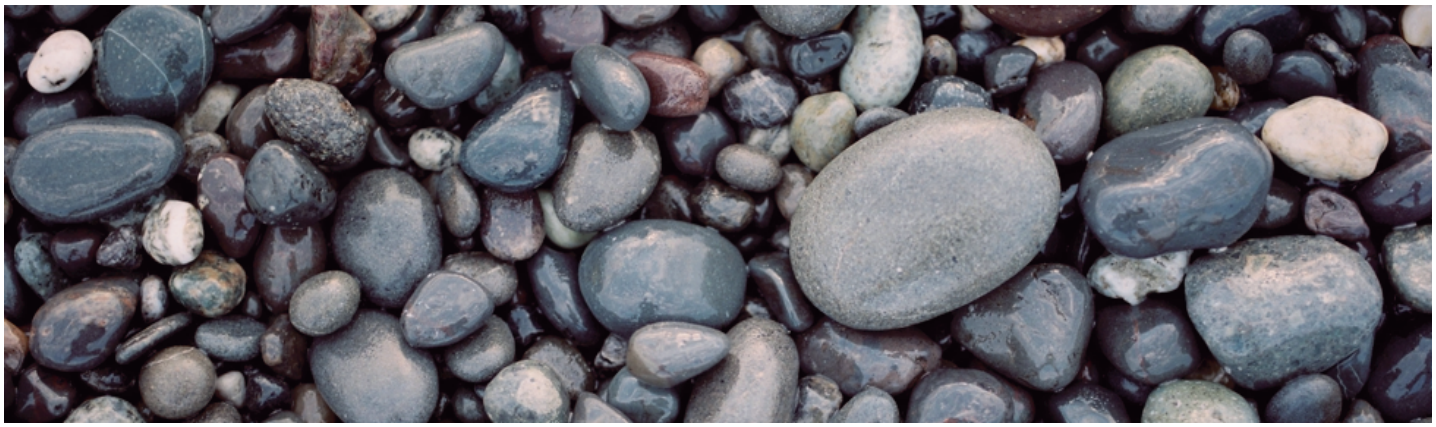
Inspired

ISSUE 27
MAY 2017

news from the EGI community



TOP STORIES

The EGI ISO certifications
page 3

EGI Services for open science
page 4

GEOSS Data and Computing Challenges
page 6

Terradue & EGI for Earth Observation
page 7

MORE

01 Computing centres: LNEC

07 The new Accounting Portal

08 OPENCoastS forecast service

09 EOSCpilot: Science Demonstrators

10 AARC is dead - long live AARC!

Advanced Computing for Research

www.egi.eu

Welcome to issue 27!

In the new edition of our newsletter, we focus our services and on the contributions of the EGI Community to the Earth Observation field.

Your feedback and suggestions are always welcome!

Send an email to Sara & Iulia at:

press@egi.eu

Save the date!
The upcoming DI4R 2017 will be in Brussels
30 November - 1 December
(Credit: Marc Ryckaert/Wikimedia Commons)

Computing centres: LNEC

Mário David shows us the LNEC data centre, part of the Portuguese Distributed Computing Infrastructure (INCD)

The LNEC data centre in Lisbon is located at the headquarters of the Laboratório Nacional de Engenharia Civil, a research facility dedicated to engineering sciences. The data centre is managed by FCCN and spans an area of 370m². The centre hosts a total of 2.500 CPU cores and 1.1 PB storage, from which about 700 TB under the Lustre distributed filesystem serving the HTC and HPC clusters, and about 400 TB under Ceph serving the Openstack cloud IaaS.

The data centre is part of the National Distributed Computing Infrastructure (INCD) - the Portuguese Digital Infrastructure. INCD is a partnership between LIP, LNEC and FCCN and represents Portugal in the EGI Federation, in IBERGRID (the Iberian Infrastructure) and in the Worldwide LHC Computing Grid (WLCG).

Your Data Centre

If you work with or at one of the +300 data centres federated in the EGI e-infrastructure, we would love to hear from you!

Send your pictures to press@egi.eu

More information

INCD - National Distributed Computing Infrastructure
http://www.incd.pt

INCD provides computing and storage services to Portuguese-based scientific communities in all domains and participates in national and international projects of strategic relevance.

LNEC - National Laboratory of Civil Engineering
http://www.lnec.pt

The EGI ISO certifications: new and improved service delivery

Yannick Legré writes about what the certificates mean for the EGI Community

The post delivery on a fine Wednesday afternoon back in March included a very special package: the official print certificates sent by TÜV SÜD to confirm that the EGI Foundation has been awarded a ISO 9001:2015 and a ISO/IEC 20000-1:2011.

This was excellent news. EGI is now the first European-wide publicly-funded e-infrastructure to be certified against ISO standards, sign of its maturity and of the competence of its actors.

The certifications demonstrate that our management systems regarding the EGI Service Catalogue follow ISO 9001:2015 and ISO/IEC 20000-1:2011 requirements. This includes all activities of planning, implementation, monitoring and continual improvement of all processes.

In short, this means that we can offer better services to the EGI consumers and more value to the EGI Council.

The long road to certification started with a strategic commitment by the EGI Council to establish efficient processes to manage EGI's service offer. The first step began in 2015 with the full implementation of FitSM, a lightweight version of the ISO 20000 standard.

Having FitSM in place and all the EGI Foundation staff trained and certified in service management according to this standard meant that EGI service management was already complying with about 80% of the ISO standards. The preparations for the two rounds of audits in October and December 2016 ensured that the remaining 20% requirements were also met.

The hard work of the past two years paid off: we can now be sure that we are operating improved, high-quality and reliable services to better serve our user base. And we do so with consistent, clear, streamlined processes that assure our stakeholders a better return on investment.

And that, for me, is the definition of a win-win!

More information

Yannick Legré is the Managing Director of the the EGI Foundation. @ylegre

About the certifications

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.

The certifications are an achievement of the EGI-Engage project, co-funded by the European Union (EU) Horizon 2020 program under grant number 654142.

EGI Services for open science

Tiziana Ferrari writes about current and future EGI service offer

State of the art

One of EGI-Engage's key achievements so far was the publication of the internal and external EGI Services Catalogue. This was an important milestone to develop a clear value proposition, increase the accessibility of available services, and make the providers from the EGI federation more visible to funders and policy makers.

The **external catalogue** includes all services available for researchers and is organized across three broad categories: Compute, Storage and Data and Training.

The **internal catalogue** offers the services necessary to federate local, regional and national infrastructures into the EGI Federation. These services are offered to the federated data centres to provide the "glue" that holds together a harmonized pan-European system for international research collaborations.

Scientific applications powered by EGI services offer community-specific capabilities such as datasets, search capabilities, data analytics, and scientific software. They are a tremendous research-enabling instrument: the current set of integrated scientific applications counts thousands of users worldwide and has provided high-impact examples of how individual researchers can benefit from EGI services, for instance:

> A team from the University of Monash used the HADDOCK portal to model the structure of iron-binding receptors in bacteria. The work was published as a paper in *Nature Communications*.

> A Swedish scientist used VIP to analyse the results of a long-term study of the effects of multiple sclerosis. The results were published in the *Multiple Sclerosis Journal*.

We are now expanding the scientific application portfolio with contributions from the NGIs, the EGI-Engage Competence Centres and any other partner who is willing to engage with EGI.

To make present and future services discoverable and accessible online, the project has been successfully prototyping a **Marketplace**, which we are planning to bring in full

Services by and for the EGI Community

External Services: Compute, Data & Storage, Training
For researchers, SMEs & industry, citizens & governments, higher education. *www.egi.eu/services/*

Internal Services: Making the federation work
For the members of the EGI Council. *www.egi.eu/internal-services/*

Scientific applications powered by EGI services
Provided by and for members and partners of the EGI Community. Currently available: HADDOCK, Chipster, PowerFit, NBIS toolkit, VIP, Jupyter Notebook, Peachnote, and many more. *www.egi.eu/use-cases/scientific-applications-tools/*

production in August. The marketplace will give the possibility to browse services by category, choose configuration options and place service orders.

New Services

DataHub will increase accessibility to third-party research data for downstream analysis via federated authentication and authorization, the bridging of preserved data and computing with the federation of distributed data repositories and the possibility of associating permanent identifiers to the output data that is generated by the processing workflows of EGI.

The DataHub is currently being validated to address two use cases: the cross-domain federation of existing storage infrastructures, and the provisioning of a distributed platform for the management of replicas of publicly data collections.

Applications on Demand (AoD) is a service designed to offer a scalable HTC environment. It delivers a whole cluster complete with job scheduler and a library of scientific applications (e.g. Statistical R, Jupyter Notebook) and generic utilities (e.g. Docker). The AoD also allows researchers to run custom scientific applications. AoD's compute and storage services are currently provided by CYFRONET, INFN, BELSPO, MTA SZTAKI, CESGA, UPV, and BIFI. All NGIs interested in making the AoD library available to national user communities can simply "plug" their national IaaS resources to the service. The service was opened for beta testing at the EGI Conference 2017.

The **EGI CheckIn** will enable access to EGI services using federated authentication mechanisms. It is a proxy service operated as a central hub, connecting federated Identity Providers (IdPs) residing 'outside' of the EGI ecosystem, and Service Providers (SPs) that are part of EGI. CheckIn was successfully integrated with the ELIXIR AAI infrastructure, which operates both an IdP and attribute provider service to manage user accounts and personal attributes for every ELIXIR user. CheckIn is in now production and services in addition to the current ones (GGUS, GOCDB and AppDB) are being incrementally added.

More information

Tiziana Ferrari is the Technical Director of the the EGI Foundation. @tferrariEGI

EGI Services
https://www.egi.eu/services

Applications on Demand
https://access.egi.eu/start

GEOSS Data and Computing Challenges

Stefano Nativi and Barbara Ryan write about a data platform for earth observation

Society is facing unprecedented challenges for food, water and energy security. Resilience to natural hazards is an increasingly important issue as is ecosystem sustainability. Population growth, pandemics and the development of a sustainable economy all have currently unknown impacts against a backdrop of climate change, which has the potential to exacerbate all these issues.

Earth observations are necessary to report and model climate change and to calculate greenhouse gas emissions, in line with the Paris Agreement 2016 on Climate. The intergovernmental Group on Earth Observations (GEO) works to connect the demand for sound and timely environmental information with the supply of data and information about the Earth that is collected through observing systems and made available by the GEO community. GEO is leading a worldwide effort to build a Global Earth Observation System of Systems (GEOSS). GEO convenes providers and users of open Earth Observation data, aiming to highlight best practice and eliminate duplication of effort to harness the Data Revolution for the benefit of humanity.

GEOSS is a set of coordinated, independent Earth observation, information and processing systems that interact and provide access to diverse information for a broad range of users in both public and private sectors. GEOSS links these systems to strengthen the monitoring of the state of the Earth. Developed over the last decade, GEOSS makes more than 200,000,000 open Earth Observation data resources accessible for better decisions on a range of areas from food security, to protection of biodiversity, renewable energy and disaster resilience. With almost 160 data providers, an important element of GEOSS is the brokering framework called the GEO DAB (Discovery and Access Broker). The GEO DAB implements a brokering pattern ensuring the necessary scalability and flexibility.

The GEOSS evolution includes big data analytics capabilities, particularly data cube functionalities, to move from data sharing to information and knowledge generation and sharing, in particular to support the UN Agenda 2030 and the Sustainable Development Goals. In keeping with the System-of-Systems principles and leveraging the implemented brokering pattern, GEOSS is looking at the EGI as a valuable e-infrastructures federation to underpin the new analytics capabilities requested by the GEO Community.

GEOSS in numbers

> GEOSS finalizes about **15,000 queries per day**.
> In 2016 GEOSS handled about **4.5 million of requests**.
> GEOSS brokers and harmonizes almost **160 GEOSS Providers** of data and information.
> The GEOSS Providers share about **200 million data granules** (i.e. downloadable files), organized in about **42 million datasets.**
> The GEOSS Portal counts about **3,000 visits per month**.

More information

Group on Earth Observations (GEO)
www.earthobservations.org/

Global Earth Observation System of Systems (GEOSS)
www.earthobservations.org/geoss.php

Stefano Nativi is co-chair of the "GEOSS Architecture and Evolution" Working Group.

Barbara Ryan is the director of the GEO Secretariat

Terradue & EGI: a partnership for Earth Observation

Pedro Gonçalves on how EGI's Federated Cloud supports Terradue's operations for the Geohazards Exploitation Platform

Earth observations from satellites produce vast amounts of data. In particular, the new Copernicus Sentinel missions are playing an increasingly important role as a reliable high-quality and free open data source for scientific, public sector and commercial activities.

ICT solutions can facilitate the handling of these large volumes of data and are nowadays modifying the expectations that organisations have on new service development and on support to Earth Observation (EO) data exploitation. Their goal is more and more to develop capacities to create added value, involving SLAs and accountability with business partners for the data products and services they bring in this process.

Terradue Cloud Platform is addressing this topic with solutions to transfer EO processing algorithms to cloud infrastructures. The platform also provides services to optimise the connectivity of the data centres with more integrated discovery and processing methods. For example, Terradue provides the engineering and operational support for the Geohazards Exploitation Platform (GEP), an ESA-funded partnership also involving private companies (TRE-ALTAMIRA), research centres (CNR IREA, CNRS ENS, CNRS EOST and INGV) and space agency (DLR EOC).

GEP and EGI

GEP offers a rich set of ready to use EO data processing services for the analysis and monitoring of earthquake, volcanoes and landslides. The platform federates the geohazards community by creating a workplace with cloud-based models of collaboration, where data providers, users and technology providers join forces to produce scientific and commercial exploitable results.

EGI supports Terradue with matchmaking services between ICT consumers and the appropriate provider(s) across the EGI Federation and beyond. The computing and storage resources from the ReCaS Bari and BELNET-BEGRID centres are used by Terradue to help the global scale systematic production of the DLR InSAR Browse Medium-Resolution Service on the GEP. With this service, the platform produces interferograms to show where earthquakes are most likely to impact society.

Currently in a ramp-up phase, which began in September 2016 covering 20% of the world seismic mask, GEP is planned to reach its peak of 50% by the 2nd quarter of 2017 with a production rate of about 320 Sentinel-1 scenes (160 interferometric pairs) per day. This production is fully ran on EGI Federated Cloud resources.

The Service level agreements (SLAs) established with EGI Foundation enabled Terradue to extend the hybrid cloud infrastructure using a new OpenNebula OCCI driver, and provided a reliable cloud infrastructure for the users of the ESA Thematic Exploitation Platforms.

More information

Terradue
http://www.terradue.com/

Pedro Gonçalves is Chief technical Officer and Founder of Terradue

The new Accounting Portal

Diego Scardaci introduces the new version of one of EGI's key tools

The EGI Accounting Portal is an operational tool that processes, summarizes and displays the Accounting Repository data, acting as a common interface to the different accounting record providers. Readers can browse views of the data displayed in a user-friendly way. The portal helps EGI members and external parties to understand the use of resources and how they serve the needs of research.

The Accounting Portal is developed and operated at CESGA and is supported by the accounting repository infrastructure operated at STFC, where records from the all EGI data centres are collected.

During the EGI-Engage project, the Accounting Portal was completely renewed adopting new web tools that offer a better user experience and its look and feel was improved through a modern interface. The portal was re-organised with an intuitive menu to follow the structure of the EGI service catalogue.

Furthermore, we developed a completely new view to retrieve accounting data for the Scientific Disciplines and new types or reports, for example the TOP 10/100 resource centres. New features are also available such as the dynamic interaction with graphics, allowing to expose further details or zoom certain parts, the contextualised online help, providing information about the meaning of the several metrics and variables used within the portal. Custom views for research infrastructures can be also easily added, as it was already done for WLCG (see the Research Infrastructure item in the main menu).

Accounting portal developments are continuing and, before the end of EGI-Engage project, more features will be made available. This will include:

> integration with the EGI Federated AAI (the CheckIn service)

> maps with the distribution of the accounting data over a geographical area and

> the possibility to choose the measurement unit to be used to show the accounting data.

What can we learn from the Accounting Portal?

How many HTC jobs were executed by NGIs this year?
From 5,600 submitted by the Croatian NGI in January to the more than 19 million sent by CERN in April.

Which data centres were the top 10 HTC resource providers in the last 12 months (normalised elapsed time)?
CERN-PROD, RAL-LCG2 (UK), INFN-T1 (Italy), FZK-LCG2 (Germany), IN2P3-CC (France), GRIF (France), DESY-HH (Gremany), RRC-KI-T1 (Russia), Tokyo-LCG2 (Japan), SARA-MATRIX (Netherlands)

How much CPU time was consumed by researchers in the EGI HTC infrastructure in the last 6 months?
About 14 billion hours, measured in normalised CPU time (HEPSPEC06)

And if you want specifics... in December 2016 how many CPU hours were consumed by researchers working in the field of Astronomy?
About 330 million (normalised elapsed time).

How many Virtual Machines were created by the researchers in the top 10 Cloud resources providers in the last 12 months?
About 190,000.

More information

The **Accounting Portal** was developed by **CESGA** and aggregates information collected in the accounting repository operated by **STFC**.

http://accounting.egi.eu

Diego Scardaci is a member of the EGI Foundation User Community Support Team

Predicting water conditions along the Iberian coasts

Iulia Popescu writes about the OPENCoastS project and its forecast mission

Seas and oceans are important drivers for the European economy and they need to be preserved and developed in a sustainable way. In support of this view, the European Commission is implementing a long-term strategy called Blue Growth to boost the progress of the European marine sector.

One of the initiatives adhering to Blue Growth is OPENCoastS, or On-demand Operational Coastal Circulation Forecast Service.

OPENCoastS is a service that builds on-demand circulation forecast systems for different users along the North Atlantic coast. OPENCoastS generates forecasts of water levels, 2D velocities and wave parameters over the region of interest for periods of 72 hours. The system is useful in anticipating natural disasters and accidents in the coast, e.g. floods and chemical spills and can help in search and rescue operations.

The service was developed by the National Laboratory for Civil Engineering (LNEC) in 2010 as WIFF (Water Information Forecast Framework) and uses the SCHISM modelling system.

Since 2010, the system has been producing 48-hour forecasts on a daily basis for the Portuguese coast and is running on High-Throughput Compute and storage resources provided by the NCG-INGRID-PT data centre, which is part of the Portuguese National Distributed Computing Infrastructure (INCD) and the EGI federation.

LNEC provides this service to Portuguese-based researchers free of charge and the regional forecasts are open to everyone via the web portal, but LNEC researchers Anabela Oliveira and Alberto Azevedo are working on making OPENCoastS a service available at an international scale.

The role of distributed computing and the future of OPENCoastS

Oliveira and Azevedo argue that the development of forecast systems requires a strong knowledge of coastal processes and IT, along with access to significant computational and storage resources. The INCD computational and storage capacity needed to run the water forecast service every day uses three typical configurations:

> Reference run: 60h CPU + 3GB storage / 72h period daily forecast

> Large run : 600h CPU + 30GB storage / 72h period daily forecast

> Very large reference run: 1200h CPU + 60GB storage / 72h period daily forecast

These values can amount to up to 2500 cores and 2TB storage per year for a total of 100 site deployments.

In the future, the platform could use even more computing resources to facilitate the access to circulation forecasts to biologists, geologists and biogeochemists, who have strong needs in understanding the impact of water dynamics and ecology.

More information

OPENCoastS portal
http://ariel.lnec.pt/

Laboratório Nacional de Engenharia Civil (LNEC)
http://www.lnec/pt

Iulia Popescu is a Communications Officer at the EGI Foundation

EOSCpilot: Science Demonstrators

Matthew Viljoen on one of the project's key work packages

EGI is proud to be a partner of the European Open Science Cloud for Research Pilot project (EOSCpilot). The two-year-long project kicked off in January 2017 and lays the foundation for the European Open Science Cloud (EOSC), as described by the European Commission.

Led by STFC and involving 33 partners across Europe, the project will engage with stakeholders to build the skills and trust necessary for an open approach to scientific research and develop a series of science demonstrators with existing communities and infrastructure.

Another aim of the EOSCpilot project is to understand and address the barriers that stop European research from fully tapping into the potential of data. To improve interoperability between data infrastructures, the project will engage different scientific and economic domains, countries and governance models and will demonstrate how resources can be shared even when they are very large and complex.

The science demonstrators developed by EOSCpilot are key to the success of the project. They represent the early adopters of EOSC and are selected to integrate diverse services and infrastructure to demonstrate interoperability in different scientific domains.

EGI is leading the Service Pilots task within the project. This task uses the expertise from existing services to analyse the requirements of science demonstrators and to identify any technical gaps. EGI is uniquely positioned to do this thanks to our team's experience in service delivery, our diverse scientific community and links with other e-infrastructures (such as EUDAT and GÉANT).

During the first year of the project, EOSCpilot has been working with the first round of scientific demonstrators, and is currently selecting the next round.

The Science Demonstrators

The first five give a good idea of the broad range of communities and technologies that the project is working with:

1) Data Preservation for High Energy Physics aims to bring the Open Data Portal, developed by CERN to provide data preservation (including software and documentation) for the High Energy Physics community, to the wider scientific community. The solution will be deployed on generic cloud services and proved to work with sample preserved High Energy Physics data before being prepared for non-HEP data.

2) PanCancer has developed the Butler genome analysis framework on the EMBL-EBI Embassy cloud and its EOSCpilot demonstrator is working to deploy it on generic cloud services. This will enable the analysis framework to be used by more communities.

3) The photon-neutron science community demonstrator is working on making its data and analysis tools available to scientists all over the world using institutional and public cloud services.

4) TextBridge is a text mining solution developed for the humanities and social sciences community at the University of Florence. This community is working with EOSCpilot to migrate their university service to a generic cloud service.

More information

EOSCpilot project
http://eoscpilot.eu/

Matthew Viljoen is a Senior Operations Officer of the EGI Foundation

AARC is dead - long live AARC!

Laura Durnford writes about the bridge between the old and the new AARC projects

The first Authentication and Authorisation for Research and Collaboration (AARC) project concluded on 30 April 2017 after two years, with many useful outputs for e-infrastructures, research infrastructures and libraries. The second AARC project, started 1 May, will build on these achievements and bring a new focus.

AARC has been creating a common framework for research and collaboration communities, with one blueprint architecture, one set of policies, and one collection of training materials that should work for everyone and allow their authentication and authorisation solutions to work together. AARC has also been working with research collaborations to pilot and improve specific technical and policy aspects.

AARC's approach means that research collaborations can spend less time and less money reinventing the authentication and authorisation wheel, and their researchers can focus on research. Safe and more reliable access for more researchers to more services, data and software, will allow greater cooperation between research collaborations and open up the possibilities for exciting new research.

What has the AARC project produced?

> **AARC Blueprint Architecture**, a set of interoperable building blocks for people designing and implementing access management solutions for international research collaborations

> **Policies** to complement the technical research plus recommendations and best practices to implement a scalable and cost-effective framework for integrated solutions:

>> **Snctfi** - identifies operational and policy requirements to help establish trust between an infrastructure and identity providers. For use by personnel responsible for management, operation and security of an infrastructure and those wishing to assess its trustworthiness.

>> **Sirtfi** – AARC was the main sponsor for work to create an assurance framework that allows organisations to cooperate in the coordination of incident response, in the event of a federated security incident.

> Pilots expanding the coverage of federated access, pilots testing the integration of AARC results.

What next?

While the goals and objectives of the second AARC project will largely remain the same, the project will take two main routes:

> AARC will expand efforts to engage with target communities to disseminate information, deliver training, gain feedback, and implement the AARC framework. The number of partners has increased from 20 to 26, helping to make this approach easier.

> AARC will shift the technical focus to the question of how to authorise permitted users to access the resources across the boundaries of different infrastructures and research collaborations.

To maximise the uptake of existing materials and any that are developed in the next phase, a fresh impetus will be given to developing and delivering training, and to disseminating via 'how to...' documents, webinars and so on. This will all be supported by a change to the project website so that it becomes a shop window for the outputs of the AARC project.

More information

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

Laura Durnford is a Senior Communications Officer at GÉANT.