

Inspired

ISSUE 31
JUNE 2018

news from the EGI community



TOP STORIES

IBERGRID looking towards the EOSC

page 1

In the spotlight: EGI Operations Team

page 3

EGI FedCloud use case: PhenoMeNal

page 4

EGI & DIRAC join forces for a new service

page 5

Virgo is using the EGI Workload Manager

page 6

MORE

- 07 GDPR and research collaborations
- 08 New EGI service: ISO 27001 Training
- 08 DI4R – Submit an abstract and register!



www.egi.eu

Welcome to issue 31!

This edition of the newsletter is focused on updates from the EGI community, new services and use cases.

Your feedback and suggestions are always welcome!

Send an email to Sara & Iulia at:

press@egi.eu



IBERGRID looking towards the EOSC

Isabel Campos and Jorge Gomes highlight the IBERGRID developments

The signature on November 8th in 2003 of the Scientific & Technology Cooperation Agreement between Spain and Portugal, marked the start of a strategic scientific cooperation in the Iberian Peninsula, which included distributed computing as one of the key topics.

IBERGRID was born out of the Iberian Common Plan for distributed infrastructures released in 2007. Since then, IBERGRID has been federating infrastructures from Iberian research & academic organisations mainly focused on grid, cloud computing and data processing. The IBERGRID infrastructure comprises 12 computing and data centers in Spain and Portugal.

A number of replicated services guarantees data integrity and resilience in the operations of the critical services.



The infrastructure has provided 984 million processing hours since 2006 to support the HEP experiments and several user communities. This includes 19 million hours on biomedical applications and ~6 million hours on computational chemistry.

Strictly on cloud support, more than 216,000 Virtual Machines have been instantiated providing more than 2 million cloud processing hours to Lifewatch ERIC in the last year.

On the R&D side, service integration activities are taking place in numerous areas. An example is OPENCoastS, a service to provide on-demand circulation forecast systems as a service for the Atlantic coasts. The service is deployed at the computing site NCG-INGRID-PT, part of the EGI Federation, but it is being integrated into EOSC-hub as a Thematic Service in collaboration with LIP, LNEC, INCD, UNICAN, CNRS, and CSIC.

On the software development side, IBERGRID is contributing in many areas. CSIC has developed OpenStack support for VOMS authorization & authentication, cloud pre-emptible instances (OPIE) as well as CPU Cloud accounting.

The Technical University of Valencia developed & maintains the Infrastructure Manager (IM), a key service to support the instantiation of tailored clusters now part of the EOSC-hub service catalogue.

Support to user-level container execution has been developed & is maintained by the IBERGRID software teams at LIP. Udocker is an extremely successful software product - more than 310 stars in GitHub: which is being explicitly recommended in many computing centers around the world as the best solution for users to execute containers, without requiring the intervention of the system administrator layer.

Software Quality Assurance has generated an enormous amount of activity in the Iberian area. LIP, CSIC, CESGA and UPVLC are in charge of ensuring the quality of the UMD software deployed by EGI. The Accounting Portal of EGI is maintained & developed by CESGA for the EGI community.

The IBERGRID community gathers around conferences in the Iberian area, and a dedicated event, the IBERGRID conference series, this year takes place in Lisbon from 11th to 12th of October.

IBERGRID looks into the future EOSC with optimism. From the user support side our main assets are a very consolidated user-base, and well-reputed user engineering and support teams.

From the technical point of view, we count on worldwide-recognised teams, with expertise & technical background to address the specific requirements from scientific communities in the EOSC era.

IBERGRID in EGI

IBERGRID is a key Operations Centre of the EGI Federation. The resources made available by IBERGRID sites have been instrumental in supporting the 4 largest scientific collaborations based at the Large Hadron Collider (ALICE, ATLAS, CMS, LHCb).

Other examples of researchers supported by IBERGRID are:

CLIC and ILC: the next adventure in High Energy Physics

The Compact Linear Collider Study (CLIC) & the International Liner Collider (ILC) are two collaborations set up to explore what happens when electrons and positrons (which are antielectrons) collide at high-energy.

CLIC and ILC use electrons & their anti-particles (instead of protons as the LHC does) to collect a new range of high-precision measurements & get a different view on high-energy physics.

(+ 760,000 CPU hours in IBERGRID)

Genetics

Biologists found that some old industrial landscapes can have surprising benefits for biodiversity.

(+18 million CPU hours in IBERGRID used by the biomed VO since 2006)

Computational Chemistry

A team is pioneering a way to produce high-quality, carbon-neutral methane gas using industrial waste carbon dioxide.

(+5 million CPU hours in IBERGRID used by the compchem VO since 2007)

The WeNMR suite of portals

WeNMR supports +9,000 structural biologists worldwide with a suite of online, user-friendly portals. The portals are powered by High-Throughput Compute resources provided via an SLA by IBERGRID and other sites of the EGI Federation.

(+330,000 CPU hours in IBERGRID since 2013)

LifeWatch ERIC

LifeWatch is a distributed Research e-Infrastructure to advance biodiversity research and to address the big environmental challenges and support strategic solutions to environmental preservation.

(Supported with IBERGRID cloud resources of +2 million hours in the last 12 months)

OPENCOSTS

Operational Coastal Circulation Forecast Services provides an on-demand circulation forecast system-as-a-service for the European Atlantic coasts.

(Currently being integrated into IBERGRID and EGI as an EOSC-hub thematic service).

More information

Isabel Campos is IBERGRID Coordinator in Spain.

Jorge Gomes is IBERGRID Coordinator in Portugal.

IBERGRID

wibergrid.lip.pt/site/

In the spotlight: EGI Operations Team

Matthew Viljoen highlights the team's activities and plans

The **EGI Operations Team** leads the work of the distributed team of National Operations Centres and regional staff in addition to teams dedicated to security response and technical experts.

The goals are:

- Service delivery in a reliable and trusted way
- Gathering requirements from end users and understanding when a new functionality is needed
- Defining and maintaining Operational Level Agreements (OLAs) from service providers
- Making sure service monitoring is working to meet obligations defined in the OLAs
- Incorporating OLAs with Service Level Agreements (SLAs) for end user communities
- Tracking usage with the EGI Accounting service, which enables the future capacity planning & knowing who used what.

Security is a paramount concern in EGI. The Operations Team is involved in the ongoing evaluation and response of security threats and the monitoring of security vulnerabilities across the federation, ensuring that patching is done sufficiently quickly.

The team supports the development and running of infrastructure services dealing with user Authentication and Authorization (AAI).

In this way, access can be granted to resources in an easy way, across the federation and globally, with other infrastructures.

This allows the exchange of best practices and paves the way for new collaborations.

In addition to these efforts, the Operations Team runs & serves a distributed helpdesk which lets users submit problem tickets or requests for new functionalities.

Software distribution is another important aspect of the EGI Operations work. The team maintains trusted channels for obtaining software and engages with early adopters to ensure that the software is ready for general usage.

Last but not least, EGI Operations ensures that the knowledge required to use, operate and maintain operations is accessible and retained in the form of adequate documentation.

Operations Plans for 2018

Over the course of 2018, the EGI Operations Team is engaging in a number of different activities to continue to improve the quality of service delivery.

The central EGI Operations team is currently leading the Service Management development of the EOSC-hub project to ensure that best practices formed during the achievement of ISO 9001 and ISO/EIC 20000-1 certification in 2017 are followed during the establishment of the European Open Science Cloud.

The team is also working to improve the security, general quality and usability of the EGI Federated Cloud service.



More information

Matthew Viljoen is Operations Manager of the central operations team at the EGI Foundation.

Core EGI Operations Team:
(in the photos)

Bruce Becker
Baptiste Grenier
Alessandro Paolini
Vincenzo Spinoso
Matthew Viljoen

Deploying PhenoMeNal virtual research environments on the EGI Federated Cloud

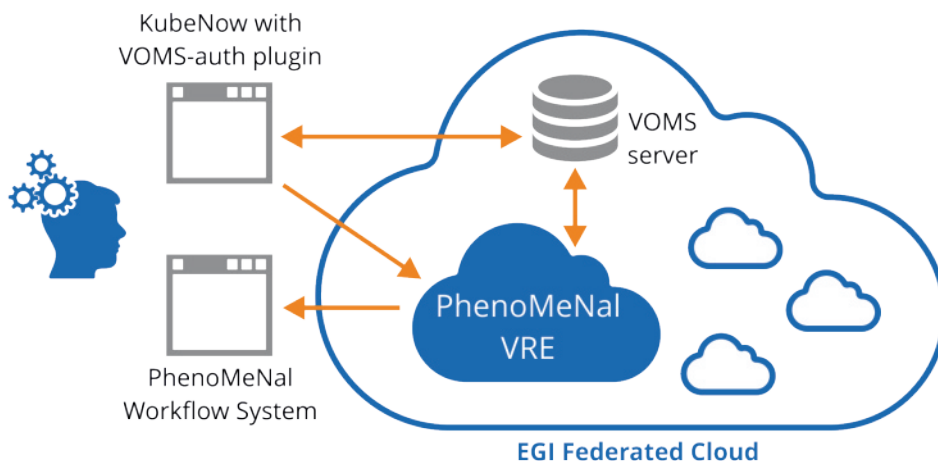
Ola Spjuth and Anders Larsson on the successes of the PhenoMeNal project

What is PhenoMeNal?

The PhenoMeNal e-infrastructure supports data processing and analysis workflows for large datasets generated in metabolomics research. The computational services provided within PhenoMeNal enable researchers to build and run reproducible data processing workflows in cloud environments. They aim to improve the understanding of the causes and mechanisms underlying health and disease, and to render the generated data & metadata more traceable as demanded by reproducible research. PhenoMeNal's services can enhance a wide range of metabolomics applications in biological, agricultural and environmental sciences.

PhenoMeNal services

The **PhenoMeNal Virtual Research Environment (VRE)** is implemented as a microservice architecture with Docker containers orchestrated in a Kubernetes environment. The containers can be connected into analysis pipelines using scientific workflow platforms, such as Galaxy. For deployment, PhenoMeNal has developed the KubeNow cloud deployment software that simplifies instantiation of complete virtual infrastructures with Kubernetes, including storage, network and other required services readily available.



The **PhenoMeNal Gateway** provides a web-based Graphical User Interface (GUI) on top of KubeNow. The use of Galaxy as the main workflow system provides a user friendly GUI for workflow authoring, and execution of tools and workflows.

The **PhenoMeNal App library** is a service catalogue of all open source metabolomics tools available through the VRE. The library also provides an online training featuring tutorials and technical documentation on the developed tools and workflows.

Deployment on the EGI Federated Cloud

PhenoMeNal has worked closely with the **EGI Federated Cloud** to harness EGI's strategy for provision of a secure infrastructure. PhenoMeNal VREs are now successfully deployed on EGI Federated Cloud resources (Cloud provider IN2P3-IRES).

This achievement was a joint effort between the two consortia and covered the extension of the PhenoMeNal deployment

engine (KubeNow) to include OpenStack Virtual Organization Management System (VOMS) authentication methods.

The work was carried out on EGI cloud resources provided by **IN2P3-IRES** within the ELIXIR-FR virtual organisation.

More information

Ola Spjuth and Anders Larsson are members of the PhenoMeNal consortium based at Uppsala University.

PhenoMeNal

phenomenal-h2020.eu/home

EGI and DIRAC join forces to develop a new service

Yin Chen and Baptiste Grenier on the achievements of the collaboration



The **8th DIRAC user workshop** was held 22-25 May in Lyon and hosted at CC-IN2P3 by CNRS.

DIRAC was created to support the data management and processing of the LHCb experiment at CERN. Today it acts as a complete solution to support grid, cloud and HPC resources, targeting various large scientific communities including LHCb, Belle II, EGI, CTA, GridPP, WeNMR, VIP, FranceGrilles, SKA. DIRAC provides complete solutions for production managements, for handling distributed large volumes of scientific data and optimising job executions.

The software package includes:

- A **data management system** that comes up with access protocols to various storages, reliable data transfer (based on FTS), file catalogues allowing metadata-based data-discovery, and optimising routine data distribution tasks.
- A **workload management** that allows to aggregate resources of different source & nature in a single system computing, and optimise the job submission.
- A **flexible transformation system** that provides automated data-driven submission of processing jobs with workflows of arbitrary complexity
- An **accounting system** to collect & store data regarding to the activities, and a monitoring system to monitor components.
- A **webportal** with an easy access user interface.

The modular organisation of the DIRAC components allows selecting a subset of the functionality suitable for particular applications or easily adding the missing functionality. These are very useful for communities to have customised environments for handling their own data. DIRAC evolves over time and in recent development, it includes new features such as supports for Cloud, HPC, multi-VO etc.

In order to better support research communities, the EGI teams collaborated with DIRAC to co-develop a new service: the **EGI Workload Manager**, also known as DIRAC4EGI.

The service is led by the EGI Foundation & operated by IN2P3 on resources provided by CYFRONET.

The EGI Workload Manager is already used in production by early adopters like WeNMR, that were able to easily switch their Science Gateways from gLite WMS to DIRAC. The work is currently being supported by the WeNMR Thematic Service under the **EOSC-hub** umbrella.

The accessibility of this service will also be extended soon: currently, authentication is based on the old X509-based model that is well known in the distributed computing for research world, but in the context of the EOSC-hub project, this service is being integrated with the **EGI Check-in service**.

This will allow users to connect using their institute's Identity Providers without having to own a personal grid certificate.

EGI has a formal procedure to include production-level services into the EGI service catalogue, including a complete business plan for service provisioning, assurance of resource capacity, full documentation, service order and help desk.

EGI is now in the final stage to sign an OLA with the technology and resource providers of the DIRAC Workload Manager. With an OLA in place, this service will be officially included into the EGI service catalogue and made accessible to the EGI community.

More information

Yin Chen is Senior Technical Outreach Expert at the EGI Foundation.

Baptiste Grenier is Senior Operations Officer at the EGI Foundation.

DIRAC

<http://diracgrid.org/>

Virgo is using the EGI Workload Manager

Alessandro Paolini reports on the new collaboration

Virgo is a giant laser interferometer designed to detect gravitational waves & located at the European Gravitational Observatory (EGO) site in Italy. Virgo is a collaboration of the French National Center for Scientific Research (CNRS) and the National Institute for Nuclear Physics (INFN) and is operated by an international consortium.

The Virgo collaboration has demanding needs to store and analyse a large amount of data recorded by the interferometer. A typical analysis run can last for about 6 months, and depending on the amount of data, hundreds of thousands of jobs can be submitted to the computing resources, with each job lasting anywhere between a few hours to a full day.

That is why Virgo chose to use the **EGI Workload Manager** – to easily dispatch jobs to computing resources and to manage the data necessary as input for the jobs or produced as output.

The EGI Workload Manager (also known as DIRAC4EGI) is an EGI service based on the DIRAC technology and suitable for users that need to exploit distributed resources in an optimised and transparent way.

The type of resources that DIRAC can support include computing resources (grid, cloud, and batch systems), storage and catalog resources. Access to DIRAC4EGI, a multi-VO DIRAC server, is provided by EGI to communities

that lack resources for installing and managing an own dedicated server or are simply looking to try the functionalities of the tool.

Virgo is now performing tests using this instance. The fact that DIRAC is already used by many communities as a mature tool was a factor in the decision.

In addition to the EGI Workload Manager, the Virgo collaboration also decided to test distributed data management solution to better understand its potential.

For that, it was agreed to set-up a dedicated DIRAC file catalog component as well, hosted at the INFN data centre in Italy.

The tests conducted so far show good performance results. For example, the catalog was populated with millions of records, and the performances were good even with a number of records similar to the real numbers that are expected to be in production.

The tests also allowed to find and fix some misconfigurations on the resource centres available in France, Italy & the Netherlands. In the following months, even more sites will be involved and there are plans to move and register the production data between the sites, using the DIRAC data transfer feature.

To help the VIRGO community understand this technology, a winter school will be jointly organised by EGI, DIRAC and VIRGO in late 2018.

The school targets new VIRGO members, postdocs, and senior researchers and will provide courses on the following topics:

- grid and cloud concepts and the EGI e-infrastructure
- high-level solutions developed by INFN for provisioning, creating, managing & accessing pool of heterogeneous computing resources (DODAS – Dynamic On Demand Analysis Service)
- use of the DIRAC system to handle user payloads running on any EGI grid- and cloud-based resources and on other computing resources.

More information

Alessandro Paolini is Operations Officer at the EGI Foundation.

Virgo

www.virgo-gw.eu/

EGI Workload Manager

wiki.egi.eu/wiki/Workload_Manager

The observation of gravitational waves by the Virgo & LIGO Scientific Collaborations paved the way to the **Physics Nobel Prize** awarded in 2017 to Rainer Weiss, Kip Thorne & Barry Barish.

Making GDPR easier for research collaborations

David Groep introduces the guidelines published by the AARC project

We want researchers to be able to use resources from multiple e- and research-infrastructures. If the infrastructures have policies in common, their services can trust each other so they can more easily exchange data.

This makes it easier for their users to accept the policies no more than once.

However, policies are presented when a researcher joins a community and they must match the requirements of e-infrastructures (such as EGI) and comply with the new GDPR. For example, if you are a Community Manager and you organise users into groups to determine who gets access to what service – our [suite of guidelines](#) can come in very handy. Many of them are co-developed by AARC and EGI.

For those of you worrying about GDPR compliance: remember it is all about striking the right balance between your legitimate need to manage your community and its resources, versus the impact on researchers.

That is what our new guidance helps you do: with federated identity management (FIM), the data about your members is minimal by design.

We have brought together the best of the guidance from each of the privacy regulators across Europe, with a focus on our research communities.

So when you set up a research community, broaden your use cases, or extend your services, you should see what the impact on privacy will be.

Is it truly “risky”? If your research itself is about people, you will likely need to do a risk assessment. If you are just using your users’ personal data to access services in the infrastructure, then the AARC guidelines will make it a lot easier.

We have 2 documents in the area: [The Data Protection Impact Assessment Guide for Communities](#) gives you a set of handles to determine whether or not you fit in the most common scenario. And if you need a specific implementation guide: why not look at the [guidance](#) we gave the Life Science community, which is scoped to community needs.

If you are in doubt, contact the AARC team and we can work jointly to analyse your needs.

The EGI structure for organising policies has been the basis of the [AARC policy starter kit](#), a ‘handbook’ and set of templates you can use to ensure that your community’s or infrastructure’s policy meets global expectations.

We will expand it to ensure all policy aspects needed for communities & infrastructures are part of that kit. But we need people to work with us so that our policies reflect their needs.



So please feel free to join the EGI security policy group, the WISE community, IGTF or REFEDS to work with us. There’s a limited number of policy experts in the world and we really need your engagement & support to make this activity a success!

More information

David Groep is the Team Leader of AARC’s Policy Team & member of the EGI Security Coordination Group.

AARC Project Policies

aarc-project.eu/policies/

New EGI Service: ISO 27001 Information Security Training

EGI Foundation welcomes a new addition to its training services: Information Security Management according to ISO/IEC 27001.

ISO/IEC 27001 is part of the ISO/IEC 27000 family of standards designed to help organisations keep information assets secure. It is complementary to FitSM, with similar principles towards implementing management systems and diving deeper into areas around information security e.g. risk management and implementing technical, physical and organisational information security controls.

Value of ISO/IEC 27001:

- Ensures that security risks are appropriately managed and prioritised
- Guards organisations against information security risks
- Protects the data entrusted to organisations
- Supports fulfillment of legal responsibilities (e.g. GDPR)

The courses are split into:

- Foundation level (2 days) focused on providing the fundamentals of Information Security and how to implement a management system (ISMS) in organisations through a combination of people, processes & IT systems.
- Professional level (3-4 days) for managers & personnel working to implement, maintain and operate information security within an organisation. Achieving the ISO/IEC 27001 Professional level provides an additional role-based certification "Information Security Officer".

EGI delivers trainings in 2 ways:

- Open Registration: for individual registrations; the training will be organised at a pre-determined date and location;
- In-House: for organisations needing several members of staff to be trained; the date and location of the training will be mutually agreed.

The first ISO 27001 Foundation training is set to take place in Amsterdam, 19 to 21 September. A formal certification will be offered to all participants of the course that successfully pass the final exam.

The training topics include:

- Introduction to ISO/IEC 27001
- Foundations of Information Security Management Systems (ISMS)
- ISO/IEC 27001 – Minimum requirements & security controls
- Selected practical recommendations & guidelines
- Related standards & frameworks
- Certification options according to ISO/IEC 27001

More information

ISO 27001 Training

egi.eu/services/iso-27001-training/

DI4R – Submit an abstract and register!

The **DI4R 2018** (9-11 October) is jointly organised by EOSC-hub, GÉANT, OpenAIRE and PRACE and this year is hosted in Lisbon by LIP – the Portuguese Particle Physics Laboratory.

Keeping with the tradition of inviting researchers to take the lead of the programme, DI4R 2018 will be chaired by two top members of the High-Energy Physics community: Sinéad Ryan, from Trinity College Dublin, & Volker Gülzow, from DESY.

"Digital infrastructures underpin research – enabling simulation, visualisation, storage & analysis & helping collaboration and innovation in an era of big data and large-scale computation," says Professor Ryan.

"No Computing – No Science! DI4R discusses new technologies but more important is getting people together across disciplines and talk to each other during workshops or over coffee," says Professor Gülzow.



"So take the opportunity and join us for a great event at DI4R 2018 in Lisbon!"

Registration is now open, and a Call for Abstracts will be online until 29 June.