

Inspired

ISSUE 30
MARCH 2018

news from the EGI community



TOP STORIES

The new EGI Operations Manager

page 1

Pollution forecasts in Bulgaria

page 3

EXTraS portal: hidden treasures in the sky

page 4

EOSCpilot Science Demonstrators

page 5

The eXtreme Data Cloud project

page 6

MORE

- 02 EGI at the first EOSC-hub week
- 07 EGI & EUDAT close collaboration
- 08 eInfraCentral introduces their service catalogue



Advanced Computing
for Research

www.egi.eu

Welcome to issue 30!

Happy Spring! This edition of the newsletter is focused on new EGI recruits, use cases and contribution to projects.

Your feedback and suggestions are always welcome!

Send an email to Sara & Iulia at:

press@egi.eu



Introducing the new EGI Operations Manager

In March 2018, Matthew Viljoen took over as EGI Operations Manager, leading the Operations team. In this article he introduces himself and outlines the team's strategic priorities for 2018.

As someone with a background in Service Delivery and Operations, I was delighted to be offered the position of Operations Manager within the EGI Foundation, and look forward to the challenges ahead as EGI continues to build on its strength into the future.

Prior to working with EGI, I had been working at the UK Rutherford Appleton Laboratory (RAL) for twelve years. Here I was initially managing and services for the UK National Grid Service and for the EGEE project. I then operated the petabyte scale data storage service for the UK WLCG Tier 1, a principal site for making CERN data available to UK high energy physicists. At RAL I also developed and moved into production a comprehensive new data storage and access service for the Diamond Light Source UK national synchrotron



which now serves multiple petabytes of storage.

In 2015, I started working at EGI and led the Platform Development work package in the EGI-Engage project, which identified missing functionality in the HTC and cloud computing as well as federated data EGI services.

A new challenge

As **EGI Operations Manager**, I will be leading the coordination and evolution of the federated operations team to ensure that it develops an ever stronger and

consistent high quality global service offering, meeting the needs of researchers regardless of their discipline.

This will require forging strong links with national e-Infrastructures and thematic research infrastructures, and streamlining all operations.

Leading the Operations Team, I will endeavour to continue to improve the quality of service delivery. This follows on from the work done in 2017 within EGI Foundation to obtain ISO/IEC

9001 and ISO/IEC 20000 certification of EGI Service Management processes.

This will ensure that we deliver the best possible services, working alongside the significant contribution and hard work from service and resource providers within the EGI Federation. We have always taken data privacy seriously and information security remains a priority for the Operations Team.

We are working to ensure that service delivery will continue in the light of the General Data Protection Regulation (GDPR) coming into force in May 2018, by making sure all requirements of the regulation are met and reflected in service delivery agreements. In preparation for this, all the EGI Operations Team members have obtained ISO/IEC 27001 certification in Information Security Management. More generally, during 2018 we will be working to improve the security, general quality and usability of the EGI Federated Cloud service.

Looking ahead

2018 also brings us into an exciting new phase of EGI as it is leading the development of **EOSC-hub**, the first steps of the **European Open Science Cloud**. This aims at harmonizing major European e-infrastructures such as EGI and **EUDAT** to one consistent service offering for all European users of publicly funded e-Infrastructures for research and academia.

The project is also integrating services developed in the **INDIGO-DataCloud project** and services operated by members of the 18 Research Infrastructures part of the project.

Within EOSC-hub, I am leading the work of coordinated operations and federated service management, creating an IT service management framework covering all aspects of service delivery within the EOSC-hub.

The EOSC-hub project itself is coordinated by the EGI Foundation and we will draw from our

extensive experience in running the largest academic e-Infrastructure to make it a success.

Achieving these goals will be challenging but the EGI Foundation Operations team includes people with extensive technical and skills, working in many domains.

Finally, I'd like to welcome the Operation's team latest recruit, **Bruce Becker**, who joined the team after a long history of close collaboration with EGI and has previously worked at CSIR Meraka in South Africa.

More information

Matthew Viljoen is the new EGI Operations Manager.

Bruce Becker is the new EGI Senior Operations Officer.

EGI highlights at the first EOSC-hub week

The first **EOSC-hub week** is just around the corner, taking place from 16 to 18 April in Málaga, Spain. The EOSC-hub week will be made of two events: **two public days** (16-17 April, open to everyone) and **one all-hands meeting** (18-20 April, open only to EOSC-hub partners).

The public days welcome the participation of service providers, representatives of the research communities and policy makers engaged in the establishment of the European Open Science Cloud.

EGI will participate to the programme with the following sessions and workshops:

- **EGI in the EOSC ERA: status, impact and future steps** - this session will outline EGI's role and next steps within the European Open Science Cloud
- **National e-Infrastructures: status reports and future plans** - this session will allow national service providers and European research organisations to highlight their success stories and future plans.

- **Federated Cloud status and evolution:** the session will be a discussion around the EGI Federated Cloud model and latest technical developments.

The draft programme of the event and instructions for registration are available online.

More information

EOSC-hub project

<http://www.eosc-hub.eu/>

Pollution forecasts in Bulgaria

A recent study powered by computing resources of the Bulgarian Academy of Sciences shows the atmospheric structure and air quality index in Bulgaria.

Air pollution has become a real concern in the past few years in Europe and it poses a growing health risk in many of our cities. According to a 2016 study by the **World Health Organisation**, a lot of respiratory diseases such as lung cancer and chronic pulmonary disease have been associated with air pollution.

Knowing more about air pollution and how it affects our day-to-day lives is key to improving our well-being and quality of living.

Ivelina Georgieva and the team of researchers at the **Bulgarian Academy of Sciences** wanted to know more about this issue and studied the atmospheric structure of Bulgaria and its capital Sofia, for a period of seven years, from 2008 to 2014. They wanted to determine the **Air Quality Index (AQI)** of Sofia and what factors contribute to the air pollution of the city.

Air Quality Index is a trait of the atmosphere that directly reflects the impact of the air pollutants on the health and quality of life of a given population.

To study the air quality of Sofia, Georgieva and her colleagues performed extensive numerical simulations of the atmospheric composition of the city, using up-to-date modelling tools.

The calculations had large computing requirements and Georgieva decided to use the resources of the **Bulgarian Institute of Information and Communication Technologies**.



"For example, 1 day simulation with meteorological model at 16 CPU needs a runtime of 3 hours and 530 MB output storage space, and 7 years simulations require 1 year runtime", Georgieva explains. "Without the HPC cluster and the High-Throughput Compute environment we would not be able to do all this – install and run the models, organise the calculations in jobs, the needed time for the results and of course the needed storage capacity."

The results of the simulations, published in the **Cybernetics and Information Technologies** journal, give an extensive description of the atmospheric composition – its behaviour, origin and health impact.

Georgieva and her team discovered that the pollutants affecting air quality the most are sulphur dioxide and fine particulate matter, with road

transportation being one of main factors that contribute to the development of air pollution.

Part of this work was conducted during the **EGI-InSPIRE project** and was based on the **Bulgarian Chemical Weather Forecast System**: a system of forecasting air quality over Bulgaria.

More information

The study was powered by **High-Throughput Compute** services provided by the **Bulgarian Academy of Sciences**.

BAS is the representative of Bulgaria in the EGI Council.

The EXTraS portal: hidden treasures in the sky

Daniele D'Agostino and Giuseppe La Rocca on the successes of the EXTraS project and how EGI is supporting it

Almost all astrophysical objects, from stars in the surroundings of the solar system to super-massive black holes in the nuclei of very distant galaxies, display a distinctive variability in terms of flux and spectral shape.

Most of the fluctuating phenomena have been discovered with modern soft X-ray observatories, but a lot of information available in data archives is still unexploited.

This is where **the EXTraS project** stands out. Like a treasure-hunt, EXTraS has been harvesting the unexplored information buried in the data collected by ESA's X-ray space observatory XMM-Newton, during 16 years of observation. The results of the project, together with the new analysis software, are now released to the scientific community. The software is important for enhancing the potential of discovery of the XMM-Newton mission, especially because it collects new data every day.

Another success of the EXTraS project is the resulting science gateway – or **the EXTraS portal** – a web platform that gives scientists the possibility to run analysis on the whole XMM-Newton science archive (XSA).

The project's ICT community was heavily involved in the development of this portal: the team focused on closing the gap between the community-specific

presentation layer, the general-purpose middleware and “fabric” layers of the science gateway. As a result, the frontend of the portal was set up with PortalTS, a modern web that is based on up-to-date technologies and architectural approaches such as the use of microservices.

With regards to the compute capabilities, the EXTraS portal relies on the **EGI Cloud Compute** service within the extras-fp7.eu virtual organisation and is freely available to all scientists.

The subsequent computational infrastructure is made of four components: the EGI Applications Database (AppDB), the eToken Server, the CERN Virtual Machine File System infrastructure (CVMFS) and the cloud providers of the EGI Federation.

All the technical details of the EXTraS portal architecture are presented in the project's latest paper: ***A science gateway for Exploring the X-ray Transient and variable sky using EGI Federated Cloud***, published in Future Generation Computer Systems journal.

Thanks to **the EXTraS project**, scientists now have access to a massive discovery space which allows them to explore new questions in the astronomy and astrophysics research fields.



More information

Daniele D'Agostino is Partner & Science Gateway Manager at EXTraS.

Giuseppe La Rocca is eToken Server Manager of the project & responsible for liaising with the EGI Federation cloud providers.

The EXTraS project is led by Andrea De Luca.

EXTraS project

<http://www.extras-fp7.eu/>

EOSCpilot Science Demonstrators: one year later

Gergely Sipos outlines the status of the science demonstrators at the end of the first project year

The **EOSCpilot** project has recently passed its mid-term review in Brussels, receiving positive feedback from the panel of experts. One of the key activities of the project was the support of science demonstrators that act as early adopters of 'to-be EOSC' services.

By the end of the first project year, 15 science demonstrators were selected from diverse scientific domains. Each received advice on available services and was provided with technical support and training to integrate scientific applications with services that fit their purpose. The demonstrators provided feedback to the project and promoted their work to various audiences beyond the consortium.

During this period, the **EGI Foundation** and several members of the **EGI federation** were involved in the selection and support of science demonstrators. Liaison with service and technology providers happened both within Europe and overseas, e.g. with Compute Canada.

Here is the status of the first 10 science demonstrators:

- **TextCrowd**: a virtual research environment for implementing the Natural Language Processing (NLP) encoding/metadata enrichment of textual archaeological reports has been set up.

The service allows researchers to store textual documents in a cloud folder, perform NLP operations, trigger the semantic

enrichment of the text and get information in RDF format.

- **PanCancer**: ported the Butler application for large scale processing of cancer genomes onto Compute Canada and sites of the **EGI Federated Cloud**.

- **Photon/Neutron**: two applications (OnDA and Crystfel) were containerized & tested on HPC & cloud platforms at DESY.

- **DPHEP**: assessed whether CERN's own preservation system could be replaced with 'out-of-the-box' services combined of CVMFS, B2SAFE and Trustworthy Digital Repository. Data ingestion/replication of small datasets from CERN to CINES was put in operation with an average speed of 800Mbps/sec. Data retrieved from CERN was ingested into CINES & replicated to CINECA for open access.

- **ERFI**: Used the **EGI Open Data Platform** to develop a data integration framework for sharing datasets between the ICOS and the IS-ENES RIs.

- **EPOS/VERCE**: Integrated three cloud providers of the EGI Federation with a Virtual Research Environment to compute a realistic scenario of earthquakes shaking using misfit calculations.

- **PROMINENCE**: Deployed SLURM clusters on the EGI Federated Cloud to run containerized MPI-based applications on a hybrid EGI-commercial cloud platform.

- **LOFAR**: Enhanced three pipelines using container technology.

Used Common Workflow Language as workflow engine to link the different pipelines. Provided access to FZJ/PSNC HPC resources to increase the variety of platforms where to execute the different pipelines.

- **CryoEM**: Extended a workflow editor environment to enable workflow import-export capabilities, which allows researchers to work more according to the FAIR principles.

- **EGA Datasets**: Gave access to the B2FIND instance to deposit metadata produced by genomics pipelines.



More information

Gergely Sipos is Customer & Technical Outreach Manager at the EGI Foundation.

Gergely is involved in the WP4 (Science Demonstrators), WP5 (Services) and WP7 (Skills) of EOSCpilot.

Introducing the eXtreme Data Cloud project

Daniele Cesini on the follow-up project of INDIGO-Data Cloud

The **eXtreme Data Cloud project** started in November 2017 as a follow-up of the successful INDIGO-DataCloud project and will last until January 2020. XDC develops scalable technologies for federating storage resources and managing data in highly distributed computing environments. The initiative is funded by H2020 under the EINFRA-21-2017 call "Computing e-infrastructure with extreme large datasets", with the consortium made of 8 partners - University of Cantabria, DESY, CERN, CNRS, AGH, ECRIN ERIC, EGI Foundation and led by **INFN**.

XDC brings together technology providers with long-standing experience in software development and research communities belonging to a broad spectrum of data-intensive scientific disciplines, such as life sciences, biodiversity, clinical research, astrophysics, high-energy physics and photon science.

The different communities provide concrete use cases and heterogeneous requirements in the field of access and management of data volumes at an unprecedented, 'extreme' scale. XDC will integrate some intelligence on top of data management functionalities to meet the specific needs of scientific communities and implement a more flexible computing infrastructure in Europe. The project's services are developed through the use of standards and protocols available on state-of-the-art distributed computing

ecosystems. They can be easily plugged into European e-Infrastructures and in general on cloud based computing environments such as EGI, the European Open Science Cloud (EOSC) and the Worldwide LHC Computing Grid (WLCG).

XDC will release open source software based on already existing components enriched with new functionalities and plugins that can be adopted by as many user communities as possible. Production quality services such as EOS, dCache, Onedata, the Indigo-PaaS Orchestrator will all be integrated into the XDC service catalogue with improved functionalities concerning policy-driven data management, quality-of-service based data movement, smart caching, metadata management, secure storage and encryption.

Services provided by XDC will be scalable to cope with extreme scale scientific experiments like those run at the Large Hadron Collider at CERN and the Cherenkov Telescope Array. Given the XDC software will be released as Open Source platforms available for general exploitation, we foresee a wide adoption of XDC functionalities by less experienced and smaller user communities as well.



EGI & XDC

EGI is contributing to the eXtreme Data Cloud project with dissemination, training & technical exploitation tasks.

EGI is also involved in the quality assurance of the project and ensures that the software outputs of XDC can be easily used on e-Infrastructures.

More information

Daniele Cesini is the Project Coordinator of XDC.

eXtreme Data Cloud

www.extreme-datacloud.eu

Making data and cloud resources interoperable using EUDAT & EGI services

Sara Garavelli on the close collaboration between EUDAT and EGI

From 22 to 25 January 2018, Porto hosted over 230 participants of EUDAT's conference **"Putting the EOSC vision into practice"**.

Attendees included policy makers, service providers and research communities representatives from 25 countries working on various data challenges.

The conference was opened by **Augusto Burgueño Arjona**, Head of the eInfrastructure Unit of the Directorate General for Communications Networks, Content and Technology (DG CONNECT), who presented EOSC as an open science instrument supporting the collaboration between e-Infrastructures & research infrastructures: *EOSC has to be an inclusive ecosystem where horizontal and thematic service providers work together to meet the user needs.*

The discussion on how to put the EOSC vision into practice was addressed with breakout sessions themed on the creation of a thriving data economy.

The topics approached were in the range of: interoperability of services, the role of research infrastructures as thematic service providers, business models and sustainability of data infrastructures, legal issues.

One of the sessions was dedicated to the results accomplished by the collaboration of EGI and EUDAT for the implementation of a production cross-infrastructure offering access to data and high-throughput computing resources.

The work of the collaboration involved concrete user communities in the design process. This helped both EGI and EUDAT to better shape their services to match real needs of their users.

Two major use cases were brought in by the **ICOS** and **ENES** research communities:

The **ICOS** use case focused on the new web-based service offered on the ICOS Carbon Portal to perform 3-dimensional Stochastic Time-Inverted Lagrangian Transport (STILT) atmospheric transport model calculations. The input data consists of meteorological air transport data (from ECMWF), data on greenhouse gas emissions (from EDGAR), and atmospheric observations (from ICOS and other sources). The output data shows time series of concentrations of greenhouse gases and their resulting footprints at selected locations. The data was successfully handled via a combination of EUDAT B2STAGE and B2SAFE services and other network file management systems, while the production model was visualised using EGI computing services.

The **ENES** use case addressed the volume increase of the climate data archive by employing the EUDAT General Execution Framework (GEF) Workflow API, in combination with EUDAT B2 services, and interfacing with the EGI Federated Cloud. Post-processing results

(e.g. data on carbon gas emissions) are sent back to be displayed and further processed at the IS-ENES platform. They can also be downloaded in different common data formats as tailored products via a simple website interface. The input data - typically Coupled Model Intercomparison Project phase 5/6 (CMIP5/CMIP6) data - is now being downloaded locally by climate impact researchers and makes room for a more sustainable data workflow.

To conclude, the session at the EUDAT conference was a great opportunity to present examples of the collaboration between EGI & EUDAT and between generic and thematic service providers in their effort to support researchers' needs.



More information

Sara Garavelli is Outreach Manager at the EUDAT CDI.

eInfraCentral introduces their service catalogue

Jelena Angelis answers frequently asked questions

eInfraCentral was set up to ensure that by 2020 a broader and more varied set of users (including industry) discovers and accesses the existing and developing e-infrastructure capacity. This is done via the creation of a common service catalogue in collaboration with Europe's leading e-infrastructures, and the development of a portal through which to access it.

Why should researchers use the eInfraCentral Portal?

The eInfraCentral Portal makes it easier for researchers to find the information about a broad range of e-infrastructure services they may be interested in, discover new ones, compare various service offerings in one place, and assess the relevance of the offered services by rating them and leaving feedback.

How does eInfraCentral support the development of professional service catalogues?

eInfraCentral is working to aggregate the information available on the service catalogues of a range of e-infrastructure and service providers. We have developed a service catalogue template, building on the experience of the five partner e-Infrastructures. The template is now being rolled out to another 30+ e-infrastructure projects and is helping to foster an alignment of service descriptions. Managers of service catalogues can use our results to improve their service presentation and management and we are organising webinars to support this process.

How does eInfraCentral portal complement existing platforms, for example, the EGI Marketplace or the future EOSC-hub marketplace?

Our mission is to enhance the visibility of such marketplaces, not replace them. eInfraCentral is a portal that gives an opportunity for researchers to compare information on e-infrastructure services from various providers and marketplaces. Once a service is selected, the portal redirects the researcher to the service provider or marketplace where the service can be accessed and ordered.

Is the eInfraCentral service catalogue a basis for the EOSC service catalogue?

Our long-run aim is to ensure the broadest coverage possible of services from European-level and national e-infrastructures. Along with EOSC-hub and OpenAIRE-Advance, eInfraCentral is one of the key building blocks of the future EOSC Portal as recognised in the European Commission's Implementation Roadmap for the European Open Science Cloud.

We will be working with colleagues from these related projects in the coming year to setup the EOSC service catalogue.

How does the project guarantee the quality of the service catalogue?

First of all, the quality is ensured via the neutrality of the eInfraCentral consortium partners – by avoiding conflict of interest, the project team guarantee the objectivity and trustworthiness of the service catalogue. Secondly, selected metrics on service level targets will also be displayed to inform users about the quality of the services available. Thirdly, metrics from the use of the portal will be fed back to the service providers to help them improve their services.

How can you register on the eInfraCentral Portal?

It will shortly be possible to register services and create your user profile via the eInfraCentral Portal. In the meantime, please contact us directly at: contact@eInfraCentral.eu for further information.

More information

Jelena Angelis is the eInfraCentral project manager. Browse the beta version of the catalogue of services:

<http://beta.einfracentral.eu/home>

