Inspired

news from the EGI community
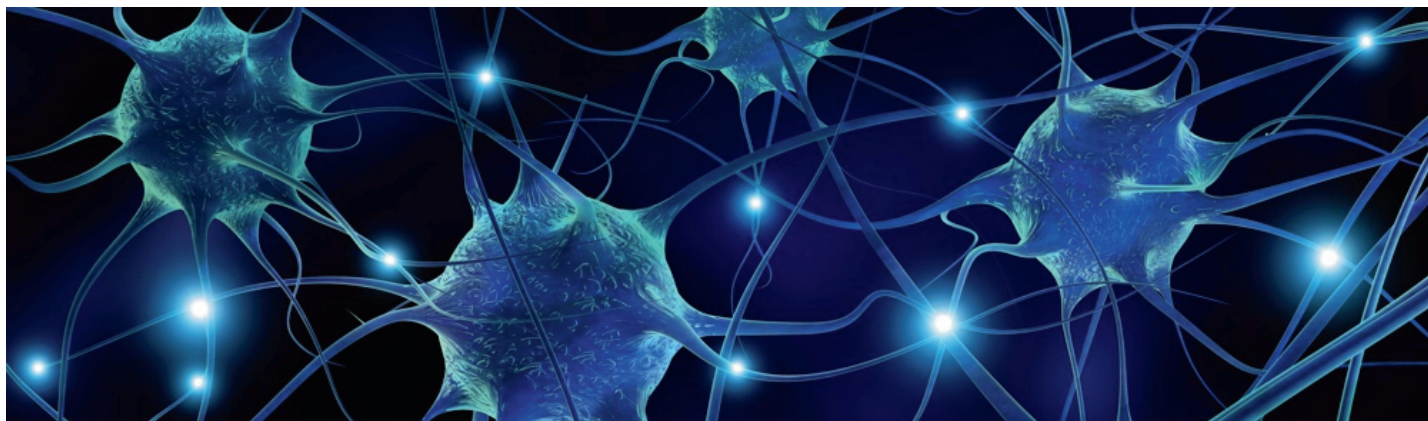
eGi

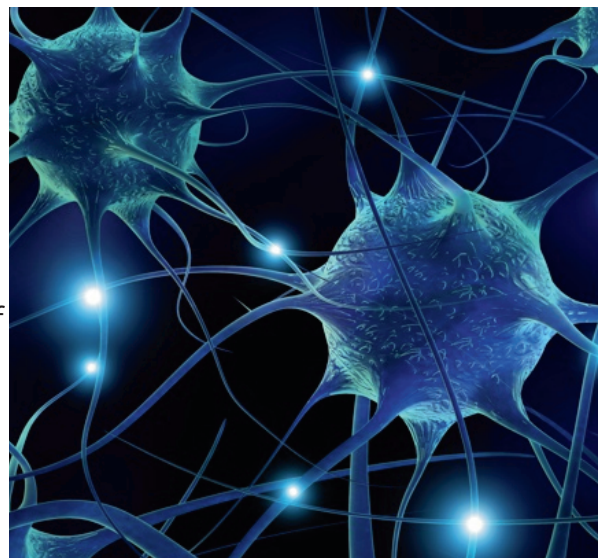**Advanced Computing for Research**

www.egi.eu

Welcome to issue 28!

In the new edition of our newsletter, we focus on the end of the EGI-Engage project and the achievements of the last 30 months.

Your feedback and suggestions are always welcome!

Send an email to Sara & Iulia at:

press@egi.eu

DI4R 2017: Connecting the building blocks for Open Science

Brussels: 30 November - 1 December 2017

The Digital Infrastructures for Research conference will take place this year in Brussels, Belgium, from 30 November to 1 December 2017.

Europe's leading e-infrastructures, EGI, EUDAT, GÉANT, OpenAIRE, PRACE and RDA Europe, invite all researchers, developers and service providers for two days of brainstorming and discussions under the theme "Connecting the building blocks for Open Science".

The 2017 edition of the DI4R conference will showcase the policies, processes, best practices, data and services that, leveraging today's initiatives – national, regional, European and international – are the building blocks of the European Open Science Cloud and European Data Infrastructure.

The main goal of DI4R 2017 is to demonstrate how open science, higher education and innovators can benefit from these building blocks, and ultimately to advance integration and cooperation between initiatives.

The event is collocated with the EOSCpilot 1st Stakeholder Engagement Event taking place on the 28 & 29 November 2017.

Open Call

The Programme Committee, chaired by Franciska de Jong from CLARIN, welcomes abstracts to be submitted to the following Topic Areas:

> Interoperability

> Data science and skills

> Impact evaluation and metrics

> Security, trust and identity

> The EOSC & EDI building blocks

> Business models, sustainability and policies

Registration

Online registration for the event is now open. Early-bird rates are available until 30 October.

More information

DI4R website
www.digitalinfrastructures.eu

DI4R Call for Abstracts
(INDICO website)
indico.egi.eu/indico/event/3455

DI4R Registration
https://www.digitalinfrastructures.eu/registration

What is the European Open Science Cloud?

Iulia Popescu on what we know so far about the European initiative

The idea of a European Open Science Cloud (EOSC) took shape in 2015, as a vision of the EC of a large infrastructure to support and develop open science and open innovation in Europe and beyond. The EOSC is projected to become a reality by 2020 and will be Europe's virtual environment for all researchers to store, manage, analyse and re-use data for research, innovation and educational purposes.

What are open science and open innovation?

In 2015, the European Commission set three goals for research and innovation policy within European Union: Open Innovation, Open Science and Open to the World.

These concepts promote the idea of opening up European research and innovation systems to move towards a reality where knowledge is created through global collaborations, where "the digital and physical are coming together" as described by Carlos Moedas, the Commissioner for Research, Science and Innovation.

One year later, the notions of open science & open innovation took a more concrete form as they become a strategic aim for Europe's scientific landscape outlined in an official EC report: "Open Innovation, Open Science, Open to the World – a vision for Europe".

Open Science

In the report, open science is defined as a new approach to scientific progress based on sharing all available knowledge using new collaborative tools and digital technologies. The outcome would be a shift in the modus operandi of doing research from the early stages and publication of results to sharing them at a larger, global scale. Open science evokes "a change in the scientific landscape towards a public funded science to be more accessible, transparent, collaborative and closer to citizens" (Europe's future: Open Innovation, Open Science, Open to the World).

Open Innovation

The premise behind open innovation is to allow knowledge to circulate more freely and create a culture of new products and markets as well as shared social and economic values.

Towards the European Open Science Cloud

The European Open Science Cloud is envisioned by the European Commission as a supporting landscape to foster open science and open innovation: a network of organisations and infrastructures from various countries and communities that

---

EOSC-related Publications

European Cloud Initiative - Building a competitive data and knowledge economy in Europe (2016) *http://go.egi.eu/eci*

Realising the European Open Science Cloud (2016) *http://go.egi.eu/hleg-eosc*

Open Innovation, Open Science, Open to the World – a vision for Europe (2016) *http://go.egi.eu/ooo*

Europe's future: Open Innovation, Open Science, Open to the World (second report, 2017) *http://go.egi.eu/ooo2*

Report on the governance and financial schemes for the European Open Science Cloud (2017) *http://go.egi.eu/OSPPrep*

The EOSC declaration *(forthcoming)*

supports the open creation and dissemination of knowledge and scientific data (Report on the governance and financial schemes for the European Open Science Cloud).

The creation of EOSC is aimed at removing technical, policy and human barriers, leading to knowledge creation and economic prosperity in Europe.

The European Commission's "European Cloud initiative" publication, issued in April 2016, set an ambitious vision for the European Open Science Cloud: "to give Europe a global lead in scientific data infrastructures and to ensure that European scientists reap the full benefits of data-driven science."

A vision in action

The European Open Science Cloud is intended to set off the ground by federating existing scientific data infrastructures that are now spread across disciplines and EU member states. This will make access to scientific data easier and more efficient (Realising the European Open Science Cloud).

The EOSCpilot project

The EOSCpilot project supports the first phase in the development of the EOSC. The project brings together stakeholders from research infrastructures and e-Infrastructure providers and will engage with funders and policy makers to propose and trial EOSC's governance framework.

The project has already selected 10 science demonstrators functioning as high-profile pilots

EGI and EOSC

EGI endorses the principles of the EOSC Declaration and commits to contribute to the implementation of the European Open Science Cloud:

Governance and funding

> Support the definition, implementation and operation of the EOSC structure with more than 300 data centres in 50 countries.

> Contribute its best practices and experience to the definition of the EOSC policies and ensuring interoperability among suppliers at a global scale.

Data culture and FAIR data

> Provide and improve implementation guidelines for FAIR services (Findable, Accessible, Interoperable, and Re-usable) in the area of advanced compute, federated identity

provisioning, authentication and authorisation.

> Develop the certification schemes and skills necessary to become users or operators of digital research infrastructures and EOSC involving multiple research communities, techno-logy experts and service providers.

Research data services and architecture

> Manage the EOSC-hub Service Integration and Management system.

> Offer advanced compute (Cloud and High-Throughput) and data services from publicly-funded and commercial organisations.

> Operate a federated identity provisioning, authentication and authorisation services for the EOSC users and service providers.

that integrate services and infrastructures to show inter-operability and its benefits in a number of scientific domains: life & earth sciences, high-energy physics, social sciences, physics and astronomy.

The EOSC-hub project

The EOSC-hub project was successfully reviewed by the European Commission and is expected to start in January 2018. The scope of EOSC-hub is to create the integration and management structure of the European Open Science Cloud.

The project will enable an open access to research resources from a myriad of scientific disciplines via a digital hub: an integration system of software and services from major European e-infrastructures and research infrastructures. The

hub will act as an entry point for researchers and innovators to discover, access, and use a variety of advanced data-driven resources.

The consortium of the project is led by the EGI Foundation and brings together more than 100 beneficiaries and linked third parties including research infrastructures, e-Infrastructure providers, SMEs and academic institutions. The project follows the EOSC guidelines recently released by the EC in the EOSC declaration.

More information

Iulia Popescu is a Communications Officer at the EGI Foundation

What is FAIR?

Gergely Sipos summarises the FAIR principles and how EGI is contributing to their implementation

Findable, Accessible, Interoperable, Reusable – FAIR: an acronym that recently became inevitable for anyone involved in research data management, or in any of the initiatives relating to the European Open Science Cloud.

Digital scientific data, tools, workflows and services are becoming available at increased speed and unprecedented scale. Unfortunately, a large segment of these digital objects remains unnoticed, unaccessed or un-used beyond their producer team, limiting our abilities of extracting maximum benefit and knowledge from these research investments.

The FAIR principle was first introduced in a workshop held in Leiden in 2014, where agroup of like-minded academic and private stakeholders met to discuss ways to overcome obstacles in data discovery and reuse.

FAIR consists of 15 elements that define the characteristics needed to enable reuse by third-parties. For example, to be **Findable (F)**, data should:

> have a persistent identifier

> be described by metadata.

Although the elements of the FAIR principles are related they are also independent and separable. The principles may be adhered to in any combination and incrementally, as providers' publishing environments evolve to increasing degrees of 'FAIRness'.

The FAIR principles precede implementation choices, and do not enforce or recommend any specific technology, standard, or implementation-solution. The principles are also not a standard or a specification. They establish a concise and measurable set that can act as a common denominator across institutes, across data and service providers and across disciplines. This means that they can be used as a guide to help data and tool owners to evaluate if their data, tools and services are findable, accessible, interoperable, and reusable.

How EGI contributes to a FAIR digital world

1. We provide technologies, tools and related training for developers who want to create FAIR services (e.g. workflows, tools, VREs), and support them in operating those services within pan-European infrastructures.

2. We facilitate the integration of FAIR data into EGI to offer those data for processing applications that harness the HTC, Cloud and Container services offered within the infrastructure.

3. We provide consultancy and training about service management and sustainability planning for scientific and developer teams to ensure their research output is professionally managed and secured for a long term.

If you want to contribute to the evolution and implementation of the FAIR principles, then a good place to start is the next FORCE11 Conference in Berlin between October 25-27, or the DI4R Conference in Brussels between November 30 – December 1.

More information

Gergely Sipos is the EGI Foundation Customer and Technical Outreach Manager

More about FAIR
Wilkinson, M. D. et al.
doi:10.1038/sdata.2016.18

The 15 FAIR principles
https://www.force11.org/

The EGI Marketplace

Diego Scardaci writes about one of the key results of the EGI-Engage

The EGI Marketplace will be launched in production in the next few weeks.

This tool will be the platform where EGI-related services, delivered by EGI providers and partners, can be promoted, discovered, shared, ordered and accessed. It will include EGI services as well as discipline and community-specific tools and services enabled by EGI and/or provided by third parties.

The EGI Marketplace is designed as an electronic market: it's a platform where services can be advertised and where customers can easily order and access them. In addition, the Marketplace will enhance visibility for resource and service providers, raising awareness of what they can provide as well as helping to promote cross-disciplinary research.

The Marketplace will also act as the main web interface to access the Applications on Demand service (AoDS). Orders for the AoDS could be submitted only from the Marketplace and will be managed via an automatic workflow guaranteeing a quick and smooth access to the applications.

The EGI Marketplace was developed using the PrestaShop technology, a free, open source e-commerce solution largely adopted in the commercial world with a wide community behind it.

Marketplace workflow

> **Authentication**: Managed through the EGI Check-in service, which allows customers to use the credentials of their home organization. Customers are required to register during their first login into the Marketplace to create a customer profile in the database. Part of the data is retrieved by the EGI Check-in service and additional data is gathered through a form.

> **Discover and order services**: The customer browses the Marketplace, finds what is needed and selects the services s/he needs to order.

> **Check-Out**: The service orders are submitted with information complemented with the customer profile. The order then follows the appropriate service order management, according to the EGI Integrated Management System (IMS) processes and procedures.

The first testing phase after the Marketplace becomes operational will include only services provided by EGI. After this is completed, the Marketplace will be opened to the whole EGI collaboration and partners, in particular for publishing thematic community services (i.e. services provided by third partied that rely on EGI services).

The EGI Marketplace will be further enhanced in the EOSC-hub project and will become a key component of the future European Open Science Cloud facilitating the discovery, the order and the access of a large set of services provided by several stakeholders. To reach this aim, several activities are already planned in four main areas:

> Technical development of interfaces to retrieve and publish service data to/from other tools (e.g. the eInfraCentral service registry);

> Publishing of thematic community services;

> Service order management automation;

> Pay-for-use: launch of first commercial offers.

EGI-Engage: a list of key exploitable results

Tiziana Ferrari lists the main achievements of the past 30 months

The EGI-Engage project (full name: Engaging the Research Community towards an Open Science Commons) ran from March 2015 to August 2018 with funds from European Union (EU) Horizon 2020 programme (grant number 654142).

The project brought together 43 partners with a mission to expand the capabilities of a backbone of federated services for compute, storage, data, communication, knowledge and expertise, complementing community-specific capabilities.

What were the key results of EGI-Engage?

Update of the strategy, governance & procurement procedures

The EGI Federation now has a new EGI strategy and a new governance model adapted to the recent evolutions of the e-Infrastructure landscape in Europe. The team also worked on an analysis of opportunities and barriers for cross-border procurement of e-Infrastructure services.

http://go.egi.eu/strategy2020

Integrated Management System and Certification

Foillowing 18 months of work to define a system to plan, implement, monitor and continually improve all business processes under its responsibility, the EGI Foundation was awarded ISO 9001:2015 and ISO/IEC 20000-1:2011 certifications.

http://go.egi.eu/cert

Security policies

The security team updated the policy frameworks to follow the technical evolution of the EGI services and also to make them more general and re-usable by other initiatives.

Thematic services integrated

Thrpoughout the project, the team worked with research communities and research infrastructures to co-design and co-develop new services. Most of these services are now offered as integrated scientific applications with EGI's e-Infrastructure services.

http://go.egi.eu/sat

Improved EGI service portfolio

The EGI service portfolio was redesigned during EGI-Engage, with improved service definitions, and divided in:

External service catalogue

Aimed at researchers, research communities and businesses, contains compute, storage and data, training and applications services provided by the EGI Federation.

http://go.egi.eu/SCpdf

Internal service catalogue

This catalogue contains tools designed to facilitate coordination and improve how the EGI Federation works together. The EGI internal services are provided for the benefit of the EGI Council members and affiliated organisations.

Tools for federated service management

The project supported technological innovation and new services in the area of Service Registry and resource allocation. During EGI-Engage, the EGI Accounting Portal was redeveloped and improved with a new user interface, new views and features. The operational tools were also continuously improved and adapted to satisfy new requirements from service providers and user communities.

Open Data platform

During the project, the team developed a platform designed to make data discoverable and available in an easy way across all EGI federated resources. The EGI DatHub will offer scalable data access and compute capabilities around scientific datasets for scientific groups at the large scale. Once in production, the EGI DataHub will enable data processing in hybrid environments like public and private clouds.

Expanded Federated Cloud computing

The EGI Federated Cloud was expanded with new IaaS capabilities. It now integrates existing commercial and public IaaS Cloud deployments and e-Infrastructures with the current EGI production infrastructures.

EGI Marketplace

A platform where EGI-related services, delivered by EGI providers and partners, can be promoted, discovered, shared, ordered and accessed. It will include EGI services as well as discipline and community-specific tools and services enabled by EGI and/or provided by third parties.

Applications on Demand

A service providing researchers dedicated access to computational and storage resources, as well as other facilities needed to run scientific applications.

http://go.egi.eu/aod

More information

Tiziana Ferrari is the Technical Director of the EGI Foundation and was project coordinatior of EGI-Engage

Introducing the EOSC-hub project

Integrating and managing services for the European Open Science Cloud

EOSC-hub is the H2020 EINFRA12 (A) project proposal submitted by a consortium of 74 partners under the coordination of EGI, EUDAT and INDIGO-DataCloud. The action was positively reviewed by the European Commission and the project is planned to start in early 2018.

The EOSC-hub mission is to contribute to the EOSC imple-mentation by enabling seamless and open access to a system of research data and services provided across nations and multiple disciplines. The project will offer these resources via the Hub – an integration and mana-gement system of the European Open Science Cloud, acting as a European-level entry point for all stakeholders. The Hub will deliver a catalogue of services, software and data from the EGI Federation, EUDAT CDI, INDIGO-DataCloud and research e-Infrastructures. The Hub builds on mature processes, policies and tools from the leading European federated e-Infrastru-ctures to cover the whole life-cycle of services, from planning to delivery. The Hub aggregates services from local, regional and national e-Infrastructures in Europe and worldwide. The Hub will act as a contact point for researchers and innovators to discover, access, use and reuse a broad spectrum of resources for advanced data-driven research. The services will include services in four broad areas: Common, Thematic, Collaborative and Federation.

The catalogue will be open and progressively extended to include data and thematic services from external partners willing to collaborate with the project. In order to do so, the project will run a network of Competence Centres involving early adopters, and a stakeholder engagement programme aiming at reaching out to new user groups and service providers. The project aims at evolving the service catalogue according to the users' requirements and the latest technological develop-ments. Through the virtual access mechanism, more scienti-fic communities and users will have access to services for their scientific discovery and collabo-ration across disciplinary and geographical boundaries.

The project will improve skills and knowledge among resear-chers and service operators by delivering specialised trainings and by establishing competence centres to co-create solutions. The project creates a Joint Digital Innovation Hub that stimulates an ecosystem of industry/SMEs, service providers and resear-chers to support business pilots, market take-up and commercial boost strategies.

Competence Centres: results in EGI-Engage

Gergely Sipos outlines the main achievements of the Competence Centres

EGI-Engage pioneered a new model of engagement and support for Research Infrastructures, based on distributed centres where national initiatives, user communities, technology and service providers join forces to collect and analyse requirements, integrate community-specific applications into state-of-the-art services, foster interoperability across e-Infrastructures, and evolve services through a user-centric development model.

We called them the **Competence Centres** – or CCs for short.

EGI-Engage launched eight CCs 2.5 years ago and while these continue to operate beyond the project, I am pleased to report that the initiative was a success and we plan to take the CC model into the new EOSC-hub project, which is due to start in January 2018.

Here is a summary of what we achieved together with the research communities:

ELIXIR – Life sciences

The ELIXIR CC aimed at evaluating, adopting and promoting technologies and resources from EGI to the wider ELIXIR research community. The team collected representative life science use cases that could benefit from EGI services and then set up a federated cloud infrastructure combining the EGI Federated Cloud with ELIXIR cloud providers and with the ELIXIR Authentication and Authorisation system to implement those use cases. The results were:

> The cBioPortal from CESNET is now ported to and hosted 24/7 to the CESNET cloud site.

> The compute-intensive part of the META-Pipe metagenomics pipeline use case from CSC and Marine metagenomics use case from EMBL-EBI were successfully ported to the federated cloud resources.

> Life scientists are now able to instantiate their own data analysis environment in the cloud (Insyght Comparative Genomics use case from CNRS IFB and PhenoMeNal project use case from EMBL-EBI).

> Users in the US and in Europe can access the same tools and run them on local clouds. (JetStream interoperability use case from the US University of Indiana).

BBMRI - Biobanking

This CC was set up to develop and pilot data processing workflows for sensitive personal data. The work resulted in:

> Expansion of the BiobankCloud platform with the authentication and authorization mechanisms to allow integration with common AAIs (e.g., BBMRI-ERIC AAI, EGI Check-in).

> A demonstrator was ported to the private cluster of the MMCI hospital (in the Czech Republic), where the data analysis workflow has been performed on data of real patients.

> The selection of the biobank workflows most suited as use cases (from CZ, NL, SE).

MoBrain – Structural biology

The CC aimed at lowering barriers for scientists to access online portals and tools for structural biology, building on the work of the WeNMR/INSTRUCT and NeuGrid4You teams. The CC:

> Implemented GPGPU-enabled web interfaces for the AMBER and DisVis online portals, providing an enhanced service by exploiting the faster performance of accelerated computing.

> The Scipion cloud framework was deployed into the EGI Federated Cloud to allow researchers to obtain 3D maps of macromolecular complexes.

> Continued to support a continuous robust use of HTC

resources - the HADDOCK portal, for example, has been sending about 10 million jobs per year.

> In collaboration with the INDIGO-DataCloud project the CC put into production two new web portals making use of the available grid GPGPU resources via Docker containers: DisVis (114 registered users) and PowerFit (79 registered users).

DARIAH – Arts and humanities

The goal of the DARIAH CC was to raise awareness of e-Infrastructures' benefits. To achieve this the CC:

> Established a VO to collect compute and storage resources for the DARIAH community provided by the EGI data centres and cloud sites.

> Developed and deployed the DARIAH Science Gateway with applications (Simple Semantic Search Engine, Parallel Semantic Search Engine, and DBO@Cloud) and three services (Cloud Access, Workflow Development, and File Transfer), and enabled a federated login.

> Established a Working Group within the DARIAH-ERIC community to provide advisory support and promote the benefits of using Cloud infrastructure and the DARIAH CC services beyond the time limits of the EGI-Engage project.

> Coordinated participation and contribution to 15 external events to promote and disseminate the achievements.

LifeWatch – Biodiversity sciences

The CC's goal was to assess and implement requirements of LifeWatch research communities for e-infrastructure services. Throughout the project, the CC:

> Integrated pattern recognition tools and data flow handlers with the IFCA cloud site.

> Compiled a LifeWatch service catalogue of 16 services covering support for ecological observatories, workflows, virtual labs and citizen science.

> Supported the deployment of services via Federated Cloud resources and used 5.5 million CPU hours during the project.

EISCAT_3D – Ionosphere and atmosphere observatory

The CC worked on the development of the EISCAT_3D user portal backed by EGI federated HTC and cloud services. This portal will provide scientists with services to discover, access and analyse (e.g. visualise, mine) data generated by EISCAT_3D.

The EISCAT_3D portal has a working access control and interfaces for data discovery and download as well as a function for analysis job submissions. Moreover, the system facilitated the development of data models and modelling tools within the EISCAT_3D community, and the applicability of operating a central portal service for scientists to interact and compute with EISCAT data.

EPOS – Earth sciences

The CC collected, analysed and compared community needs with EGI technical offerings, resulting in three pilots:

> AAI: demonstrated interoperability between the EPOS AAI and the EGI Check-in service; the prototype was developed based on the UNITY IDM technology and interfaced with Check-in.

> Earthquake simulation (MISFIT): showed how an existing seismology application can be improved by integration with the EGI Federated Cloud.

> Satellite Data: set up an environment with EGI to support the development of new services for satellite data processing. The pilot deployed an EPOS service on top of the Geohazard Thematic Exploitation Platform by Terradue from the satellite data TCS, and linked it to the EGI Federated Cloud to exploit its computing and storage resources.

Disaster mitigation

The CC worked to develop customised IT services to support climate and disaster mitigation researchers in Asia and produced:

> Two web portals to simulate tsunami wave propagation (iCOMCOT) and weather conditions (WRF).

The two portals provide standalone and ease-of-use simulation tools for the entire lifecycle of a tsunami event and numerical weather prediction.

A new interface for the EGI Federated Cloud

Enol Fernández describes the new features of the EGI Applications DataBase

The EGI Applications Database (AppDB) has recently expanded its core functionality with a new dashboard: the Virtual Machine Operations (VMOps) dashboard. VMOps provides a Graphical User Interface (GUI) that performs Virtual Machine (VM) management operations on the EGI Federated Cloud.

The dashboard introduces a user-friendly environment where users can create and manage VMs along with associated storage devices on any of the EGI Federated Cloud providers. It provides a complete view of the deployed applications and a unified resource management experience, independently of the technology driving each of the resource centres of the federation.

Users can create new infrastructure topologies, which include a set of VMs, their associated storage and contextualization, a wizard-like builder that guides them through the selection of the virtual appliances, virtual organisation, resource provider, and the final customisation of the VMs that will be deployed.

Its tight integration with the AppDB Cloud Marketplace allows for an automatic discovery of the appliances which are supported at each resource provider Once a topology has been created, VMOps allows management

actions to be applied both on the set of VMs comprising a topology and on fine-grained actions on each individual VM

The dashboard removes the need for users to own X.509 certificates. They can now log in to the EGI Check-in service and use their institutional credential to see details about their membership to virtual organisations.

VMOps then accesses resources on behalf of the user by employing temporary credentials, which are obtained via the RCAuth MasterPortal, or by employing Per-User-Sub-Proxy technologies, depending on the level of integration of each VO.

The VMOps dashboard also integrates with the EGI ARGO Monitoring Service and with the EGI GOCDB to present the status of the providers and any scheduled downtimes on a single view, allowing users to select the most appropriate providers.
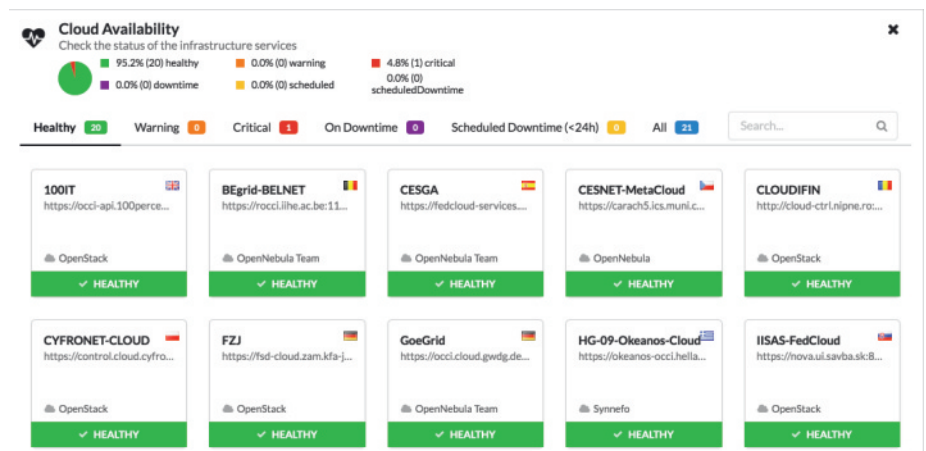
The VMOps dashboard is developed and hosted by the Institute of Accelerating Systems and Applications (IASA) and is built on top of the EGI Federated Cloud. VMOps interacts with providers via the TOSCA standard and uses the Infrastructure Manager (IM) as an IaaS federated access tool.

It has been designed with a scalable architecture composed by a front-end and several back-ends for load balancing and provides a RESTful API which other services can use for integration.

More information

Enol Fernández leads the cloud development activities at the EGI Foundation.

VOMps dashboard

https://dashboard.appdb.egi.eu /vmops

Outcomes of the INDIGO-DataCloud project

Davide Salomoni and Giacinto Donvito on how INDIGO lives up to the Better Software for Better Science motto

INDIGO-DataCloud is an EU-funded project that ran with the objective of developing a new cloud software platform for the scientific community. With this in mind, the team developed tools to facilitate the exploitation of distributed cloud and storage resources through public or private infrastructures.

The project's 30 months were exciting and ripe with results. We believe that the foundations laid by INDIGO will continue to find proper development and adoption in a wide variety of fields, public and private, at the service of science and for the benefit of the overall general public. The key achievements are:

1) The **involvement of scientific user communities** to define and track their requirements: the INDIGO team categorised requests, identified requirements and classified them into three areas: storage, computational and infrastructural.

2) The **identification of technology gaps** linked to concrete use cases. These gaps helped the team to validate the technical implementations and to define the INDIGO technical architecture: a modular framework, fully based on open standards, covering all areas of the cloud stack (IaaS, PaaS, SaaS).

3) **Two major software releases**: MidnightBlue (announced in August 2016) and ElectricIndigo (in April 2017). ElectricIndigo now consists of about 40 open modular components, 50 Docker containers and 170 software packages, all supporting up-to-date open operating systems. This result was accomplished by exploiting key European know-how, reusing and extending open source software and contributing code to upstream projects.

4) The release of **two service catalogues**: a short one, with a high-level description of the INDIGO solutions, and a longer version, with details about components and reports of sample applications.

5) The creation of **two large distributed testbeds** to support development activities and pre-production applications. The testbeds allowed the communities to integrate INDIGO components into scientific applications now deployed in production over public or private infrastructures. So far, scientific communities that integrated INDIGO components belong to the domains of life sciences, physics, structural biology, earth sciences, physics and cultural heritage, among others.

6) The establishment of **collaborations with IBM, ATOS and T-Systems**: The INDIGO team worked with industry leaders to facilitate the adoption and enhancement of INDIGO components.

7) The **participation to the EOSC-hub project**: The INDIGO team will nominate the project's Technical Coordinator. INDIGO will also contribute to the EOSC-hub service catalogue with many of its components: identity and access management, token translation, virtual filesystems (Onedata), advanced IaaS services, neutral access to heterogeneous Cloud resources (Infrastructure Manager), web frontend services and user-level containers.

8) The positive evaluation of **two spin-off projects**: eXtreme-DataCloud and DEEP-HybridDataCloud, due to start in late 2017 / early 2018. These projects will continue to develop many INDIGO components in areas such as data lifecycle management, smart caching, flexible metadata management for big data sets, PaaS-level access to HPC resources and real-time, streaming-based data ingestion and processing. We expect that these developments, once matured to production level, will eventually find a place in the European Open Science Cloud service catalogue, to further enhance and facilitate the work of scientists and resource providers.

More information

Davide Salomoni and **Giacinto Donvito** led the INDIGO-DataCloud project

https://www.indigo-datacloud.eu/

The EDISON community for data science enthusiasts

Themis Athanassiadou introduces the Data Science Pro portal

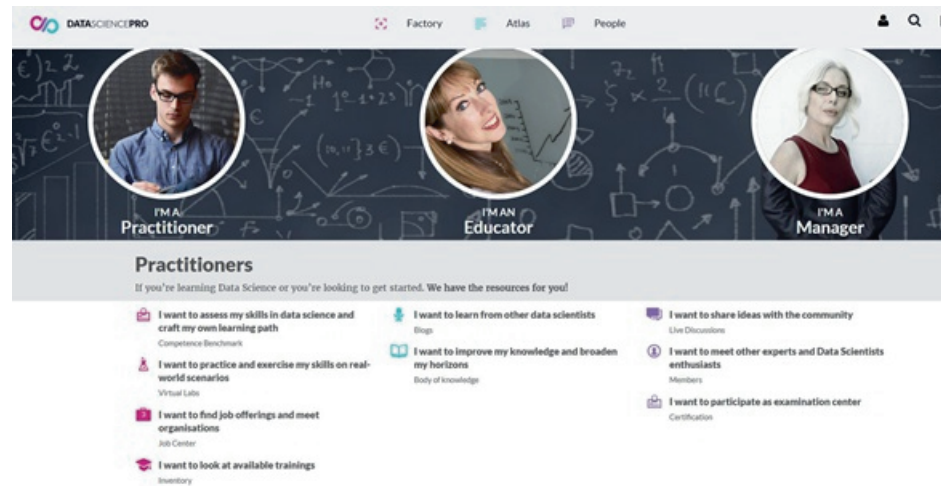Homepage of the Data Science Pro website

The EDISON project, where the EGI Foundation is a consortium member, wants to accelerate the establishment of the data scientist profession, which is defined as an expert who can extract meaningful value from the data collected and also manage the whole lifecycle of data, including supporting scientific data e-Infrastructures.

One of the project's goals is to bring prospective data scientists closer to industry and training organisations and to create an interactive environment where these groups can collaborate and share their knowledge and experiences.

To make this happen, the EDISON team created the Data Science Pro - a community portal for data science enthusiasts. The portal's vision is based on the EDISON Data Science Framework, a collection of documents and guidelines set by EDISON members to define the data science profession.

What does Data Science Pro do?

Data Science Pro gathers a collection of data science courses offered by universities, professors and experts in the field and is built as a dynamic marketplace. It's a place where educators can create trainings, universities can post their programs adhering to EDISON, employers can search for professionals with the right profiles, and job seekers can improve their skills whilst getting the training they need for the jobs they want.

The community portal will be launched in September 2017 with support from the EGI Federated Cloud and the D4Science e-infrastructure. The Data Science Pro portal will provide access to a Virtual Labs environment hosted by EGI community partners and built around practical data sets with education purposes. This is possible by integrating the EGI AAI to allow researchers to use their organisational IDs to access the portal and EGI Federated Cloud resources.

As of February 2017, the implementation phase was finalised with the support of UKIM, a member of MARGI (the Macedonian arm of EGI), The Research Institute for Telecommunications and Cooperation and Engineering Ingeneria Informatica.

The portal is now being tested and validated by a set of users including the EDISON project partners, EDISON Liaison Group members, and selected universities.

Through the Data Science Pro, the EDISON team hopes to bring forth the project's value proposition, which is tailoring each offering according to the supply and demand needs of the Data Science profession and further promoting the EDISON legacy.