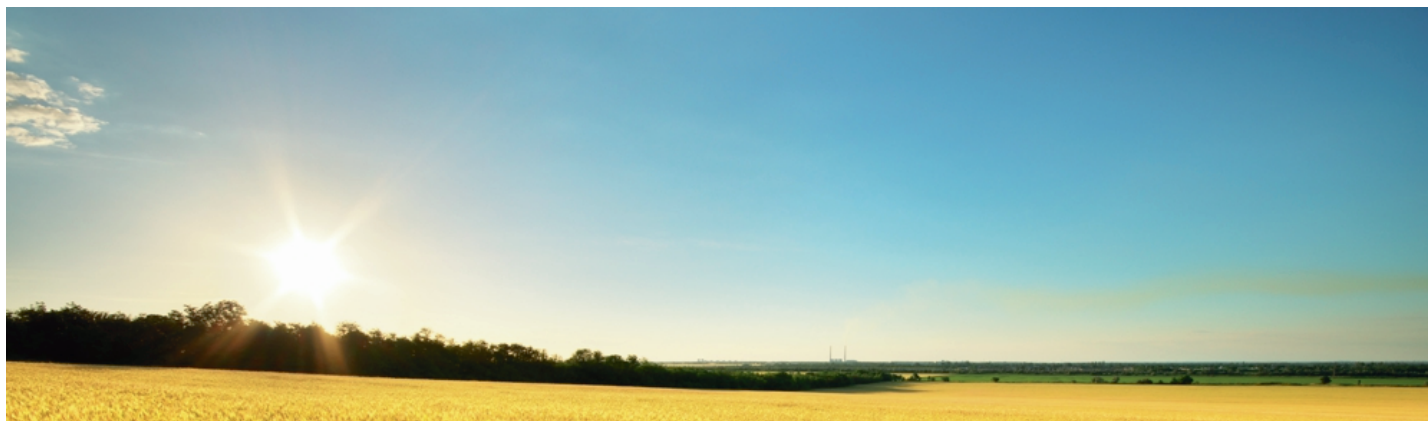


Inspired

ISSUE 26
FEBRUARY 2017

news from the EGI community



THE NEW PROJECTS ISSUE

EOSCPilot: getting the EOSC started

page 3

RISCAPE: e-infrastructure landscapes

page 4

NextGEOSS: Earth Observation

page 5

AGINFRA+: food & agriculture research

page 6

OTHER STORIES

AARC's attribute management pilot

page 7

EGI Federated Cloud architecture

page 8

MORE

01 Computing centres: ReCaS Bari

09 Next EGI Conference: Catania 9-12 May



Advanced Computing
for Research

www.egi.eu

Welcome to issue 26!

In the new edition of our newsletter, we focus on the new projects starting in 2017 and the contributions of the EGI Community.

Your feedback and suggestions are always welcome!

Send an email to Sara & Iulia at:

press@egi.eu



Computing centres: ReCaS Bari data centre

Giacinto Donvito

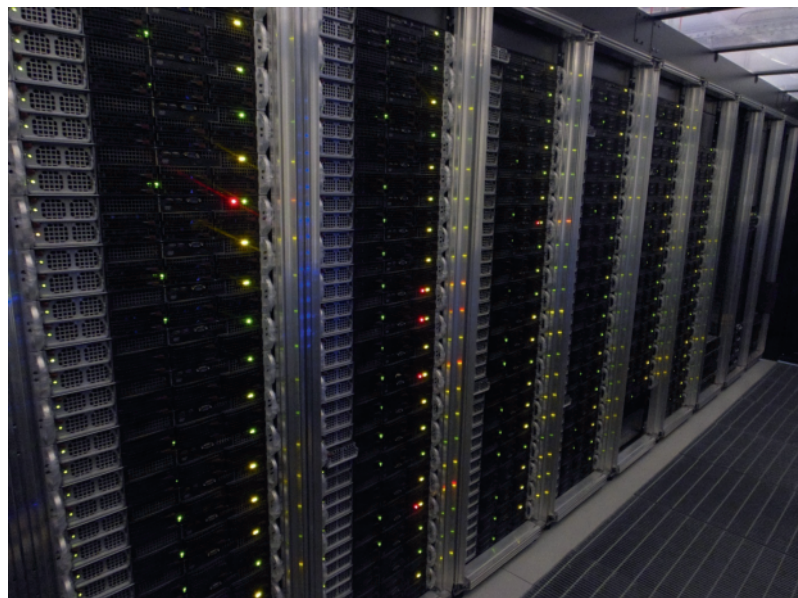
The Bari ReCaS data center has been built by the University of Bari Aldo Moro and the National Institute of Nuclear Physics within the framework of the ReCaS project.

The data centre was completed in July 2015 and inaugurated on July 9, 2015.

The ReCaS-Bari data center is home to 128 servers (equipped with AMD processors) for a total of about 10.000 cores. The data centre has a storage capacity of about 5PByte of disk and 2.5 PByte of tape.

This data centre is serving many communities with diverse technologies: OpenStack, HTCondor, and a typical HPC cluster with infiniband and GPU.

ReCaS Bari is a data centre fully integrated in the EGI federation and is also an EGI Federated Cloud provider. Besides serving scientists at a European level, the data centre also supports about 160 local users from high-energy physics, theoretical physics, medical physics and biosciences.



More information

Bari ReCaS DataCenter

<https://www.recas-bari.it/>

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

EGI data centre helps the IBM research lab to model an exascale computing system

Giuseppe La Rocca tells how Poland's PSNC will support the IBM team to create a new software model

The IBM research team in Zurich set up a project to develop a methodology for estimating the performance, power consumption and cost of exascale systems. The project is called Algorithms and Machines (A&M) and is part of DOME, a joint program with the Netherlands Institute for Radio Astronomy (ASTRON).

The main objective of this collaboration is to develop technologies to support the Square Kilometre Array (SKA), the world's largest radio telescope currently being developed.

The A&M team set out to model an exascale computing system required by the SKA data processing pipeline. This system and the software running on it may allow an early and fast design-space exploration. To construct the software model, the A&M methodology used a platform-independent software analysis tool that measures software properties (such as available scalar and vector instruction mix, parallelism, memory access patterns and communication behaviour). As the software models are extracted at application run-time, they can only be collected on current systems which are orders of magnitude smaller than exascale. To predict the software models at exascale, the methodology used an extrapolation tool which employs advanced statistical techniques. Once extrapolated, the software model was then combined with

a hardware model that captures the performance constraints and dependencies of a computer system. The mathematical formulas allow for a fast exploration of a large design-space of hardware processor- and network-related parameters.

To validate the analytical performance estimates, the A&M team required access to systems with different network topologies, (e.g., fat-tree and dragonfly). The team contacted EGI for support to obtain service access to such systems. EGI identified the **Poznan Supercomputing and Network Center (PSNC)** in Poland as a provider to offer such an environment and kicked started the collaboration.

PSNC offered access to Orzel / Eagle, a supercomputer with a performance of 1.4 PFlops computing power and a fat-tree network interconnect fabric. The A&M team then ran MPI applications of different problem sizes and number of MPI processes on the system, using configurations of two and three-

level fat-tree topologies. The first validation results for the MPI-simple implementation of Graph 500 (a MPI benchmark for analytics workloads) showed that the analytical methodology can estimate the time performance with an accuracy of 82%, which is a very encouraging result. In the future, more MPI applications will be analysed to validate the A&M methodology.

This work was done in the context of the joint ASTRON and IBM DOME project and was funded by the Dutch Ministry of Economic Affairs and the Province of Drenthe. The IBM A&M team is very pleased with the involvement of specialists from PSNC and thankful to EGI for its efforts of intermediating the partnership with PSNC.



More information

Giuseppe La Rocca is part of the EGI Foundation Technical Outreach team.

EOSCpilot: getting the European Open Science Cloud off the ground

Tiziana Ferrari summarises what we learned at the project's kick-off meeting

The EOSCpilot project will support the first phase in the development of European Commission's flagship initiative, the European Open Science Cloud (EOSC): a trusted and open environment for the European scientific community to share, store and re-use scientific data.

The project brings together stakeholders from research infrastructures and e-Infrastructure providers and will engage with related funders and policy makers to create an open environment to use research data, knowledge and services.

EOSCpilot is coordinated by STFC and involves 33 participants in total, including the EGI Foundation on behalf of the EGI community.

The kick-off meeting of the EOSCpilot took place in Amsterdam, from the 17th to the 18th of January.



Goals of the EOSCpilot

The main objectives are:

- > to develop a number of pilots that integrate services and infrastructures to demonstrate interoperability in a number of scientific domains
- > to establish the governance framework for the EOSC and contribute to the development of European open science policy
- > to engage with stakeholders and research communities for an open approach to scientific research

EGI's contribution to the project

EGI will co-lead the WP5 (Services) contributing to the definition of the EOSC federation model, of its overall architecture and the rules of participation as service provider. Within WP5 Services EGI will lead the service pilots task participated by European e-Infrastructures and Research Infrastructures.

EGI will also bring its experience in community engagement, community application integration and training within WP4 (Science Demonstrators) and WP7 (Skills).

EGI is involved in five initial pilots:

Pan-cancer analysis of whole genomes

EOSCpilot goal: accelerate genomic analysis on the EOSC

How: improve the computing competencies and include VM deployment and data management to be easily used

What is the European Open Science Cloud?

The European Open Science Cloud (EOSC) aims to accelerate and support the current transition to more effective Open Science and Open Innovation in the Digital Single Market. It should enable trusted access to services, systems and the re-use of shared scientific data across disciplinary, social and geographical borders.

in *Realising the European Open Science Cloud*
<http://go.egi.eu/hleg-eosc>

The photon-neutron community

EOSCpilot goal: improve the community's computing facilities

How: by creating a virtual platform for all users (e.g., for users with no storage facilities at their home institutes)

The ICOS infrastructure

EOSCpilot goal: enable a comparable data access across multiple research communities

How: by working on data integration and harmonised access

The Parthenos project

EOSCpilot goal: enable an advanced text-based service based on shared semantics

How: by developing new software to enable a semantic enrichment of text sources and make it available on the EOSC.

The High-Energy Physics community (HEP)

EOSCpilot goal: a long-term preservation and larger-scale use of physics data

EGI involvement in the EOSCpilot

WP4 Science Demonstrators

Objective: to develop Science Demonstrators in domains that will show the relevance of the EOSC Services and how they enable data reuse. EGI is involved in 5 use cases (see text for details).

WP5 Services

Objective: to develop service pilots to underpin the science demonstrators and serve as reference implementations for future EOSC services. EGI is involved in the definition of the EOSC service portfolio and the service management framework.

WP7 Skills

Objective: to develop an EOSC education and training strategy and coordinate its delivery. EGI will be involved in developing training material and workshops on the EOSC services.

How: deployment of HEP data in the EOSC via a webportal and open it up to other research communities.

The European Open Science Cloud for Research will bring together the European IT infrastructure and will create an open network for sharing research data and knowledge. The EOSCpilot project is the first step towards reaching this goal.

More information

Tiziana Ferrari is the technical director of the EGI Foundation.

EOSCpilot website
<http://www.eosc-pilot.eu/>

RISCAPE: e-infrastructure landscape

Roberta Piscitelli gives an overview of the project

The RISCAPE project is a three-year project set up to provide an international landscape analysis report on the position and complementarities of the major European research infrastructures. RISCAPE builds on the European Research Infrastructures (RIs) in the ESFRI landscape report (2016) and on the analysis done or currently underway in the H2020 cluster projects.

A landscape analysis involves identifying the key players in a field, sector or geography and classifying them by characteristic

(e.g., type of organization, target beneficiary). This helps nonprofits to understand the broader context in which they are operating, and design their strategy accordingly to maximize their impact.

EGI's role will be to identify international e-Infrastructures in different geographical areas, to examine their common technical features and the areas of societal challenges that they focus on. The EGI team will analyse the e-Infrastructures via interviews and desk research.

The analysis will result in the identification of mechanisms for cooperation including the use of international agreements.

More information

Roberta Piscitelli is part of the EGI strategy and policy team

RISCAPE website
www.riscape.eu/

NextGEOSS: Earth Observation

Sy Holsinger writes about how the EGI Federated Cloud will support the new project

The Group on Earth Observation (GEO) is a partnership of more than 200 national governments and another organisations that envisions a future where decisions and actions for the benefit of humankind are informed by coordinated and sustained Earth observations.

The GEO community is creating a Global Earth Observation System of Systems (GEOSS) to integrate observing systems and share data by connecting existing infrastructures. There are more than 200 million open data resources in GEOSS from over 150 national and regional providers (such as NASA and European Space Agency).

NextGEOSS is a 3.5 year project serving as a European contribution to GEOSS by developing the next generation hub for Earth Observation (EO) data where users can access data and deploy EO-based applications.

The NextGEOSS project consortium is made of 27 organisations from 13 European countries, who recently met in Lisbon from 16 to 18 January to kick-off project activities.



EGI and NextGEOSS

EGI will contribute to NextGEOSS with computing resources made available through the EGI Federated Cloud allowing the project to connect data and cloud computing resources to user communities and enable an integrated network of application support.

This will initially be demonstrated through a number of scientific and business oriented pilots where EGI will offer technical advice and consultancy to identify the best solutions to get the applications up and running on an integrated cloud platform.

NextGEOSS will also stimulate data exploitation by commercial enterprises with a number of business-oriented pilots, therefore EGI will support the defining formal agreements for long-term business relationships beyond the life of the project.

NextGEOSS objectives

- > Implement a single access point, federated data hub using state-of-the-art data mining and discoverability techniques;
- > Implement Quality of Service and community feedback mechanisms on the data hub;
- > Access to the most relevant data sources for Europe, across all major Earth Observation domains.

More information

Sy Holsinger is part of the EGI strategy and policy team

NextGEOSS website
<http://nextgeoss.eu/>

AGINFRA+: food and agriculture research

Gergely Sipos describes how EGI will support the new initiative

AGINFRA+ is a three-year project set up to support the community working on agriculture and food research. AGINFRA+ will further develop the resources and services of the AGINFRA project's research data e-infrastructure.

The AGINFRA+ project builds on the experience and work of its partners: Agroknow, the Alterra Institute of the Wageningen University & Research Center, the National Agronomic Research Institute of France (INRA), the National Institute for Risk Assessment of Germany (BfR), the National Research Council of Italy (CNR), the National and Kapodistrian University of Athens (UOA), Pensoft Publishers (PENSOFT) and EGI.

AGINFRA+ will evolve the AGINFRA data e-infrastructure by using core e-infrastructures such as EGI to provide a sustainable channel addressing user communities around agriculture and food. The project will demonstrate how these scientific communities may carry out rapid development and deployment of innovative applications and workflows, powered by e-infrastructures. This will illustrate the value of AGINFRA as a virtual research environment for the domain of agriculture and food.

The kick-off meeting of the AGINFRA+ project was held from 16 to 17 January at the INRA headquarters in Paris.



EGI's role

EGI will support the uptake of the e-infrastructures in the agriculture domain and will do that via three use cases:

Food security

Goal: investigate what makes strategically important crops such as wheat, maize and rice resistant to extreme weather conditions.

How: integrate high-throughput plant phenotyping data with e-infrastructures, deploy the PHIS webservice and new software to extract information from data in order to find the most resistant types.

Agro-climatic & Economic modelling

Goal: Generate a crop yield forecasting model

How: scale up the computing capabilities to improve forecasts of crop and soil types.

Food safety risk assessment

Goal: support risk assessment using mathematical models

How: improve the current software and existing tools to

allow for repeatability. This will be used to predict the growth of microbes, for instance.

EGI will also participate in the collection and analysis of e-infra requirements coming out of the use cases, offering services from the EGI catalogue and bringing in relevant technologies from the EGI network. EGI will work strongly with CNR, member of the project and provider of the d4Science system, on supporting the use cases.

EGI will also contribute to the sustainability activities in the project. One of the goals of this task is to facilitate business engagement and establish an AGINFRA association that would ensure continuity for the results.

More information

Gergely Sipos leads the EGI Foundation Technical Outreach team.

AGINFRA website
<http://aginfra.eu/>

AARC's attribute management pilot

Paul van Dijk on the work behind AARC's latest two deliverables

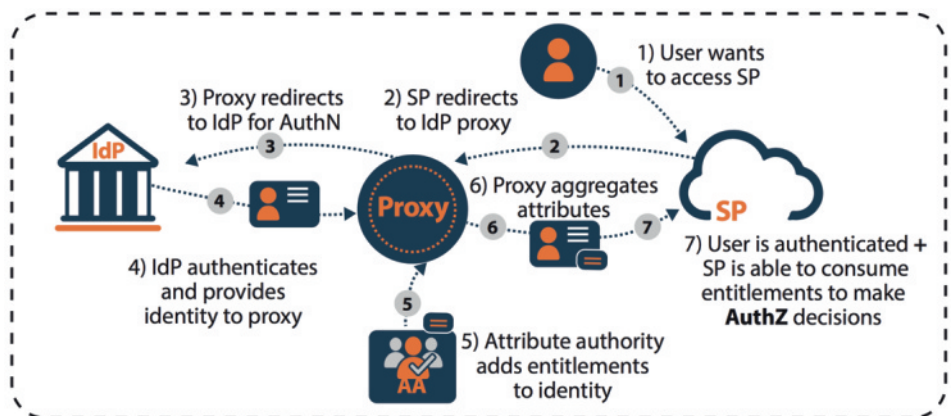
The goal of the AARC project is to guide research communities, cloud providers or commercial service providers to navigate their way through the galaxy of complex technologies that are used for Authentication and Authorisation (AAI). The project is an opportunity for communities to work together, to harmonise AAI approaches and to find suitable components to manage access to shared resources.

Activities

During the lifetime of AARC, we started a comparative analysis among the AAI components used in the e-infrastructures in Europe, we looked at the assurance aspects among these e-infrastructures, and then drafted a blueprint AAI architecture and piloted the integration of several AAI components in production infrastructures. All with the aim of gluing AAI components together and providing a stepping stone for research communities and e-infrastructures to manage access to their shared resources in a scalable way.

Attribute management pilot

One AARC pilot task is led by EGI and focuses on the deployment of components for attribute management and consumption



Piloted, proxy based, flow for authentication and authorization in e-infrastructures

in a federated environment. The use case for this pilot applies to the needs of e-infrastructures (like EGI and EUDAT), and to research infrastructures supporting multiple communities to manage access in a federated setting. And why federated? Because components such as identity providers (IdPs), service providers (SP) and attribute authorities (AA) are typically operated by separate entities.

The picture below provides a simplified view on the attribute management pilot setup where the e-infrastructure or research community can use externally managed attribute authorities (such as CManage and PERUN), aggregate these attributes from different sources in a central proxy (SimpleSAMLphp), and forward the enriched set of attributes in such a way that they can easily be consumed by service providers (such as OpenStack Liberty) to make authorisation decisions.

The latest deliverables

The work of the attribute management pilot led to two new AARC deliverables:

- > Pilots to support guest users solutions, and
- > First report on the pilots deployed by SA1

Both are available for online consultation.

More information

Paul van Dijk is Community Manager for Research at SURFnet

AARC project
<https://aarc-project.eu/>

The EGI Federated Cloud architecture

Enol Fernández gives an overview of the federation model of the EGI Cloud

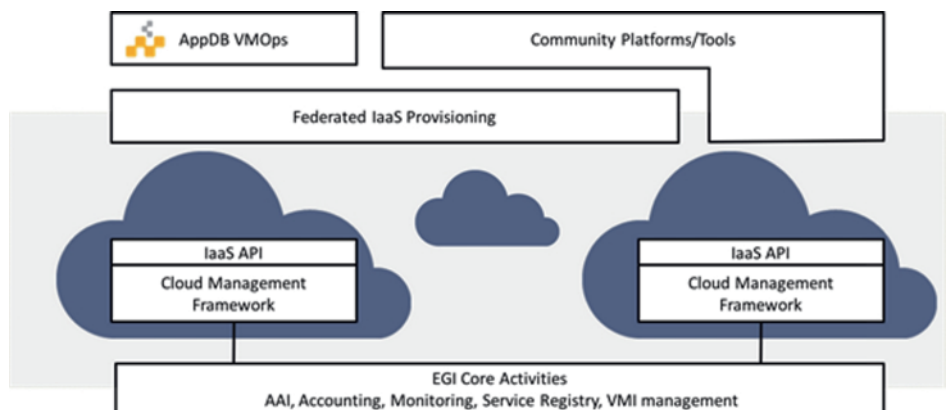
The EGI Federated Cloud integrates community, private and/or public clouds into a scalable computing platform for data and/or compute-driven applications and services. The original architecture was put into production in May 2014.

The EGI community has refined the initial concept and evolved its architecture according to emerging user demands.

The architecture is based on the concept of an abstract Cloud Management Framework (CMF) that supports a set of cloud interfaces to communities.

Each resource centre of the infrastructure operates an instance of this CMF according to its own technology preferences and integrates it with the federation by interacting with EGI core components:

- > **Service registry** for configuration management of federated cloud services.
- > **EGI AAI** for authentication and authorisation across the whole cloud federation.
- > **Accounting** for collecting, and displaying usage information.
- > **Information discovery** about capabilities and services available in the federation.
- > **Virtual Machine image catalogue** and distribution, replicating VM images as needed by the user communities in a secure way.
- > **Monitoring**, performing service availability monitoring and reporting of the distributed cloud service end-points.



This integration is performed by using public interfaces of the supported CMFs, thus minimising the impact on site operations.

Providers are organised the **Open Standards** and **OpenStack** realms, each realm exposing a homogeneous interface.

The realms use different interfaces to offer IaaS capabilities to the users: the Open Standards Realm uses OCCI standard (supported by providers with OpenNebula, OpenStack or Synnefo Cloud Management Frameworks), while the Open Stack Realm uses OpenStack native APIs (support limited to OpenStack providers). Open Stack was introduced in the federation in November 2015 and can co-exist with the Open Standards Realm within the same resource provider.

Users can interact with cloud providers in several ways:

- > Directly using the IaaS APIs of the resource centres to manage individual resources.
- > Leveraging federated IaaS provisioning tools that allow managing and combining resources from different providers and enable the portability of appli-

cation deployments between them. The EGI Federated Cloud task force is currently in the process of evaluating and selecting the best tools for this task.

> Using the AppDB VMops dashboard, a web-based GUI that simplifies the management of VMs on any provider of the EGI infrastructure. AppDB VMops relies on the Infrastructure Manager, a Federated IaaS Provisioning tool developed within INDIGO-DataCloud.

Community Platforms are built on top of the federation, either by using IaaS APIs or Federated IaaS provisioning, and provide community-specific data, tools and applications which can be supported by one or more realms. New realms can be defined by agreeing with the providers on which interfaces and EGI core services to use for the federation.

More information

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

Next EGI Conference: Catania 9-12 May

Iulia Popescu on our next event, to be held alongside the INDIGO Summit 2017

The EGI Conference 2017 and the INDIGO summit 2017 will take place this year in Catania, Italy from 9 to 12 May. The events are hosted by the INFN Catania, part of the Italian National Institute for Nuclear Physics.

The EGI Conference 2017 is the EGI Community's main event of 2017.

The conference programme is focused on the technical road-mapping of EGI, with days dedicated to authorisation and authentication, compute services both HTC and cloud, as well as storage & data services. Sessions on e-infrastructure governance, procurement and business models are also included in the programme.

The INDIGO Summit 2017 will be the flagship event of the INDIGO -DataCloud project, with a focus on user engagement and the INDIGO service catalogue.

This summit explores the solutions provided by the INDIGO software, applying them to concrete use cases brought forward by scientific communities



and resource providers. The event will also be an opportunity for service providers or service integrators to understand how the INDIGO solutions and services are appealing for industry.

About Catania

The two events will be held in Catania and the venue is Le Ciminiere, a modern museum complex housed in a converted sulphur refinery. Catania is a delightful city, with a rich culture and history, hosting many museums, restaurants, churches, parks and theatres and it's very well known for its street food.

The programme for both events

is online on the conference's website and includes joint sessions, for example the EGI-INDIGO workshop on community application support.

The registration for the events is open and more information is available on the event's site.

More information

Iulia Popescu works on EGI-Engage & INDIGO DataCloud communications

EGI Conference 2017 +
INDIGO Summit 2017
<http://go.egi.eu/cf17>

Programme at a glance

Tuesday 9 May	Wed 10 May	Thu 11 May	Fri 12 May
EGI Conference Plenary	EGI Federated Cloud	INDIGO Summit plenary	INDIGO Data Ingestion
EGI AAI Services	Joint Services	EGI NGI Roadmaps: Engagement & Ops	INDIGO brainstorming & conclusions
Competence Centres	EGI Security	EGI Data Services	EGI technical boards (closed)
EOSC implementation	Earth Observation	INDIGO Open Forum, Solutions, Integration & perspectives	
EGI-INDIGO workshop	EGI-INDIGO workshop		