Inspired

news from the EGI community

TOP STORIES

Sergio Andreozzi joins the EC's Open Science Policy Platform

The EGI DataHub prototype

Use case: Cherenkov Telescope Array

Resources committed in 2016

EVENT REPORTS

MORE

Advanced Computing for Research

www.egi.eu

Welcome to issue 25!

In this Autumn edition of our newsletter, we focus on the impact assessment of the first half of the EGI-Engage project, the latest SLAs and EGI Use Cases.

Your feedback and suggestions are always welcome!

Send an email to Sara & Iulia at: press@egi.eu

Computing centres: Torino

Luciano Gaido presents the Torino data centre

The data centre of INFN in Torino hosts an OpenNebula-based IaaS cloud infrastructure comprising about 1700 computing cores and about 2PB of disk storage.

A number of different applications are operated on top of the cloud infrastructure, including:

> a WLCG Tier-2 grid site and a virtual analysis facility for the ALICE experiment at the CERN Large Hadron Collider,

> a Tier-2 grid site for the BES-III Experiment in Beijing and several other smaller virtual computing clusters.

More information

INFN-Torino *http://www.to.infn.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

Sergio Andreozzi joins the EC's Open Science Policy Platform

...and writes about representing EGI in an exciting new expert group

Open Science and international cooperation are top policy priorities for Carlos Moedas, the European Commissioner for research, science and innovation.

A draft of the **European Open Science Agenda** was published earlier this year identifying five broad lines of actions:

fostering open science, removing barriers, developing research infrastructures for Open Science, mainstream Open Access to research results and embedding Open Science in society.

This was later followed by an EC Communication on the European Cloud Initiative explaining the idea of the European Open Science Cloud and by the FAIR (Findable, Accessible, Interoperable and Reuseable) data management guidelines for Horizon 2020 projects. And last month, the High Level Expert Group on the EOSC announced its final report

2016 has definitely been an intense year for Open Science in Europe. Developing Open Science policies is a continuous effort and to foster their implementation, the Commission established the **Open Science Policy Platform (OSPP)**, a group of 25 experts from different stakeholder categories collaborating with the EC to co-develop an Open Science Policy Agenda for Europe.

I was nominated member of the OSPP and this is a clear recognition of the EGI's alignment with the intents of the EC, with a particular focus on digital infrastructures for compute- and data-intensive research and collaboration. In fact, EGI has Open Science at the core of its strategy and is committed to its development. This is clearly visible in our flagship H2020 project EGI-Engage, designed as an instrument to contribute to the implementation of an Open Science Commons.

The OSPP group, chaired by Johannes Vogel, met for the first time on 19 September in Brussels to scope and actions. In a round table, Commissioner Moedas also asked questions about the development of new business models for open access publications, making Open Data 'FAIR' and fostering scientific integrity.

The next meeting is scheduled for 9 December and will focus on the European Open Science Cloud: a federated, globally accessible environment where researchers, innovators, companies and citizens can publish, find and re-use each other's data and tools for research, innovation and educational purposes. Many building blocks are already available in Europe and the goal is to turn them into a collaborative ecosystem able to support participatory production and dissemination of knowledge.

How can you participate?

As a member of the OSPP, I aim to bring the voice and position of the EGI community as inputs to the work of the platforms.

If you are interested in contributing to the development of open science policies and their implementation in Europe:.

subscribe to the ospp-discuss@mailman.egi.eu mailing list to receive updates and participate in the preparatory discussions.

More information

Open Science Policy Platform
http://go.egi.eu/OSPPhome

Sergio Andreozzi leads the Strategy & Policy activities of the EGI Foundation.

Introducing the EGI DataHub prototype

Matthew Viljoen gives an overview of an upcoming EGI service

EGI is proud to announce the launch of the EGI DataHub, a service prototype designed to make data discoverable and available in an easy way across all EGI federated resources.

The EGI DataHub allows users to make their data available using different levels of access: from completely unrestricted open access to open data to authenticated access to closed data sets. This is possible as a result of the seamless integration with the EGI AAI service.

The data hosted on the EGI DataHub can be readily accessible by cloud Virtual Machines (VMs) or running grid jobs thanks to full integration with EGI Federated Cloud and High-Throughput compute resources. The use of protocols such as POSIX and web services guarantees easy and scalable access to data from cloud and HTC applications. This ensures maximum compatibility with existing applications and minimum hassle for developers and users alike.

The EGI DataHub is built on top of the EGI Open Data Platform using Onedata technology to connect a wide range of existing storage services, regardless of their underlying technology (e.g. Lustre, Amazon S3, Ceph, NFS, or dCache).

In complementing existing HTC (High Throughput Computing) EGI resources and the EGI Fedcloud (the existing cloud federation offered by EGI) with the new EGI DataHub, exciting new possibilities are opened for data intensive computing:

> Cloud computing and HTC resources using fast and scalable datasets

> Tiered storage: fast local cache and access to large remote data providers

> Complex orchestration and workflows involving computing and data discovery, ensuring locality of data if needed for efficient processing

A first preview of the EGI DataHub was launched at Krakow, Poland during the DI4R event. This included a training event of the technology using the EGI Fedcloud, ran by Diego Scardaci from EGI. This session proved to be popular and will be repeated at future events.

If you are interested in understanding how the EGI DataHub can help your project then please get in touch with us. We are also keen to talk to you if you have open data or repositories that you wish to benefit by the EGI DataHub making your data more discoverable and accessible to EGI computing resources.

The EGI DataHub, EGI Open Data Platform and the new functionality on the EGI FedCloud that enables integration with these new technologies are being developed as part of the EGI-ENGAGE project and the INDIGO DataCloud project.

More information

Towards European Open Science Commons: The EGI Open Data Platform and The EGI DataHub. Viljoen et al 2016. *Procedia Computer Science* http://go.egi.eu/PCS

Matthew Viljoen is senior operations officer at the EGI Foundation

EGI use case: CTA, the world's leading gamma-ray observatory

High-Throughput Compute and storage solutions for CTA's computational challenges

Credit: Akihiro Ikeshita, Mero-TSK, International

The **Cherenkov Telescope Array (CTA)** brings together 1350 scientists and engineers from 32 countries with the goal of building the world's largest and most sensitive very high-energy gamma-ray observatory.

The CTA will be used to understand the role of high-energy particles in the most violent phenomena of the Universe and to search for annihilating dark matter particles.

The computational challenges

The CTA will be a distributed array of more than 100 telescopes built in La Palma in Spain and the European Space Observatory site in Paranal, Chile, and is expected to produce up to 27 PB per year for long-term archive.

The computational challenges start right there: how to transfer all this data from the telescopes to scientists across the world? Then, the CTA infrastructure needs space to archive the data and enough processing power for data reduction and large-scale Monte Carlo simulations.

CTA aims to be the first ground-based gamma-ray public observatory. This means that a fraction of the observation time will be opened to the whole scientific community. Any scientist in the world will be able to submit an observation proposal to CTA and will have access to the corresponding data, which after a proprietary period, will become public.

This collaborative model, with the participation of the scientific community, implies other major challenges: "we need to make sure we have the capability to provide a unified and efficient access to data, which will follow the common standards of the Virtual Observatory", explains Luisa Arrabito, the CTA computing grid technical coordinator.

Why EGI services?

CTA relies on **High-Throughput Compute** and **Online Storage** services to manage its computational challenges during the project's preparatory phase. The benefits are:

> The possibility to aggregate resources from several sites and benefiting from opportunistic resources

> The long-term reliability of storage resources

> Common and transparent data access for all CTA members

The compute and storage services are provided to the consortium via the **CTA Virtual Organisation** one of EGI's most active user groups.

Since 2012 the consortium has used EGI services to guide the choice of the best two sites to host CTA telescopes and additional simulations to determine the optimal geometry for the telescope array.

Computing & storage resources committed in 2016

Małgorzata Krakowian rounds up one year of service level agreements

At the beginning of 2016, the EGI Foundation established a new framework to mediate service level agreements (SLAs) between user communities and service providers of the federation.

One of the main purposes of an SLA is to create a reliable, trust-based communication channel between researchers and providers to agree on services and their levels and types of support.

Throughout 2016 we worked with a number of research communities from different scientific fields and we are proud to announce the following agreements:

Peachnote

Peachnote is a music score search engine and lets users discover performances and improve their score reading. Peachnote will use EGI's Cloud Compute to give visitors of their website access to a massive amount of music data.

Discipline: Arts and Humanities

Total: 104 virtual CPU cores, 162 GB memory and 8.8 TB storage

LSGC

The Life-Science Grid Community (LSGC) is a virtual research community that supports life-science applications on the EGI federation. LSGC will use EGI High-Throughput Compute and Online Storage for the community's research activities.

Discipline: Life Sciences

Total: 25.9 TB storage and 46.2million CPU hours

Terradue

Terradue is leading the technical development of the European Space Agency's geohazards and hydrology thematic exploitation platforms.

Terradue will use EGI Cloud Compute resources and services to help researchers tackle new data challenges.

Discipline: Earth Sciences

Total: 360 virtual CPU cores, 800 GB memory and 10 TB storage

EXTraS

The EXTraS project is on a mission to develop software tools important for enhancing the discovery potential of ESA's X-ray space observatory XMM-Newton.

EXTraS will use EGI Cloud Compute services for the project's activities.

Discipline: Astronomy and Astropshysics

Total: 50 virtual CPU cores, 200 GB memory and 1 TB storage.

DARIAH

The DARIAH Competence Centre is broadening the usage of e-Infrastructures in the field of the Arts and Humanities.

DARIAH will access EGI's Cloud Compute to help advance this scientific domain.

Discipline: Arts and Humanities

Total: 30 virtual CPU cores, 70 GB of memory and 2TB of storage

DRIHM

The DRIHM project is a prototype research infrastructure for simulating extreme hydro-meteorological events such as flash flooding.

DRIHM will rely on EGI services to better plan research activities and programmes.

Discipline: Earth Sciences

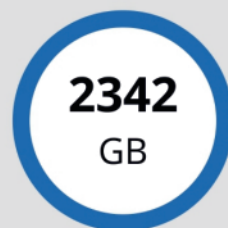Total: 50.5 million CPU hours, 326 GB memory and 10.5 TB storage.
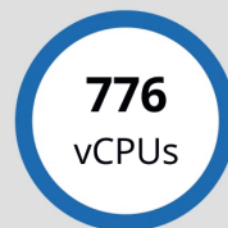
Resources committed in 2016

151.2 million — Computing hours
128.2 TB — Storage
2342 GB — RAM memory
776 vCPUs — Virtual CPUs

as of November 2016

MoBrain

The main objective of the MoBrain competence centre is to lower barriers for scientists to access modern e-Science solutions from micro to macro scales. EGI is contributing to MoBrain's research programmes with HTC and Online Storage services.

Discipline: Life Sciences

Total: 55 million CPU hours, 384 GB memory and 61 TB storage

BILS

The Swedish Bioinformatics Infrastructure for Life Sciences (BILS) is a distributed national research infrastructure proving bioinformatics support to researchers.

BILS relies on Cloud Compute and Online Storage services from EGI.

Discipline: Life Sciences

Total: 172 virtual CPU Cores, 400 GB memory and 9 TB storage.

Why SLAs?

The SLA framework to systematise the commitment of resources between providers and users brings numerous benefits for all parties involved:

For resource providers

> Direct communication with user communities and clarity on expectations;

> Clear responsibilities and rules/policies concerning usage of the resources;

> Recognition and greater visibility to role of the provider by requiring an explicit acknowledgment.

For research communities

> Better communication and clarity on expectations;

> Increased confidence that services will be delivered;

> Easier planning of research activities.

For the EGI Foundation

> Promoting the EGI service value with funding agencies and policy makers at national and European level;

> Fulfilling the EGI mission of bringing scientists closer to the advanced computing services they need for research and innovation

More information

If you would like to know how to better ensure resources and service levels, please contact sla@mailman.egi.eu

Małgorzata Krakowian is a senior operations officer at the EGI Foundation

EGI Federated Cloud training for application & software developers

Giuseppe La Rocca describes the jOCCI-API focussed event

During the September DI4R conference, EGI organized a first training course targeted at developers of high level, cloud-based services The course was attended by 25 software developers who work with research communities to create 'Platform as a Service'(PaaS) and 'Software as a Service' (SaaS) environments on top of the EGI Federated Cloud system.

The main part of the training course focused on the use of the jOCCI-API, a Java library implementing transport functions for rendered OCCI (Open Cloud Computing Interface) queries. OCCI is a RESTful Protocol and API designed to facilitate an interoperable access to cloud-based resources across multiple resource providers and heterogeneous environments. The formal specification is maintained by OGF's OCCI-WG.

This Java library is built on top of jOCCI-core and currently provides HTTP transport functionality with authentication methods and basic requesting interface to easily communicate with OCCI servers.

The EGI Federated Cloud providers

By using this new jOCCI library, PaaS and SaaS developers as well as scientific applications developers have now an alternative to the command line client to interact with EGI Federated Cloud providers. The jOCCI library presented during the course allows the user to interact with EGI Federated Cloud providers.

The participants of the training were introduced to a way of using REST technology to fetch data from the EGI Application Database. With this technology, it is now possible to retrieve a list of virtual appliances, cloud providers and other information programmatically.

The source codes used during this advanced course are available on GitHub. The training material is accessible on the Federated Cloud user support page.

The **EGI Federated Cloud** is an open cloud system that enables the federation of institutional clouds into a scalable platform for data and/or compute driven applications and services. The EGI Federated Cloud, launched in 2014, is currently deployed on 22 institutes across Europe and one SME in the UK and offers an 'Infrastructure as a Service' capabilities for research and industry.

More information

https://mvnrepository.com/artifact/cz.cesnet.cloud/jocci-api
http://occi-wg.org/

Giuseppe La Rocca is part of the technical outreach team at the EGI Foundation.

Design your e-infrastructure workshop

Gergely Sipos highlights the outcomes of the training and consultancy event

The Design Your e-infrastructure workshop was organised on September 27 in Kraków, Poland, by four major European e-infrastructures: EGI, EUDAT, GÉANT and OpenAIRE.

The workshop was a mixed training and consultancy event for research projects, communities and infrastructures, put together to facilitate the uptake of e-infrastructure services.

The afternoon was filled with break-out groups, one for each use case, where researchers had the opportunity to engage directly with e-infrastructures to analyse their use cases, identify suitable services for their use cases and to define implementation plans.

Slides and presentations

http://go.egi.eu/design16

The use cases and a few workshop outcomes were:

Data repository for design research

Challenge: Sille Julie Abildgaard, Research Assistant at the Copenhagen Business School explained that her institute needs to find a way to store, share and analyse the video and audio data recorded during product design.

Outcome: EUDAT and OpenAIRE experts reviewed the requirements with Sille Julie and concluded that the use case requires a tailor-made solution. This will be further developed in a future project proposal.

Multi-scale complex Genomics (MuG VRE)

Challenge: Modesto Orozco and Josep Gelpí, on behalf of the MuG VRE H2020 project, needed a plan to use EUDAT long-term data storage and EGI cloud services for computing and service hosting.

Outcome: The workshop helped them to refine the concept design and to establish connections with service providers. It was also useful for the providers to help them understand gaps in their offerings (e.g. lack of support for graph data in EUDAT B2Share).

OpenDreamKit

Challenge: Dima Pasechnik presented the OpenDreamKit VRE project.

Outcome: The workshop identified the EGI Federated Cloud, the EUDAT B2FIND (metadata catalogue) and the OpenAIRE research community dashboards as relevant services for Dima's project.

Extreme Light Infrastructure (ELI)

Challenge: Tamás Gaizer presented the ELITRANS project, who is helping the ELI sites in the Czech Republic, Hungary and Romania to operate as a single distributed international laser facility of pan-European dimension.

Outcome: The participants identified ELI's e-infrastructure requirements - such as storage services (online and archive), computational resources (HPC, HTC), network connectivity and user portals - and fit them into the scope of the common ELI-wide data management service layer.

The workshop gave me the opportunity to attain updated knowledge within the field of digital infrastructures. It was a great way to locate and establish contact with potential collaborators, service providers and research groups.

Sille Julie Abildgaard

Chronic Lymphocytic Leukemia (CLL)

Challenge: Kostas Stamatopoulos and Fotis Psomopoulos presented the CLL ERIC institute, devoted to improving the lives of CLL patients. They need solutions to establish a persistent repository for patient and clinical trial data, and the compute services to analyse the data.

Outcome: The experts helped Kostas and Fotis to design a demonstrator setup with EGI's cloud services, EUDAT's long-term storage solutions and OpenAIRE's repository services.

EGI at the OpenStack Summit Barcelona 2016

Enol Fernández reports from the event

The OpenStack Summit was held in Barcelona 21-24 October 2016.

OpenStack is an open source software to create private and public clouds. Its the driving technology of most of the EGI cloud resource providers.

The OpenStack summit gathers together thousands of IT leaders, cloud administrators, app developers and contributors on a single event. I went to Barcelona on behalf of the EGI community and with me were colleagues representing the INDIGO-DataCloud project.

I presented the EGI Federated Cloud infrastructure and how it leverages OpenStack and other technologies to support the computational needs of researchers all across Europe. The presentation increased our visibility on the OpenStack community and triggered discussions with the user groups and developers about our federation and how we can collaborate to make OpenStack a better fit for our needs.

Science and Keystone meetings

I also joined the Scientific Working Group and the Keystone Design Sessions.

The Scientific Working Group attracts research and academic cloud providers using OpenStack join to provide requirements and share best practices on the use of the software for their academic users.

Keystone is the authentication and authorization service of OpenStack and plays a crucial role for the EGI federation: it enables the federated users to login at each of our providers. We provided input for new

features that would improve the current support for federated users in Keystone.

EGI & OpenStack

The summit was also a good place to learn about OpenStack and the cloud world. Interoperability and multi-clouds applications, a field where EGI Federated Cloud has worked since its inception, is gaining interest from the community. Containers and container management services are a clear trend for application development and deployment and some of the more packed talks featured these topics.

EGI offers a Cloud Container Compute service that can be enhanced with some of the technology developments presented at Barcelona. Orchestration and TOSCA, a standard for orchestration, are also getting traction and could be adopted by the EGI cloud to provide a better user experience when porting applications to our infrastructure.

The summit increased our involvement in the OpenStack community. Now we are known

to the Scientific Working Group and have joined the Federation, Scientific Datasets & Data Analytics, and GPU activity areas within the group to shape the future developments and share our experience in those areas.

We are already following the specifications of the upcoming release of OpenStack and in collaboration with INDIGO-DataCloud contributing to new ones that will improve our federation.

Next OpenStack summit will be held in May 2017 in Boston. We are already looking forward a bigger participation from EGI!

More information

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

OpenStack
http://www.openstack.org/

Barcelona Summit
https://www.openstack.org/summit/barcelona-2016/

EGI-Engage interim impact report

Tiziana Ferrari answers an EC survey with figures from March 2015-August 2016

In September, the European Commission's DG-CNECT sent an evaluation questionnaire to all e-infrastructure projects. The questionnaire defines five dimensions of impact: relevance, effectiveness, efficiency, coherence and EU added value.

Here is a summary of EGI-Engage's contribution to this initiative.

Map of the EGI e-Infrastructure. The depth of green is proportional to the number of data centres in a country.

Relevance

Q: Provide measures of the usage of the services provided by your e-infrastructure and improvement over time. For example, traffic growth on the networks, usage of supercomputing facilities etc.
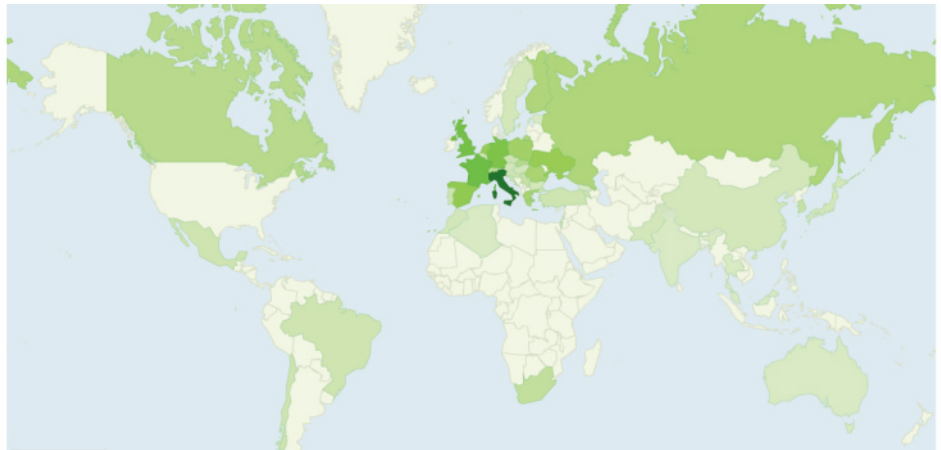
A: Cloud and Storage service capacity federated in EGI have been growing exponentially during the last years. In the last 12 months the use of compute services has increased by +26%.

Compute and Storage capacity KPIs (2016 QR3)

> 770,000 HTC cores (+26% yearly increase)

> 6,646 Cloud cores (in production since May 2014)

> 285 PB of Online storage (+12% yearly increase)

> 281 PB of Archive storage (+42% yearly increase)

Users

> EGI serves an international user community with researchers from all regions of the globe. The active registered users are estimated to be 48,000 in 2016 QR3 (+26% yearly increase).

Effectiveness

Q: Provide benchmarks comparing your e-infrastructure against e-infrastructures in other scientific powerhouses such as USA, China and Japan.

A: EGI is by far the largest High-Throughput Compute (HTC) federation in the world, with +300 data centres across 56 countries. The infrastructure is operated locally primarily at universities and research laboratories in Europe but also in Africa and Arabia, the Asia Pacific region, North and South America.

> The largest HTC infrastructure in the USA is the NSF-funded Open Science Grid with about 100 distributed data centres. XSEDE, the collaborative federated HPC infrastructure for researchers in the US, federates access to 214,000 cores.

Efficiency

Q: Were things built and stand idle or are they used?

A: The EGI usage figures speak for themselves:

> HT Compute: 1,640 Billion CPU hours (last 6 months, 2016 QR3), +26% yearly increase

> HT Compute: 1.7 million jobs/day on average (last 12 months)

> Cloud Compute: 256,000 instantiated VMs (last 12 months)

Q: Does your e-infrastructure assist or harm private business development in an area?

A: EGI supports business development through a programme launched in 2015 aimed at:

> Access to compute, storage and research data services: EGI provides computing capacity to test workflows, models, and applications with dedicated support and consultancy.

> Platform integrators: e.g.

Terradue (developer of the ESA geohazards and hydrology ESA exploitation platform) and CloudSME (provider of SaaS services for engineering and manufacturing).

> Providers of Infrastructure as a Service: To date the EGI Federation includes one cloud SME located in UK (100%IT).

Q: Some e-infrastructures are more domain-specific than others – how are infrastructure needs of different sciences treated?

A: EGI is multidisciplinary and addresses all science disciplines according to the national research priorities. With Compute, Storage and Data Management services it supports 14 fully integrated Research Infrastructures and communities, including for example the LHC - Large Hadron Collider, CTA - Cherenkov Telescope Array or LifeWatch - European Infrastructure for Biodiversity and Ecosystem research.

Coherence

Q: To what extent your e-infrastructure project supports other activities (e.g. internationa-lization by making infrastructures global, open access policies by proving services for open data and open science, etc.). Do you consider your e-infrastructure neutral with respect to other research priorities?

> EGI services are developed, deployed and supported by open source software, to allow sharing of scientific applications in the form of cloud virtual appliances.

> EGI has a service-oriented

approach in which research communities are not only consumers but also providers of innovation and services for community-specific capabilities in one integrated catalogue.

> Co-design is supported through Competence Centres involving Research Infrastructures and research communities that contribute to the development of human networks of experts, services providers and technology providers.

EU Added Value

Q: Could you provide examples about how your e-infrastructure adds value at EU level?

> EGI federates national e-Infrastructures and resources from international research organizations and provides open access to advanced research computing tools, in support of the European research Area (ERA).

> EGI supports the Open Science Commons, an overarching policy designed to overcome the barriers preventing the implementation of the ERA. This concept builds on the principles of Open Science, or Science 2.0, supported by the European Commission and others and also the commons principle for

management of shared resources. EGI is seeking other e-Infrastructures, research infrastructures and stakeholders from the digital research community to collaborate in constructing the Open Science Commons.

> EGI contributes to digital innovation by providing access to the digital capabilities or resources available to the research sector for the exploitation of commercially viable research data, and to support testing and experimentation. The most advanced collaborations are those involving industry and SMEs doing business through the exploitation of Earth Observation data including Copernicus data.

More information

The EGI-Engage Interim Impact Report (March 2015-August 2016) is available online: http://go.egi.eu/iir

Tiziana Ferrari is the EGI Foundation's Technical Director