

# Inspired

ISSUE 20  
JULY 2015

news from the EGI community



(Tango7174 / Wikimedia Commons)

## TOP STORIES

### Data-driven economy - what is it?

*page 2*

### The INDIGO-DataCloud project

*page 3*

### Competence Centres: first updates

*page 4*

## MORE

- 01 EGI is looking for Champions!
- 05 The Leishmaniasis Virtual Laboratory
- 06 The Civic Epistemologies project
- 07 Scientific computing for physics research in Romania
- 07 EGI Community Forum 2015 in Bari



**European Grid  
Infrastructure**

[www.egi.eu](http://www.egi.eu)

# This Issue

In this edition of *Inspired*:

- > **Sergio Andreozzi** explains the 'data-driven economy' buzzword and how can EGI contribute
- > **Davide Salomoni** introduces the INDIGO-DataCloud project created
- > **Gergely Sipos** reports the results of the first few months of the Competence Centres

Also:

- > **Leishmaniasis Virtual Laboratory**: single access points for researchers and public health workers
- > **Civic Epistemologies**: e-networked citizen science
- > **Scientific computing for physics research in Romania**: challenges ahead

Send your feedback and suggestions to:  
sara.coelho@egi.eu Thanks!



Our next event will be held in the port city of Bari in the Italian coast of the Adriatic Sea. (Image: Tango7174 / Wikimedia Commons)

---

## EGI is looking for Champions

*We invite scientists to apply for travel support today!*

EGI launched a Champions Programme to offer travel support to researchers that use EGI's computing services for their work.

EGI Champions are practicing scientists currently engaged in computationally demanding research projects. They know from first-hand the benefits of working with EGI.

They are enthusiastic early-career scientists keen to go to conferences and spread the word about EGI.

In exchange for their enthusiasm, EGI will contribute to travel expenses and provide networking opportunities with other like-minded scientists across Europe.

EGI is looking for scientists and researchers that:

- > Use computing services provided by the EGI Federation in their research (e.g., grid jobs, cloud, science gateways)
- > Want to be part of a network with a shared interest in distributed computing
- > Can talk passionately to fellow scientists about the benefits of using EGI's computing services
- > Are willing to travel to conferences to present their scientific results

If you have these characteristics, you can apply for travel support today.

The call for EGI Champions will remain open until May 2016.

### More information

**About the Champions programme**  
<http://go.egi.eu/champs>

**Application form**  
<http://go.egi.eu/apply>

**Any question?**  
[press@egi.eu](mailto:press@egi.eu)

# A data-driven economy: what is it? And what can EGI do?

*Sergio Andreozzi explores the role EGI can have in the wider European context*

One year ago, the European Commission launched a new strategy on Big Data with the aim to support and accelerate the transition towards a data-driven economy in Europe. As the EGI ultimate goal is to empower researchers to carry out data and compute-intensive science and innovation, we want to explore what synergies can take place.

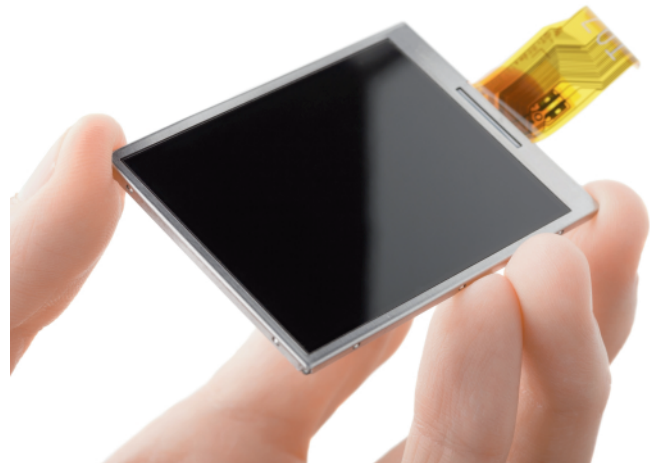
## What is a data-driven economy?

According to the EC, a data-driven economy should have at least the following features:

- 1) availability of good quality, reliable and interoperable datasets from (e.g. scientific experiments or healthcare) with no geographical restrictions;
- 2) availability of infrastructures to reach the data (fast Internet) and to manage/process them (e.g. with grid or cloud resources);
- 3) a strong skills base and trained people to meet the demands of the labour market;
- 4) cooperation between public and private sector, to ensure technology and knowledge transfer;
- 5) a public sector that adopts and stimulates new data services and digital goods.

## How does EGI relate to this?

Research communities served by EGI create, maintain and analyse large datasets. In many cases, there are opportunities that would be unlocked if these data would be made publicly available to a wider audience (e.g. to other research



communities or to the business sector) together with the tools and services to analyse them. For instance, earth observation, climate or life science data. The EGI services and solutions used by researchers (e.g. software, online tools, e-infrastructure services), could be reused by the private sector, especially in the pre-commercial phase, so to stimulate innovation. EGI, as a community, also has the skillset needed to handle big data applications, thus also offering collaboration opportunities in the area of training and skills development.

## How can EGI contribute to a data-driven economy?

With the start of the EGI-Engage project, we are investing more in developing business relationships with industry and SMEs. This provides a platform where industry and SMEs can explore big data technologies with general purpose compute and data services that foster reuse of research data. The goal is to maximise the impact of results from the research and to improve the competitiveness of private companies that may benefit from accessing a big data ecosystem.

## Who are our key partners?

For the last three years, EGI has been collaborating with the Helix Nebula Initiative to ensure the development of hybrid cloud services in Europe. Recently, EGI.eu became a member of the Big Data Value Association to contribute to the activities on behalf of the EGI community. By participating in BDVA task forces, we will promote an active role of EGI members in the various areas of the big data value so to increase the opportunities to transfer of technology, knowledge and services to the private sector that can benefit from big data sets, applications, services or expertise. On the other side, collaborating with the BDVA can help us to expand our network of contacts, understand the requirements from the private sector better and discover different approaches to big data challenges.

## More information

Data-driven economy  
<http://go.egi.eu/ddec>

BDVA  
<http://www.bdva.eu/>

# INDIGO-DataCloud: open source software for science

*Davide Salomoni introduces the project created to set up an open source data and computing platform*

INDIGO-DataCloud (INDIGO for short) is a project led by the Italian National Institute for Nuclear Physics (INFN) with 26 beneficiaries in 11 countries, including e-Infrastructure providers, resource centers, scientific communities and industrial partners. EGI-Engage is also a partner and a key player in collecting and elaborating the use cases INDIGO is targeting to work on.

The mission of INDIGO is to develop an open source data and computing platform, aimed at scientific communities, able to be deployed over hybrid e-infrastructures.

The main ingredients for success are: the full participation of European scientists from multiple disciplines, and the presence of developers, resource centers and infrastructure providers with experience in distributed computing solutions. While many offerings for the provisioning of Cloud-based resources exist today, especially IaaS, the INDIGO consortium has identified several issues that prevent full exploitation by scientific communities. INDIGO aims to fill these technological gaps by developing the missing services from concrete use cases selected by the partner scientific communities.

The first months of activity focused on INDIGO's software architecture. The project will extend the capabilities of existing IaaS solutions, considering for example how to exploit diverse storage architectures and how to efficiently use hardware

resources, e.g. through simplified adoption of containers. INDIGO will then extend existing PaaS solutions, allowing the integration of grid and cloud services provided by EGI, WLCG, EUDAT, PRACE and by the HelixNebula Initiative. These will be made available through AAI services compliant with GEANT's policies. Finally, it will provide a flexible and modular presentation layer connected to the IaaS and PaaS extensions developed by the project, supporting reusable portal components and the adoption of dynamic workflows.

The scientific communities involved in INDIGO include LifeWatch (life science), WLCG (the Worldwide LHC Computing Grid), EuroBioImaging (medical imaging biobanks), INSTRUCT (molecular dynamics), the astrophysics experiments LBT and CTA, ENES (climate modeling), ELIXIR (biological and medical science), EMSO (seafloor observatories), DARIAH (big data in arts and humanities) and other communities linked to libraries, museums and the EGI Competence Centers. INDIGO also has a strong link with a number of INFRADEV projects, key infrastructure players such as EUDAT2020 and PRACE, large research data centers across Europe and industrial partners. Connections and interviews with these communities resulted in more than 100 technical requirements collected by WP2 (Network Activity); these are currently being evaluated, ranked into different categories centered on Computational,



**INDIGO - DataCloud**

*"INDIGO aims to fill technological gaps by developing the missing services from concrete use cases selected by scientific communities."*

Storage or Infrastructural needs, and discussed at the technical level, focusing on the development of the INDIGO IaaS, PaaS and SaaS/portal components.

By the end of the project, INDIGO will release an open source software suite capable of supporting many scientific communities in both private and public clouds, fostering advancement and exploitation of distributed services across a European backbone for science.

## More information

**Davide Salomoni** is INDIGO's project coordinator

**Full name**  
INtegrating Distributed data Infrastructures for Global Exploitation

Funded by the EC under the Horizon 2020 call EINFRA1-2014

**Start/end:** Apr2015-Sep2017

*url:* [www.indigo-datacloud.eu/](http://www.indigo-datacloud.eu/)

# First updates from the Competence Centres

*Gergely Sipos reports the results of the first few months of work*

## ELIXIR

The ELIXIR CC is working with various EGI product teams to customise EGI services for evaluation and inclusion within the ELIXIR Compute Platform. These services are expected to provide key contributions to the platform: service registry, Virtual Machine Image marketplace, service availability and reliability monitoring service, usage accounting system, cloud computing resources and user certificate generation and management service.

So far, initial tests have been carried out and recommended developments and customisations are ongoing. Use cases for more profound tests and applications will be selected during Autumn 2015 as the ELIXIR Compute Platform matures.

## MoBrain

MoBrain shared requirements concerning the discovery of GPGPU resources within EGI with the team working on the implementation of related features in the production system. Discussions have started between the WestLife VRE project, the NeuGRID4you community and the WeNMR community about the concept and architecture of an integrated Virtual Research Environment that would enable multi-scale simulations 'from molecules to brain'. Investigations to host the MoBrain portal service on the EGI Federated Cloud have begun.

## DARIAH

The CC members are in the process of establishing an official Working Group within the DARIAH community to integrate the CC activities into the core DARIAH work streams. This Working Group will collect requirements from digital humanities researchers and will use these to fine-tune the implementation of the integrated science gateway that will be setup for the support of storage, data discovery and data analysis/simulation needs of the community.

## LifeWatch

LifeWatch has deployed and is currently integrating a site into the EGI Federated Cloud (OpenStack cloud deployment in Sevilla).

They aim to be able to offer it as a resource to the LifeWatch community soon.

A technical meeting is being organised by the CC for 2-4 September in Santander.

## EISCAT\_3D

CC members are working on a document to gather the requirements for the EISCAT-3D portal and to capture the design specification of its functional components and applied data model.

The document will drive the portal implementation work of the CC contributing to the effort in EISCAT\_3D to establish an integrated e-infrastructure to handle, share and use big scientific data from incoherent scatter radars.

## EPOS

The CC members agreed to focus the work on two use cases:

- 1) an AAI use case dealing with delegation and incorporating data retrieval, transfer and linkage, and
- 2) a computational use case dealing with the integration of computational resources and services for EPOS. The use cases are documented in a public report. The CC expanded to include the EPOS Satellite Community (EPOS WG14) who will contribute to the specification and implementation of the computational use case.

## Disaster mitigation

The CC is working on the definition of data collection and preparation use cases for tsunami and weather simulation. This requires internal training within the CC about tools, models, analysis and data existing at the partners, as well as engagement with research institutes, modelling groups and data providers in the region. The next DMCC face-to-face meeting will be held in Kuala Lumpur, Malaysia in August.

## BBMRI

*Work starts in October 2015.*

### More information

The Competence Centres (CCs) are supported by EGI-Engage's task SA2 lead by Gergely Sipos.

<http://go.egi.eu/SA2>

# Leishmaniasis Virtual Laboratory: single access points for researchers and public health workers

## *The LeishVL team tells the story behind the portal*

The Leishmaniasis Virtual Laboratory (LeishVL), focuses on surveillance of the *Leishmania* parasite and its vectors, and contains a number of online collaborative and research tools for advancing the current knowledge about the disease. The virtual laboratory allows public health workers, researchers, international organisations and pharmaceutical companies to access and to supply information or data about the *Leishmania* parasite.

LeishVL is the result of the joint effort of European and Brazilian research teams involved in the EUBrazil Cloud Connect project. Leishmaniasis is one of the most neglected diseases in the world. It affects the poorest of the poor in developing countries with 350 million people considered at risk of contracting leishmaniasis and 1.5-2 million new cases every year. Risk factors include environmental and human-made changes such as urbanization of rural areas, immune status (due to co-infection with HIV), malnutrition or drug resistance.

LeishVL includes data collections owned by the Spanish Instituto de Salud Carlos III and by the Brazilian Centro de Pesquisas René Rachou/Fiocruz (COLFLEB and CLIOC). Leishmaniasis-related information available in GenBank and PubMed is also included in the LeishVL. Records are curated and enhanced with additional annotations that include geospatial coordinates of where the DNA samples were

collected, better taxonomical classification, and gene identification.

The main objective of the LeishVL is to provide a database of quality data composed of DNA sequences, journal citations, and other clinical and biological data related to Leishmaniasis, as well as the parasite (*Leishmania* spp.) that causes the disease and phlebotomine sandflies that spread it. The LeishVL contributes to improve the quality of these datasets and make them available for *Leishmania* and sandfly species identification and for other tasks that are of paramount importance for the prevention and control of Leishmaniasis.

Features include:

- > Data collections curated by experts
- > Connecting communities as the means to success
- > Augmented metadata
- > Data provenance maintenance
- > Continuous updates
- > Open access & interoperability

Data access and pipeline execution are provided in the LeishVL through a RESTful JSON API that supports open standards like GeoJSON, W3C PROV and implements OAuth 2.0 as the authentication mechanism. LeishVL services are built on top of the EUBrazilCC infrastructure, which adopts the EGI Federated Cloud model by supporting open standards: OCCi for IaaS management, and VOMS for

resource level authorisation management.

The Leishmaniasis Virtual Laboratory is an open, collaborative environment for research in leishmaniasis is available on the EGI Virtual Appliances Marketplace.



### EUBrazil Cloud Connect

The EUBrazilCC project focuses on an interoperable-by-design approach to manage a heterogeneous infrastructure, which includes private clouds, super-computing and opportunistic desktop resources, using different middleware to manage the IaaS resources. The aim is to create an intercontinental federated eInfrastructure for scientific usage.

### More information

The LeishVL team is:  
Erik Torres, Israel Cruz,  
Gabriel Melim, Jacek Cala  
and Ignacio Blanquer

The Virtual Laboratory:  
<http://go.egi.eu/leishvl>

LeishVL in the EGI  
Application Database:  
[http://go.egi.eu/LVL\\_AppDB](http://go.egi.eu/LVL_AppDB)

# Civic Epistemologies: e-networked citizen science

*Antonella Fresa and Claudia Pierotti on the plans for a roadmap for digital cultural heritage research*

Civic Epistemologies is a coordination action funded by the European Commission under FP7 from August 2014 until November 2015. The project aims to produce a roadmap for an open e-Infrastructure to support the participation by European citizens in digital cultural heritage and humanities research. Artists and the creative sector can stimulate this process.

The roadmap is expected to benefit - and contribute to the technology advancement of - many other sectors in addition to humanities.

The Civic Epistemologies' roadmap rests upon the concept of "citizen science", which refers to initiatives, led by academics, where citizens are engaged in scientific research on a voluntary basis, contributing to data gathering (generally referred as "crowdsourcing") as well as participating in the creation of new knowledge. It takes into account results from discussions with stakeholders (policy makers, cultural/academic institutions, researchers, citizens and citizen association, creative enterprises, e-Infrastructure communities) and outputs from the project's activities:

> **Three international workshops and three focus groups** were organised by the project, with the participation of stakeholders from all over Europe and New Zealand, to analyse workflows and requirements of citizen science practices, highlighting the need for combining user centred tools together with customised software/apps in

order to enable real cooperation between researchers, cultural institutions and citizens.

> **A pilot study** in Ireland explored the interaction of teenage volunteers with place names and place-based heritage through interviews with senior citizens, conducted by the youngsters as part of their school curricula. The pilot resulted in the design of an Intergenerational Digital Toolkit, a website including audio segments of the interviews and guidelines on how to set up an oral history inquiry.

> **Two case studies** in Coventry (UK) examined how art workers and homeless people from a social enterprise in Oxford engaged with cultural heritage via digital technologies and to explore how the visitor app realised for the Coventry City Council fits the needs and expectations of its potential stakeholders.

The project's website has a numerous collection of examples of best practices and successful initiatives in the domain of citizen science and digital cultural heritage.

Project partner EGI.eu contributed with a market analysis, looking at alternative revenue streams that may cover, alone or

in a mix model, the costs of a citizen science focused e-Infrastructure. A related implementation strategy is upcoming.

An outlined Strategic Research Agenda is under development to pinpoint the research activities aimed at the development of citizen science practices in the domain of the digital cultural heritage and humanities, powered by dedicated e-Infrastructure services.

A set of learning resources is offered as a reference point for the development of new skills enabling heritage professionals, humanities researchers and citizens to interact in joined researches.

All of this flows into the roadmap, whose intermediate versions are available on the project's website for comments and contributions. The roadmap's definitive version will be presented at the project's final conference in Berlin, on 12-13 November 2015.



## More information

Civic Epistemologies  
[www.civic-epistemologies.eu](http://www.civic-epistemologies.eu)

# Scientific computing for physics research in Romania: challenges ahead

*Alex Nicolin*

Romania's recent full membership at CERN as well as the Extreme Light Infrastructure - Nuclear Physics (ELI-NP) research facility and numerous other smaller-scale research projects, are shaping a very active community dedicated to scientific computing.

Grid computing, in particular, has been the focal point of the computing support offered within the Worldwide LHC Computing Grid (WLCG) to three of the CERN experiments, namely, ALICE, ATLAS and LHCb, in which Romania is involved. Along with high-energy physics, the grid offers efficient computational solutions both for communities with well-established computing needs, such as the biophysics, solid-state physics, and the nuclear physics communities, to name only a few, and emergent communities such as the one formed around ELI-NP. This new

infrastructure is currently under construction in Magurele, within the premises of Horia Hulubei National Institute for Physics and Nuclear Engineering (IFIN-HH), and is expected to be operational in 2018.

ELI-NP will bring a new opportunity to study the fundamental processes of light-matter interaction using the most intense lasers ever built. This facility and its extreme lasers will contribute to advancements in the frontiers of fundamental physics, new nuclear physics and astrophysics as well as applications in nuclear materials, radioactive waste management, material science and life sciences.

On the computing side ELI-NP raises a series of significant data and computing requirements that cover high-performance and high-throughput computing, and storage of experimental data in the petabyte regime. The

envisaged computing activities go from the transfer of legacy codes to new languages and new hardware infrastructures, as well as the optimization and parallelization of current codes, to small- and large-scale testing of hardware and software solutions. Given the computing needs of ELI-NP, it is expected that the local community of computational scientists, jointly with the Magurele High Tech Cluster, will establish partnerships within the industrial sector as to eventually co-develop new hardware and software solutions on data storage and scientific computing at large.

## More information

*Alex Nicolin* is the NGI International Liaison for the Romanian NGI

## EGI Community Forum 2015: Bari 10-13 November

The next EGI Community Forum will take place in Bari, Italy between 10-13 November 2015. The event will be organised by EGI.eu in collaboration with the partners of the Italian National Grid Initiative (INFN, INAF and INGV) and hosted by INFN.

Bari is a beautiful city in the region of Apulia. It has an active cultural life and a proud multicultural history with influence from its Norman, Swabian, French, Spanish and Russian rulers that can be

admired even in its diversified architectural structure and some peculiar historical buildings.

Bari's international airport is well connected to the major cities in Europe by direct flights and is linked to the city via a direct train.

The Conference will be held at Villa Romanazzi Carducci an ancient fully modernized villa, surrounded by beautiful gardens near the centre of the city and with a very comfortable hotel inside the same area.

*We look forward to see you in Bari!*

## More information

CF 2015 Indico pages  
<http://go.egi.eu/CF2015>

Conference website  
<http://cf2015.egi.eu>