# Inspired

**ISSUE 16**JULY 2014

news from the EGI community



## **TOP STORIES**

How do I see the future of EGI?

What can the FedCloud do for you? page 3

Transnational access across EGI page 4

Four years of EGI-InSPIRE

Perun: id and access management



## **European Grid Infrastructure**

www.egi.eu

## **MORE**

- O1 Call for Participation: EGI Pay-for-use
- **08** E-infrastructures for agriculture: time to grow!
- **09** Wrap-up of the Community Forum in Helsinki

## This Issue

In this issue...

- > Yannick Legré writes about where he sees the future of EGI
- > Salvatore Pinto describes what can the EGI Federated Cloud do for scientists
- Malgorzata Krakowian presents the latest figures of transnational access across EGI
- > Tiziana Ferrari looks back at four years of accomplishments through the EGI-InSPIRE project
- > Michal Prochazka introduces Perun an identity and access management system
- > Andreas Drakos & Robert Lovas say that it's time to grow an e-Infrastructure for Agriculture
- > Sy Holsinger opens the Call for Participation in the EGI Payper-use scheme
- > ... and a wrap up of the user-oriented sessions at the Helsinki Community Forum

Please send your comments, feedback and suggestions to: sara.coelho@egi.eu Thanks!



Cloud Towers, by Akseli Gallen-Kallela (1865–1931) (source: wikicommons)

## Call for Participation EGI Pay-for-Use

Sy Holsinger invites resource providers and user communities to get involved in refining the recently developed pay-for-use model

Following on from a policy approved by the EGI Council to explore alternative business models, EGI has been investigating the implementation of pay-foruse mechanisms to complement free-at-point-ofuse models. This has led to the development of a proof of concept that brings together technical and non-technical solutions to provide services on-demand and allow the users to pay for the delivered service according to their usage. This new model will be useful for researchers who have dedicated budgets to procure their ICT capabilities "as a service".

During the initial phase of implementation, 21 resource providers across 12 countries defined prices and advertised them in the EGI service registry (19 for grid-related services and 7 for cloudrelated services). Now, EGI is inviting both resource providers and user communities to help test, refine and advance the processes and functionality. The team will provide Resource Providers with an opportunity to improve the model's pricing information, help to expand the pool of providers and contribute feedback to the current framework.

Participating user communities will be given the ability to consume an amount of services related to a defined monetary value and be asked to provide feedback on the consumer side of the process.

#### More information

If you are interested in joining or would like more information, please contact egi-pay-for-use@mailman.egi.eu by 5 September 2014.

More information can be found at: https://wiki.egi.eu/wiki/EGI\_Payfor-Use\_PoC

## How do I see the future of EGI?

After starting in February as EGI.eu's Managing Director, Yannick Legré talks about

the future of EGI

been interesting. Stepping in at a very delicate time for EGI while the community prepares for the Horizon 2020 programme has been a real personal challenge. But also very rewarding! After eleven years of development and production activity, EGI is now a mature e-Infrastructure for science in Europe. EGI's first duty is to serve and support research communities and infrastructures, especially those using big data and facing big science challenges. EGI will be a key provider of IT-services for the European Research Area, based

The least I can say about the last few months is that they have

I will lead EGI.eu with this mission informing all my decisions. I am confident that, as a community, EGI will be able to connect researchers with the reliable and innovative ICT services they need for their work.

on an open federation of reliable

services, which provide uniform

access to computing and data

storage resources.

Building an e-infrastructure of this quality has cost money and will continue to cost money. This is not just about buying computing resources but also operating and maintaining the system. The human element especially cannot be downplayed here, much of the money will go on employing and training up people with IT expertise, supporting the knowledge economy and facilitating excellent science.

This investment represents real value for money and the last decade has taught us many lessons about how to manage

research infrastructures and how to support big science. This body of knowledge cannot be acquired overnight and is one of EGI's biggest assets.

Any country or ESFRI looking at building a computing solution for their researchers with EGI is being offered not only computing resources but also our expertise and knowledge.

We already know how to operate, manage and maintain 'must have' services such as accounting, monitoring, authorisation & authentication. Any research infrastructure with this type of requirement doesn't need to reinvent the IT wheel.

We also need to raise our voice as a community and be more inclusive when looking at the future of e-Infrastructures in general. That is why along with EUDAT, GEANT and PRACE we are working to build an e-Infrastructure commons so that we are open and accessible, and able to adapt to the changing requirements of researchers and seizing any new technological opportunities.

The European Grid Infrastructure

is not merely a set of tools, technologies and applications. EGI is the people, people from across Europe coming together with a common goal, to support research. EGI is also not a monolith - it is built on the individual NGIs. They are EGI and EGI is the NGIs, the successes of the infrastructure are the successes of the NGIs and vice versa.

The EGI community's strength is its diversity and so far I have concentrated on working with the council on a new governance structure. EGI, in its entirety, from the small NGI to the largest of the EIROs, has a great potential to deliver value. I believe the best way to make the most of this asset is to create a new paradigm for collaboration -transparent, inclusive and based on mutual trust, open discussions and collaboration. This will be one of my personal objectives at the head of EGI.eu: to lead EGI to a position where we can accelerate science in Europe, where we can work better together and where all stakeholders can be heard and represented.



## What can the EGI Federated Cloud do for you?

Salvatore Pinto explains how researchers can benefit from the EGI Federated Cloud

The EGI Federated Cloud is a federation of institutional private clouds, offering cloud services to scientists in Europe and worldwide. If you are a researcher, you can access the Federated Cloud via a unique interface, with single credentials and with the possibility to distribute the load across multiple providers.

## What is the Cloud and what the cloud can do for you?

The cloud is an IT provisioning concept that allows users to access remote virtual services which can be rapidly provisioned and released on-demand with minimal management effort. In other words, the cloud provides resources that can be used from anywhere via the Internet. These services are based on flexible virtualisation technologies and can be obtained ondemand upon request. Some examples:

- > If you want to store file X, simply put it in the cloud and you can then download it from anywhere. This is called Storage-as-a-Service (StaaS).
- > If you want to use an application, you just need to press a button and you will have an instance of the application running for you to use. This is Software-as-a-Service (SaaS).
- > If you want to run an algorithm over a big amount of data, upload it to the cloud and run it (Platform-as-a-Service, PaaS).
- > If you want access to a web server, tell the cloud how much memory, CPU and disk you need

and we will provide it as part of the Infrastructure-as-a-Service (laaS) role.

#### The EGI FedCloud as a laaS

As of May 2014, the EGI Federated Cloud offers operationally Infrastructure-as-a-Service which can be used by scientists everywhere, while the other services are available mainly for testing and piloting.

The FedCloud laaS provides a virtual web server on demand. These machines are very similar to a normal PC with an internet connection. You can start and stop them using a client or an API. When you start a server, you can specify the amount of memory, CPU and disk you want, picking up one between the flavours available.

You can specify the Operating System (OS), selecting one of the Basic OS registered (eg. CentOS, ScientificLinux, Ubuntu) or uploading your own VM (e.g. Microsoft Windows-based). You can install any software or dependencies you need or customise the OS as you desire. You can assign to the server a public internet IP (e.g. to host a web portal) or a private one



© Morriander / wikicommons / CC-BY-SA 3.0

(only for internal access). You can connect to the server from anywhere, using protocols such as Secure SHell (SSH).

With the Federated Cloud IaaS, you can exploit additional capabilities derived from the flexible and on-demand nature of the cloud concept. You have the possibility to dynamically start new machines, whenever the previous ones are not enough for your needs. You can dynamically add additional storage disk, if and when you feel there is not enough disk space into your virtual server. You can use scripts to set your credentials to access the server or automatically install or configure your application on start-up.

## How to get started?

If you are interested in the EGI Federated Cloud services, visit the EGI Federated Cloud Wiki (https://wiki.egi.eu/wiki/Fedcloud-tf:Users#Getting\_access).

For more technical in-deep information, take the "Porting applications to the EGI Federated Cloud" webinar (http://www.youtube.com/watch?v=HlJjLkS2Xj8).R

## Transnational access across EGI

## Małgorzata Krakowian presents the landscape

One of EGI's greatest assets is the contribution it makes to Europe- and world-wide collaborations. This is well aligned with the European Commission's priorities for the European Research Area (ERA), as outlined in a document called *A Reinforced European Research Area Partner-ship for Excellence and Growth*, published on 17 July 2012:

"Optimal transnational cooperation and competition defining and implementing common research agendas on grand-challenges, raising quality through Europe-wide open competition, and constructing and running effectively key research infrastructures on a pan-European basis".

In the same year, EGI setup a dedicated virtual team to gather requirements for accounting

reporting that would demonstrate the transnational usage of resources across the infrastructure. Thanks to development efforts, EGI is now able to show how researchers within each national infrastructure are using the resources of other national infrastructures. This accounting capability highlights how important it is to support collaboration between researchers worldwide and in Europe.

Based on Certification Authorities (CA) that provide the user DNs (Distinguished Name) of certificates published in the accounting data, EGI has calculated for the period between May 2013 and April 2014 that, excluding CERN consumption, 32% of CPU usage is provided by resource centres in other EGI countries. With

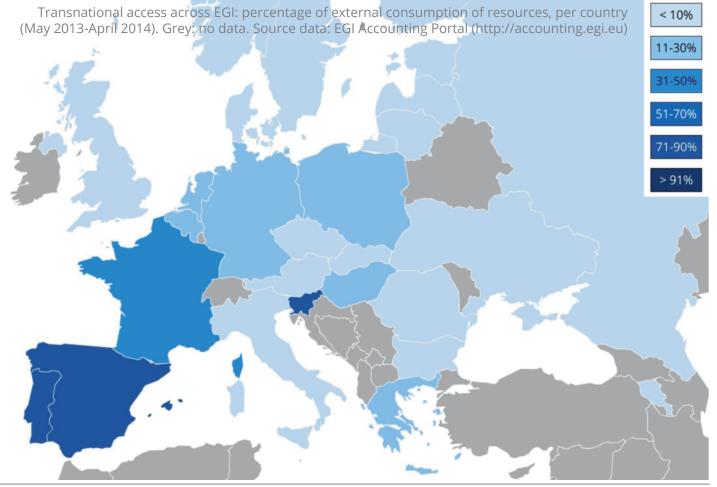
CERN, the transnational usage of the infrastructure raises to 64%.

This means that at least one third of the work done by scientists on EGI is supported by resource providers located in other countries.

These values, reported in the Accounting Portal, should be interpreted considering the following factors:

- > A small fraction of resource centres are not willing to publish DNs of their users outside of their country for policy reasons.
- > Users are mapped to a country of his/her certificate CA, which is not necessarily the country of the user.

The map below shows the distribution of NGIs with different levels of users who benefit from the international nature of EGI.



# Four years of accomplishments through the EGI-InSPIRE project

At the beginning of the fifth year of EGI-InSPIRE, Tiziana Ferrari looks back at past achievements and future challenges

EGI-InSPIRE is the EC project created to support the setup and evolution of EGI. Four years after its inception in 2010, it is time to look back at the results we achieved and to assess how well the project's objectives have been met.

The first major change was the definition of a new governance for distributed High Throughput Computing (HTC) in Europe. The provisioning and operations of HTC services evolved from a set of activities delivered through projects entirely supported by the European Commission, to a new self-sustainable landscape of National Grid Initiatives (NGIs) responsible for national HTC services, under the coordination by EGI.eu. Since then, EGI evolved into a stable federation spanning 57 countries relying on 37 Operations Centres, with increasingly strong partnerships in Europe and worldwide with other e-Infrastructures, user communities and technology providers.

The governance of EGI, grounded on the principles of openness and cooperation, is today stably based on the EGI.eu Executive Board and the EGI Council at a policy and strategic level, and technically on several managerial and advisory boards including the Technology Coordination Board and the UMD Release Team, the Operations Management Board, the User

Community Board, and the NGI International Liaisons.

During the four years of InSPIRE, EGI has progressively developed and established a vision, mission and core values that go beyond the original objectives by expanding them into a long-term strategy to support the European Research Area.

From an infrastructure driven by the needs of few well-established user communities, EGI is now evolving towards the Open Science Commons, to let researchers from all disciplines have easy and open access to the digital services, data, knowledge and expertise they need to collaborate and perform excellent research.

EGI will follow this vision to ensure that researchers will be able to:

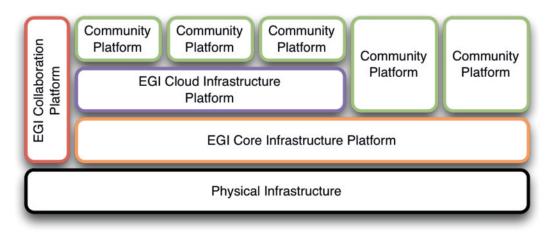
- > Access a single point of contact for obtaining the necessary ICT services (integrated and interoperable) the related capacity and support from the various e-Infrastructures (including commercial providers),
- > Connect to the best expert consultancy to understand the services they need or to support developing new solutions to perform their digital research,
- > Freely discover, share, use, reuse research outputs (publications, data, software, workflows).

**Project Objective 1** required EGI to operate the largest distributed

European infrastructure while evolving towards an increasing self-sustainable model. This is being accomplished by developing a portfolio of services and solutions, the engagement with European Research Infrastructures, the establishment of payper-use services. We are also addressing new customer segments with the new EGI Cloud Platform, which is attracting the interest of SMEs and industries thanks to its open standards and the level of trust that can be ensured by publicly funded cloud operators. In the last four years the

## EGI in numbers at the end of EGI-InSPIRE year 4

- > Resource providers (including integrated infrastructures): 355
- > CPU Cores (including integrated and peer providers): 593,424
- > Installed capacity including integrated providers (HEPSPEC million): 4.054
- > Installed storage capacity:
  - 286.0 PB (Disk)
  - 117.6 PB (Tape)
- > Usage
  - jobs/day (million): 1.54
- CPU Wall clock hours (million hours/day): 4.64



The EGI Platform Architecture (MS518)

infrastructure has increased its technical stability, and this has ensured continued and improved support to user communities in all sectors from the long-tail of science (**Project** Objective 2) to the heavy user communities (Project Objective 3). To expand on this work, EGI developed an Engagement Strategy to help EGI.eu, the NGIs and EIROs and the user communities in targeting technical support activities towards new large research infrastructures. And these research infrastructures demonstrate how existing solutions can be successfully adopted as a foundation layer and used to extended support to many pan-European research infrastructures (Project Objective **4**). As a reflection of this, in four years 41 multinational and 26 national Virtual Organisations were created, five of which to support testing activities by ESFRI projects.

EGI policies, procedures and operational tools are now technology agnostic and the operations architecture is easily extendible to any type of platform, including cloud, HPC and desktop grids, as demons-trated in May 2014 with the launch of the Federated Cloud solution (**Project Objectives 5 and 6**).

From a monolithic infrastructure based on a single internally-

sourced grid technology, EGI changed into a landscape of integrated platforms delivering different and complementary services to the end-user. The current platform architecture – comprising a core infrastructure, the cloud, high-throughput data analysis and the collaboration and community Platforms – is now easily extensible.

With this architecture in place, we can now assert that EGI is an infrastructure capable of federating any type of ICT service at a pan-European level to enable transnational access.

We are very proud of the European added value demonstrated by EGI, thanks to the support of the European Commission: in the last 12 months on average 28% of the resources in EGI are provided to support usage by scientists from other countries in Europe and world-wide. This, together with over 2,000 scientific publications made possible by EGI, truly demonstrates the international crossborder nature of research and the primary leading role of the European Commission as enabler of this potential.

## The six objectives of EGI-InSPIRE

- > PO1: Ensure the continued operation and expansion of today's production infrastructure and transition to a governance model and operational infrastructure that can be increasingly sustained outside of specific project funding
- > PO2: Provide continued support of researchers within Europe
- > PO3: Help supporting the current heavy users
- > PO4: Engage with new user communities including new potential heavy users of the infrastructure from the ESFRI projects through new Interfaces that expand access to new user communities
- > PO5: Provide mechanisms to integrate existing infrastructure providers in Europe and around the world
- > PO6: Extend the infrastructure to integrate new Distributed Computing Infrastructure technologies (e.g. cloud, desktop, HPC)

## Perun – an identity and access management system

Michal Prochazka writes about an application already making a difference at the Czech NGI



Perun is an application developed to manage users, groups and access to the services in highly distributed environments. It gives virtual organisations (VOs) that consume grid- and web-based services the ability to manage access to such services.

Web-based services usually deploy federated login solutions that cannot be easily combined with the digital certificates used in EGI. A central system to manage users and their different digital identities like Perun can solve such a challenge.

We used two proven concepts to deliver a reliable and robust solution. First concept is VOs - a well-established organisational unit in grid environments. The second concept implemented in Perun is a push mechanism used to deliver access control lists to the end services.

Perun covers the whole user lifecycle, including user enrolment, VO and groupings organisation, assigning groups to the services and membership expirations.

Because users may have existing digital identities, we tried to consolidate them instead of creating new ones. This allows Perun to put users from different organisations into one place, regardless of the type of authentication mechanism they are using (federated identity, digital certificate or just simple login).

We have designed Perun to be deployed into existing infrastructures in order to consolidate data about the users and resources spread across various systems. Data consolidation is done by integrated import plugins that are able to get information from various sources. Perun can also export stored information using different export plugins, so the integration can be done step by step. A key feature is delegation of rights - this allows entities to be managed by the users who are formally responsible for them.

While Perun can consolidate various users' identities and store other user related information, it can also be used as an attribute authority (AA) within identity federations. AA gives not only basic information about the users, but also theirs group and VO membership information. We have made a successful pilot installation for DocuWiki, which uses federated authentication and authorisation based on groups derived from the AA managed by Perun.

The Czech National Grid Infrastructure is entirely managed by Perun. Some challenges remain while managing EGI VOs - for example, how to create user accounts on all the machines and services; something that cannot be done manually. Users' account expiration and suspension

must be done automatically in such environment. For this, Perun prepares user accounts on the UI machines, feeds VOMS servers with required data and manages mailing lists for VO users. For FedCloud VOs, Perun enhances VOMS with additional functionalities such as notification of which users are banned or expired, which allows managers to stop or remove their virtual machines.

#### Perun in numbers

- > 1.5 years in production
- > Currently managing >3000 users
- > Configures access to 1800 services
- Manages 135 national and international VOs (including EGI, FedCloud and the South African Grid Infrastructure)

#### More information

Michal Prochazka is a researcher at CESNET

Perun is open source project supported by CESNET (www.cesnet.cz) and CERIT-SC (www.cerit-sc.cz).

Perun web pages: http://perun.cesnet.cz

Perun on GitHub: https://github.com/CESNET/perun

## E-Infrastructure for Agriculture: it's time to grow!

Andreas Drakos and Robert Lovas invite you to join the planned Agricultural

Sciences Virtual Team

Agriculture-related science domains are gaining more and more attention. Chris Rawlings from Rothamsted Research gave a great, comprehensive overview of the topic at the EGI Community Forum, on the same day that we organized a networking session around the same theme.

## What have we learned during the agricultural sciences networking session?

At international level we gave a special focus to the solid results obtained by the FP7 agINFRA project. This project deals with open linked data and high-level integrated services for data management, processing and visualisation of agriculture data with the participation of the United Nations' Food and Agriculture Organisation.

The new grid and cloud computing based components in agINFRA provide added value services to the existing repositories and directories, such as CIARD RING that registers 1000+ datasets and information services in agriculture from institutes worldwide. The agINFRA solution provides easy-to-use APIs, highlevel user interfaces, science gateways with workflow support, and uses EGI resources and EGI suppor-ted technologies as well in an agINFRA VO managed by IPB (NGI\_AEGIS).

At the national level, Agrodat.hu was presented as a project that aims to establish an agricultural knowledge centre and decision-support system based on data gathered by an innovative, complex sensor system in Hungary and also from international open repositories using agINFRA-empowered services.



© Myrabella / wikicommons / CC-BY-SA 3.0

## Is agricultural science ready for Big Data?

The Big Data challenge has reached agricultural sciences and new actions are under preparation. The agINFRA services, already available through the EGI Application Database, are just a first step to cover the growing needs of this community. This is why we are proposing a new Virtual Team to bring together the agriculture research community and EGI resource and technology providers. The VT's main goals are:

> Promote sharing of data and services:

Very often the difficulties experienced by this community are due to the lack of information about best practices. The VT will join forces with well-established networks and initiatives as the aglNFRA and Agrodat.hu, to support the 'Open AGRI-Gate' campaign, an open agricultural data campaign designed to promote open data, register agriculture-related datasets and data services in one place and make them accessible and reusable by everyone.

> Gather special requirements:

The VT will bring together stakeholders on the agricultural field and gather specialised requirements that will serve as base for further collaboration e.g. in H2020.

> Sustaining existing solutions:
Sustainability of European
projects' results usually depends
on users, funding agencies, base
technology and solution providers.
Since EGI-related activities and
sustainability plans might be
important for this community as
well, the mutual harmonisation,
sharing best practices and
experiences would be of high
value to both EGI and initiatives.

## More information

Andreas Drakos, Agro-Know drakos@agroknow.gr

Robert Lovas, MTA SZTAKI (Hungarian NGI) Robert.lovas@sztaki.mta.hu

agINFRA: http://aginfra.eu Agrodat.hu: http://agrodat.hu

## Wrap up of the Community Forum in Helsinki

Eleni Katragkou gives an overview of the environmental and climate science sessions

On Tuesday 20 May 2014, the Environmental science on grids and clouds session brought together members of the environmental science community, who presented their experiences on existing services, requirements and trends around the use of EGI grid and cloud solutions within this domain. The session showcased successful applications in the fields of atmospheric modelling and biodiversity studies as well as discussed the technical challenges and ideas for exploitation with participants. The session closed with the announcement of the 'Atmospheric Modeling and Services' Virtual Team and an open invitation to participate.

On Wednesday 21 May, a networking session on Proceeding European Joint **Efforts on Climate Service Products** was organised by the Climate Service Center to discuss the development and impact of upcoming climate service products and initiate a network for a potential Horizon 2020 proposal. During the session, plans for joint activities to develop climate service prototypes were presented based on the following questions: How to effectively collect prototype ideas from the community without duplication of efforts? How to organise the transition from prototype to product operationalisation?

On Thursday 22 May, a training session on Web-processing services for climate data was also organised by the Climate Service Center. During the training, both technical providers and user communities interested in climatic data followed a seminar through a user-friendly, web-based interface to assess, visualize and combine climate data processes. The interface provided access to the Earth System Grid Federation climate data archive allowing for tailored, end user data processes and web processing services.

Eleni Katragkou is an EGI Champion

## Fotis Psomopoulos and Afonso Duarte summarise the Life Sciences session

Grid and cloud computing are now a fundamental element in all analysis and experiments in Life Sciences. The importance is evident: from the impact of Big Data as viewed in most current research approaches, to the highly complex applications and protocols required to compute new ideas.

During the conference, a workshop was held to further understand the state-of-the-art of grid and cloud computing in EU Life Science research.

The workshop brought together researchers active in Life Sciences from around Europe and facilitated networking between them under the auspices of EGI.

The main body included talks from key Life Science infrastructures and projects such as WeNMR, LifeWatch, BioMed Bridges, ELIXIR and LSGC among others. Presentations were complemented by technical talks on the advances in ELIXIR, the use of the Galaxy interface, and HPC and biodiversity to name a few. The workshop concluded with a roundtable discussion covering several key topics:

- > Training and education on new computational tools in Life Sciences as part of established curricula, such as pre-set VMs for training purposes aiming at a wider audience
- > Better define user requirements and data analysis

pipelines for providing standards and best practice approaches

- > Develop a high-level platform, for easy use of existing and further implementation of new Life Science tools
- Address the shortage of experience on dealing with Life Science Big Data

Fotis Psomopoulos and Afonso Duarte are EGI Champions

## Claudio Prandoni rounds up the e-Infrastructures and services for data preservation and curation workshop

The workshop, organised by the DCH-RP project, contributed to reinforce the fruitful cooperation started among projects and initiatives working in the domain of DCH, e-infrastructures and digital preservation (such as APARSEN and Open Planets Foundation) towards the establishment of a common registry of services and tools to facilitate the storage, access and preservation of digital data.

The main outcomes of the workshop can be summarised as follows:

Discussed about possible options to ensure the - of the registry of services and tools developed in DCH-RP. The aim of the registry is to help a broad range of DCH communities, institutions and projects to plan the implementation of their digital preservation processes and to understand the commonly available options. The registry contains descriptions of common user scenarios as well as information about the tools that may be used to implement them. In addition it supports searching the tools database using several criteria and it will also provide assessments and reviews of most relevant tools.

Collected feedback on the interim version of the roadmap for the long-term preservation of digital cultural content, which aims to help policy makers and programme owners to plan ahead and assist managerial teams of cultural institutions in taking decisions related to digital preservation and to support cultural heritage institutions in defining practical action plan with a realistic time frame for its implementation.

This feedback is now being analysed and will be taken into account in the final version of the roadmap, which is due by the end of September.

Claudio Prandoni works for Promoteur s.r.l.

## Diego Scardaci tells us what happened during the FedCloud demo session

During the networking session, EGI.eu organised a demonstration to show the most interesting EGI Federated Cloud use cases. To increase the interest of the visitors we decided to adopt a new format for this demo session, without formal presentations. Instead, the presenters were distributed in a unique room and showed their demo at the same time.

After a short introduction of the demos, the participants were free to go around the room, choose what was most interesting for them, ask questions and discuss with the demo teams. The format allowed the participants to have more time for questions and more interaction with the demo teams.

The demonstrations presented were:



- > Biovel OpenModeller/COMPS, BioSTIF and Openrefine on top of Biovel portal. Presenters: Daniele Lezzi, Giacinto Donvito, Yuliya Fetyukova, Karl-Heinz Sylla and Matthias Obst
- > Catania Science Gateway Framework as cloud application broker and infrastructure broker. Presenter: Giuseppe La Rocca
- > ENVRI/EISCAT\_3D. Presenter: Salvatore Pinto and Ingemar Haggstrom

- > Peachnote OCR analysis in music scores. Presenter: Salvatore Pinto
- > SCI-BUS/WS-PGRADE.
  Presenters: Zoltan Farkas and
  Peter Kacsuk

The demos are available at http://go.egi.eu/fcdemos.

Diego Scardaci is a Technical Outreach Expert at EGI.eu