



# COMMUNITY FORUM

19<sup>TH</sup> - 23<sup>RD</sup> MAY 2014 | HELSINKI, FINLAND

ESGI COMMUNITY FORUM | BOOK OF ABSTRACTS

HELSINKI 2014



---

# BOOK OF ABSTRACTS

---

First published in 2014 in the Netherlands by EGI.eu  
Science Park 140  
1098 XG Amsterdam  
The Netherlands  
www.egi.eu  
contact@egi.eu

ISBN 978 90 816927 3 1

Copyright © 2014 by Members of the EGI-InSPIRE Collaboration

This work is licenced under the Creative Commons  
Attribution-Noncommercial 3.0 Licence.

To view a copy of this licence, visit:  
<http://creativecommons.org/licenses/by-nc/3.0>

or send a letter to Creative Commons, 171 Second Street,  
Suite 300, San Francisco, California, 94105, USA.

The work must be attributed by attaching the following reference to the copied  
elements: Copyright © Members of the EGI-InSPIRE Collaboration, 2014.

EGI-InSPIRE (European Grid Initiative: Integrated Sustainable Pan-European  
Infrastructure for Researchers in Europe) is a project (INFSO-RI-261323) co-funded  
by the European Commission as an Integrated Infrastructure Initiative within the  
7th Framework Programme. EGI-InSPIRE began in May 2010 and runs for 4 years.

This book contains abstracts submitted online, through a call for contributions, that  
were accepted as a part of the Forum programme. It does not include abstracts  
submitted by invitation from workshop and session conveners during the final phase  
of the programme's preparation.

The Organisers and the Publisher have made every effort to provide accurate  
and complete information in this book. However, changes or corrections may  
occasionally be necessary after publication.



Editor: Sara Coelho  
Design: Rory J Murphy

# **EGI COMMUNITY FORUM 2014**



19<sup>th</sup> - 23<sup>rd</sup> MAY 2014  
• HELSINKI •

---

---

# WELCOME

---

---

**Dear Delegate,**

On behalf of the European Grid Infrastructure community, the University of Helsinki and CSC - IT Center for Science Ltd, we would like to welcome you to the EGI Community Forum 2014.

The week marks the start of EGI's fifth year and the in preparation for the Horizon 2020 programme, the event will focus on the EGI contribution to advance excellent science in the European Research Area, through the use of innovative services for data and computing.

During the course of the week there will provide opportunities to existing and new user communities to present their requirements, report on success stories, get support in porting their application environment and network with technical providers and other partners.

For the first time the Community Forum features two new ways for the community to interact, Networking sessions and Hackathons. We expect both to be successful and an ongoing part of future EGI events. The Networking sessions will provide an easy way to share ideas, receive feedback, and be exposed to the views, perspectives and ideas of other participants and engage in a constructive dialogue with them, complemented by collaborative, hands-on Hackathons.

We also welcome the delegates from the co-located events, the European Globus Community Forum, the NorduGrid Conference 2014, the Nordic e-Infrastructure Collaboration (NeIC) meeting and the joint meeting between the Research Data Alliance Europe (RDA), EUDAT, OpenAIRE and the Finnish National Research Data project (TTA).

We hope you all have a productive and enjoyable week.

**Yannick Legré**  
EGI.eu  
*Director*

**Kai Nordlund**  
University of Helsinki  
*Vice Dean, Faculty of Science*



# TABLE OF CONTENTS

<b>WELCOME</b> .....	<b>2</b>
<b>MONDAY</b>	<b>8</b>
<hr/>	
<b>WORKSHOPS</b> .....	<b>10</b>
<b>SESSIONS</b> .....	<b>10</b>
Helix Nebula workshop: interoperability among e-Infrastructures and commercial clouds .....	10
Tools for operating research infrastructures .....	10
Supporting users and crowd computing with desktops grids .....	11
Advancing accounting services for users .....	11
<b>WORKSHOPS</b> .....	<b>11</b>
E-infrastructure services for earth science .....	12
<b>TUESDAY</b>	<b>20</b>
<hr/>	
<b>WORKSHOPS</b> .....	<b>22</b>
Going beyond the grid to enable Life Science data analysis .....	22
e-Infrastructures for data preservation and curation .....	22
EGI and XSEDE tools and common needs for the computational chemistry, molecular modelling and materials science community .....	23
ENVRI workshop .....	23
Software Vulnerability Handling and practical incident recognition .....	24
<b>SESSIONS</b> .....	<b>26</b>
Environmental science on grids and clouds .....	26
Porting new applications to EGI .....	30
Applications on federated clouds .....	34
EGI services for global software and common data analysis .....	36

**WEDNESDAY 40****WORKSHOPS ..... 42**

Business development and pay-for-use .....42

Open access to EGI research outputs .....42

GPGPU integration and user application support.....43

**SESSIONS ..... 44**

New services for astronomy and Astrophysics ..... 44

Science gateway frameworks .....48

Evolution of the EGI authentication and authorisation structure .....52

New data management solutions for EGI .....56

**NETWORKING SESSIONS ..... 60**

Pre-commercial procurement on digital preservation.....60

Networking Life Sciences and EGI towards H2020.....61

DIRAC 4 Life Science .....61

Grid and cloud services for the agricultural community.....62

Proceeding European joint efforts on climate service products.....62

**THURSDAY 64****WORKSHOPS ..... 66**

OpenNebula in science/HPC and cloud federation.....66

Monitoring availability and reliability in EGI-inSPIRE and beyond.....66

**SESSIONS ..... 68**

Community building and engagement .....68

Dirac virtual research environment for EGI .....72

**TUTORIALS ..... 78**

Web processing services for climate data with examples for impact modelers .....78

Foundation training in federated IT Service Management according to

FitSM-1:2013 .....78

Managing virtual servers.....79

Introduction to Stoxy - CDMI-compliant lightweight storage proxy .....79

Training on data preservation .....80

OpenStack hands-on.....80

**FRIDAY****82**

---

<b>TUTORIALS .....</b>	<b>84</b>
Building an OpenNebula cloud on the EGI Federated Cloud.....	84
Hands-on DIRAC user-friendly job management.....	84
Globus technology training .....	85
Best practices for cloud application architecture.....	85
Federated Cloud image lifecycle management.....	86
 <b>INDEX OF AUTHORS &amp; CONTACT DETAILS .....</b>	 <b>88</b>

The EGI Community Forum is organised by:



European Grid Infrastructure

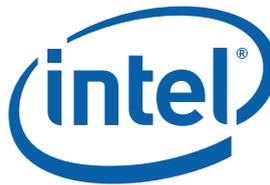


CSC - IT Centre For Science



University of Helsinki

With the support of:



Intel Corporation  
*Gold sponsor*



International Science Grid This Week  
*Media partner*



Federation of Finnish Learned Studies  
*Media Partner*

---

---

# MONDAY

---

---

## **WORKSHOPS ..... 10**

- Helix Nebula workshop: interoperability among e-Infrastructures  
and commercial clouds..... 10
- Tools for operating research infrastructures ..... 10
- Supporting users and crowd computing with desktops grids..... 11
- Advancing accounting services for users..... 11

## **SESSIONS ..... 12**

- E-infrastrucuture services for earth science ..... 12



## Helix Nebula workshop: interoperability among e-Infrastructures and commercial clouds

Helix Nebula brings together leading European scientific organisations, commercial cloud suppliers, technology providers and publicly funded e-Infrastructures to provide a cloud platform based on open standards and a marketplace. The Helix Nebula platform will provide cutting edge services able to satisfy the requirements of complex use cases by scientific communities, public authorities, private research labs, and users working in the 'long tail' of science.

The Helix Nebula EC-funded project is moving towards its conclusion and the suppliers in Helix Nebula's initiative have shifted to the production phase with the set up of an initial

service platform. In this context, this workshop aims to assess progress made in interoperability among public e-Infrastructures and commercial cloud providers, and to discuss emerging interoperability challenges within a service delivery platform based on a hybrid cloud environment in the context of Helix Nebula's future strategy.

The objectives of this workshop include: presenting the interoperability roadmap highlighting progresses and open actions according to Helix Nebula's future strategy, informing the EGI user communities and stakeholders about services offered by the Helix Nebula marketplace and its governance in order to gather

input for future developments, and continuing the discussion of roles of public e-Infrastructures in the Helix Nebula marketplace in a future perspective.

---

**Carmela Asero** *EGI.eu*

**Sergio Andreozzi** *EGI.eu*

## Tools for operating research infrastructures

EGI's operational tools are part of the Infrastructure and Collaboration Platform. They are technology-agnostic and they can be easily extended to meet the operational needs of any distributed Research Infrastructures. The adoption of EGI operational tools, deployed and developed within the EGI-InSPIRE project, allows the reuse of existing solutions that address the typical needs of any distributed RI. This workshop is aimed at NGIs, Research Infrastructure operators and user communities.

---

**Diego Scardaci** *INFN*

# Supporting users and crowd computing with desktops grids

The International Desktop Grid Federation is working closely together within an EGI virtual team to make it as easy as possible for scientists to use desktop grid resources, including unused computing capacity from university computers (local desktop grids) or from volunteer citizens at home (crowd computing).

The workshop will discuss:

1. How to use computing resources from desktop grids while staying in the comfort zone of EGI. Using the EGI-connected portals, job submission and monitoring tools to get access to additional thousand or ten-thousands of processors. We

will also explain to community organisers what is needed to make their community 'desktop grid ready'. EGI and IDGF have formed a virtual team that can assist.

2. How to organise a local desktop grid within a university or institute by connecting the organisation's computers into one huge computing engine. Port applications to this engine and provide additional number-crunching to scientists. IDGF has a support team that can help.

3. Crowd computing. It is exciting to see thousands of enthusiastic citizen volunteers donating computing time to your applications. Not only can you

get your work done more quickly, but also it provides an excellent opportunity to spread knowledge about your research to society. IDGF has a special, supported, crowd computing grid dedicated to EGI applications.

---

**Robert Lovas** *MTA SZTAKI*

**Jozsef Kovacs** *MTA SZTAKI*

**Tamas Kiss** *University of Westminster, London*

**Ad Emmen** *EDGI, DEGISCO, e-IRGSP3 projects*

# Advancing accounting services for users

There have been many recent developments in the accounting service: updates to existing software, addition of new accounting types and research into how to provide accounting services. This workshop aims to introduce the new developments along with instructions on how to configure the necessary software so sites can take part. Additionally, it will introduce some upcoming developments and seek discussion on the plans for taking them forward.

With the recent end of support for EMI2, it has become critical that sites update their software to the latest version of the APEL accounting software. Doing so involves more than a simple upgrade of packages. This

workshop aims to guide the attendee through the process of migrating their data to the new system and installing and configuring the software components.

---

**Stuart Pullinger** *STFC*

**John Gordon** *STFC*

# E-INFRASTRUCTURE SERVICES FOR EARTH SCIENCE

*This session highlights some of the more recent applications deployed on EGI, or national cloud and grid infrastructures to produce scientific results in the broad field of Earth Sciences.*

*Applications belonging to the broad field of EarthSciences have been successfully using European e-Infrastructures for their research for a long time. Applications from various different domains like atmospheric chemistry, seismology, meteorology, hydrology, etc. have been deployed and used to produce scientific results.*

*Solutions implemented on the basis of different middlewares and in combination with data infrastructures such as proposed/provided by EUDAT and domain-specific infrastructures will be presented.*

**Convenors:** Monique Petitdidier, CNRS | Horst Schwitenberg, Fraunhofer

## Increased computational requirement of meteorological and crisis applications

In this talk, we present an application we developed for the domain of meteorology and crisis management. The IMS Model Suite is a complex software system designed to forecast weather and hazardous weather phenomena, assess environmental pollution and predict the consequences of accidents. We discuss the computational requirements and our experiences on how to meet them by high performance computing. The process of a weather related hazard assessment and prediction of its consequences results in complex dataflows and work-flows among databases, models and simulation tools. A regional sandstorm pollution

assessment and prediction requires building a dust source and strength database, running a 3D meteorological model, and running a dispersion model performing the simulation of the release transport and deposition of the pollutant with respect to the numerical weather prediction data, released material description, topography, land use description and user defined simulation scenario. Several post-processing options can be selected according to particular situations. Sandstorm modelling requires both high resolution and large scale meteorological modelling. High resolution enables precise modelling of wind in sand/dust uplift areas (local conditions) while large

scale of simulation is needed to capture sand transport, which can be thousands of kilometres in a few days.

---

**Ladislav Hluchy** *Institute of Informatics SAS, Bratislava*

# Grid, cloud and mobile exploitation in Earth Sciences

Access to distributed computing and storage infrastructures based on grid and cloud paradigms has recently been improved thanks to the spread of science gateways and the support for mobility.

In this contribution, we present the results obtained in the context of two European projects funded by the European Commission, both under its Seventh Framework Program and both belonging to the Earth Sciences' domain: ei4Africa and EarthServer.

The ei4Africa project aims to boost the Research, Technological Development and Innovation (RTDI) potential of African e-Infrastructures, and also to support policy dialogues and Euro-African

cooperation in the framework of the joint Africa-EU Strategic Partnership on 'trade, regional integration and infrastructures' (JAES Partnership 3). In the presentation we will show a use case where simulations of the Weather Research and Forecasting (WRF) Model are simultaneously and seamlessly executed on grid and clouds, including the EGI Federated Cloud.

The EarthServer project aims at establishing open access and ad hoc analytics on extreme-size Earth Science data, based on and extending leading-edge array database technology. In the presentation we will show some of the applications integrated in the EarthServer science

gateway mobile and the services developed to allow federated authentication and WCS/WCPS standards exploitation.

---

**Riccardo Bruno** *INFN*

**Giuseppe La Rocca** *INFN*

**Roberto Barbera** *University of Catania and INFN*

**Antonio Calanducci** *INFN*

**Marco Fargetta** *INFN*

**Torleif Markussen Lunde**  
*University of Bergen*

**Marco Pappalardo**

*Software Engineering Italy*

## Atmospheric applications on the European grid

The civil and public society ask for certainties i.e. precise values with a small error range for short, medium and long term predictions in all Earth Science domains; but science mainly answers only in terms of probability of occurrence. To improve the answer and/or decrease the uncertainties:

1. New observational instruments are deployed in order to have a better geographical coverage and more accurate measurements.
2. New algorithms and simulations have been developed, leading to intensive computing, addressed to a variety of architectures.

The European grid with its large number of available processors has played a significant role to treat huge collections of non-parallel independent jobs in the following examples. In the first case, the satellites surveying surface ozone need new instruments, IASI-NG and MAGEAQ-TIR, via OSSE (Observing System Simulation Experiments) to create millions of pseudo-observations used to test the new instrument performances. The second and third cases are focused on aerosols and clouds important to survey the earth radiative budget. EGI permits to reduce aerosol data from the PARASOL at full scale (6km) to be compared with data from CALIPSO and MODIS,

and also to run 3D Monte Carlo simulation for a variety of cirrus in order to study the cloud heterogeneities and to assess the limits of the operational cloud parameter algorithms.

---

**Monique Petitdidier** *CRNS*

# Distributed trace processing on cloud resources

The EU-funded project VERCE aims to address specific seismological use cases employing resources spanning available e-Infrastructures on the basis of requirements elicited from the seismology community. It provides a service-oriented infrastructure to deal with the challenges the researchers face in the data-intensive and high-performance computations employed in modern seismology. In particular, the implementation is driven by two major use cases. The first is the computationally intensive forward and inverse modelling of Earth system models, which is implemented with support for multiple waveform simulators running on HPC systems and x86 clusters,

and the data-oriented seismic wave cross-correlation.

In this talk, we present a service that is intended to complement the services of the VERCE projects infrastructure, a service that uses cloud resources to offer simplified execution of data pre- and post-processing scripts. It offers users access to the ObsPy seismological toolbox for processing data with the Python programming language, executed on virtual cloud resources in a secured sandbox.

The solution encompasses a front end with a modern graphical user interface, a messaging infrastructure as well as Python worker nodes for background processing. All components are deployable in

the cloud and have been tested on different environments based on OpenStack and OpenNebula. Deployments on commercial, public clouds will be tested in the future.

---

**Steffen Claus** *Fraunhofer SCAI*  
**Lion Krischer** *Ludwig-Maximilians-Universität München*  
**Andre Gemuend** *Fraunhofer*

icclim/ocgis: a generic climate impacts indices calculation package  
 interfaced with the tailored IS-ENES web portal climate4impact WPS

Within the FP7 European projects IS-ENES/IS-ENES2 on climate model data infrastructure, a web portal tailored for climate change impact communities is being developed. A first prototype version has been released in 2013. To support the impact communities, a package (icclim) to calculate climate indices (starting with the ECA D indices) is currently developed. Several constraints had to be considered: full integration with the climate4impact WPS (pyWPS); fast performance for near-real time 'on the fly' calculations, notably with the integration of icclim within OpenClimateGIS (ocgis) which features OpenDAP time and spatial sub-setting as well as

data chunking, but also with the optimisation of the code through an integrated Python C shared-library. icclim also ensures that NetCDF output files are CF-compliant and preserves existing metadata.

---

**Natalia Tatarinova** *CERFACS*

# Migrating climate models to grid: from EGEE to EGI experiences

Grid computing is nowadays an established technology, which offers an alternative to traditional HPC resources. The grid has proven to be a great infrastructure to perform climate experiments that involve large amounts of independent simulations, such as ensemble predictions and sensitivity analysis. But, the heterogeneity and distributed nature of the grid poses new challenges to climate applications willing to exploit them.

Over the last ten years, the Santander Meteorology Group has been involved in developing frameworks that allow climate models to make efficient use of grid resources. This work started in the EELA project, where a

framework prototype was used to simulate El Niño phenomenon with the Community Atmosphere Model (CAM) on the EGEE infrastructure. An evolved version of this first prototype was used to create WRF4G (EGI application), which allows to run the Weather Researcher and Forecasting (WRF) regional model on distributed infrastructures. WRF4G can be executed on both grid and HPC infrastructures and, today, it contributes to international initiatives such as CORDEX and European FP7 projects such as SPECS and EUPORIAS.

This work is partially funded by the Spanish PLAN NACIONAL de I+D+i 2008-2011 (WRF4G, Ref.# CGL2011-28864) and the

European Regional Development Fund (ERDF).

---

**Antonio S. Cofiño** *University of Cantabria*

# Seismology data management in VERCE

Seismic data is processed and analysed by researchers across the world. Each researcher deals with terabytes of data (raw and processed) which is hard to manage, process and share when using only local facilities. In order to handle such a big volume of data, a robust and scalable data management platform is required, allowing storage and discovery of large and heterogeneous datasets. A shared catalogue should enable users to search standard and community defined metadata, including user annotations and provenance.

Moreover, we aim at minimising the transfer of data among the computational clusters within a distributed processing scenario

via the implementation of a federated architecture and standard protocols.

When it comes to sharing data and resources, privacy of data is a main concern. Each institute requires retention of complete control over the access permissions on the archived data, and establishment and application of data sharing policies in agreement with the research partners.

We will illustrate in this work the approach of the VERCE project to tackle these challenges.

---

**Visakh Muraleedharan** *IPGP*

## Two recent seismological applications implemented on the EGI infrastructure

This work presents two applications that were implemented in EGI:

1. We use data from an industrial seismic network made up of 2,320 short-period sensors installed on the seafloor above an oil reservoir. We perform 24,210,360 cross-correlations between each unique pair of sensors, a three-month long computation. Using cross-correlations, it is possible to recover the impulse function of the medium between each sensor pair. With good enough sensor coverage, one can produce high-resolution images of the subsurface; this is called ambient noise surface-wave tomography. Using 6.5 hours

of noise data from the Valhall network, we implement two types of tomographic imaging; we are also able to extract the local azimuthal anisotropy.

2. As datasets grow in size, it becomes necessary to implement automated detection algorithms to produce complete event catalogues. We perform a matched-filter search that uses a template, made up of seismic waveforms recorded at several stations, to search for other events in time that originate from the same source as the template. This search calculates the correlation coefficients between the template event and the seismic dataset every tenth of a second. The dataset we

use covers 10 seismic stations, recording at 100 Hz, over a two and a half year time period. Given the thousands of template events to analyse as well as the scalar nature of the processing, this analysis takes full advantage of the EGI and was computed in under three weeks.

---

**William Frank** *Institut de Physique du Globe de Paris*

**Aurelien Mordet** *Institut de Physique du Globe de Paris*

**David Weissenbach** *UPMC*



# User training - getting started with the European Grid Infrastructure

This session provides introduction to the concepts of distributed computing, and to the architecture, components and user-facing services of the European Grid Infrastructure (EGI).

The tutorials will consist of short talks and demonstrations that introduce the:

- Main services of the European Grid Infrastructure
- How to get access to the EGI production system
- Federated Cloud from EGI
- Complex jobs, high-level tools for application development and access

The course does not require prior knowledge of grid computing, or

of EGI. The session aims to be an introduction to EGI for new members of the community, and demonstrate the value of EGI for data- and compute-intensive e-Science projects. Attendees will have the opportunity to become familiar with the structure and components of the European grid and how these can advance e-Science applications.

---

**Gergely Sipos** *EGI.eu*

## EGI-CSIRT- improving security in a cloud environment

Managing security in a cloud environment is a challenge. The focus of this session is on security monitoring technologies and how to use them in a cloud environment.

In order to help assess VM images from a security point of view we suggest a couple of checks and tools that can be done. The intended audience is either the cloud provider administrator, VM endorser, or VM operator, mostly running Linux-based OS. We will demonstrate how a VM can be checked to detect vulnerable packages installed on a VM file system and will show how these checks could be persistently installed in the image. It is also important to have a

centralised log management available; therefore we discuss possibilities of how to store these logs from within the VM, either during run-time or at least offline. There are also a couple of other rather simple precautions that can foster security of a node on the public Internet, e.g. to disable password-based authentication, preventing common directory and/or brute-force attacks. We will also discuss these possibilities and demonstrate how they can be enforced.

Another central problem in distributed computing environments is how to efficiently suspend identities found in activities misusing the infrastructure. In this session we will show how a central

user suspension framework is deployed in the European Grid Infrastructure.

---

**Sven Gabriel** *NIKHEF*  
**Daniel Kouril** *CESNET*  
**Leif Nixon** *LIU*

MONDAY

WORKSHOPS

SESSIONS

NETWORKING SESSIONS

TUTORIALS

---

---

# TUESDAY

---

---

## **WORKSHOPS ..... 22**

Going beyond the grid to enable life science data analysis .....	22
E-infrastructures for data preservation and curation.....	22
EGI and XSEDE tools and common needs for the computational Chemistry, molecular modelling and materials science community .....	23
Sustainability and common services of environmental research infrastructures .....	23
Software vulnerability handling and practical incident recognition.....	24

## **SESSIONS ..... 26**

Environmental science on grids and clouds .....	26
Porting new applications to EGI .....	30
Applications on federated clouds .....	34
EGI services for global software and common data analysis .....	36



# Going beyond the grid to enable Life Science data analysis

Grid and cloud computing are now a fundamental element in all analyses and experiments in Life Sciences. This is becoming even more evident in the world of Big Data and the Horizon 2020 work programme. This workshop aims to bring together researchers active in Life Sciences from around Europe and facilitate the networking between them within the context of EGI. The first part of the workshop will motivate participants by including targeted talks from key infrastructures and projects within the Life Sciences community, such as ELIXIR. Furthermore, the participants will have a chance of introducing their experience and expertise in a round of flash presentations. Finally, the discussion phase will

aim to provide clear proposals for possible collaborations and suggestions for future EGI actions.

---

**Fotis Psomopoulos** *Institute of Applied Biosciences, Centre for Research and Technology Hellas*  
**Afonso Duarte** *ITQB-UNL*

## e-Infrastructures for data preservation and curation

The need for novel, more efficient and affordable solutions for data preservation and curation is increasing in the Social Science and Humanities. In particular the Digital Cultural Heritage (DCH) sector is producing a large volume of digital content that needs to be safely stored and curated, permanently accessed, and easily shared and re-used by researchers.

At the same time, the so-called 'hard sciences' are demonstrating that research capabilities can really be advanced by the use of e-Infrastructures and that a shared implementation of common e-Infrastructure layers could be beneficial and cost effective. In the last years, e-Infrastructures

and DCH communities have entered a dialogue and now several data-infrastructure projects exist, whose results can be adapted and re-used in the DCH domain. To this aim, to test the services and tools developed to facilitate the storage, access and preservation of digital data and to provide feedback for the further improvement of such services, fruitful cooperation started among these projects.

This workshop aims to:

- Present the first concrete results of the cooperation between different projects
- Attract new projects and initiatives working in the domain of DCH,

e-Infrastructures and digital preservation to find synergies and discuss opportunities for cooperation, starting from concrete use cases.

---

**Antonella Fresa** *Promoter SRL*  
**Claudio Prandoni** *Promoter SRL*  
**Michel Drescher** *EGI.eu*

## EGI and XSEDE tools and common needs for the computational chemistry, molecular modelling and materials science community

The evolution of computing technologies towards networks of remotely accessible top level machines (High Performance Computing, HPC), such as the US XSEDE, meets the needs of intensive number crunching and big data handling applications of computational scientists of various disciplines. Compute time on such machines, however, is preferentially awarded as grants offered to few record breaking ambitious scientific projects selected by means of an evaluation of the proposals submitted after a centralised call. This discourages multiannual development projects in addition to not explicitly fostering integration with local resources. Opposite to the grant system

is the opportunistic model adopted by the European Grid Infrastructure (EGI), in which members of the research communities are offered on demand access to a shared platform made of lower level resources (High Throughput Computing, HTC). But the opportunistic model, while allowing multiannual projects and the integration of several heterogeneous local resources, reduces the possibilities of using large-scale applications.

The workshop will discuss the pros and cons of the model proposed by the COMPChem Virtual Organisation, which combines the use of HTC and HPC machines and fosters both a better long term research

planning and a combination of computational approaches. The analysis will focus on the use for that purpose of XSEDE and EGI platforms and related middleware.

---

**Antonio Laganà** *University of Perugia*

## ENVRI workshop

European Research Infrastructures are facing growing demands for managing, storing, processing and analysing increasing amounts of research data. For these purposes they need advanced e-Infrastructures. Now in its third year, the ENVRI project, a collaboration in the ESFRI Environmental Cluster, has been set up to assist the research infrastructures in this effort. The ENVRI workshop will be divided into two parts.

ENVRI: Dealing with complexity in Environmental Research Infrastructures

This workshop gives an overview of the ENVRI project and its achievements. Emphasis is on

the developed Reference Model a conceptual model that is used to describe computational aspect of research infrastructures. The ENVRI Reference Model provides common language and understanding, promotes technology and solution sharing, and improves interoperability.

ENVRI: Sustainability of Research Infrastructures

This workshop discusses challenges and opportunities in sustained operations of research infrastructures, especially concerning the solutions for data management and processing. The European e-infrastructure are a potential source of providing long-term services for research

infrastructures. The perspectives from ENVRI Sustainability planning activity are reported and discussed together with views from service providers.

---

**Antti Pursula** *CSC*  
**Yannick Legré** *EGI.eu*  
**Alex Hardisty** *Cardiff*  
**Samuel Keuchkerian** *CNRS*

# Software Vulnerability Handling and practical incident recognition

This session will focus on two topics. Software Vulnerability handling and practical incident recognition.

The first part will review the Software Vulnerability Group (SVG) activities and the changes needed to take this activity to the Cloud. The main focus of SVG continues as ever to be to “eliminate existing software vulnerabilities from the deployed infrastructure and prevent the introduction of new ones, thus reducing the likelihood of security incidents”. This will include what anyone (including a user) should do if they spot a potential software vulnerability.

In the second part we will look at a scenario when it all went wrong and you suddenly have

a VM that does more than the things you expect when starting it up. As a hands-on exercise we will provide you with a Cloud-VM, that has several settings/installed software that you probably don't really want there. You have to find them.

---

Sven Gabriel *NIKHEF*



# ENVIRONMENTAL SCIENCE ON GRIDS AND CLOUDS

---

*This session aims to bring together those members of EGI who work with environmental science research communities to exchange experiences on existing services, requirements and trends concerning the use of EGI grid and cloud solutions within this domain.*

**Convenor:** Eleni Katragkou, AUTH

## A virtual Biodiversity lab

e-VirtualBiodiversityLab is a project with a Galaxy interface enabling researchers to run a series of classical tools in molecular biodiversity studies (e.g. alignment, phylogenies, and distances between sequences). A special emphasis has been put on taxonomic annotation from molecular data, i.e. annotate an unknown sequence with the label of the closest known reference in a database. With the first tool, we study the 'shape' of the reference database (organised as clusters of taxa, or continuously varying pattern), and use this information to give an estimate of the quality of the taxonomic annotation. With the second tool, a file with a large number of unknown sequences coming

from an environmental sample (e.g. diversified eukaryotic community in a lake) is compared with a reference database, and an inventory is produced. The amount of sequences produced by Next Generation Sequencers requires work on scalability of these tools. Years ago, the magnitude was a few thousand unknown sequences per run some; current figures are now closer to tens of millions. This task can be easily distributed, and EGI production grid is an ideal infrastructure for sharing the tool. A new layer has been added on the Galaxy server where the job can be launched on the grid, using DIRAC as middleware. This permits the use of an e-lab for biodiversity on the grid. The next

steps are to diversify and enrich those pipelines, and connect it with an IRODS implementation, for sharing data.

---

**Alain Franc** CNRS

# Modelling the impact of climate change on the air quality of Bulgaria using the grid environment

The main goal of the joint research by scientists from the NIMH and IICT in Bulgaria, was to determine climatic values for the most important meteorological variables and to estimate the impact of climate changes on air quality at several key points. The computational work was carried out via a grid-based workflow. Three different scenarios were investigated: 1960-2000 (Control Run, CR), 2020-2050 (Near Future, NF) and 2070-2100 (Far Future, FF), following the IPCC scenario A1B. The calculations were data and compute intensive and required the use of substantial grid resources from the Bulgarian NGI. Using the climatic version of the operational weather

forecast model ALADIN, the team created a meteorological database at resolution of 10 km for all three periods. These data were used as input for the US EPA Models-3 System (MM5, CMAQ and SMOKE). Using the TNO emission inventory for 2000, the workflow computations attempted to determine the impact of climate change, Abstracting out the variability in emissions, which are difficult to predict at the current stage. AUTH-Greece provided the chemical boundary conditions for Bulgaria. The dispersion calculations were made for the last 10-year periods for CR, NF and FF, and are presented and discussed, drawing conclusions about the expected climate

change impact under the internationally recognised scenario. The computations produced high amount of data for each grid job and used MPI. This work was supported by EGI-InSPIRE and SuperCA++ projects.

---

**Emanouil Atanassov** *IICT-BAS*

## Homogenisation of access to DCIs

In most science fields, the complexity of computing models is growing and access to Distributed Computing Infrastructures (DCI) has become essential. Additionally, there are available a large variety of DCIs such as workstations, clusters, grid services or clouds with different Distributed Resource Managements (DRM) as well. This issue can represent a significant barrier for scientists. Therefore, tools to homogenise the access to these DCIs are one of the most important challenges facing.

To deal with this issue, several approaches have been proposed from a resource admin point of view. In this work is introduced a new solution to give

homogenous access to DCIs but without resource admin intervention. This solution, named DRM4G, aims to expand the possibilities of scientists using it for their computing models. One example of that is the WRF4G framework used to simulate large-scale climate experiments. This framework takes advantages of DRM4G to provide a homogeneous access between WRF4G users and DCIs. In order to achieve this, DRM4G enables the submission and management of jobs to DRMs such as FORK, PBS/Torque, SGE, LSF, LoadLeveler and SLURM and grid EGI services such as Globus and CREAM.

This work is partially funded by the Spanish PLAN NACIONAL

de I+D+i 2008-2011 (WRF4G, Ref.# CGL2011-28864) and the European Regional Development Fund (ERDF).

---

**Carlos Blanco** *University of Cantabria*

## BioVeL: A Biodiversity Laboratory exploiting EGI computing infrastructure

BioVeL is a virtual e-laboratory that supports research on biodiversity issues using large amounts of data from cross-disciplinary sources. BioVeL offers the possibility to use computerised 'workflows' to process data, be that from one's own research and/or from existing sources.

A researcher can build his own workflow by selecting and applying successive 'services' (data processing techniques), or by re-using existing workflows available from the BioVeL's library. This virtual e-laboratory cuts down research time and overhead expenses. BioVeL also provides access to a worldwide network of experts.

Within the project several

activities have been carried out to successfully exploit the computing resources that EGI infrastructure is able to provide. We will show the development done in order to exploit standard grid computational resources within a Taverna workflow, together with services deployed in the EGI Federated Cloud task force.

---

**Giacinto Donvito** *INFN*  
**Salvatore Pinto** *EGI.eu*  
**Carole Goble** *University of Manchester*  
**Nuno Ferreira** *EGI.eu*  
**Gergely Sipos** *EGI.eu*

## Web-Processing Services for climate data - with examples for impact modellers

Impact modelling forced by climate data is often connected with big data processing, but a frequent problem is that impact modellers are not optimally-equipped with hardware (computing and storage facilities) nor programming experience for software development. Web Processing Services (WPS) can close this gap and offer a valuable practical tool to process and analyse big data. WPS represents an interface to perform processes over the HTTP network protocol, enabling users to trigger processes through a website. The appropriate processes are predefined, together with access to the relevant data archives where appropriate data is stored. In the case of the WPS we

present here, the data archive of the Earth System Grid Federation (ESGF) is connected with a search process and provides access to the climate data archive of the ESGF for CMIP5 and CORDEX data. Furthermore, the WPS we present conforms with the standardisation defined by the Open Spatial Consortium, allowing combination with WPS from other institutions to establish a network of computing providers. Besides several general processing operations realised with climate operator commands (CDO), a range of specific processes can also be performed within a WPS.

---

**Nils Hempelmann** *Climate Service Center*



# PORTING NEW APPLICATIONS TO EGI

*This session will provide a forum to discuss and explore the latest use cases of application porting to grids and clouds.*

**Convenor:** Karolis Eigelis, EGI.eu

## Non-commutative simulations on the grid

Here we present a first implementation of a set of simulations in the field of non-commutative geometry in the grid framework, using the parametric job tool. Such calculus (typically Monte Carlo-like simulations) are usually done on local resources such as local clusters, our goal here is to open the way to the grid paradigm to such models or even to the 'fuzzy' community.

We will study a quantised  $\varphi^4$  scalar field theory approximating field with  $N \times N$  matrices. We are interested in particular in the phase transitions of the theory as we change the parameters of the action. The quantities of interest are susceptibility and specific heat, as well as other

order parameters.

It is worthwhile to compare the present method with the simulations of such theories on the lattice. In general the simulations of scalar theories on fuzzy spaces are slower than in their lattice counterparts, since the fuzzy models are intrinsically non-local due to the self-interaction term  $\varphi^4$ , and for higher power of the self-interacting the number of operations to calculate each Monte Carlo step  $\delta S$  grows even faster. However, we can expect some advantages with the simulations of other symmetric field theories.

---

**Bernardino Spisso** *Federico II Napoli and INFN*

## Virtual research environment 'optimal engineering design as a service'

Porting engineering applications to the grid and cloud platform can be based on the paradigm of Service-Oriented Computing (SOC) which utilises services as the basic constructs to support the development of rapid, low-cost and easy composition of distributed applications, even in heterogeneous environments.

It is proposed to develop a services repository (platform supporting services and application support services) from which users will be able to composite their design route in the particular engineering field.

The platform supporting services offer the standard operations for service management and hosting (e.g. cloud hosting, event processing and management,

mediation and data services, services composition and workflow, security, connectivity, messaging and storage).

The application support services are created by investigating the generalised engineering design process and selecting its loosely coupled stages and procedures for subsequent transfer to the forms of standardised web-services. It is possible also to analyse the existing mathematical modelling and optimal design software for the possible re-use of the best algorithms and design procedures implementation in creating the repository of application support services.

---

**Anatolii Petrenko** *National Technical University of Ukraine*

## Scaling MATLAB applications to the bwHPC project in the state of Baden-Württemberg, Germany

MATLAB is a popular programming environment for algorithm development, data analysis, and visualisation. MATLAB is used as a fundamental tool for research and development throughout institutions and universities worldwide.

The bwClusters are part of the state-wide bwHPC project in Baden-Württemberg, Germany. bwHPC is the successor of the bwGRiD project, which was a distributed aggregation of computer nodes operated by eight state universities that ended in 2013. Besides the utilisation of the universities clusters, the focus of bwGRiD laid on the transparent centralisation of distributed computers into a

single grid that offered easy-to-use HPC services. Because bwClusters were homogeneous, they did not fit the needs of all research groups. Five of the clusters are still operational and will soon be substituted by the new bwForClusters, which are heterogeneous and offer special research clusters that are tailored to the needs of some of the bigger research groups.

This presentation will show how MATLAB users can take advantage of the bwGRiD and bwHPC to solve computationally and data-intensive problems. In particular, it will show how researchers can easily prototype parallel applications on their desktops and then scale to the clusters without any code

changes. It will also discuss the technical and licensing challenges that have been resolved so that scientists can take advantage of the clusters for their collaborative research activities.

---

**Silvina Grad-Freilich** *The MathWorks*

# Multicore job management in the Worldwide LHC Computing Grid

Two years after the very successful first run of the Large Hadron Collider finished, data taking is scheduled to restart in early 2015. The experimental conditions for this second run include higher collision energy and beam luminosities, both leading to increased data volumes and event complexity. In order to process the data generated in such a scenario, and also best exploit the multicore architectures of current CPUs, the LHC experiments have been developing parallelised data analysis and simulation software. However, workload scheduling in these conditions becomes a complex problem in itself, as computing jobs with a broad range of resources requirements

have to be efficiently distributed across the multiple sites which make up the Worldwide LHC Computing Grid (WLCG). A WLCG Task Force has been created with the purpose of coordinating the joint effort from experiments and WLCG sites. This contribution will present the activities of the Task Force, including the experiences from sites on how to best use the different batch system technologies, the development of advanced workload submission tools by the experiments and the real-size scale tests of the different proposed strategies.

---

**Antonio Perez-Calero**  
**Yzquierdo CIEMAT**  
**Alessandra Forti** *University of Manchester*

## Application porting and web services for the molecular and materials sciences and technologies community

The increasing availability of computer power on grid platforms has prompted the implementation of complex computational codes on distributed systems and, consequently, the development of appropriate visual interfaces and tools able to minimise the skills required by the final user to carry out massive grid calculations. In the present work we analyse the procedures adopted to implement some computational chemistry codes making use of the IGI web portal on the infrastructure of the Italian Grid Initiative (IGI). The applications described here are those considered as use cases during the Italian Grid Training Workshop held in

Rome in January 2014: CRYSTAL (ab initio quantum chemistry for calculations on crystals, slabs and polymers), Quantum Espresso (electronic structure and materials modelling at the nanoscale) and VENUS (chemical dynamics for classical trajectory simulations). The three use cases cover the key situations met when running computational molecular and materials sciences and technologies applications on distributed computing infrastructures (DCI) and in particular high-throughput computing, MPI runs and distributed data management.

---

**Alessandro Costantini** *INFN*

# EU-Brazil Cloud Connect: integrating services for heterogeneous infrastructures

EU Brazil Cloud Connect (EUBrazilCC), a project from the second coordinated call of research projects of the European Commission and the Brazilian Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), aims at developing a heterogeneous integrated e-Infrastructure for research.

EUBrazilCC leverages a set of components for the use of supercomputing, private cloud and cloud opportunistic resources in desktops. EUBrazilCC will expose these resources through programming frameworks and scientific gateways, easing the adaptation and deployment of the applications that use data and computing resources on both

sides of the Atlantic.

EUBrazilCC has three use cases demonstrating the feasibility of using distributed and heterogeneous resources: epidemiology, advanced medical systems simulation and climate change.

The integration of the components takes into account existing standards to maximise interoperability with provisioning systems and existing infrastructures.

This abstract describes the components and access models for the different types of resources. Tools for transparent access to supercomputers such as CSGrid are combined with execution frameworks to deploy applications on multiple private

and public cloud resources (COMPSS) and with complex workflow managers that run on top of different infrastructures through e-Science Central. EUBrazilCC services include a parallel data analysis service (PDAS) for big data analytics and the mc2 framework for scientific gateways.

---

**Ignacio Blanquer** *UPVLC*

# High-performance computing for nanoscale simulations

The Slovak Grid NGI will set up an NGI virtual centre of excellence for application development and consultancy to researchers from various application domains to improve their adoption of EGI e-Infrastructure services. Our work has been focused on grid application activities such as analysis of case studies and solving of either ready simulation programs or development of software packages. Researchers and involved research communities have varying requirements on the e-Infrastructure services and resources and so far 11 case studies have had their details worked out in details, all differing in specific scientific demands and nanotechnology applications.

From the selected case studies, the case study 'Nanoscale modelling and structure optimisation' which uses OOMF and MagPar simulation SW packages will be presented more in detail. New phenomena and exotic magnetic arrangements like vortices or skyrmions are investigated in this case study for practical applications like information storage. Simulations of dynamical processes in magnetic devices are quite CPU time-consuming and new algorithms are developing for increasing simulation speed and efficiency. Related to this case study, the nanotechnology portal is already developed as a pivot step to make easier access for end-users toward

high performance applications on the grid.

---

**Viet Tran** *UI SAV*

# APPLICATIONS ON FEDERATED CLOUDS

---

*This workshop invites interested parties to congregate and discuss the next steps for advanced cloud use cases beyond what is currently requested and supported.*

*For example, advanced topics such as automated scaling of virtual appliances (VA), orchestrated VA deployment and provisioning, integrating software defined networks (SDN), Messaging, DNSaaS etc. are welcome to be discussed.*

**Convenor:** Michel Drescher, EGI.eu

## An introduction to cloud computing

This talk will serve as an introduction to the concepts, technologies and ideas that underlie the cloud computing paradigm and its uses by research and industry. It will include a discussion on EGI's offering in this area, the EGI Federated Cloud, covering standards, technology decisions and added benefit to the EGI community. The talk will provide a foundation that will be built on by the other talks within the session.

---

David Wallom Oxford

# A cloud-based simulation platform for manufacturing, engineering, and beyond

Simulation software is widely used in both industry and academia. However, large scale simulation require extensive amount of resources. Cloud computing offers the potential for scalable, on-demand access to resources that can be used to speed up simulation and make it available from custom web interfaces.

CloudSME is an FP7 European research project that develops a cloud-based one-stop-shop solution to provide a scalable platform for small or larger scale simulations and enable the wider take-up of simulation technologies, especially in manufacturing and engineering SMEs. The CloudSME Simulation Platform supports end user SMEs

to utilise customised simulation applications in the form of Software as a Service (SaaS) based provision. Simulation software service providers and consulting companies also have access to a Platform as a Service (PaaS) solution that allows them to quickly assemble custom simulation solutions in the cloud for their clients.

The CloudSME Simulation Platform is being built on existing and proven technologies enabling the project to deliver its results faster. These baseline technologies, the gUSE/WS-PGRADE framework and the CloudBroker platform, are already widely utilised by EGI user communities. This presentation will provide an overview of the

CloudSME Simulation Platform and the major industry use-cases currently being built. It will also outline plans and opportunities integrating the CloudSME platform with the EGI Federated Cloud.

---

**Tamas Kiss** *University of Westminster, London*

# Successful running of Fermilab experimental simulation on federated grids and clouds

Fermilab neutrino experimental analysis and simulation needed to run mostly at Fermilab due to access to a NAS server for their code and data. Fermilab staff worked with the experimenters to move their application to the CVMFS distribution and then run their application on several distributed grid sites, FermiCloud, and Amazon AWS.

We found both internal and commercial cloud service to give performance and reliability similar to an opportunistic Open Science Grid site.

We will present comparative execution times and efficiencies, work that was done, and suggested best practices for porting and running applications on external grids and clouds.

---

**Steven Timm** *Fermi National Accelerator Laboratory*

# EGI SERVICES FOR GLOBAL SOFTWARE AND COMMON DATA ANALYSIS

---

*The CernVM-FS is firmly established as a method of software distribution for the LHC experiments at the WLCG sites. Use of CernVM-FS outside WLCG is growing steadily, its advantages being acknowledged by other HEP and non-HEP communities.*

*The workshop will explain why CernVM-FS makes it easier for VOs to manage their software and run jobs at sites and also how this technology addresses the problem of application software installation at sites and what are the costs involved.*

*The workshop also compares CernVM-FS with the Frontier Database Caching System which uses the same http proxy cache infrastructure as CernVM-FS but it is optimised for distributing common information from a database rather than software.*

**Convenor:** Catalin Condurache, STFC

## Application characteristics that best fit the CernVM Filesystem and the frontier distributed database caching system

The CernVM Filesystem (CernVM-FS) and the Frontier Distributed Database Caching System (Frontier) are both based on using local http proxy caches provided by grid sites. This approach enables a high volume of simultaneous read requests from similar grid jobs with a small number of servers. CernVM-FS is optimised for distributing software, and Frontier is optimised for distributing data stored in databases. Both systems work great for the applications they're designed for, and could be put to good use in more applications, but they're also both subject to limitations. This presentation discusses the application characteristics that work best with these two

systems and the characteristics that don't work well. Recent and planned features for both systems are included.

---

**Dave Dykstra** *Fermi National Accelerator Laboratory*

# CernVM Filesystem : the tool for software distribution

The Cern VM Filesystem (Cern VM-FS) is firmly established as a method of software distribution for the LHC experiments at the WLCG sites. Use of CernVM-FS outside WLCG is growing steadily, its advantages being acknowledged by other HEP and non-HEP communities.

The presentation will give a status update on the current work carried by the EGI CernVM-FS Task Force to establish a CernVM-FS infrastructure facilitating the efficient distribution of software across the EGI computing resources. It will also highlight the Task Force commitments to help sites to support CernVM-FS at batch farm level and achievements in assisting Virtual Organisations

to migrate their software distribution to the CernVM-FS technology.

---

Catalin Condurache *STFC*

# Managing multidisciplinary software repositories for grid with CernVM-FS

The Finnish Grid Initiative (FGI) has been using CernVM-FS for software distribution since late 2011. Since both the grid and the VO ([fgi.csc.fi](http://fgi.csc.fi)) are both national and multidisciplinary there is a big spread of software, from compilers, math libraries, high-energy physics software, genomics software to physical chemistry software.

A new approach is for CSC, the IT center for science in Finland, to start distributing Chipter, a toolkit for genomics analysis, via CernVM-FS. The toolkit consists of software and a base set of genomics databases that tally in at around 130 gigabytes compressed. The problems of distributing a dataset this big to multiple places in the world make

CernVM-FS a perfect match, especially since few end users will use all parts of the toolkit at once.

For deploying software in both repositories CSC has developed a helper script that makes software installation easy as a normal user on a worker node. Using this, software can be written onto `/cvmfs/` directly, removing the need for path changes later.

---

Ulf Tigerstedt *CSC*

# Software integration and distribution in the catch-all SAGrid VO with Jenkins CI and CernVM-FS

The South African National Grid (NGI\_ZA) operates a catch-all Virtual Organisations (VO) which all South African researchers can use to access computing and data resources. Since it is not a domain-specific VO, there are several widely varying uses for the applications supported by this VO, which have been requested by the users themselves. Members of the VO with the Software Manager role usually did porting and deploying the applications in the past. A great portion of the time taken to port applications consisted in identifying and satisfying the dependencies of the applications, which could not be guaranteed to be available on the worker nodes. This manual process, while

successful, resulted in a long lead-time between identifying new applications and their actual usage once deployed.

This issue has been resolved in the SAGrid VO using two pieces of now mature technology, CernVM-FS and the Jenkins Continuous Integration system. By providing a 'clean' testing environment that replicates the worker node for various architectures and operating systems, the user themselves (with no special roles), can attempt the porting of the application. This is beneficial both to the user and the VO Software Manager, since the former knows their application well while the latter may have no prior knowledge. Once

applications have passed compilation and execution tests, the VO Software Manager merely has to make a few manual checks before staging it to the SAGrid VO CernVM-FS repository, which is mounted on all the sites' WN.

---

**Bruce Becker** *South African  
National Grid*



---

---

# WEDNESDAY

---

---

## **WORKSHOPS ..... 42**

Business development and pay-for-use .....	42
Open access to EGI research outputs .....	42
GPGPU integration and user application support.....	43

## **SESSIONS ..... 44**

New services for astronomy and .....	44
Astrophysics .....	44
Science gateway frameworks .....	48
Evolution of the EGI authentication and authorisation structure .....	52
New data management solutions for EGI .....	56

## **NETWORKING SESSIONS ..... 60**

Pre-commercial procurement on digital preservation.....	60
Networking Life Sciences and EGI towards H2020 .....	61
DIRAC 4 Life Science .....	61
Grid and cloud services for the agricultural community.....	62
Proceeding European joint efforts on climate service products .....	62



# Business development and pay-for-use

Sustainability of EGI requires a multifaceted approach. One aspect of EGI's strategy is the increase of business development activities and the potential addition of pay-for-use models. To achieve such a result will depend on community engagement from a wide range of both technical and non-technical competencies to implement. Advancing from an initial exploratory report approved by the EGI Council in 2013 is a proof of concept that will take place during 2014. This workshop serves as a checkpoint of ongoing activities and an opportunity to discuss key points moving forward.

---

Sy Holsinger *EGI.eu*

# Open access to EGI research outputs

The open access policy to research will be mandatory for all publicly funded projects in Europe within Horizon 2020. Future projects will need to define how to make their research outputs publicly available. As this will be a general concern, EGI has established a collaboration with OpenAIRE to define processes, tools and policies for the EGI community to comply with the open access policy. The goal of this workshop is to explain what new projects need to do to comply with the policy and what services are available within EGI through OpenAIRE to implement it.

---

Sergio Andreozzi *EGI.eu*

# GPGPU integration and user application support

The integration of new computational services, such as General Purpose Graphical Processing Units (GPGPUs), into the European Grid Infrastructure has to be handled in such a way that access to these new services is provided in a manner consistent to the current grid job-submission mechanism.

The GPGPU virtual team was initially established to determine the (then) current and future scale of GPGPU resource deployed at EGI Resource Centres, and also to gauge the impact that these resources would have on users. Subsequently, several EGI workshops presented a set of technical integration challenges in providing such a service.

Furthermore, since September 2013 the GPGPU virtual team has been investigating how to address these problems. The result of this work is the first presentation of a prototype GPGPU service. We shall demonstrate a variety of applications from several scientific disciplines, such as computational biology, molecular dynamic, and astrophysics.

We shall also discuss the remaining strengths and weaknesses of this approach, and will also focus on how this prototype work can be further developed and expanded upon in the future.

---

**John Walsh** *Trinity College,  
Dublin*

# NEW SERVICES FOR ASTRONOMY AND ASTROPHYSICS

---

*The aim of this workshop is to provide an opportunity for the Astronomy and Astrophysics community to meet and share experiences, problems and successful stories, to present new users, new projects and the work done during recent years.*

**Convenors:** *Giuliano Taffoni, INAF | Caludio Vuerli, INAF*

## The data transfer for the MAGIC telescopes

The MAGIC telescopes produce about 150 TB of data per year in the observation site at the isle of La Palma in the Canary Islands (Spain), including raw and processed data at the observation site. This figure is likely to increase by a factor of two or three in 2014, after an update of the readout hardware. The data is transferred to the data centre at the Port d'Informació Científica (PIC) in Barcelona, where it is stored and served to the whole MAGIC collaboration. The data is required to be available at the data centre within 24 hours of the observations. For the high-level processed data, the requirement is to be available before 12:00 (CET) of the following day, so it can be used

in the scheduling of observations. A custom data transfer software has been developed around FTS to meet these requirements and to be robust and reliable, and has been running for about two years. It also includes an easy management interface and a monitoring web page that is open to all MAGIC users. Here we present this software development, a review of its performance and an evaluation on its usability and shortcomings.

---

**Roger Firpo Curcoll IFAE**

# Computing requirements for a several cubic kilometre-sized underwater neutrino telescope

We present the computing requirements of KM3NeT, a future European deep-sea research infrastructure, which will host a distributed network of neutrino telescopes with a volume of several cubic kilometres at the bottom of the Mediterranean Sea. KM3NeT is part of the ESFRI roadmap and will open a new window on the Universe. The telescope will search for neutrinos from distant astrophysical sources like gamma-ray bursts, supernovae or colliding stars and will be a powerful tool in the search for dark matter in the Universe. The KM3NeT computing model is similar to the one of the CERN experiments with different levels of tiers. For the different

steps of data processing (simulation, filtering, calibration, reconstruction and analysis) several software packages are used. The computing requirements of the KM3NeT span from serial to multi-parallel or GPU-optimised jobs. The collaborative nature of the infrastructure demands very frequent WAN data transfers and data sharing among individuals, groups and the public.

---

**Christos Filippidis** *NCSR*  
*Demokritos*

# Workflows for propagation models in Heliophysics

The Sun is responsible for Coronal Mass Ejections - events that can be damaging to our technological infrastructure. There is, therefore, an effort to understand the cause of these eruptive events and how they propagate from Sun to Earth. As the physics governing their behaviour is still not well understood, there is a need to develop theoretical descriptions known as a Propagation Models. It is often difficult to define a single propagation model capable of correctly describing the physics of these events, and it is even harder to implement it reliably and precisely. Finally the validation using real observational data is often challenging on its own. We envisage that workflows can

offer viable solutions to overcome the theoretical and practical difficulties in studying, designing, implementing and validating propagation models. Accordingly, in the ER-FLOW project, we have developed workflows that can be used by the users as 'building blocks' to implement models that accommodate their research requirements.

---

**Gabriele Pierantoni** *Trinity*  
*College Dublin*

## An innovative proposal for the CTA science gateway

The CTA science gateway prototype presents a graphical user interface based on a workflow-oriented framework (gUSE/WS-PGRADE). It is equipped with a flexible SSO (SAML based) Authentication and Authorisation access control. An interactive desktop environment the Astronomical and Physics Cloud Interactive Desktop (ACID) is available through the science gateway. The user will be able to exploit the native graphical interface of the tools provided by the ACID environment. A cloud data service shares and synchronises data files and output results between the user desktop and the science gateway.

---

Alessandro Costa *INAF*

## STARnet gateway federation

Nowadays, the Astrophysics community typically deals with big data coming from real-world observations or large numerical simulations that are impractical to be processed on standalone computers, thus requiring distributed computing infrastructures (DCIs). Science Gateway (SG) technologies are offering a framework to realise and operate DCI applications with little concern for the technical details on how and where the underlying computation actually takes place. An SG allows researchers to store, manage, catalogue, and share data collections or evolving apps which otherwise would be confined in specialised institutes. STARnet Gateway Federation

is a network of SGs operated by different EU Institutes that bring together their expertise. The focus of the federation is to foster research collaboration and advance scientific research by means of developing services, such as authentication, infrastructure access, handling of big data archives and workflow repositories.

---

Ugo Becciani *INAF*



# SCIENCE GATEWAY FRAMEWORKS

*This session will focus on the latest developments on science gateways and workflow systems applied to research problems and user requirements.*

**Convenor:** Nuno Ferreira, EGI.eu

## Development of a new advanced version of the iort\_therapy application in the GARR Science Gateway

To fully address clinical requirements we have improved and refined iort\_therapy, a GeANT 4-based tool for IORT. Nowadays the application models the IORT beam line and calculates the dose in a realistic phantom derived from patient CT scan, improving the accuracy of simulation dosimetric results.

The use of the iort\_therapy application is characterised by an extensive consumption of hardware resources, mainly in terms of RAM (2-6 GB) and management of big-size output files (up to many GB), posing technical issues when executed in a general purpose grid environment.

For these reasons, we are going to implement a dedicated

Computing Element (CE) supporting iort\_therapy development and testing prior its use on the grid. In the future, the implemented nodes could be integrated and used by the whole grid community and thus accessible by researchers working on similar high hardware consumption applications.

---

**Carlo Casarino** *IBFM-CNR-LATO*

## SCI-BUS gateways for a large set of European scientific user communities

SCI-BUS has developed a gateway framework and a customisation technology by which this framework can be turned into a scientific domain-specific gateway. The framework technology is getting more and more popular inside and outside Europe. By January 2014 more than 13,000 downloads from more than 70 countries have happened from sourceforge and more than 80 science gateways have been set up based on this core technology. Eleven science gateways have been created and operated as production services by SCI-BUS project partners, six have been installed by SCI-BUS subcontractors and seven by SCI-BUS associated partners. In the talk a survey of

these science gateways will be shown emphasising how the given communities turned the gateway framework into their application domain-specific science gateway, what features of the framework is mostly used and what is their experiences of using the framework and the experiences of their users using these gateways.

---

Peter Kacsuk MTA SZTAKI

## InSilicoLab science gateway framework

The InSilicoLab framework emerged as an answer to a need for tools that would be able to expose the power of distributed computing infrastructures to scientists in a way that resembles their usual work. A science gateway created with this framework allows researchers to use it as a workspace that organises all data relevant to their research and allows for complex computations in a way specific to their domain of science.

The processing model of InSilicoLab puts the focus on preserving the same functionality and usage patterns the users have on their personal computers. On the visual side, the gateways are tailored to a specific domain of science, or

even a class of problems in that domain - so that the users are presented with an interface that is as familiar to them as possible. The core of the framework provides mechanisms for managing the users' data - categorising it, describing with metadata and tracking its origin - as well as for running computations on distributed computing infrastructures. Every InSilicoLab gateway instance is built based on the core components, but is provided with data models, analysis scenarios and interface specific to the actual domain it is created for.

---

Joanna Kocot CYFRONET  
Tomasz Szepieniec CYFRONET

# The Virtual Environment for a Superior neuro-PsychiAtry: the VESPA project

The VESPA project aims to provide a virtual environment for qualitative and quantitative evaluation, and rehabilitation of motor and cognitive diseases such as mental or linguistic retard, and Alzheimer's Disease. The VESPA system is built on top of a fully immersive, virtual reality system, and is remotely supervised. The system is designed to be deployable not only in rehabilitation centres but also in hospitals, schools, and rest-homes. The Health system will be taken advantage of by VESPA specialists who are enabled by the integrated videoconferencing system to support, monitor and train patients and local support technicians to exploit the power

of the VESPA system. The VESPA project will exploit the power, scalability, and flexibility of clouds and grids to feed a theoretically unlimited number of installations sites through a centralised repository of evaluation and rehabilitation scenarios. The automatic capabilities of continuous monitoring and daily updating of rehabilitation plans, as well as a science gateway based portal available to doctors, VESPA specialists, patients and their families are a plus. VESPA acquires technologies deriving from diagnostic services developed by the DECIDE FP7 project. The VESPA system will enable numbers of children and elderly to carry out their daily

rehabilitation motor-cognitive tasks at schools, homes, and rest-houses assuring continuity into education and assistance, saving parents and relatives from daily trips to rehabilitation centres.

---

**Marco Pappalardo** *Software Engineering Italia SRL*



# EVOLUTION OF THE EGI AUTHENTICATION AND AUTHORISATION STRUCTURE

*A federated e-infrastructure needs to harmonise user authentication solutions and provide tools for attribute management. This session collects contributions from users and providers on Authentication Authorisation Infrastructure (AAI), plus a discussion segment.*

**Convenor:** *Peter Solagna, EGI.eu*

## e-Science gateways with external attribute authority

Higher Education External Attribute Authorities (HEXAA) is a GN3+ open-call project, building upon SAML attribute request facility, relying on the results of the eduGAIN project and considering the requirements of worldwide research communities and federations.

The model currently used in federated attribute handling derives from the concept that Identity Providers (IdP) are the sole sources of the required attributes of the Service Providers (SP). However, this model cannot accommodate real-world requirements.

Using HEXAA will not only protect users' personal attributes but also facilitate the release of attributes via standardised

methods with the users' full consent and control. This software solution will enable the virtual organisation (VO) administrators to maintain a list of supplementary attributes for specific services (i.e., attribute release policy); it will also enable users to maintain some of their own attributes (i.e., user profile); and finally it will enable VO administrators to maintain VO-specific user attributes (i.e., users' VO profiles). We believe that this model will also significantly increase the ability of IdPs and SPs to join federations.

One of the key objectives of HEXAA is to support e-Science applications with federated identity management and

external attribute authorities. We will present the advantages of the integration of WS-PGRADE/gUSE-based science gateways and HEXAA.

**Mihály Héder** *MTA SZTAKI*

# A solution for access delegation based on SAML

During this presentation we would like to show a solution to access delegation which we have implemented, based on 'SAML Condition for Delegation'. In particular, a message flow protocol has been designed and RESTful web services, based on OpenSAML library, have been implemented to provide service oriented delegation.

Access delegation enables a user or process to act on behalf of another user or process: in other words, a user or process obtains a subset of the privileges of another user or process, preserving their identity. Several solutions are currently applied to implement it for example, proxy certificates, OAuth 2.0, and KeyStone-bearer token. According to the analysis

performed, the first solution has poor performances in grid environment, and the second one is not totally secure, even if it is very popular in the mass market. The third one does not preserve the identity of the delegate, so it is not a full delegation technique, but an 'impersonation' process which is much less secure.

SAML is a very popular solution for identity federation: OASIS specification 'SAML 2.0 Condition for Delegation' extends SAML Tokens to obtain delegation support. This solution has been chosen after some comparisons with the other mentioned technologies. It has been considered the best one because it is based on a consolidated standard, it

implements full delegation, it is very secure because messages are signed, and does not introduce a strong performance degradation.

**Ciro Formisano** *Engineering  
Ingegneria Informatica*

# The Grid IDentity Pool: an example of worldwide 'cross-domain' identity federation

Although identity federations belonging to the education and research world gather more than 17 million users from about 1,500 organisations, and more than 2,100 services (figures from REFEDS), there are large areas of the planet, especially in the developing regions, where organisations do not have identity providers in place and where scientists thus have less chances to access web-based services that could effectively bridge the gaps in e-Science adoption and e-Infrastructure exploitation. In order to address and overcome these limitations, and as a framework to promote the establishment of identity federations and federated identity services in various

regions and within different virtual research communities, the Grid IDentity Pool (GrIDP) was created and has been in production for about two years. GrIDP is an identity federation supporting cross-institutional e-Infrastructure services and communities and providing federated authentication to its members. Through GrIDP, users can access high-level services, such as science gateways, either using the credentials provided by the organisations they belong to or those released by catch-all IdPs.

In this contribution we present the GrIDP federation, its identity and service providers and the results of the work done in the context of projects such as

CHAIN-REDS and eI4Africa to promote the establishment of federated services in the Arab countries, in sub-Saharan Africa and in Latin America.

**Roberto Barbera** *University of  
Catania and INFN*  
**Marco Fargetta** *INFN*

## Improved resilience and usability for science gateway infrastructures via integrated virtual organisations

The workflow-enabled MoSGrid science gateway is especially tailored to perform advanced molecular simulations in a user-friendly way. It employs a distributed science gateway infrastructure applicable for science gateways in general. A key focus is the high usability by creating an intuitive virtual environment. Since the underlying distributed computing infrastructure (DCI) including job, workflow and data management is complex, it becomes a constant challenge to ensure high availability. The goal of our work is to improve the resilience and usability of the MoSGrid science gateway. This was achieved regarding the security by easing the registration of users

and lowering the complexity of their authentication. Without the following solution, users had to apply for a membership in a virtual organisation (VO) and the application was manually handled via a VO manager. In our solution, a user can register via an intuitive process and after his account is unlocked he automatically becomes a member of the MoSGrid VO. The distinguished names (DNs) of these are mapped to the respective cluster-specific accounts. To enable this, a series of modules was created to allow the extraction of user information, make it available via secure download, and finally use it to configure the DCI user database. The efforts made it

possible to depreciate a service, removing a potential source of error and simplifying the registration procedure.

---

**Richard Grunzke** *Technische Universität Dresden*



# NEW DATA MANAGEMENT SOLUTIONS FOR EGI

*This session is focused on data management. It will cover applications, workflows, resources and projects.*

**Convenors:** *Helmut Heller, LRZ-DRG | Sasu Tarkoma, Helsinki University | Balasz Konya, Lund University*

## Long term data preservation for CDF at CNAF

After the end of data taking in 2011, CDF is now facing the challenge to both preserve the large amount of data produced during several years and to retain the ability to access and reuse it in the future. The CDF Italian collaboration, together with INFN-CNAF computing centre, has developed and is now implementing a long term future data reservation project. The project comprises the copy of all CDF raw data and user level n-tuples (about 4 PB) at CNAF and the setup of a framework that will allow access and analysis of the data in the long term. In this talk we first illustrate the difficulties and technical solutions adopted to copy, store and maintain CDF data at

CNAF. We then describe how we are exploiting virtualisation techniques to build the long term future analysis framework, and the validation tests under development to check data integrity and software over the time.

---

**Luca dell'Agnello** *INFN*

## Data Avenue: a flexible data transfer service among various DCIs

Globus Online is a popular service to transfer data among Globus grid storages. Data Avenue is a similar service but with much more flexibility. It is a file commander tool for data transfer, enabling easy data moving between various storages services (such as grid, cloud, cluster, supercomputers) by various protocols: http, https, sftp, gsiftp, srm.

With Data Avenue users can up- and download their data to storage services for scientific computation. Additionally, they can copy, move and delete files as well as they can create and copy folders.

The Data Avenue family consists of three members, Data Avenue Blacktop, Data Avenue and Data

Avenue@SZTAKI

Data Avenue Blacktop is the core service that provides unified access to different types of storages. It offers a web service interface for performing different operations on the supported storages (SFTP, GSIFTP, HTTP, HTTPS, SRM, Amazon), and an HTTP server for convenient data up- and download. The Data Avenue core service can be accessed with an easy-to-use Java API making integration with existing software components an easy task.

Data Avenue is the user interface that makes data transfers easy and user friendly.

Data Avenue@SZTAKI is the SZTAKI installation of the Data Avenue service. It can be used by

those who choose not to install their own service or just would like to see how it works!

Data Avenue is also realized as a Liferay portlet and hence every science gateway based on Liferay can use this portlet to provide the service for their users.

---

**Peter Kacsuk** *MTA SZTAKI*

**Zoltan Farkas** *MTA SZTAKI*

**Akos Hajnal** *MTA SZTAKI*

## A CHAIN-REDS solution for data workflows

This work is presented on behalf of the CHAIN-REDS project and proposes a solution for data workflows. The project has developed several tools that provide an easy and intuitive access to repositories hosted worldwide:

The Knowledge Base displays information about open access document repositories and data repositories including name, organisation that manages it, country it belongs to, scientific domain it covers and direct links to access it.

The Semantic Search Engine allows any user searching for a specific term in the previous repositories and also in those managed by the ENGAGE platform.

A Science Gateway for executing unattended jobs on grid, cloud and local clusters.

Profiting from all these tools and the technological development underneath and by assigning Persistent Identifiers to specific datasets, a workflow will be presented that fully covers the common research cycle: search of specific term or data (repositories); retrieval of the associated raw data; use of these data as input for a specific application or code; execution of the latter; production of new data (results); and, upload and storage of the new publication and dataset into repositories already displayed in the Knowledge Base in order to start the cycle again.

---

**Rafael Mayo-Garcia** *CIEMAT*

**Manuel Rodriguez-Pascual**

*CIEMAT*

## WebFTS: a graphical frontend for file transferring with FTS3

We present a web-delivered file transfer and management solution which allows users to invoke reliable, managed data transfers on distributed infrastructures. The fully open source solution offers a simple graphical interface through which the power of the FTS3 service can be accessed without the installation of any special grid tools.

Created following simplicity and efficiency criteria, WebFTS allows the user to access and interact with multiple storage elements. Their content becomes browsable and different filters can be applied to get a set of files to be transferred. Transfers can be invoked and capabilities are provided for checking the

detailed status of the different transfers and resubmitting any of them with only one click.

The 'transfer engine' used is FTS3, the service responsible for distributing the majority of LHC data across WLCG infrastructure. This provides WebFTS with reliable, multi-protocol (gridftp, srm, http, xrootd), adaptively optimised data transfers.

---

**Andres Abad Rodriguez** *CERN*

## Big data analytics in the EUBrazil Cloud Connect project

The EUBrazil Cloud Connect (EUBrazilCC) project is a first step towards providing a user-centric, cross-Atlantic test bed for European & Brazilian research communities. It includes the implementation of three multidisciplinary & highly complementary scenarios, covering Epidemiology, Medical Systems simulation, Biodiversity & Climate Change. This contribution is strongly related to the 3rd use case which has a special focus on big data analytics and it is built on a close collaboration among European & Brazilian excellence centres. It deals with multiple multidimensional and heterogeneous data sources such as output of global and regional

climate models simulations and different types of satellite data. To address the use case requirements, the EUBrazilCC project is providing a cloud framework for big data analytics (named PDAS) joining novel storage models, HPC and parallel database management solutions. The framework carries out (near) real time data analytics tasks (e.g. data reduction, time series analysis, data slicing) on large multidimensional datasets, exploiting parallel data operators (MPI/OpenMP based). The PDAS exposes a WS-I interface (GSI/VOMS enabled) to interoperate with the EGI infrastructure. The implementation of a cloud interface to enable elastic

resource provisioning as well as the interaction with the COMPSs framework to provide workflow-based analytics on massive volumes of data are major goals to be addressed during the EUBrazilCC project implementation.

---

**Sandro Fiore** *Euro Mediterranean Center on Climate Change (CMCC)*

# iRODS: setup and use of a national data management system in the French NGI

Scientific communities are acquiring more and more data and the need for tracking and managing of data in a simple framework is essential for providing scientific results in a timely manner. In France, several laboratories have gained experience with iRODS (Rule-Oriented Data Management System), which acts as a data grid by providing transparent access to data which can be spread over different physical locations on heterogeneous storage technologies.

Five of these laboratories (CCIN2P3 in Villeurbanne, IPHC in Strasbourg, LPSC in Grenoble, MCIA and CBIb in Bordeaux) have combined their efforts by providing a single

iRODS infrastructure within the framework of the French NGI.

This system is intended for hosting users from any scientific domain in need of storage or data management, but with little experience or dedicated personnel in the domain of data management. In this contribution, an overview of the iRODS features is given, followed by the description of the national iRODS instance.

The hosting of end users and the early adopter experience are also presented, as well as the development of the project towards building a national pool of expertise for the use and the administration of the iRODS product.

---

**Catherine Biscarat** CNRS

**Jerome Pansanel** CNRS

## Onedata global data storage

The increasing number of powerful computing environments is creating problems more difficult than the increase in data volumes. Researchers need to take into account aspects such as the kind of data and the processing speed required to use the large amount of information. The different requirements of multiple groups of users make it necessary to provide different storage systems and service that process data. It is usually not possible to provide this all at one site so many users have to process their data in multiple data centres. However, data management in such a distributed environment is too complicated for many of them. The authors present a novel

solution, called Onedata, that simplifies user's work in organisationally distributed environments by provision of a uniform and coherent view of all data stored in all the data centres used. It also supports work in groups, data sharing/publication and serves data efficiently.

However, simplification of the system from the user point of view results in increase of number and difficulty of management tasks that have to be done by administrators or automatons. Hence, Onedata provides functionalities that also simplify administrators work: automatic rule-based data management, the infrastructure state monitoring data gathering and visualising and data

protection from unauthorised access.

This paper describes Onedata in terms of its architecture, current implementation status and presents exemplary use cases.

---

**Lukasz Dutka** CYFRONET



# Networking Life Sciences and EGI towards H2020

The increase in complexity of studies and techniques within Life Sciences (LS) lead to an explosion in the demand of computational resources. From Big Data to High Performance Computing (HPC), from the development of tailored solutions to general use applications, the LS community is one of the target communities for future utilisation of computational resources. The European Commission has recently presented the work programmes of Horizon 2020, where multidisciplinary approaches are a key factor for success. This is particularly true for LS calls.

This networking session will bridge technology providers with users having the H2020

framework as a background.

This session will stimulate discussion of ideas and requirements from both EGI and LS communities, thus identifying common grounds for future collaboration, boosting the success of proposals to be submitted.

The session should be an open meeting where delegates can share their thoughts and is closely related to Tuesday's 'Going beyond grid to enable life science data analysis' Life Sciences workshop.

---

**Afonso Duarte** *ITQB-UNL*  
**Fotis Psomopoulos** *IAB, CRT*  
*Hellas*

## DIRAC 4 Life Science

DIRAC has started a collaboration with EGI.eu to offer an open service to the European e-Infrastructure. This service will be piloted by several communities of different fields (e.g. Astrophysics, Life Science, Engineering) and will then be offered to others as single access point to all kind of distributed computing and storage resources in Europe and beyond.

The Life Science communities are offered a dedicated networking session with the DIRAC 4 EGI team to analyse their use cases and requirements to join the initiative.

This networking session will be followed by a hackathon session in which some selected use cases will be further explored

for its online incorporation to the initiative.

---

**Ricardo Graciani Diaz** *University of Barcelona*  
**Andrei Tsaregorodtsev** *CNRS*  
**Tom Fifield** *UNIMELB*

# Grid and cloud services for the agricultural community

The European Grid Infrastructure provides (among others) grid and cloud powered services to user communities in order to cover their infrastructure needs for data processing. One of the communities that over the past years have been increasing its needs for access to such services for powerful and large-scale data processing is the agricultural community.

Initiatives such as the FP7 agINFRA project have been trying to fill this gap, by creating an e-infrastructure and integrated services specified for the agriculture community. Under this scope, this networking session will try to bring together stakeholders to discuss and exchange ideas

regarding the needs of the agricultural community from e-Infrastructures (such as EGI), and outline some available solutions and joint plans. The expected outcome is to receive feedback from the needs of such specific community as well as gather the perspectives of the participants through a constructive dialogue. This networking session will provide useful input for a planned EGI virtual team targeting specific needs of the agriculture community.

---

**Robert Lovas** *MTA SZTAKI*  
**Andreas Drakos** *Agro-Know Technologies*

## Proceeding European joint efforts on climate service products

Climate service is obliged to develop and provide information products, data and advices based on past, present and future climate for the business community, policies and public institutions. The information about climate, climate change and its impacts on natural and human systems is tailored to suit specific user needs including user specific guidance of how to use the information.

There are several European organisations involved in climate services. Although institutional arrangements differ, the climate service objectives and climate service customers are rather similar. There are various on-going incorporating activities for climate related communities

with different focus e.g. the Joint Program Initiative (JPI) on climate or the European Climate Service Initiative. Still, with regard to Horizon 2020, the climate service community requires a first network for a potential proposal in 2015 and to discuss the development and impact of upcoming climate service products.

The session will be informal but is intended to precede plans for joint activities in the development of prototypes of climate service products in regards to following questions:

- How to collect ideas of prototypes from the community in an effective

manner without duplication of efforts?

- How to organise the transition from a prototype to the phase of operationalisation of a product?

---

**Daniela Jacob** *Climate Service Center*



---



---

# THURSDAY

---



---

## **WORKSHOPS ..... 66**

- OpenNebula in science/HPC and cloud federation ..... 66  
 Monitoring availability and reliability in EGI-inSPIRE and beyond ..... 66

## **SESSIONS ..... 68**

- Community building and engagement ..... 68  
 Dirac virtual research environment for EGI ..... 72

## **TUTORIALS ..... 78**

- Web processing services for climate data with examples for  
 impact modelers ..... 78  
 Foundation training in federated IT Service Management according to  
 FitSM-1:2013 ..... 78  
 Managing virtual servers ..... 79  
 Introduction to Stoxy - CDML-compliant lightweight storage proxy ..... 79  
 Training on data preservation ..... 80  
 OpenStack hands-on ..... 80



## OpenNebula in science/HPC and cloud federation

OpenNebula is being used by many supercomputing and research centres to build high performance computing and science clouds for hosting large-scale virtualised computing and data processing environments. These science-optimised clouds are providing service-driven, on demand access to scientific and technical computing capabilities to solve complex problems and drive innovation. OpenNebula is also one of the main cloud management platforms used in several federated cloud e-Infrastructures such as The EGI Federated Cloud.

The workshop will serve as a meeting point for users, operators, and researchers of OpenNebula clouds in

science and HPC, and a unique opportunity for discussion and collaboration with related projects and federated cloud e-Infrastructures. This event will focus on:

- Presenting use cases and deployment experiences
- Introducing new integrations and developments
- Discussing limitations and potential enhancements
- Collaborating with other projects and communities

---

**Ignacio M. Llorente**

*OpenNebula/C12G and UCM*

**Ruben Montero** *OpenNebula Project*

**Daniel Molina** *OpenNebula Project*

## Monitoring availability and reliability in EGI-inSPIRE and beyond

In this workshop we will present the current status and the future of the availability and reliability monitoring service, which is now operated and further developed by GRNET, CNRS and SRCE.

Starting from May 2014 the central SAM services are going to be operated by a consortium comprised of GRNET, CNRS and SRCE. The consortium has taken over the task of maintaining and further developing the existing SAM framework. The first component of the framework that was re-designed and re-implemented from scratch was the Availability Computation Engine. Based on the new framework that in the heart of the new engine, we have planned to re-design and implement the rest

of the monitoring framework in order to make it more lightweight, sustainable and robust.

In this workshop we will present the current status of the SAM services. We will show case the new Availability and Reliability Computation Engine and we are going to discuss the roadmap for the new developments that will take place in the next year.

---

**Christos Kanellopoulos** *GRNET*



# COMMUNITY BUILDING AND ENGAGEMENT

*EGL's sustainability plans have become increasingly coupled with its long-term strategy to connect researchers from all fields of science across the whole European Research Area (ERA) with the reliable and innovative ICT services from EGL that they need to undertake their collaborative world-class and world-inclusive research.*

*The EGL Outreach activity integrates and mobilises experts, tools and support from the community to engage with communities and help them evolve into active and self-sufficient users of e-infrastructures.*

*This workshop will provide opportunity for scientific communities to present and discuss the status and plans of their use of EGL, and to discuss and resolve any outstanding issue with the EGL community.*

**Convenor:** Gergely Sipos, [EGL.eu](http://EGL.eu)

## Teaching grid to the masses: the SURFsara & EGL-InSPIRE grid computing MOOC

Massive Open Online Courses (MOOCs) allow people from various backgrounds to receive education on almost any topic, and have recently become very popular because of this. In this EGL-InSPIRE project, a MOOC was created to teach grid computing to a wider audience. For this purpose, lectures were recorded, animations were created and grid users were given the opportunity to demonstrate how grid computing impacts their science. Participants were given the opportunity to test their new knowledge using quizzes, which were mandatory to complete the course, and practical assignments. At the end of the course a final assignment needed to be handed in.

The MOOC itself was hosted on a platform developed by the University of Amsterdam and the videos were made available on YouTube. The course attracted around 300 participants, of which 10% managed to finish the course.

---

Jan Bot SARA  
Anatoli Danezi SARA  
Jeroen Schot SARA  
Tijs de Kler SARA

## The Finnish Grid Infrastructure computing environment and tools

The Finnish Grid Infrastructure (FGI) consortium started its activity in late 2011. This ten-site distributed computing infrastructure, funded by Finnish Science Academy, CSC - IT Center for Science, Ltd. and nine Finnish Universities, replaced the national materials science grid project (M-Grid).

In this presentation we will take an overview of FGI focusing on its computing environment and tools.

The majority of users on FGI run third-party applications provided through modular Runtime Environments and made available by a CernVM-FS instance that guarantee a common and up-to-date software pool for all clusters. The availability of a

large software selection through Runtime Environment definitions has been one of the key factors for FGI's success.

Arcrunner, an in-house developed grid job management tool, automatizes job submission, monitoring and results retrieval tasks and provides a simple way to process large parallel computing tasks. This tool is being used, for example, on analysis tasks with massive R and MatLab.

Arcrunner is also used in application specific grid interfaces, especially for bioinformatics applications like BLAST, InterProScan and Exonerate, where all steps have been embedded into a simple interface that hides the

complexity of grid middleware from the user.

Some usage statistics, lessons learned and future projects like the Finnish grid and cloud infrastructure will be presented as well.

---

Luis Alves CSC

## Towards a Material and Molecular Science Research Community

The Chemistry, Molecular & Materials Science and Technology (CMMST) community has developed a set of activities, software and human services aimed at supporting Computational Chemistry applications on the European Grid Infrastructure (EGI).

Central to this work is the synergy established between the MOSGrid and ScalaLife projects and the COMPCHEM, GAUSSIAN, IBERGRID Virtual Organizations (VO) of EGI. They have formulated a proposal for the setup a specific CMMST Virtual Research Community (VRC) in EGI. The proposal is a result of the EGI-InSPIRE Virtual Team (VT) project "Towards a CMMST VRC" promoted by the

members of the COMPCHEM VO.

Central aspects of the document are the creation of a library of interoperable codes, the implementation of a set of tools facilitating the selection of appropriate computing elements, the design of a credit system rewarding the work done on behalf of the community and the adoption of de facto standards for quantum chemistry and molecular dynamics data formats.

---

Antonio Lagana UNIPG

# Developing new GridPP user communities: a case study with CERN@school

We present an account of the GridPP Collaboration's efforts to engage with and build up a new community of grid users for the CERN@school project. CERN@school is a research programme initiated at the Simon Langton Grammar School for Boys in Canterbury, UK. It brings detector technology developed by the Medipix Collaboration into schools. The fundamental premise of CERN@school is that students should drive their own research programmes and pursue their own lines of scientific enquiry. However, without access to computing resources capable of storing, processing and sharing the data collected with the CERN@school equipment, this goal becomes difficult to realise.

As part of its commitment to engaging new communities with grid technologies, GridPP has worked with the Langton Star Centre to capture the data storage, processing and access requirements of the CERN@school scientific programme. The fruits of this collaboration, as well as the lessons learned and plans for the future, are reported in this session for wider dissemination and discussion.

---

**Tom Whyntie** *QMUL*

## The ITaaU method for fostering collaborative multidisciplinary utilisation of IT in the digital and post-digital economy

IT as a Utility Network+ has been running since the summer of 2012 under the Research Councils UK Digital Economy theme. The network's purpose is to foster deeper understanding of new technologies and services through multidisciplinary research activities for diverse user communities. Such understanding can be gleaned through focused workshops, innovative pilot projects and cross-sector secondments. The user communities that have engaged with the ITaaU Network to date include: user interaction designers, food security experts, technologists for emerging economies, trust and security experts, accessibility experts, librarians and other information

specialists. As a result of the projects, secondments and workshops funded to date, the network has produced a wide and rich portfolio of collaborative research from teams that might not otherwise have been able to help each other.

In addition to the research elements of these activities, we have been pleased to see the resulting innovation leading to adoption in critical situations. These have included the coordination of mountain rescue teams, government archives and potentially, for improving security in the food chain.

At the heart of these activities has been a drive to further our collective understanding of how open innovation can benefit large

and small enterprises, start-ups, academic labs and government agencies in the delivery of IT utilities that best meet the needs of users.

---

**Steve Brewer** *University of Southampton*

# Platform of national e-Science centres in Europe

The advancement of science has been strongly stimulated by the very existence of advanced e-Infrastructures, such as provided by European and national back bone networks and the resource infrastructures on top of that (PRACE, EGI, EUDAT and ESFRI-facilities). Gradually, however, the focus on provisioning e-Infrastructures and ICT is shifting towards their innovative deployment in science. Enhancing science and so facilitating new discoveries by (optimal) use and re-use of techniques, software, tools and methodologies across disciplines is what the future of science, research and development will depend on. e-Science is the discipline that addresses this

shift in focus that goes well beyond the Big Data wave, however much this wave floods our present thoughts about conducting modern science. In order to strengthen the position of e-Science as a domain *per se*, the aspirations of the field, the skills level of the scientists working in that field and the broader educational aspects a European platform of centres is being implemented (provisional name PLAN-E for Platform of National e-Science Centres in Europe), to bundle present knowledge and expertise across Europe and to define a practical work program towards close cooperation between centres involved in conducting e-Science.

---

Patrick J.C. Aerts *NLeSC*

## Belle II computing model in relation to the EGI infrastructure

The Belle II experiment is a next-generation B factory experiment at the KEK laboratory, Tsukuba, Japan, scheduled to start operating in 2016. Belle II is expected to collect about fifty times more data than its predecessor, the Belle experiment. By 2022, the data collected volume is expected to be comparable to the data volume of the LHC experiments. Processing and analysing such a huge data sample requires a new computing model for Belle II, updated from the one used for Belle where all computing resources were provided by KEK to a distributed computing scheme. The most efficient method to handle this challenge is by using a system,

which exploits a combination of different types of computing resources. The designed Belle II computing system is based on DIRAC, which provides an interface to large dedicated clusters, grid infrastructures and cloud resources, and AMGA for the management of files metadata. A common software framework is used in the entire chain from the data acquisition up to the data analysis stage. It is designed in a modular way, steered via python files and supports parallel execution on multi-core nodes.

The Belle II experiment is an international scientific collaboration with more than 600 scientists and institutions from several European countries.

Good EGI infrastructure support for the Belle II project could win important support from institutions in European countries, thus enabling the Belle II collaboration to achieve its scientific goals.

---

Marko Bracko *JSI*

# DIRAC VIRTUAL RESEARCH ENVIRONMENT FOR EGI

*This session is dedicated to report on the latest activities related to the DIRAC 4 EGI pilot service installation.*

*DIRAC pilot for EGI is a prototype for a full scale DIRAC installation of an open Virtual Research Environment (VRE) service for any interested research community (or individual researchers) requiring access to distributed e-Infrastructure (grid, cloud or others).*

**Convenors:** *Ricardo Graciani Diaz, University of Barcelona | Andrei Tsaregorodtsev, CNES*

## The HADDOCK WeNMR portal: from gLite to DIRAC submission in three hours

The HADDOCK web portal is a widely used scientific portal for the modelling of biomolecular interactions. It makes use of the EGI grid infrastructure for job submission. Users interact through a user-friendly web interface. Each user submission translates into several hundred individual grid jobs that are handled by the complex workflow beyond the portal. To date HADDOCK counts over 3,600 registered users worldwide and has resulted in the last year in over 860,000 grid jobs for a total of over 150 CPU years (normalised CPU time (kSI2K), as reported by the EGI Accounting Portal).

Until recently, the grid-enabled portal was only making use of

standard gLite-based submission and retrieval of jobs via a user-interface. In collaboration with the DIRAC 4 EGI team, we implemented a DIRAC client in a few hours, adapting very easily the submission machinery of the portal. The first tests indicate a very efficient submission process, with a larger number of jobs being handled within a defined time window compared to standard gLite-based submission. The DIRAC client also has the advantage that it does not require any local EMI middleware installation, which reduces the burden of local system administrators.

With the DIRAC service in place, we now have a clone of our HADDOCK portal, which is

expected to increase its overall output and directly benefit our end-user community.

---

**Alexandre Bonvin** *eNMR/WeNMR (via Dutch NGI)*

# Towards harmonised workload management for the biomed VO with DIRAC

Biomed has been the most active Life Science Virtual Organisation (VO) in EGI for the last three years, representing 77% of the normalised CPU time consumed for Life Sciences on the infrastructure. For various reasons, workload management was never centrally coordinated in the VO, which resulted in two issues:

1. Central monitoring is lacking.
2. Computing resources are not efficiently exploited.

To address these issues, we are investigating the adoption of DIRAC as a central workload management solution in the VO. In addition to central monitoring,

DIRAC provides an efficient pilot-job mechanism to balance the load among computing elements. However, much remains to be done for biomed to be able to exclusively rely on DIRAC for workload management. In particular:

- One should make sure that DIRAC can actually be used by all existing users, tools, frameworks and portals in the VO.
- User-level documentation should be widely available and maintained.
- A production-level DIRAC instance able to support the whole VO activity should be deployed and operated.

- Once this is addressed, we are confident that using DIRAC at a VO-level would globally improve workload management, as already reported by large international collaborations.

That would also open other perspectives in terms of data management and access to different resource types.

---

**Tristan Glatard** *CNRS*

**Franck Michel** *CNRS*

**Johan Montagnat** *CNRS*

**Jerome Pansanel** *CNRS*

**Ricardo Graciani Diaz** *University of Barcelona*

**Andrei Tsaregorodtsev** *CNRS*

## DIRAC 4 EGI: report on the experience

DIRAC 4 EGI pilot service was proposed at the end of 2013 to provide open access to the European distributed e-Infrastructure. This presentation will report on the experience from the point of DIRAC:

- the requirements of the different communities joining
- how we approached the different use cases
- usage of the service so far (e.g. number of users, number of executed tasks, used resources)
- response from the resource providers

---

**Ricardo Graciani Diaz** *University of Barcelona*

**Andrei Tsaregorodtsev** *CNRS*

# Optimisation of diesel injection using grid computing

Despite the knowledge of the turbulence as one of the factors inducing jet atomisation, improving the mixing of air and fuel, and therefore combustion, there are still many unresolved questions about the interaction between the turbulence and cavitation, and its effects on the development of diesel jet.

This abstract shows the results of a study using grid computing to model the internal flow of diesel jets in cavitating conditions and including the effects of turbulence by RANS and LES methods.

This study involved the execution of over 500 RANS and LES simulations, evaluating the influence of the position of the needle on the internal flow for

three levels of injection pressure and 21 discharge pressures.

The execution requires the use of OpenFOAM, which requires a complex installation and configuration, being an interesting case for IaaS. The performance obtained using directly WMS-enacted jobs is being compared with the use of DIRAC framework for pilot jobs.

---

**Miguel Caballer** *UPVLC*

# Open access to hybrid clouds with DIRAC

DIRAC project has been involved in some use cases in the EGI Federated Cloud (Fedcloud) for testing purposes. It has demonstrated the ability to aggregate different IaaS providers in an interoperable and transparent manner to support SaaS in smart science. Recently, DIRAC for EGI pilot portal is deploying an open access to e-Infrastructures for several use cases. In addition to traditional EGI grid resources there is the possibility to connect with the EGI Fedcloud, which is the simplest deployment option for some of the use cases because particular platform requisites can be easily integrated in a virtual machine. Furthermore, technical collaborations between

DIRAC and commercial clouds opens the possibility of using DIRAC as a cloud broker of an hybrid nature, which depending on user requirements can choose private cloud, federated cloud or commercial cloud in a transparent manner.

---

**Victor Mendez** *UAB*

# DIRAC 4 EGI: roadmap towards H2020

DIRAC 4 EGI pilot initiative is intended to prove the feasibility of a pan-European DIRAC installation providing open access to e-Infrastructures to any interested community or individual.

After collecting the feedback from the initial providers and users of the service, its long term operation and strategy for the incorporation of new communities will be discussed. The vision of the proponents, EGI.eu and the DIRAC, is that this service will provide a simple and powerful, single entry point to the available e-Infrastructure. At the same time, DIRAC offers to its users the possibility to customise their experience building their own virtual

research environment using the same DIRAC framework or connecting their own solution to the provided service.

The strategies to fund these activities in the context of Horizon 2020 will be presented.

---

**Ricardo Graciani Diaz** *University of Barcelona*

## Application management framework to dynamically exploit grid and cloud computational infrastructure: JST and DIRAC

Nowadays, the execution of a huge number of concurrent applications is a common problem for several small and medium user communities (e.g. Bioinformatics, Biomedicine, Astrophysics). Moreover, the available computational technologies able to exploit both grid and cloud infrastructures require a step forward in terms of abstraction and flexibility of the tools used to manage application submission. This work presents a detailed comparison in terms of functionalities and performance between JST and DIRAC, which provides standard web services interfaces (REST or SOAP). In particular, capabilities of such tools are described, in order to outline the possibility to submit

job from standard high-level application, such as workflow management systems (Taverna, LONI pipeline, Galaxy), exploiting different kind of computational resources in a seamless way. The work conducted in order to exploit both JST and DIRAC within standard web portals technologies like Liferay is discussed.

---

**Giacinto Donvito** *INFN*  
**Giorgio Maggi** *INFN-Bari*  
**Vincenzo Spinoso** *INFN*  
**Stefano Nicotri** *INFN-Bari*  
**Pasquale Notarangelo**  
*INFN-Bari*

# Report on the experience of hosting and operating DIRAC 4 EGI pilot instance

An instance of DIRAC, hosted by CYFRONET and the Polish NGL, has been set up early 2014 as a pilot for many user communities within EGI. From an operational point of view, this pilot instance allows to implement and test at a large scale a collaborative operations model involving NGLs, user communities and DIRAC developers, which is largely based on previous experiences of multi-disciplinary DIRAC instances, such as the one operated by France Grilles, the French NGL.

The presentation will report on the first months of service operation, both from a technical and organisational point of view. It will present the setup and architecture of the service, as

well as the different procedures put in place to operate and monitor it.

---

**Gilles Mathieu** *CNRS*

**Tomasz Szepieniec** *CYFRONET*

**Ricardo Graciani Diaz** *University of Barcelona*

**Andrei Tsaregorodtsev** *CNRS*



# Web processing services for climate data with examples for impact modelers

Impact modelling forced by climate data is often connected with big data processing. But impact modelers are often not equipped with appropriate hardware (computing and storage facilities) or appropriate programming experience.

Web Processing Service (WPS) is an open standard defined by the Open Spatial Consortium (OGC). It is an interface to perform processes over the HTTP network protocol.

This tutorial is an introduction to the ClimDaPs project, which is using WPS for climate data processing. It is based on the PyWPS implementation of WPS and provides additionally a simple web-based user-interface to access and combine climate

data processes.

It provides access to the climate data archive of the Earth System Grid Federation (ESGF) for CMIP5 and CORDEX data. Performing simple processes of climate data up to complex impact models are already available within ClimDaPs. One can also visualize climate data and processed results.

Besides the introduction of existing processing possibilities, we will show how you can add your own climate data processes to ClimDaPs and other WPS services.

---

**Nils Hempelmann** *Climate Service Center*

## Foundation training in federated IT Service Management according to FitSM-1:2013

This foundation training course in federated IT Service Management provides training in the fundamentals of service management and introduces some of the specific challenges faced when managing IT services across complex and federated communities.

The training is carried out over one afternoon and the following morning, culminating in a short exam. Successfully passing the exam will grant participants a Foundation Certificate in Service Management for Federated IT Infrastructures, provided by the internationally recognised standards organisation TÜV SÜD. The course is structured around the FitSM-1:2013 standard (see [www.fedsm.eu/fitm](http://www.fedsm.eu/fitm) for details),

which is compatible with ITIL and ISO/IEC 20000 but is tailored to work in federated environments such as Grids and federated clouds. The FitSM standard and the training course are produced and run by the FedSM project, which is funded by the EC to bring improved service management to several infrastructures, including EGI.

---

**Thomas Schaaf** *Ludwig-Maximilians-Universität, München*  
**Michael Brenner** *RRZN, Leibniz Universität Hannover*

# Managing virtual servers

The aim of this training session is to provide a quick introduction to OCCI (Open Cloud Computing Interface), describe basic elements defined by the standard, and demonstrate its practical applications while focusing on virtual machine management in IaaS-based clouds as it pertains to the EGI Federated Cloud. It will also describe, explain, and demonstrate the concepts behind virtual machine contextualization and its applicability in various use cases. Attendees will get a chance to apply these concepts in a hands-on part of the session utilizing tools used within the EGI Federated Cloud environment, namely the rOCCI client and various server-side OCCI implementations.

**Boris Parak** *CESNET*

---

**Enol Fernandez** *CSIC*

# Introduction to Stoxy - CDMI-compliant lightweight storage proxy

Tutorial will present Stoxy (STORage proXY) - a CDMI-compliant storage service developed as part of the EGI Federated Cloud. It offers a standard HTTP-based interface for managing data in multiple back ends. During the tutorial an overview of Stoxy will be given along with the development roadmap, typical scenarios and deployments. Attendants are very welcome to bring their own laptops to install/test out how it works.

---

**Ilja Livenson** *KTH*

# Training on data preservation

Digital encoding has become the dominant way in which we create, shape and exchange information.

Digital Preservation is an emerging field for research and development and needs reflection in education and training.

This session provides training customised to the needs of research communities in the following areas:

1. Data Management planning
2. Trusted archives, audit and certification.
3. Meta-data packaging of representational information for preservation.
4. Value generation through data re-use.

5. Business models for data preservation.

Training will be provided by the Network of Excellence APARSEN - Alliance Permanent Access for the Records of Science in Europe Network, whose mission is to help to train the next generation of digital preservation experts by designing different types of curricula.

---

**Matthias Hemmje**  
*FernUniversität, Hagen*

# OpenStack hands-on

This hands on tutorial will introduce the OpenStack cloud middleware, focusing on the compute project (nova) we will describe its architecture, how all the components interact and what is required to run a fully functional OpenStack cloud (i.e. compute, network, image, volume storage and identity services).

The tutorial will cover the following areas:

- Overview of the OpenStack project and timeline.
- OpenStack Architecture and services.
- OCCl integration within OpenStack.
- VOMS integration within OpenStack.

The attendees will learn how to setup their own OpenStack cloud environment, covering also the installation of VOMS authentication and the support for the Open Cloud Computing Interface (OCCI).

---

**Alvaro Lopez Garcia CSIC**



---

---

# FRIDAY

---

---

<b>TUTORIALS .....</b>	<b>84</b>
Building an OpenNebula cloud on the EGI Federated Cloud.....	84
Hands-on DIRAC user-friendly job management.....	84
Globus technology training .....	85
Best practices for cloud application architecture.....	85
Federated Cloud image lifecycle management.....	86



# Building an OpenNebula cloud on the EGI Federated Cloud

This hands on tutorial will give an overview of how OpenNebula can be used to build and operate private clouds. The attendees will build, configure and operate their own OCCI-compatible OpenNebula cloud compliant to the EGI Federated Cloud standards.

---

**Ignacio M. Llorente**

*OpenNebula/C12G & UCM*

**Daniel Molina** *OpenNebula Project*

**Carlos Martin Sanchez**

*OpenNebula Project*

**Boris Parak** *CESNET*

# Hands-on DIRAC user-friendly job management

Users, even from a single research community, have very different degrees of programming expertise. Providing a single access point to a distributed infrastructure that suits many communities is a big challenge. The DIRAC project provides a general purpose framework for building distributed computing systems. There is a large interest from smaller user communities to have a simple tool for accessing grid and other types of distributed computing resources. 'DIRAC as a Service' could play this role. To demonstrate accessing facilities of DIRAC, the present tutorial is focused in a portal job management by the final users. The tutorial has a brief introduction and previous

dummy tests of every user environment. Then, two blocks of about 1H each, in the first one, participants will launch prepared jobs with previous existing templates, monitor the runnings, getting the outputs using output sandbox. The second block is about advancing job submission using parametric jobs and third-party storage, as well as retrieval of the outputs.

---

**Ricardo Graciani Diaz**

*University of Barcelona*

# Globus technology training

The goal of this contribution is to give scientific end-users users and administrators of e-Infrastructures a hands on experience with the Globus Toolkit and Globus Online services (GO) for reliable, high-performance, and secure file transfer (<http://www.globusonline.eu>). Data sharing will also be demonstrated. Moreover, the workshop shows the benefit and usage of preconfigured Globus Appliances.

At the end of the training, the participants are expected to be able to manage their own endpoints with Globus Online, to submit, monitor their data transfers, and share them with other users.

---

**Ioan Muntean** *UTC*

# Best practices for cloud application architecture

Many user groups trying to bring their applications into the cloud choose VM images as the 'packaging format'. Depending on the structure of the application and the intended use cases that are to be run on cloud resources, there may be alternative ways of packaging the application, thus keeping images small and avoiding problems that may arise from the need to update individual assets within images. This will ultimately lead to optimisations in application delivery and start up, presenting a better experience to the user. This session will start with considerations about available IaaS cloud features and how applications can make the best use of them.

Main services of the European Grid Infrastructure:

- How to get access to the EGI production system
- Federated Cloud from EGI
- Complex jobs, high-level tools for application development and access Main services of the European Grid Infrastructure
- How to get access to the EGI production system
- Federated Cloud from EGI
- Complex jobs, high-level tools for application development and access

---

**Bjoern Hagemeyer** *JUELICH*  
**Boris Parak** *CESNET*  
**Miroslav Ruda** *CESNET*

# Federated Cloud image lifecycle management

The EGI Applications Database has evolved as the *de facto* EGI Federated Cloud Virtual Appliance Marketplace, offering the easiest way to automate distribution of heterogeneous cloud infrastructures for scientists for managing virtual machine lifecycles within the EGI Federated Cloud infrastructure. In this context, the main mission of the service has been enhanced to:

- Store and provide virtual appliances and their metadata.
- Act as a distribution mechanism between the submitters/community and the resource providers.
- In order to accomplish this, the HEPiX VMCAster/VMCAatcher

technology has been developed to provide secure and coherent distribution of virtual appliance to multiple cloud infrastructures.

The user, submitter and/or resource provider will be able to learn about:

- What is a Virtual Appliance?
- What is an image-list? Trust and the security model.
- How to register a Virtual Appliance
- How to upload VM image lists either using the AppDB Web interface or the VMCAster tool
- Publishing an image list
- 'Propose' images to be included into a VO-wide image bundle

- Subscribing to an image list and select an image
- Blocking an identity and blocking an image
- Integrating resources with the federated cloud

---

**Marios Chatziangelou** IASA  
**Owen Synge** DESY (HH)



# INDEX OF AUTHORS & CONTACT DETAILS

## A

<b>Andres Abad Rodriguez</b> <i>andres.abad.rodriguez@cern.ch</i> .....	58
<b>Patrick J.C. Aerts</b> <i>p.aerts@esciencecenter.nl</i> .....	71
<b>Luis Alves</b> <i>luis.alves@csc.fi</i> .....	69
<b>Andreozzi, Sergio</b> <i>sergio.andreozzi@egi.eu</i> .....	10, 42
<b>Asero, Carmela</b> <i>carmela.asero@egi.eu</i> .....	10
<b>Emanouil Atanassov</b> <i>emanouil@parallex.bas.bg</i> .....	27

## B

<b>Roberto Barbera</b> <i>roberto.barbera@ct.infn.it</i> .....	13, 53
<b>Ugo Becciani</b> <i>ugo.becciani@oact.inaf.it</i> .....	46
<b>Bruce Becker</b> <i>bbecker@csir.co.za</i> .....	38
<b>Catherine Biscarat</b> <i>biscarat@in2p3.fr</i> .....	59
<b>Carlos Blanco</b> <i>blancojc@unican.es</i> .....	27
<b>Ignacio Blanquer</b> <i>iblanque@dsic.upv.es</i> .....	33
<b>Alexandre Bonvin</b> <i>a.m.j.j.bonvin@uu.nl</i> .....	72
<b>Jan Bot</b> <i>jan.bot@surfsara.nl</i> .....	68
<b>Marko Bracko</b> <i>marko.bracko@ijs.si</i> .....	71
<b>Michael Brenner</b> <i>brenner@rvs.uni-hannover.de</i> .....	78
<b>Steve Brewer</b> <i>s.brewer@soton.ac.uk</i> .....	70

<b>Riccardo Bruno</b> <i>riccardo.bruno@ct.infn.it</i> .....	13
---	----

## C

<b>Miguel Caballer</b> <i>micafer1@upv.es</i> .....	74
<b>Antonio Calanducci</b> <i>antonio.calanducci@ct.infn.it</i> .....	13
<b>Carlo Casarino</b> <i>carlo.casarino@polooncologicocefalù.it</i> .....	48
<b>Marios Chatziangelou</b> <i>marios.chatziangelou@cern.ch</i> .....	86
<b>Steffen Claus</b> <i>steffen.claus@scai.fraunhofer.de</i> .....	14
<b>Antonio S. Cofino</b> <i>antonio.cofino@unican.es</i> .....	15
<b>Catalin Condurache</b> <i>catalin.condurache@stfc.ac.uk</i> .....	36, 37
<b>Alessandro Costa</b> <i>alessandro.costa@oact.inaf.it</i> .....	46
<b>Alessandro Costantini</b> <i>alessandro.costantini@cnaf.infn.it</i> .....	32

## D

<b>Anatoli Danezi</b> <i>anatoli.danezi@surfsara.nl</i> .....	68
<b>Luca dell'Agnello</b> <i>luca.dellagnello@cnaf.infn.it</i> .....	56
<b>Giacinto Donvito</b> <i>giacinto.donvito@ba.infn.it</i> .....	28, 75
<b>Andreas Drakos</b> <i>drakos@agroknow.gr</i> .....	62
<b>Michel Drescher</b> <i>michel.drescher@egi.eu</i> .....	22, 34
<b>Afonso Duarte</b> <i>aduarte@itqb.unl.pt</i> .....	61
<b>Lukasz Dutka</b> <i>lukasz.dutka@cyfronet.pl</i> .....	59

**Dave Dykstra**  
*dwd@fnal.gov* ..... 36

**E**

**Karolis Egelis**  
*karolis.egelis@egi.eu* ..... 30

**Ad Emmen**  
*ad@almeregrid.nl* ..... 11

**F**

**Marco Fargetta**  
*marco.fargetta@ct.infn.it* ..... 13

**Zoltan Farkas**  
*zoltan.farkas@sztaki.mta.hu* ..... 57

**Enol Fernandez**  
*enolfc@ifca.unican.es* ..... 79

**Nuno Ferreira**  
*nuno.ferreira@egi.eu* ..... 28, 48

**Nicola Ferro**  
*ferro@dei.unipd.it* ..... 60

**Christos Filippidis**  
*filippidis@inp.demokritos.gr* ..... 45

**Sandro Fiore**  
*sandro.fiore@unile.it* ..... 58

**Roger Firpo**  
*roger.firpo@gmail.com* ..... 44

**Ciro Formisano**  
*ciro.formisano@eng.it* ..... 53

**Alessandra Forti**  
*Alessandra.Forti@manchester.ac.uk* ..... 32

**Alain Franc**  
*alain.franc.33@gmail.com* ..... 26

**William Frank**  
*frank@ipggp.fr* ..... 16

**Antonella Fresa**  
*fresa@promoter.it* ..... 60

**G**

**Sven Gabriel**  
*sveng@nikhef.nl* ..... 18

**Andre Gemuend**  
*andre.gemuend@scai.fraunhofer.de* ..... 14

**Tristan Glatard**  
*glatard@creatis.insa-lyon.fr* ..... 73

**Carole Goble**  
*carole.goble@manchester.ac.uk* ..... 28

**John Gordon**  
*john.gordon@stfc.ac.uk* ..... 11

**Ricardo Graciani Diaz**  
*graciani@ecm.ub.edu* ..... 61, 72, 73, 75, 76, 84

**Silvina Grad-Freilich**  
*silvina.grad-freilich@mathworks.com* ..... 31

**Richard Grunzke**  
*richard.grunzke@tu-dresden.de* ..... 54

**H**

**Bjoern Hagemeier**  
*b.hagemeier@fz-juelich.de* ..... 85

**Akos Hajnal**  
*ahajnal@sztaki.hu* ..... 57

**Alex Hardisty**  
*hardistyar@cardiff.ac.uk* ..... 23

**Mihály Héder**  
*mihaly.heder@sztaki.hu* ..... 52

**Helmut Heller**  
*helmut.heller@lrz.de* ..... 56

**Matthias Hemmje**  
*Matthias.Hemmje@FernUni-Hagen.de* ..... 80

**Nils Hempelmann**  
*nils.hempelmann@hzg.de* ..... 28

**Ladislav Hluchy**  
*hluchy.ui@savba.sk* ..... 12

**Sy Holsinger**  
*sy.holsinger@egi.eu* ..... 42

**J**

**Daniela Jacob**  
*daniela.jacob@hzg.de* ..... 62

**Borje Justrell**  
*borje.justrell@riksarkivet.se* ..... 60

**K**

**Peter Kacsuk**  
*kacsuk@sztaki.hu* ..... 49, 57

**Christos Kanellopoulos**  
*skanct@grnet.gr* ..... 66

**Samuel Keuchkerian** ..... 23

**Tamas Kiss**  
*t.kiss@westminster.ac.uk* ..... 11, 35

**Tijs de Kler**  
*tijs.dekler@sara.nl* ..... 68

<b>Joanna Kocot</b> <i>ymkocot@cyfronet.pl</i> .....	49	<b>Daniel Molina</b> <i>dmolina@opennebula.org</i> .....	66, 84
<b>Balasz Konya</b> <i>balazs.konya@hep.lu.se</i> .....	56	<b>Johan Montagnat</b> <i>johan.montagnat@cnsr.fr</i> .....	73
<b>Daniel Kouril</b> <i>kouril@ics.muni.cz</i> .....	18	<b>Ruben Montero</b> <i>dmolina@opennebula.org</i> .....	66
<b>Jozsef Kovacs</b> <i>smith@sztaki.hu</i> .....	11	<b>Aurelien Mordet</b> <i>mordret@ipgp.fr</i> .....	16
<b>Lion Krischer</b> <i>krischer@geophysik.uni-muenchen.de</i> .....	14	<b>Ioan Muntean</b> <i>ioan.lucian.muntean@cs.utcluj.ro</i> .....	85
<b>L</b>		<b>Visakh Muraleedharan</b> <i>visakh@ipgp.fr</i> .....	15
<b>Antonio Lagana</b> <i>lagana05@gmail.com</i> .....	69	<b>N</b>	
<b>Giuseppe La Rocca</b> <i>giuseppe.larocca@ct.infn.it</i> .....	13	<b>Stefano Nicotri</b> <i>stefano.nicotri@ba.infn.it</i> .....	75
<b>Yannick Legré</b> <i>yannick.legre@egi.eu</i> .....	23	<b>Leif Nixon</b> <i>nixon@nsc.liu.se</i> .....	18
<b>Bert Lemmens</b> <i>bert@packed.be</i> .....	60	<b>Pasquale Notarangelo</b> <i>pasquale.notarangelo@ba.infn.it</i> .....	75
<b>Ilja Livenson</b> <i>livenson@kth.se</i> .....	79	<b>P</b>	
<b>Alvaro Lopez Garcia</b> <i>aloga@ifca.unican.es</i> .....	80	<b>Jerome Pansanel</b> <i>jerome.pansanel@iphc.cnrs.fr</i> .....	59, 73
<b>Robert Lovas</b> <i>rlovas@sztaki.hu</i> .....	11, 62	<b>Marco Pappalardo</b> <i>marco.pappalardo@softwareengineering.it</i> .....	13, 50
<b>Bjorn Lundell</b> <i>bjorn.lundell@his.se</i> .....	60	<b>Boris Parak</b> <i>256254@mail.muni.cz</i> .....	79, 84, 85
<b>M</b>		<b>Antonio Perez-Calero Yzquierdo</b> <i>aperez@pic.es</i> .....	32
<b>Giorgio Maggi</b> <i>giorgio.maggi@ba.infn.it</i> .....	75	<b>Monique Petitdidier</b> <i>monique.petitdidier@latmos.ipsl.fr</i> .....	12, 13
<b>Torleif Markussen Lunde</b> <i>torleif.lunde@cih.uib.no</i> .....	13	<b>Anatolii Petrenko</b> <i>tolja.petrenko@gmail.com</i> .....	31
<b>Carlos Martin Sanchez</b> <i>cmartin@opennebula.org</i> .....	84	<b>Peter Pharow</b> <i>phw@idmt.fraunhofer.de</i> .....	60
<b>Gilles Mathieu</b> <i>gilles.mathieu@in2p3.fr</i> .....	76	<b>Gabriele Pierantoni</b> <i>gabriele.pierantoni@cs.tcd.ie</i> .....	45
<b>Rafael Mayo-Garcia</b> <i>rafael.mayo@ciemat.es</i> .....	57	<b>Salvatore Pinto</b> <i>salvatore.pinto@egi.eu</i> .....	28
<b>Victor Mendez</b> <i>vmendez@caos.uab.es</i> .....	74	<b>Claudio Prandoni</b> <i>prandoni@promoter.it</i> .....	60
<b>Franck Michel</b> <i>fmichel@i3s.unice.fr</i> .....	73	<b>Fotis Psomopoulos</b> <i>fpsom@issel.ee.auth.gr</i> .....	61
<b>Ignacio M. Llorente</b> <i>imllorente@opennebula.org</i> .....	66, 84	<b>Stuart Pullinger</b> <i>stuart.pullinger@stfc.ac.uk</i> .....	11

## R

<b>Manuel Rodriguez-Pascual</b> <i>manuel.rodriguez@ciemat.es</i> .....	57
<b>Miroslav Ruda</b> <i>miroslav.ruda@cesnet.cz</i> .....	85

## S

<b>Diego Scardaci</b> <i>diego.scardaci@ct.infn.it</i> .....	10
<b>Thomas Schaaf</b> <i>schaaf@nm.ifi.lmu.de</i> .....	78
<b>Jeroen Schot</b> <i>jeroen.schot@surfsara.nl</i> .....	68
<b>Horst Schwitenberg</b> <i>horst.schwichtenberg@scai.fraunhofer.de</i> .....	12
<b>Gergely Sipos</b> <i>gergely.sipos@egi.eu</i> .....	18, 28, 68
<b>Peter Solagna</b> <i>peter.solagna@egi.eu</i> .....	52
<b>Vincenzo Spinoso</b> <i>vincenzo.spinoso@ba.infn.it</i> .....	75
<b>Bernardino Spisso</b> <i>nispisso@tin.it</i> .....	30
<b>Owen Syngé</b> <i>owen.syngé@desy.de</i> .....	86
<b>Tomasz Szepieniec</b> <i>t.szepieniec@cyfronet.pl</i> .....	49, 76

## T

<b>Giuliano Taffoni</b> <i>taffoni@iasfbo.inaf.it</i> .....	44
<b>Sasu Tarkoma</b> <i>sasu.tarkoma@helsinki.fi</i> .....	56
<b>Natalia Tatarinova</b> <i>natalia.tatarinova@cerfacs.fr</i> .....	14
<b>Ulf Tigerstedt</b> <i>ulf.tigerstedt@csc.fi</i> .....	37
<b>Steven Timm</b> <i>timmm@fnal.gov</i> .....	35
<b>Viet Tran</b> <i>viet.ui@savba.sk</i> .....	33
<b>Andrei Tsaregorodtsev</b> <i>atsareg@in2p3.fr</i> .....	61, 72, 73, 76

## V

<b>Roberto Valentini</b> <i>valentini.roberto@ba.infn.it</i> .....	75
<b>Caludio Vuerli</b> <i>vuerli@oats.inaf.it</i> .....	44

## W

<b>David Wallom</b> <i>david.wallom@oerc.ox.ac.uk</i> .....	34
<b>John Walsh</b> <i>john.walsh@scss.tcd.ie</i> .....	43
<b>David Weissenbach</b> <i>david.weissenbach@upmc.fr</i> .....	16
<b>Tom Whyntie</b> <i>t.whyntie@qmul.ac.uk</i> .....	70

# NOTES



