EGI-InSPIRE

TRAINING AND DISSEMINATION EVENT FOR ALL SHARED SERVICES AND THE OTHER TASKS WITHIN THE ACTIVITY

EU MILESTONE: MS619

Document identifier:	EGI-MS619-FINAL
Date:	**30/4/2013**
Activity:	**SA3**
Lead Partner:	**EGI.eu**
Document Status:	**FINAL**
Dissemination Level:	**PUBLIC**
Document Link:	https://documents.egi.eu/document/1690

Abstract

This document provides a summary of the training events that have been organised under the EGI Community Forum 2013 umbrella and that provided training and/or dissemination for the shared services and other tasks within the EGI-InSPIRE SA3 activity. The activity is titled 'Heavy User Communities'. The events that are reported in this document have been all held in Manchester, UK between 8-12th April 2013. The summaries about the training events have been collected from the community via the online abstract submission system of the EGI Community Forum.

I. COPYRIGHT NOTICE

II. DELIVERY SLIP

	Name	Partner/Activity	Date
From	**Gergely Sipos**	**EGI.eu/NA2**	**17/4/2013**
Reviewed by:	**Steven Newhouse**	**EGI.eu**	**20/4/2013**
Approved by:	**AMB & PMB**		**30/4/2013**

III. DOCUMENT LOG

Issue	Date	Comment	Author/Partner
1	2/4/2013	First internal draft	Nuno Ferreira/EGI.eu
2	17/4/2013	Draft for review	Gergely Sipos/EGI.eu

IV. APPLICATION AREA

This document is a formal deliverable for the European Commission, applicable to all members of the EGI-InSPIRE project, beneficiaries and Joint Research Unit members, as well as its collaborating projects.

V. DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the authors. The procedures documented in the EGI-InSPIRE "Document Management Procedure" will be followed:
https://wiki.egi.eu/wiki/Procedures

VI. TERMINOLOGY

A complete project glossary is provided at the following page: http://www.egi.eu/about/glossary/.

VII. PROJECT SUMMARY

To support science and innovation, a lasting operational model for e-Science is needed − both for coordinating the infrastructure and for delivering integrated services that cross national borders.

The EGI-InSPIRE project will support the transition from a project-based system to a sustainable pan-European e-Infrastructure, by supporting 'grids' of high-performance computing (HPC) and high-throughput computing (HTC) resources. EGI-InSPIRE will also be ideally placed to integrate new Distributed Computing Infrastructures (DCIs) such as clouds, supercomputing networks and desktop grids, to benefit user communities within the European Research Area.

EGI-InSPIRE will collect user requirements and provide support for the current and potential new user communities, for example within the ESFRI projects. Additional support will also be given to the current heavy users of the infrastructure, such as high energy physics, computational chemistry and life sciences, as they move their critical services and tools from a centralised support model to one driven by their own individual communities.

The objectives of the project are:

1. The continued operation and expansion of today's production infrastructure by transitioning to a governance model and operational infrastructure that can be increasingly sustained outside of specific project funding.
2. The continued support of researchers within Europe and their international collaborators that are using the current production infrastructure.
3. The support for current heavy users of the infrastructure in earth science, astronomy and astrophysics, fusion, computational chemistry and materials science technology, life sciences and high energy physics as they move to sustainable support models for their own communities.
4. Interfaces that expand access to new user communities including new potential heavy users of the infrastructure from the ESFRI projects.
5. Mechanisms to integrate existing infrastructure providers in Europe and around the world into the production infrastructure, so as to provide transparent access to all authorised users.
6. Establish processes and procedures to allow the integration of new DCI technologies (e.g. clouds, volunteer desktop grids) and heterogeneous resources (e.g. HTC and HPC) into a seamless production infrastructure as they mature and demonstrate value to the EGI community.

The EGI community is a federation of independent national and community resource providers, whose resources support specific research communities and international collaborators both within Europe and worldwide. EGI.eu, coordinator of EGI-InSPIRE, brings together partner institutions established within the community to provide a set of essential human and technical services that enable secure integrated access to distributed resources on behalf of the community.

The production infrastructure supports Virtual Research Communities (VRCs) − structured international user communities − that are grouped into specific research domains. VRCs are formally represented within EGI at both a technical and strategic level.

TABLE OF CONTENTS

1 INTRODUCTION

This document provides a summary of the training events that have been organised at the EGI Community Forum 2013 from activities funded from within the EGI-InSPIRE project and from outside it. Although, established to provide a training and/or dissemination opportunity for the shared services and other tasks within the EGI-InSPIRE SA3 activity coming from the 'Heavy User Communities', the training opportunity has spread across the whole EGI ecosystem encompassing fundamental software engineering skills to specific training in particular middleware technologies or science applications. The summaries about the training events have been collected from the community via the online abstract submission system of the EGI Community Forum.

2 TRAINING EVENTS

2.1 *Globus Online Training*

2.1.1 Summary

The goal of this contribution is to give scientific end-users users and administrators of e-Infrastructures a hands-on experience with the Globus Online service (GO) for reliable, high-performance, and secure file transfer. New features such as data sharing will also be demonstrated. At the end of the training, the participants are expected to be able to manage their own endpoints with Globus Online, to submit and monitor their data transfers.

2.1.2 Description

The training contains: presentations, short demos, and hands-on exercises. The session is organized around the following topics:

- getting a GO account;
- managing identities (e.g. based on X.509 certificates);
- managing private endpoints;
- Globus Connect – your gate to Globus Online transfers from your laptop;
- using GO to transfer data between endpoints (e.g. grid systems, participant's laptop);
- data sharing with GO;
- using the IGE GO Clients.

The participants are expected to have their own grid certificate. If necessary, temporary training certificates will be allocated to participants upon request.

2.1.3 Impact

Upon completion of the training, the following impact is anticipated, on the target categories of participants:

- end-users: manage their data in friendly, reliable, secure, high-performance manner
- e-Infrastructure site administrators: understand how grid resources could be published for usage with Globus Online
- application developers: get some insights on how applications can be enabled to work with the Globus Online service

2.1.4 Details

Event link: http://egitraining.esc.rl.ac.uk/content/globus-online-training
Duration: 90 minutes
Contact: Dr. MUNTEAN, Ioan (UTC).

2.2 Globus Tools for Grid Users and Administrators

2.2.1 Summary

The tutorial will provide first-hand practical information about Globus Toolkit v.5 and a suite of community tools, of great value to resource providers, computing site administrators, end-users of the grid, scientific application developers. The focus of the tutorial is set on the latest version 5 of the toolkit: GridFTP, GRAM5, MyProxy etc. Globus services included in the Unified Middleware Distribution of EGI will be presented as well. Furthermore, the Globus Online cloud service for reliable file transfer will be demonstrated, together with its latest features. A selection of community tools are included in the training, such as GridWay, GridSAM, AdHoc-VOMS, GridSafe, LCAS/LMAPS security integration, Globus Online Client.

2.2.2 Description

The tutorial will provide first-hand practical information about Globus Toolkit v.5 and a suite of community tools, of great value to resource providers, computing site administrators, end-users of the grid, scientific application developers. The tutorial targets both grid administrators and users of GT, offering/providing first-hand information and experience sharing from European Globus specialists.

The focus of the tutorial is set on:

- The latest version 5 of the toolkit: GridFTP, GRAM5, MyProxy etc. Globus services included in the Unified Middleware Distribution of EGI will be presented as well.
- Furthermore, the Globus Online cloud service for reliable file transfer will be demonstrated, together with its latest features.
- A selection of community tools are included in the training, such as GridWay, GridSAM, AdHoc-VOMS, GridSafe, LCAS/LMAPS security integration, Globus Online Client.

Virtual machines for the tutorial participants and remote resources for the training will be provided by the IGE project as part of the IGE Testbed. The IGE project is a consortium of Globus experts that coordinates Globus activities and developments in Europe.

2.2.3 Impact

It is expected that the participants understand how this selection of Globus services and tools work, how these satisfy common requirements of e-Infrastructures (e.g. accounting, administration), and how end-users can benefit at most from e-Infrastructures providing Globus-based services.

2.2.4 Details

Event link: http://egitraining.esc.rl.ac.uk/content/globus-tools-grid-users-and-administrators
Duration: 90 minutes
Contact: Dr. MUNTEAN, Ioan (UTC)

2.3 Computing on a big scale with OpenMOLE

2.3.1 Summary

OpenMOLE is based on a blackbox approach to embed application based on very different technologies: java / scala / C / C++ / fortran / scilab / octave / netlogo... Once embedded in the

---

platform, OpenMOLE automatically distributes numerous executions of the application on a distributed environment specified by the user, such as a cluster or even on the 100,000+ cores available on the European computing grid. The platform deals with software installations, file transfers, job failures, rendering the distributed execution entirely transparent for the user.

2.3.2 Description

The workflow paradigm is particularly well suited to describe distributed executions. Several platforms leverage this representation to fill the gap between scientific processes and high performance computing environments. In this training session we propose to advertise OpenMOLE (Open MOdel Experiment), a workflow platform that exposes several advantages when compared to other mainstream workflow systems:

- it delegates transparently the computing load from the user computer directly to the remote execution environment following a zero-deployment approach and makes it convenient to embed third party software components in a workflow,
- OpenMOLE workflows are environment agnostic and the workload can be submited on EGI as well as on other distributed computing environment such as clusters, remote servers or local desktop grid,
- the dataflow is statically typed and formally checked prior to the runtime, minimizing the risk of error when designing complex workflows.

This platform has already been used to execute huge amount of computation and is now scalable, reliable and mature. In November and December 2012 the major part of the computing time provided by the biomed VO has been orchestrated by OpenMOLE workflows.

2.3.3 Impact

The training session will show how to design OpenMOLE workflows. Attendees will learn how to:

- embed an external application in OpenMOLE,
- compose task with each-other to design a workflow,
- design large scale workflows generating several thousands of jobs,
- gather / aggregate / store produced data,
- delegate the computing load to EGI,
- create an ad-hoc desktop grid.

At the end of the session attendees should all be able to design a large scale distributed application. During the session people will be provided with formation certificates or use their own certificate. They will work on the VO of their choice.

The example applications will mostly concern model exploration since OpenMOLE has been primarily designed to explore simulation models. However this training can be beneficial for enabling distributed computing for other purposes. For instance the bioemergences platform that process 4D morphogenesis images on EGI is based on OpenMOLE.

2.3.4 Details

Event link: http://egitraining.esc.rl.ac.uk/content/computing-big-scale-openmole
Duration: 90 minutes

Contact: REUILLON, Romain (CNRS)

2.4   EMI Training: EMIR, ARGUS and EMI-ES

2.4.1   Summary

The European Middleware Initiative (EMI) is a close collaboration of the three major middleware providers, ARC, gLite and UNICORE, and other specialized software providers like dCache. Besides harmonising the middleware products, EMI is putting a strong accent on interoperability promoting, via adoption of standards, development of interfaces common to the three middlewares. In this regard, this session will provide training introducing some of the key developments brought by the EMI third release, MonteBianco, like the EMI Registry, ARGUS and the EMI Execution Service. The session is aimed to experts within the NGI, either system administrators or advanced users eager to know in advance more details about these new, production ready, features.

2.4.2   Description

This training sessions aims to provide a deep understanding of usage of new or recently developed EMI products, not yet fully deployed in large scale production infrastructures as EGI.

Each of the middleware constituting EMI, has developed throughout the years their own 'Computing Element', i.e. a grid service providing access to the computing resources localized in a site, generally managed with a Local Resource Management System, most known of which are LSF, Condor, PBS-Torque, GE. The agreement among gLite, UNICORE and ARC, has brought the definition of common, standard based specification language and interfaces, constituting the EMI Execution Service (EMI-ES) that provides a gateway to the different Computing Element flavours, allowing the transparent execution of jobs regardless of the different middleware backends.

ARGUS is the solution developed by EMI to address the need for a common, standardized authorization service. The service is based on the XACML standard, and uses authorization policies to determine if a user is allowed or denied to perform a certain action on a particular resource. Authorization policies are managed via the Policy Administration Point (PAP), and distributed to the Policy Decision Point (PDP) for evaluation, which are then used by the Policy Enforcement Point (PEP) to issue authorization decision for the controlled services.

EMI's EMIR provides a federated, distributed, hierarchical service registry, enabling, via its REST-ful interfaces, discovery of services within a Distributed Computing Infrastructure. EMIR has been designed keeping in mind the decentralised nature of NGI, furthermore based on heterogeneous Grid middleware. In this way, single point of failure or bottlenecks are eliminated, still allowing controlled indexing of service, offering for instance the freedom of choice of which service have to be seen from the higher levels of the hierarchy.

2.4.3   Impact

This session is mainly oriented to advanced user and system administrators, providing them a deep understanding and concrete hands-on for some of the most promising EMI new products, allowing a guided evaluation of their characteristics before their adoption. Both users' and point of views will be stressed, granting a complete overview of the product presented, that will be completed with some tips on deployment scenarios.

2.4.4 Details

Event link: http://egitraining.esc.rl.ac.uk/content/emi-trainings-emir-argus-emi-es
Duration: 90 minutes
Contact: GIORGIO, Emidio (INFN)

2.5 Software Carpentry Workshop

2.5.1 Summary

EGI.eu, together with the UK Software Sustainability Institute, plans to run a one-day Software Carpentry workshop (or boot camp) for researchers.

2.5.2 Description

A boot camp is an in-person, example-driven workshop which covers useful software development skills to enable researchers to be more productive programmers. These skills can include developing maintainable software, version control, automation and testing. A boot camp is specifically tailored to the needs of researchers and demonstrates how software development skills contribute to correct, reproducible and reusable research.

Short tutorials alternate with hands-on practical exercises, and participants are encouraged both to help one another. Participants usually work on their own computers (typically laptops), using either native software or a standard set of packages running in a virtual machine. This ensures that they have a working environment when the boot camp is done. Software Carpentry provides a wealth of online material for self-directed follow-up learning.

2.5.3 Impact

Software Carpentry's experience with boot camps is consistent with recent research: students learn best in a blended environment that combines directed in-person instruction with self-directed online learning. As well as improving learning, boot camps solve two other recurring problems. Researchers are busy people who often cannot make time for a semester-long traditional course. However, most can find two or three days to get started (and to get past installation and configuration hurdles if they are working on their own machines). At the same time, it can be hard to stay motivated when working in isolation; by bringing people together, boot camps create help peer support communities in selected disciplines or geographic regions.

Boot camp "highlights"

EGI.eu, together with the UK Software Sustainability Institute, will run three Software Carpentry sessions for researchers at the EGI Forum on Thursday 11th April. These sessions are drawn from Software Carpentry's highly successful and popular boot camps which teach software development skills to researchers, enabling them to do more, in less time and with less pain. The three sessions are:

- Using version control to record provenance and collaborate more easily. This session includes an introduction to Git and GitHub.
- Using testing to help ensure your software, and results, are correct.
- Data management using files and NoSQL databases to manage your data more easily. This session includes an introduction to the NoSQL database MongoDB.

Attendees at the EGI Forum are free to attend any or all of the sessions, up to a maximum of 40 - first-come first-served.

For all 3 sessions, you should have some familiarity with using a command-line (e.g. a Linux shell or terminal window, or the DOS prompt). For the testing and data management sessions, you should also have knowledge of a programming language (e.g. C, FORTRAN or Java) or scripting language (e.g. Perl, Bash or MATLAB). We'll be using Python but we'll provide a quick introduction to the key syntax in each case.

You'll also need to bring your laptop. You'll need an internet browser and you'll also need access to software tools we'll be using in the workshop. You can either use an SSH client to access a VM we'll provide for you (we'll give you the login on the day), or install VMWare and run a VM image we've prepared, or, you can install all the tools yourself.

For more information on the tools you need, and a link to our VMWare VM, please see: http://github.com/swcarpentry/boot-camps/blob/2013-04-egi-forum/README.md

2.5.4 Details
Event link: http://egitraining.esc.rl.ac.uk/content/software-carpentry-workshop
Duration: 3*90 minutes
Contact: Dr. JACKSON, Mike (EPCC)

2.6 Molecular Dynamics Simulations on Grid and HPC systems

2.6.1 Summary
The training event includes tutorials on advanced usage of two major packages for Molecular Dynamics simulations, GROMACS and AMBER, with focus on their application to modelling of biomolecular systems. In the following sessions, attendees will be presented two portals for automated submission of jobs developed by the organizing EU projects WeNMR and ScalaLife.

2.6.2 Description
Molecular Dynamics is a heavily used method for investigating the dynamic properties of systems, and has become an indispensible tool in many research areas, in particular in Life Sciences. GROMACS and AMBER are two applications that are most widely used by the scientific community. The training sessions will include:

- Advanced GROMACS and AMBER usage on HPC systems
- Hybrid-MPI/OpenMP execution
- Hybrid-CPU/GPU execution
- Using GROMACS and AMBER on Grid resources – WeNMR portal
- Preparing parameters for MD simulations of small organic molecules
- Writing and running your own custom scripts for MD on the grid
- Personal workspaces for structure preparation, MD simulation and trajectory analysis - MDWeb portal

2.6.3 Impact

GROMACS and AMBER are very powerful and versatile packages but their efficient usage requires expertise that is usually gained after a long experience with them. The tutorials will greatly help experienced and advanced users to take full advantage of the capabilities of the codes, and thus accelerate their research. At the same time, the job submission portals presented here offer user-friendly tools for the most common usage scenarios that would be of great benefit for beginners in particular.

2.6.4 Details

Event link: http://egitraining.esc.rl.ac.uk/content/molecular-dynamics-simulations-grid-and-hpc-systems

Duration: 90 minutes

Contact: APOSTOLOV, Rossen (KTH), DIJK, VAN, Marc (WeNMR), GIACHETTI, Andrea (CIRMMP).

2.7 EGI CRM Training

2.7.1 Summary

The EGI-InSPRIRE project is under the process of promoting an optimised management of relationships between EGI.eu, NGIs and their potential users. Under that context, a CRM system has been deployed by the project for the EGI community in March 2012. The CRM system simplifies the process of identifying, contacting, following-up and evaluating new user community needs and mapping those needs to EGI services.

This "EGI CRM Training" session at EGI Community Forum 2013 is targeting EGI.eu and NGI members who work in the community outreach and technical outreach areas. During the training, through use case-oriented "hands-on" exercises, attendees will perform real outreach activity and learn how to continue this activity from their home institute using the EGI CRM system.

2.7.2 Description

Customer Relations Management (CRM) are technological solutions used by organisations to track the involution of important relations with their clients, share information within the organisation members, and identify new business opportunities. CRM tools are considered valuable assets for any organisation since they ease the process of identifying, contacting, following-up and evaluating client needs and satisfaction.

The EGI-InSPIRE project needs an optimized management of relationships between EGI.eu, NGIs and their potential users because:

- EGI partners and their "clients" are distributed and the conversations happen at different places, context and time;
- EGI partners have to share the information they gather, and pass it around to build a proper support network;
- EGI-InSPIRE project needs to construct a portfolio of prospective clients and explore the opportunities.

A CRM system has been deployed by Ibergrid in the EGI-InSPIRE project in March 2012. CRM target those members of the EGI community who are focused on gathering information regarding scientific organisations, their contacts, the scientific projects they are participating in, and what needs and requirements do they present. Gathering this information is a fundamental step to identify the problems faced by the user communities, and understand how EGI may provide the proper solution, and bring the user community on board.

This tutorial is targeted to EGI.eu and NGI staffs, which are the main users of the system. It will consist on "hands-on" exercises, focused on precise use-case scenarios, which will help NGI member to:

1) Record user community contacts and any associated information based on interview and discussions;
2) Gather intelligence about scientific communities, institutes and projects and about topics of collaboration with them;
3) Monitoring activities and progress in community outreach topics
4) Help NILs and EGI.eu staff to locate expertise inside NGIs

2.7.3 Impact

The "EGI CRM Training" session at the EGI Community Forum 2013 is targeted to provide EGI.eu and NGI members the proper information and support using the CRM system along precise use case scenarios expected to be faced by NGI members during EGI-InSpire lifecycle. It will contribute to decrease the CRM learning curve steepness, and boost the CRM use within the EGI community. This is a fundamental step to promote the usability of the tool, and prove its added value for the project.

2.7.4 Details

Event link: http://egitraining.esc.rl.ac.uk/content/egi-crm-training
Duration: 90 minutes
Contact: BORGES, Goncalo (LIP), ANTONIO TOMASIO PINA, Joao (LIP), BAYONNE SOPO, Enrique (UPVLC).

2.8   Contrail cloud middleware support for EGI community platforms

2.8.1   Summary

Operators and users of EGI Community Platforms are always looking for better tools that help the researchers of their platform, and tools that make administration of the platform easier. However interoperability with the EGI core infrastructure developments has to be maintained. The Contrail project has developed several tools that can be used by Community Platforms. This includes a Cloud filesystem (XtreemFS), a Platform-as-a-Service service (ConPaaS) and a federated identity management toolkit. Several use cases show real-life application of these tools. Contrail experts are involved in EGI infrastructure development work groups so they know about the EGI developments and requirements.

2.8.2   Description

Presentations (each 20 minutes):

- Overview of Contrail offerings for EGI Virtual Research communities

This presentation will concentrate on providing an overview of the Contrail tools, use cases and support. It will do so from the perspective of Virtual Research communities. Emphasis will be on well tested tools that are directly usable.

- Cloud file system (XtreemFS)

In scientific collaborations, shared use of data is important. However data must cross administrative Cloud boundaries, be made available under certain security and performance conditions. The XtreemFS Cloud filesystem is well suited for usage as part of a community platform. The presentation will introduce XtreemFS and show how it can be used.

- ConPaaS

Deploying complex applications fast and in a scalable manner is an important part in e-Science applications. There are many cases where researchers need to do several big bursts of computation, change their applications a bit and do another set of runs. What they need then is an environment in which that is possible, without having to set up a Cloud environment for each run in cooperation with system administration. ConPaaS offers a Platform-as-a-Service service.

- Federated identity management

Single sign-on in Cloud federations is a topic in which solutions pioneered by Contrail experts are now being discussed in for instance EGI and EUDAT federated environments. This presentation will concentrate on the needs of Community Platforms. What tools are available and useful?

- Business case examples

Contrail software tools are put into action in several use cases. In this presentation we will concentrate on a Scientific Data Analysis use case and show the advantage of using the new Contrail software tools.

- Wrap-up and questions

In the wrap-up we will introduce the open source community around Contrail. The software repository where Contrail tools can be downloaded and the support that is available. Then there is room for questions.

2.8.3 Impact

The tutorial aims to give an overview of Contrail tools, the support available for these tools and examples of applications in an EGI context. As a result the participants will get a good overview of whether one or more of the tools could be applied in their situation to improve the efficiency of their community platform.

Because of the presented business & use cases they can more easily see the potential benefits, or conclude it is not something for them.

The participants will receive a good feeling for what type of support is available and whether that is suitable for them.

2.8.4 Details

Event link: http://egitraining.esc.rl.ac.uk/content/contrail-cloud-middleware-support-egi-community-platforms
Duration: 90 minutes
Contact: JENSEN, Jens (STFC), EMMEN, Ad (Contrail project)

3 CONCLUSIONS

The lack of participation from the shared services and tools activity funded within SA3 in the EGI Community Forum's training activities is a concern, but given the end of the SA3 activity in PY3 and the end of TCD's participation in SA3 in 2012, the lack of engagement is understandable.

It is however worth noting:

- The activity coming from the Biomed and CompChem domains in two of the training sessions.
- The potential seen by many partners within the EGI Ecosystem (i.e. through projects external to EGI-InSPIRE) of the value of using the Forums as an interactive showcase of their work.