



## EGI-Engage

# EGI training plan (March 2015 – February 2016)

### M6.1

---

<b>Date</b>	27 May 2015
<b>Activity</b>	SA2
<b>Lead Partner</b>	EGI.eu
<b>Document Status</b>	FINAL
<b>Document Link</b>	<a href="https://documents.egi.eu/document/2482">https://documents.egi.eu/document/2482</a>

---

### Abstract

The EGI-Engage project provides foundational training services and coordination to training activities across the whole EGI collaboration. The main goal of this activity is to operate a framework that enables members of EGI community as well as external partners to effectively create, deliver, share, reuse and benefit from training services in the context of e-infrastructures and e-science. The prime objective is to provide 'glue' and also facilitation for training activities conducted by Competence Centres, National Grid Initiatives and partner projects (including e-infrastructure and Research Infrastructures), maximising the effectiveness and impact of training across countries and communities. This is achieved in EGI-Engage by providing e-infrastructure services, data, software and other online resources, modules, a marketplace and webinar tools for trainers and trainees, and by organising and contributing to the organisation of high-impact training events.

This milestone document defines the EGI training plans for the first 12 months of EGI-Engage (March 2015-Feb 2016), describing the structure and timeline for establishing the training framework, and for utilising it within EGI-Engage and partner projects for specific training purposes. This document is a formal milestone delivered to the European Commission in May 2015 and will be regularly updated until the next formal release (May 2016) based on feedback from the EGI community.



EGI-Engage is co-funded by the Horizon 2020 Framework Programme of the European Union under grant number 654142. <http://go.egi.eu/eng>

---

 Title of the Document / Number if required
**COPYRIGHT NOTICE**

This work by Parties of the EGI-Engage Consortium is licensed under a Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>). The EGI-Engage project is co-funded by the European Union Horizon 2020 programme under grant number 654142.

**DELIVERY SLIP**

	<i>Name</i>	<i>Partner/Activity</i>	<i>Date</i>
<b>From:</b>	Gergely Sipos	EGl.eu-SZTAKI / SA2	28/05/2015
<b>Reviewed by:</b>	Genevieve Romier	CNRS / NGI-FR Int. Liaison	29/05/2015
<b>Reviewed by:</b>	Diego Scardaci	EGl.eu-INFN / JRA2	29/05/2015
<b>Reviewed by:</b>	Yin Chen	EGl.eu / JRA2	01/06/2015
<b>Reviewed by:</b>	Alexandre Bonvin	Uni. Of Utrecht / WeNMR	02/06/2015
<b>Approved by:</b>			

**DOCUMENT LOG**

<i>Issue</i>	<i>Date</i>	<i>Comment</i>	<i>Author/Partner</i>
<b>v.1</b>	28/05/2015	First issue for external review	G. Sipos / EGI.eu-SZTAKI
<b>v.2</b>	01/06/2015	Small updates based on further input from projects and communities	G. Sipos / EGI.eu/SZTAKI
<b>v.3</b>	02/06/2015	Update based on 3 reviewers' feedback	G. Sipos / EGI.eu/SZTAKI
<b>v.4</b>	02/06/2015	Update based on 4 <sup>th</sup> reviewer's feedback	G. Sipos / EGI.eu/SZTAKI

**TERMINOLOGY**

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

## Contents

1	What is the role of training in EGI? .....	5
2	How will EGI training fulfil its role? .....	7
2.1	E-infrastructure for training .....	8
2.2	Training resources .....	9
2.3	Training modules .....	9
2.4	Training marketplace .....	12
2.5	Access control system .....	13
2.6	Webinar and/or e-learning system .....	14
2.7	High-impact training events .....	14
2.8	Evaluation process .....	16
2.9	Optional services .....	16
2.9.1	Certification programme .....	16
2.9.2	Massive Open Online Courses (MOOCs) .....	17
3	Training plans and needs within EGI communities .....	18
3.1	BBMRI CC .....	18
3.2	DARIAH CC .....	19
3.3	EISCAT_3D CC .....	19
3.4	ELIXIR CC .....	19
3.5	MoBrain CC .....	20
3.6	Disaster mitigation CC .....	20
3.7	Life Science Grid Community (LSGC) VRC .....	20
4	NGIs, EIROs – Training status and plans .....	22
4.1	Bulgaria .....	22
4.2	Czech Republic .....	22
4.3	France .....	22
4.4	Italy – INFN Padova .....	23
4.5	Hungary .....	23
4.6	Portugal .....	23
4.7	Spain – BSC .....	23

---

Title of the Document / Number if required

4.8	Spain – CETA CIEMAT .....	24
5	Training collaborations to explore .....	25
5.1	EDISON .....	25
5.2	Bio-Linux.....	27
5.3	EOS Cloud .....	27
5.4	JetStream .....	27
5.5	D4Science .....	28
5.6	SoBigData .....	29
6	Summary – Activity plan until February 2016 .....	31
Appendix I.	Training module development steps.....	33

# 1 What is the role of training in EGI?

Science today is no longer exclusively produced in single research labs or within national boundaries. Modern scientific challenges call for integrated solutions, cross-country collaborations and computing power with flexible usage to analyse vast amounts of data. E-infrastructures allow scientists to share information securely, analyse data efficiently and collaborate with colleagues worldwide.

EGI operates one of the largest, collaborative e-infrastructures in the world. EGI supports the digital European Research Area (ERA) through this pan-European infrastructure, its innovative technological building blocks, and related support teams and networks for users. These all together offer reliable ICT services, which provide uniform, cost effective, user oriented and collaborative access to computing and data storage resources in more than 30 countries. EGI's mission is to help scientists to make the most of the latest computing technologies through the EGI Solution Portfolio that integrates various technological, software, hardware and human processes across NGIs and projects into:

- Solution for high-throughput Data analysis
- Solution to build and/or use Federated Clouds
- Solution to perform Federated Open Data Processing
- Solution to Operate Federated Infrastructures
- Community Driven Innovation and Support

EGI's main goal is "to empower researchers from all disciplines to collaborate and to carry out data and compute intensive science and innovation"<sup>1</sup>, therefore the efficient delivery and the user-driven evolution of the EGI solution portfolio is a critical element of success. EGI training is a key contributor here with the role to:

1. **Demonstrate to research communities** how to collaborate and carry out data and compute intensive science and innovation. → Training external communities about current EGI Solutions. (technology-push)
2. **Support knowledge exchange within the EGI community** on solutions emerging from within the NGIs/EIROs and related projects. → Training EGI members about new solutions emerging internally.
3. **Intensify knowledge exchange between EGI and the broader research area** on needs and that require collaborative effort to be tackled. → Content exchange and project facilitation between EGI and external communities. (demand-pull)

These activities must target researchers, IT support teams within research collaborations, research project coordinators, managers of research infrastructures and technical staff at service/resource

---

<sup>1</sup> EGI Strategy: <http://go.egi.eu/strategy>

---

Title of the Document / Number if required

provider institutes. Providing and integrating the right type of content for them, continuously monitoring the efficiency of training activities, collecting and evaluating feedback from trainers, trainees and self-paced learners are critical elements for success. The next sections provide details on these will be achieved in the first year of EGI-Engage.

## 2 How will EGI training fulfil its role?

To fulfil its ‘Demonstrator, Exchanger, Integrator’ role, EGI Training – at the community level - must operate a training platform, create training content, and support training delivery at events and online. These activities require:

1. **E-infrastructure:** A virtualised infrastructure on which training resources can be ported to, combined with each other and then offered in the context of specific courses or modules to support specific training objectives.
2. **Training resources:** Datasets, computing applications, workflows, science gateways, Virtual Machine Images, etc. that can be used by trainers and trainees on the training e-infrastructure during organised events and self-paced courses to carry out specific training or demonstrational tasks.
3. **Training modules:** Self-contained sets of training resources and training materials that are documented and shared with trainees and/or with self-paced learners for specific training events or courses. An event or course can include multiple modules possibly with dependences among them.
4. **Training marketplace:** An online portal through which training resources, training modules and other building blocks can be shared, advertised, discovered, reused or even discussed. The marketplace facilitates interactions among trainers, trainees and training service providers.
5. **Webinar and/or e-learning system:** An online service through which attendees of specific events or self-paced learners can access and follow courses. Webinar systems are suited for presentations that are followed real-time, e-learning systems are suited for self-paced learning modules.
6. **Access control system:** A system that provides authentication and authorisation services for trainers and trainees so they can access the training e-infrastructure, the training resources, training modules, e-learning, webinar systems and the training marketplace in a harmonised way. A harmonised access control system can facilitate building of a training community.
7. **High-impact training events:** EGI members should organise or contribute to the organisation of high-impact training events that can reach large number of new users, or can bring the use of e-infrastructures within established communities to the next level.
8. **Evaluation process:** A realistic and transparent process by which feedback can be collected from trainers, trainees and collaborators about EGI training services.

T6.1 task of EGI-Engage WP6 (Knowledge commons) has 37 person months effort for over 30 months, spread across 7 partners with the following focus (where defined):

Title of the Document / Number if required

Partner	Effort	Scope
EGI.eu	15 PM	Coordination
CESNET	1 PM	AAI
CSIC (EGI.eu UCST position)	15 PM	Cloud training coordination
NIKHEF	2 PM	Security
LIP	1 PM	Security
STFC	1 PM	Security, PRACE partnership
CERN	2 PM	Incident handling

37 PMs is insufficient to implement all the above mentioned activity areas of EGI training within EGI-Engage. Despite some other elements in EGI-Engage can provide additional effort for some areas – primarily for content development– training efforts in EGI-Engage must be focused on recognising valuable contributions from across the broader e-infrastructure and European/global training landscape, and integrating these into a coherent EGI training portfolio that can strengthen EGI’s unique and innovative position. The next subsections provide overviews on the current status within each of the above listed six areas, and outlines future plans from the first year of EGI-Engage.

## 2.1 E-infrastructure for training

Between 2006-2010 the EGI community (at that time called EGEE community) was operating a distributed e-infrastructure, called GILDA, for training. The infrastructure consisted of approximately 10 sites, each offering High-Throughput compute and storage services, complemented by central services required for access (such as a Certification Authority, Workload Management System, File Catalogue). The infrastructure was operating the same grid middleware services that was available on the production infrastructure and acted as a parallel infrastructure for training events and self-paced learners. Despite some of the sites of the GILDA infrastructure are still available, these are not maintained anymore and do not provide the latest, cloud-based access services of EGI which recently attract most of the new users to the infrastructure.

To serve both internal training (i.e. training users about the EGI Solutions) and external training (i.e. RIs training users about their domain tools) during EGI-Engage, EGI needs to have a virtualised training e-infrastructure in EGI-Engage. The virtualised infrastructure should be compatible with the EGI solutions, and must be flexible enough to host external resources, primarily datasets, applications, VMs, tools developed/used by RI communities for big data analytics.

During the next period EGI will pursue the establishment of such a training infrastructure in two phases:



---

Title of the Document / Number if required

1. Setup a training infrastructure in the form of a new, specialised Virtual Organisation on the EGI Federated Cloud. The environment will consist of volunteer sites, complemented by dedicated access control services and environments for user (trainer, trainee) access. The setup is currently under specification and will be completed during June-July with the prime goal to serve the following two EGI Federated Cloud training tutorials:
  - a. Software Carpentry workshop, 15-17. July, SAP Offices in Feltham, UK
  - b. HPCS Conference, 20-24. July, Amsterdam, NL
2. Seeking long-term arrangements within or beyond EGI Resource Providers for sustaining the virtualised training e-infrastructure so it can operate continuously and can serve both specific events and self-paced learners. It is expected that the resources within the virtualised infrastructure will change over time, however stability is required for the core services, such as VO information in Operations Portal, Access control services, Monitoring and incident handling service, etc.

## 2.2 Training resources

Training resources are datasets, computing applications, workflows, science gateways, Virtual Machine Images, presentation slides, hands-on guidances, web media and any other types of software, documentation and configuration settings that can be used by trainers and trainees on the training e-infrastructure to carry out specific training tasks. The exact set of training resources that a given training event requires depends on the scope and exact agenda of the event. Given the ‘Demonstrator, Exchanger, Integrator’ role of EGI training, it’s expected that the following types of training resources will be needed:

1. Resources that demonstrate the main capabilities of EGI solutions to external audiences (i.e. to people outside of EGI, e.g. to researchers).
2. Resources required to update EGI members about solutions emerging from within the NGIs.
3. Resources that should be brought into the EGI production infrastructure but first need to be demonstrated and explored in training context.

Suitable resources will be identified and ported into the EGI Training context primarily from EGI-Engage WPs and Task, for example from the Federated Open Data task (JRA2.1), Federated Cloud (JRA2.2), Accelerated Computing (JRA2.4), Competence Centres (TSA2.3-10).

## 2.3 Training modules

Training modules integrate and package training resources, services and content for specific training objectives. For example a training module about the ‘EGI Federated Cloud’, targeting potential users from academic research communities could consist of:

- One entry in the Training Marketplace, containing basic information about the module
  - Learning objectives
  - Pre-requisites for usage

---

Title of the Document / Number if required

- Expected length of completion
- Link to further elements of the module
- High level introductory guide (e.g. a webpage in the EGI Wiki)
- Materials that explain the main capabilities of the federated cloud (e.g. presentation slides stored in the EGI Documents Database)
- VM images ready for deployment in the EGI Federated Cloud (e.g. in the training VO)
- A dataset that's available on a specific cloud site of the fedcloud.egi.eu VO (and can be read and processed by the VM image after instantiation)
- Hands-on exercises that provide specific guidelines to instantiate the VMs in the training VO (e.g. Webpage in the EGI Wiki)

During the first year of the project EGI-Engage will focus on developing training modules on topics that can strengthen EGI's engagement with scientific communities, i.e. best support the implementation of the EGI Engagement Strategy<sup>2</sup>. The following modules are planned to be developed in the first year in forms that enables reusability across the NGIs:

1. Under leadership of EGI.eu UCST: a training module about the 'EGI Federated Cloud' solution, demonstrating the cloud computing concept through the VM management and data access interfaces of this platform. The module would guide the user through the following process:
  1. Login to the training front-end machine (which will be one/more VMs deployed and started in advance by the tutor in the cloud)
  2. Browse the EGI Applications Database VM image repository via its Web interface, choose a specific basic image and a cloud site where this can be instantiated (the image has been prepared and registered by the tutor in advance for those sites that support the training)
  3. Instantiate the chosen image on the chosen site (using contextualisation)
  4. Check that the VM is alive and works (through the interface the VM content offers)
  5. Create a disk volume, attach it to the VM
  6. Demonstrate the use of the content of the VM and its attached disk (through the interface the VM content offers)
2. Under leadership of EGI.eu UCST and EUDAT2020 project: a joint training module with EUDAT, demonstrating the use of EUDAT services and the EGI Federated Cloud platform for the analysis of big data. The module would guide the user through the process of
  1. Authenticating to the EGI infrastructure (Federated Cloud)
  2. Instantiating a VM in the EGI environment and performs preliminary configuration (account, network, firewall, tools, libraries, compilers, applications, ...) through contextualisation
  3. The virtual environment is started up (temporary user's credentials are copied into the VM)

---

<sup>2</sup> EGI Engagement Strategy: <http://go.egi.eu/engagementstrategy>

---

Title of the Document / Number if required

4. Staging data from EUDAT to a local storage area in the EGI Cloud for performance and locality reasons, using authentication obtained in step 1
  - data staging via GridFTP or HTTP (CDMI compliant)
  - data discoverability via B2FIND
5. The user launches the computational job in the cloud
6. Any intermediate result is shared with the public via the EUDAT B2DROP service (either results or inputs)
7. Final results are ingested back onto EUDAT for being preserved in the long term (B2SAFE, registered via PID)
3. By members of the 'long-tail of science' platform developer team (CYFRONET, CNRS, SZTAKI, INFN Catania, Poznan and possibly others): a module about the use of the EGI long-tail platform, which is a mixed HTC-cloud VO with a User Registration Portal and with a DIRAC, WS-PGRADE, CSGF, QosCosGrid web environments for users to interact with the infrastructure. The module would guide the user through the process of
  1. Requesting access to the platform.
  2. Deciding about the environment to use for using the allocated capacity.
  3. Logging into the web environments to develop and use HTC/cloud applications.
  4. In case of an exhausted resource allocation grant, get in contact with the national user support team or move on directly to a domain-specific VO.
4. **By EGI and PRACE members (STFC, SURFsara): A module about security solutions and security interoperability solutions in the EGI and PRACE e-infrastructures.**
5. By CERN, CESNET, LIP, NIKHEF: Virtual Machine preparation and certification in the EGI Federated Cloud. The module will target those who prepare and those who endorse VM images for scientific/training purposes. The module will define the characteristics of 'secure' VM images, will teach how to prepare such images, and will demonstrate how to assess the security characteristics of such images.
6. By CERN, CESNET, LIP, NIKHEF: Incident handling in the EGI Federated Cloud. The module will target the users and administrators of EGI Federated Cloud services and will train them about the processes and tools that are available to report security incidents and to respond to those incidents.

Additional modules that will be also explored in the second part of the first year through specific EGI-Engage tasks and interested NGIs:

1. Open Data Processing solution from EGI (by members of task JRA2.1).
2. GPGPU-computing in EGI (by members of task JRA2.4 and NGI-BG).
3. Deploying clouds and federated clouds for scientific and educational purposes (by members of the Federated Cloud collaboration).
4. iRODS training (in collaboration with NGI-FR and the DIRAC team).
5. Joint modules with the RIs that have partnership with EGI (e.g. Competence Centres, joint projects, MoUs).

## 2.4 Training marketplace

The EGI community, in the context of the EGI-InSPIRE project has established an EGI Training Marketplace which is available at <http://go.egi.eu/training>. The Training Marketplace is an online registry to register (advertise) and to view (browse and search) training events, Webinars, Materials, Resources and University Degrees that relate to EGI or e-science. The service supports cooperation between trainers and trainees in different localities and projects by connecting the groups through the stored items that are advertised in the Training. The Training Marketplace is typically used together with the EGI Document Database (<http://documents.egi.eu>) where files (training presentations, tutorial documents, etc.) are stored and then indexed from the Marketplace. The Training Marketplace offers Web gadgets for integration with third party websites. This facility is suitable e.g. to embed a listing of training materials into an NGI website. The Training Marketplace also provides commenting and content validation facility on each stored item:

- Any user (after registration and login) can leave a comment on a registered item.
- The user who registers an item can re-validate the item on a regular basis (e.g. once a year), indicating to those who browse the item that it's still up-to-date and valid.

The content that's currently available in the marketplace needs to be reviewed by EGI.eu and updated (ideally by those who registered these) where necessary. Further development of the Marketplace is envisaged within the EDISON H2020 project, primarily to be able to serve the training needs of 'data scientists'. The EDISON project will start in September 2015 with EGI.eu member of the project consortium. (See further information about the EDISON project in Section 5.1.

Title of the Document / Number if required

Title	Description	Last Reviewed
<a href="#">CVMFS infrastructure for EGI VOs - EGI Webinar by Mr C. Condurache/STFC</a>	This is an MP4 format recording of the EGI webinar delivered on 5 Sept 2013 and described beneath: <a href="https://documents.egi.eu/document/1932">https://documents.egi.eu/document/1932</a>	September, 2013
<a href="#">DIRAC Distributed Computing Services - EGI Webinar</a>	Dr. Tsaregorodtsev will present DIRAC (Distributed Infrastructure with Remote Agent Control)- a service which provides an efficient access to various distributed computing and storage resources for users organized in Virtual Research Communities. What makes DIRAC different with respect to other existing grid or cloud systems is that it allows to aggregate in a transparent way resources from different infrastructures using different middleware and managed within different administrative hierarchies.	August, 2013
<a href="#">EGI service offerings for ELI-HU</a>	This slideset provides an overview of EGI and the service offerings that EGI has for Research Infrastructures. The slides have been prepared by Gergely Sipos, member of the EGI.eu User Community Support Team for a teleconference meeting with ELI-HU, the Hungarian pillar of the Extreme Light Infrastructure.	February, 2013
<a href="#">EGI Overview</a>	This slideset provides an overview of the EGI community and the key EGI entities and processes. Special emphasis is given to the user community support activities and user-facing services. The slides have been used by Gergely Sipos, member of the EGI.eu User Community Support Team, at a meeting that took place on the 25th of January 2013 and involved members of EGI.eu and the Dutch e-Science Centre (NLeSC).	January, 2013
<a href="#">e-FISCAL Tutorial: Methodologies and their Uses</a>	This tutorial gives the readers an overview of the basic skills needed to initiate, execute and analyse the results of simple cost assessment exercises. It covers the key aspects of the different cost assessment methodologies and their optimal application areas. The end has been reserved for a tutorial on how e-FISCAL costing model works in practice as well as on basic cost considerations related to cloud adoption.	January, 2013
<a href="#">Conquering Data-Intensive BioSciences Problems</a>	Live Webinar 6th November 2012 <a href="http://info.sgi.com/genomics/webinar.aspx?s=sgi_MV">http://info.sgi.com/genomics/webinar.aspx?s=sgi_MV</a> to register	September, 2013
<a href="#">GISELA Training Materials</a>	GISELA project training material	September, 2012
<a href="#">Workflows &amp;</a>	These slides were created for the educational purposes, to show the platform of scientific workflows and workflow systems within EGI with a workflows use cases	July, 2012

Figure 1. 'Browse Materials' view in the EGI Training Marketplace

## 2.5 Access control system

Access control to training items (e-infrastructure, resources, modules, marketplace, webinar/e-learning) have different requirements and constraints than similar items in the production environment. Access should be as open and as easy as possible – recognising the need for access by large number of people with diverse knowledge.

The different training services and items of EGI are currently accessible via separate, disconnected authentication and authorisation systems. While an integrated, role-based access control system would be convenient (for example to grant access for a new trainer across the whole spectrum of resources), such system is not critical and the development is not expected within EGI-Engage. However, through the e-infrastructure pilot development activity EGI-Engage will explore access control for training e-infrastructure in the following way:

---

Title of the Document / Number if required

- EGI.eu User Community Support Team (UCST) to request a robot certificate from an IDGF.
- UCST registering this robot certificate in the training VO to be established in the EGI Federated Cloud.
- UCST to generate short-term proxy certificates from this robot for specific training events, and for specific users of these events (for both trainers and trainees).
- Trainers and trainees of the event to use these short-term proxies to interact with the cloud resources, VMs, datasets and User Interfaces during the course.

## 2.6 Webinar and/or e-learning system

Given the spread of the EGI federation across countries and continents, Webinars and e-learning platforms provide the only cost effective solution for many of the training needs that the community requires. EGI.eu has a subscription to the Webex teleconference platform and can host up to two Webinars at a time. EGI Webex<sup>3</sup> can be booked by members of EGI after obtaining the necessary account from the EGI.eu office ([support@egi.eu](mailto:support@egi.eu)). Participants can join EGI Webinars via a Web browsers (for free) or via a phone (pay as you go). Through the Italian NGI, EGI members have access to an Adobe Connect system, which offers an alternative hosting platform for Webinar events. While Webex rooms have currently the limit of 25 on the number of attendees for any Webinar event, the Adobe rooms can host until 200 participants. Experience shows that Webex performs better for interactive events with intense discussions among participants, Adobe serves better for presentation-style events with little/no feedback from members of the audience.

The Webex and Adobe systems will continue to serve EGI during EGI-Engage and the webinar events are advertised in the EGI Training Marketplace. 3 Webinar events have been organised by the EGI-Engage project since the 1<sup>st</sup> of March (AAI in EGI; EGI Federated Cloud; Data management solutions in EGI grids and clouds). These – as well as previous webinars – have been recorded and the recordings are freely available on the EGI Youtube channel.

EGI currently has no platform for e-learning and should explore arrangement for such through the NGIs and partner projects – for example through the Slovenian NGI which is connected to the Slovenian ELIXIR node, hosting a MOODLE deployment for ELIXIR.

## 2.7 High-impact training events

Members of the training task, as well as members of the EGI collaboration continuously monitor the landscape of events to look for training event opportunities that EGI should contribute to. The below table summarises those events that have been recognised so far by EGI-Engage as good opportunities to contribute to. The table will be updated on a regular basis and items will be evaluated to decide about the feasibility and the most suitable form of contribution in each case.

---

<sup>3</sup> <http://egi.webex.com>

## Title of the Document / Number if required

Event	Date and venue	Suggested Contribution AND Action Owner	Value of attending and possibly contributing to the event	Status
ELIXIR Data Carpentry workshop	June 22-25, 2015, Utrecht, NL	The event will include a two-day hackathon to develop and improve teaching materials for computational methods in life sciences. EGI representatives could contribute with the development of a module that's focussed on performing computational methods on EGI e-infrastructures. (Gergely Sipos)	Success EGI contribution could reach large number of scientists who work with computational methods in life sciences and could contribute to building a sustainable user base for e-infrastructures from life sciences.	Under discussion
Software Carpentry workshop	15-17. July, SAP Offices in Feltham, UK	Accepted training Tutorial on the EGI Federated Cloud. To be delivered by Diego Scardaci (INFN-EGI.eu) and Gergely Sipos (SZTAKI-EGI.eu).	Reaching SW developers various science disciplines and promote the EGI Federated Cloud to them.  Pilot repeatable course on EGI Federated Cloud.	To be delivered by EGI.eu UCST
HPCS Conference	20-24. July, Amsterdam, NL	Accepted training Tutorial on the EGI Federated Cloud. To be delivered by Enol Fernandez (CSIC-EGI.eu) and Yin Chen (EGI.eu).	Reaching researchers from HPC, HTC, cloud and big data domain and promote EGI Federated Cloud to them.  Pilot repeatable course on EGI Federated Cloud.	To be delivered by EGI.eu UCST
National e-infrastructure for science and its role within the research infrastructure roadmap (Romania)	Sept 10-11, 2015.	Alexandru Nicolin (NGI International Liaison of Romania)	Support the NGI institutes in engaging with national user communities in order to build a national roadmap for e-, and research infrastructures, as well as a stronger and more sustainable NGI. Potentially co-locate a focussed training course here.	Under discussion
EGI Community Forum	10-13 Nov, Bari, Italy	The next key community event for the EGI Community, good opportunity for knowledge	Training for NGI staff, Training for Competence Centres	Programme is under preparation

Title of the Document / Number if required

		exchange events, and short, focussed courses targeting research infrastructures and projects		
Final BioMedBridges Symposium: Open bridges for life science data	17-18 Nov, EBI, Hinxton, UK	An e-infrastructure workshop with the interested CCs, and with EUDAT? (Gergely Sipos, SZTAKI-EGI.eu)	Expose recent e-infrastructure achievements from life science to the biomedical RIs and build joint workplans with them.	Under discussion

## 2.8 Evaluation process

Measuring the impact of EGI Training and how the EGI-Engage effort benefits trainers and trainees within and beyond the project is critical for success. The training marketplace, as a one-stop shop for trainers and trainees, can offer means for us to capture metrics, such as:

- Number of up-to-date training modules in the marketplace
- Number of training events delivered (and thus registered in the training marketplace)
- Number of downloads from the marketplace per item category and per individual item
- Direct feedback on the registered items (in the form of comments and ratings)

Besides these metrics the project must collect feedback about the trainers and trainees to see whether EGI provides the right type of services, and how much impact these make inside each community/NGI/group. This can be best achieved by surveys thus the project will

- Define and run surveys regularly among those who act as trainers within the community
- Collect feedback from the attendees of EGI-related events through the trainers. To facilitate data collection and integration, the project will prepare a training survey template and will promote this to EGI trainers.

The feedback collected through the various channels will be evaluated by the training activity at least every 6 months.

## 2.9 Optional services

### 2.9.1 Certification programme

Certification can help EGI standardise training content, guarantee quality of training, and can also make the use of training resources, modules and events more attractive for both trainers and trainees. Depending on interest the EGI training activity can explore the establishment of a lightweight certification framework, e.g. one based on Mozilla Open Badges.



---

Title of the Document / Number if required

### **2.9.2 Massive Open Online Courses (MOOCs)**

A MOOC was prepared in 2014 under the coordination of EGI.eu and SURFsara in the context of EGI-InSPIRE. The content was based on the HTC solution of EGI and compatible technologies, such as specific science gateways. The MOOC was delivered as a one-time event at the University of Amsterdam, but the presentation videos are still available through the EGI YouTube channel. The currently available effort does not make it possible for EGI-Engage to repeat this MOOC, or to promise the preparation of a new MOOC. At the same time the project will liaise with the NGIs – especially where there are strong links to universities – to make arrangements at these institutes for the reuse of this MOOC and for the preparation of new MOOCs.

## 3 Training plans and needs within EGI communities

EGI exists as a community of communities. Many of these are linked to scientific projects or research infrastructures each with its own plans for training. While the previous section described what EGI training and especially the EGI-Engage training task needs to do to fulfil its role, this section provides a summary of ongoing training developments within the EGI-related research infrastructures and scientific projects. This section is based on input received from the Competence Centres (WP6) of EGI-Engage and from the structured user communities of EGI, who are represented in the EGI User Community Board.

Besides the reported activities each of the EGI-Engage Competence Centres will perform training in some shape or form during the EGI-Engage project. These activities will be primarily: (1) creation of new training modules about technical services that the Competence Centres develop, and (2) delivery of f2f or online training events based on these modules. These activities will mostly start in the second year of EGI-Engage and in the first project year the centres are largely focussed on infrastructure testing, technology evaluation, software/system specification and development.

### 3.1 BBMRI CC

The CC members have internally discussed this and it seems that we are fine with the common webinars and educational and training tools that are available in EGI at the moment.

BBMRI-ERIC core activities currently are focused on Common Services for ELSI (Ethical, Legal, Societal). These activities are already running Webinars and offering Web based resources for training on topics important for BBMRI nodes. The 'Hands on Biobanks' is the annual event that offers networking opportunity and training for biobankers and where f2f training is conducted.

In the RItrain project, flagship training curricula will be developed for RI managers to enable them being successful in setting up and operating RIs. Pilot courses will be delivered based on these curricula. The BBMRI-ERIC is one of the partners in this project. Primary target groups of the training are:

- Members of those RIs that are members of the consortium.
- Members of additional RIs that are still in the planning/preparatory phase.

In the CORBEL project one of the activities (WP9) will be focussed on educating RI operators. There will be four cluster areas here: Data management, Integration physical access, Ethics, Innovation. WP9 will define competences, develop courses, and will conduct staff exchange programmes.

---

Title of the Document / Number if required

That we can jointly perform with EGI is the development of training modules for biobankers, in which EGI can for example contribute with content on deploying IaaS cloud and operating services in those clouds. Priority topics are foreseen in CORBEL training are:

- Omics data analysis on clouds
- Deployment of private clouds in biobanks

## 3.2 DARIAH CC

Work within the DARIAH RI is carried out within Working Groups (WGs). DARIAH has a WG dedicated to Training & Education. The mission of this WG is to “Provide a training programme for researchers in the methods, tools and approaches needed to engage with the digital environment, including DARIAH services, tools and content. The WG focuses in particular on the development and delivery of international summer school programs, development of collaborative, consortium-wide online training materials; and activities that foster a better understanding of teaching digital humanities across disciplines, institutional, linguistic and cultural borders”<sup>4</sup>. One of the key activities of the WG these days is setting up a reference training curriculum for digital humanities. The WG is open to EGI to explore the involvement an e-infrastructure-related content in this reference curriculum.

## 3.3 EISCAT\_3D CC

Within the EISCAT\_3D CC training and e-learning have not been discussed at depths yet. EISCAT runs courses regularly on working with the radars and radar data. These courses usually give a background to the radar measurements, followed with hands on exercises to run the radars (selecting modes, executing, analysis and evaluation). Most of this could possibly be done via the Internet. But with remote attendance one might lose the focus on the course.

What EISCAT has not done so far is to have training for users at somewhat more advanced levels. That would be how to do make own radar modes, advanced analysis and data searches/mining. These could be perfect to run together with e-Infrastructures at somewhat regular intervals and levels. The CC can contribute to the setup of such events in the second project year.

## 3.4 ELIXIR CC

ELIXIR training is represented in the EGI Community by Brane Leskosek from Slovenia. The focus is on eLearning. The e-learning system in ELIXIR is based on the MOODLE environment and the first modules to be developed will be on Galaxy and on CHIPSTER.

- Galaxy is a popular, open, web-based platform for data intensive bioinformatics research.

---

<sup>4</sup> From the WG poster presented at the 4<sup>th</sup> DARIAH VCC in Ljubljana in April 2015.

---

Title of the Document / Number if required

- CHIPSTER is a package of analysis tools in bioinformatics. CHIPSTER has been recently ported to the EGI Federated Cloud platform.

LMS system and module setup have started. Authentication is done with EduGAIN, pipelines to be developed in Galaxy, then respective PHP scripts to be developed for MOODLE that orchestrate the pipelines. The underlying infrastructure is envisaged to be the EGI Federated Cloud.

Potential collaboration between EGI and ELIXIR e-Learning environment will focus on integrating the MOODLE – Galaxy – CHIPSTER components with the EGI Federated Cloud. The credential translation mechanism that is developed in the context of the EGI long-tail platform seems like an important component to achieved EduGAIN to X509 translation for this setup.

### 3.5 MoBrain CC

WeNMR is regularly running workshops in which virtual machines are used that are pre-packed with the necessary data and software for the participants. This currently runs on cloud(s) outside of EGI, making sure that students can use the images also on their laptops even without networks connection. (This is also a back-up plan for tutorials as one can never know how good the connectivity might be at a training session.)

In the past WeNMR got a grant from SURFsara to provide students each with a cloud Linux system to run a particular VM over two months, part of a bachelor course.

Finally WeNMR has recordings of lectures and tutorials recorded during a recent tutorial in Taipei. The videos are available on the WeNMR channel on YouTube.

During university courses the community is currently using blackboard as teaching / communication environment.

Further the MoBrain CC partners are regularly speaking at international conferences in their respective fields of research, showcasing results obtain using the EGI infrastructure (mainly grid so far) and advertising its resources. A recent example of this is the HADDOCK workshop held in Poznan Poland during which the EGI and in particular the Polish NGI grid resources were highlighted (<http://dhmg.amu.edu.pl/index.php/haddock>).

### 3.6 Disaster mitigation CC

Input to be added

### 3.7 Life Science Grid Community (LSGC) VRC

LSGC is represented in EGI through the User Community Board. The community is one of the most active user of the HTC solution and do this through a few Virtual Organisations from which the Biomed VO is the largest. Users interact with these VOs typically with high level environments, primarily the Virtual Imaging Platform (VIP) and DIRAC. LSCG would benefit from a standard Federated Cloud training module, and suggests EGI to include a DIRAC module in its training

---

Title of the Document / Number if required

portfolio because this makes the HTC platform usable for this community, so potentially for many other communities too. A VM-DIRAC element could be part of an EGI cloud module too.

## 4 NGIs, EIROS – Training status and plans

National Grid Initiatives (NGIs) and EIROS – members of the EGI Council – also can be considered as communities of their own within EGI. These also often have their own training infrastructure, internal training plan and modules which must be considered when defining the high-level EGI training plan. This section provides further details on the known training plans and priorities of the NGIs and EIROS.

### 4.1 Bulgaria

The Bulgarian NGI is currently engaged in the actual put in operation of the new 400 TFlops system that has been delivered in March and trainings will be oriented to introduce users and communities to the capabilities of the system and the various interfaces that are available to access it and the data management options. The first training is planned for October/November 2015 and will be an inductor level training, in collaboration with the BG CLADA community (CLARIN+DARIAH) and with the scientists from 3 Bulgarian national NSF projects. Audience is mainly composed of researchers and Ph.D. students from Bulgarian Grid and HPC community.

The NGI will also work on training materials. The main topic is the use of coprocessors and most notably Xeon Phi accelerators, with which we have working experience and which provide most of the computational power of the new machine. They are interested in joining a task force if it is related to the production of training material for the optimal use of computational accelerators like Xeon Phi.

The NGI would be interested if trainers from EGI could participate in some of its training events as lecturers and also they could use training content that is being developed. It can offer some of its lecturers to be trainers, especially related to the use of accelerators, GPUs or other more HPC-oriented features.

### 4.2 Czech Republic

During the first project year is planning to organise hands-on OpenNebula training for scientific user groups, focusing on HPC use of the Cloud. The NGI would be interested in joint training activities where training materials or approaches are being developed. The NGI might also be interested in train-the-trainers events.

### 4.3 France

France Grilles plans to deliver the following training sessions during the first year of EGI-Engage in French:

---

Title of the Document / Number if required

- User oriented training: DIRAC - 2 days, iRODS - 1 day (we already have training material and training infrastructure and a specific VO)
- Site administrator training: Puppet, IPV6, (Quattor possible)
- User oriented cloud: 28th of May in Strasbourg and another session will probably be organized.

The NGI's training web pages are here: <http://www.france-grilles.fr/Frances-Grilles?lang=en> and one can find links to Indico pages of past events at <http://www.france-grilles.fr/Formations-programmees?lang=en>. Our trainers are "internal" (members of the NGI).

Several of our trainers may be interested to work for other institutes who want to conduct EGI/NGI-related training: DIRAC (DIRAC4EGI or FG-DIRAC), iRODS, Cloud (OCCI / OpenStack)

We are interested to participate in the development of content of user oriented cloud training in EGI. We are interested if possible to get support for the content or through a webinar or with the help of trainers for an IPV6 training session (target: grid and cloud sites administrators).

## 4.4 Italy – INFN Padova

The site is planning to deliver a training course about OpenStack cloud for technicians working at the University of Padova. For this we are preparing training material for the following cloud topics: (1) Use of OpenStack, (2) OpenStack installation and configuration, (3) Storage in OpenStack. These materials can be shared within EGI, but the site has no effort to join European task forces on training.

## 4.5 Hungary

The NGI is already affiliated with two universities (at Szeged and Miskolc), where Cloud Computing courses are held for B.Sc. and M.Sc. students. A similar course will be introduced later this year at a third university (Óbuda). The NGI and EGI can collaborate on cloud computing training modules.

## 4.6 Portugal

The NGI do not have dedicated budget for training and currently performs activities on best effort basis. In the medium term the NGI is planning to deliver training for both system administrators related to cloud topics (like utilisation of Neutron service in Open Stack) and for scientific end-users on both cloud and grid technologies. There are links also within the NREN to universities (FCCN). Getting further support from EGI Both content and trainers would be of major important together with webseminars. We already have our own infrastructure for training.

## 4.7 Spain – BSC

BSC organizes training linked with PRACE PATC courses. Next one on COMPSs will be held around February 2016 and will include Cloud training. The audience is mainly composed of master

---

Title of the Document / Number if required

students and researchers (users) of BSC departments. An example of training on COMPSs can be found at of PATC <http://www.bsc.es/patc-comps-2015>. At this link you can find a way to profile the participants an properly organize the course <https://events.prace-ri.eu/event/327/registration/register#/register>.

We are interested in the development of content contributing our training on the COMPSs platforms and other BSC tools. There is the possibility to use the infrastructure provided by PATC if the course is organized through that channel. We can host lecturers and develop webinars using EGI tools. We will also need EGI support for the dissemination of BSC training courses. We are available to provide BSC trainers for events organized by EGI and organize events here (with trainers from other institutions) in line with BSC training program.

## 4.8 Spain – CETA CIEMAT

Currently there is no training plan designed in our site (we are new in the Federated Cloud User Support group). We are open to collaborate with any tutorial, course or training in this area. This collaboration, in function of the funding, could be offered online and/or on-site.

Last year we offered a BigData training (OpenStack + Hadoop) of 15 hours in the University of Extremadura. We have the material (in Spanish), and it is likely to hold the training during this year again.

It would be convenient to have contents, webinars and training infrastructures from EGI.



## 5 Training collaborations to explore

Given the large number of new initiatives in training, it's crucial for EGI training to monitor the broader context in which it needs to operate, and to establish partnership with initiatives that can help EGI create sustainable value-added training services. This partnership development activity has already started and reached an important milestone at the EGI Conference 2015 event, where a 2x90 minutes long session was organised with guest speakers. The session title was 'Distributed platforms for e-Learning: How can educators and scholars benefit from the Open Science Commons?'<sup>5</sup>. The presented initiatives/projects, together with partnership opportunities are described in the next subsections.

### 5.1 EDISON

The EDISON project will focus on activities to establish the new profession of 'Data Scientist', following the emergence of Data Science technologies (also referred to as Data Intensive or Big Data technologies) which changes the way research is done, how scientists think and how the research data are used and shared. This includes definition of the required skills, competences framework/profile, corresponding Body Of Knowledge (BoK) and model curriculum. It will develop a sustainability/business model to ensure a sustainable increase of Data Scientists, graduated from universities and trained by other professional education and training institutions in Europe.

To achieve this, EDISON will work with the major Data Science stakeholders from academic, research communities and industry, as well as with the professional community to help them to obtain proper education and training and/or formal certification for already practicing self-made Data Scientists, grown from the advanced research projects who want to build a new career in Data Science.

Consistent Data Science education and professional training requires besides theoretical knowledge access to real scientific data infrastructure and real large data sets to acquire practical experience and develop data centric thinking. For this, EDISON will leverage on EGI infrastructure and community/activities, as well as products from the APARSEN project, to create a supporting infrastructure for Data Science education and training that will include both example datasets and virtual labs which will allow the students or trainees to work with real data sets, infrastructure and tools.

EDISON will facilitate the establishment of a Data Science education and training infrastructure at major European universities by promoting experience of 'champion' universities involving them into coordinated development and implementation of the model curriculum and creation of cooperative educational and training infrastructure.

---

<sup>5</sup> <http://indico.egi.eu/indico/sessionDisplay.py?sessionId=95&confId=2452#20150522>

---

Title of the Document / Number if required

The project will be coordinated by the University of Amsterdam and will involve partners from universities (University of Stavanger, Univ of Southampton), Research Institute for Telecommunication and Cooperation (FTK), European Grid Initiative (EGI), industry (Engineering, Italy), and SME InMark.

EDISON defines the following objectives:

- **Objective 1:** Promote the creation of Data Scientists curricula by an increasing number of universities and professional training organisations. Increasing the number of universities offering Data Science programs is essential in order to respond to the increased demand of the European research, industry and public sectors.
- **Objective 2:** Provide conditions and environment for re-skilling and certifying Data Scientists expertise to graduates, practitioners and researchers throughout their careers. To provide conditions for a sustainable increase in the number of Data Scientists in Europe, EDISON will also support the training by professional organisations of the self-made expert, in the light of reaching the target number of graduates that gradually will satisfy the level of research and industry demand in coming 5-8 years.
- **Objective 3:** Develop a sustainable business model and a roadmap to achieve a high degree of competitiveness for European education and training on Data Science technologies, provide a basis for the formal recognition of the Data Scientist as a new profession, including supporting the future establishment of formal certification for practicing “self-made” Data Scientists.

The project will establish the **EDISON Liaison Groups** as a key channel for facilitating structured dialogue between industry, researchers and users. The purpose of performing a programme of innovation Focus Groups is to launch a dialogue between the EDISON team and experts from DS job market. These groups will discuss in prospective terms, alignment of education to market needs, as well as exploitation potential and business models to deploy EDISON results. These EDISON Liaison Group will include representatives for each of the stakeholder domains of university teachers, students, employer sectors, and data experts. The membership of these groups will grow during the project and may change in time. The foreseen roles of the EDISON Liaison Group are:

- **Employer sectors.** Visionary persons invited from academia, public sector (including government, health and libraries) as well as the private sector. The latter will include professional sectors such as accountancy and finance, transport and logistics, telecom, and the entertainment industry. These groups will reflect on the required skills and attitudes of academic trained data scientists, and comment on considered developments in academic training.
- **Data experts:** Inform about current and expected scientific and technical developments. Provide feedback about relevant existing curricula in view of expected scientific and technical developments.
- **Academic scholars and officers:** professors and leading practitioners from universities with experience in data science training to exchange views and discuss the requirements for educational training with respect to knowledge and skills goals, educational methods, examination objectives, and European cooperation.

---

Title of the Document / Number if required

## 5.2 Bio-Linux

Bio-Linux is a comprehensive, free bioinformatics workstation based on Ubuntu Linux and Debian Med. It has 10 years & 8 major releases, Over 7000 users from 1600 locations. Bio-Linux includes over 200 bioinformatics packages, including big integrative tools such as QIIME, Galaxy Server, PredictProtein, EMBOSS.

Potential activity in EGI:

Analyse the available Bio-Linux distribution and test its compatibility with the EGI Federated Cloud IaaS. Create a training module on how to perform bioinformatics analytics on the EGI Federated Cloud using Bio-Linux and reach out to specific bioinformatics groups to validate the module and to facilitate take-up within biomedical science research communities. Focus should be on uptake within Research Infrastructure communities, and involvement of them in the early testing phase can increase chances for sustainability.

## 5.3 EOS Cloud

Currently EOS cloud is a tenancy in the JASMIN Unmanaged Cloud, at STFC in the UK. EOS offers a training environment for UK based researchers. The Web interfaces of EOS are based on JASMIN custom IaaS Software, and EOS users and VMAdmins are registered as JASMIN users. During training events, each user receives two VMs: Bio-Linux and an Ubuntu Docker hosting environment. The users are with total responsibility for instantiated system and access the environment through standard remote desktop tools.

Potential training collaboration between EGI and EOS cloud on:

- The operators of EOS are planning to move this environment to the EGI Federated Cloud in order to share this with other NGIs/countries, and to benefit from the resources of the EGI Federation.

## 5.4 JetStream

JetStream is a project that got funded recently by the National Science Foundation (NSF) in the US with the goal to establish a national science and engineering cloud. JetStream is targeting the long-tail of science in science and engineering research, with a strong focus on education (partly, because this is the first cloud project by NSF targeting science and engineering in general.)

Envisaged components of JetStream: VM library, custom VMs and 'private computing environment' (sort of VPN in the cloud). The system is built on OpenStack with CentOS operation system. User authentication is foreseen via CILogon, from the InCommon identity federation.

Targeted science domains:

- Biology: iPlant and Galaxy VMs
- Earth science

---

Title of the Document / Number if required

- Field station research
- GIS
- Network science
- Social sciences

Main use cases:

1. Delivery of pre-packaged, 'lightweight' VMs to under-resourced campuses for local use.
2. Enable the execution of licenced code with the user using his/her own licence for this tool.
3. Teach how to prepare VMs, how to deploy and operate services from VMs. E.g. publish VM containing analysis tools, data, scripts. Publish services via RunMyCode (or other) interfaces. Make VMs discoverable and downloadable with Globus.

Timeline for establishing the infrastructure:

- Test gear arrives 2015 Q2.
- Production gear in Q3.
- Friendly user mode before SuperComputing 2015.
- Advanced scenarios in 2016.

The Principal Investigator of JetStream is Craig Stewart at Indiana University. Potential training collaboration between EGI and JetStream will be explored on:

1. Establishing VM preparation processes and guidelines for trainers that enable VMs that are reusable in both JetStream and in the EGI Cloud based training environment.
2. Sharing VM images (and respective usage guidelines) between JetStream and EGI, by integrating the VM image catalogues of the two infrastructures.
3. Joint development of practices and tools for researchers about how to access scientific datasets from within VMs running in JetStream/EGI clouds.

## 5.5 D4Science

D4Science is at the same level as the EGI Federated Cloud – an e-infrastructure with its own Resource Providers and access mechanism on which higher level tools and applications can be built. D4Science provides support for geospatial data, biodiversity data, statistical data, to implement the 'typical researcher data pipeline':

Data registration → Harmonisation → Generation of metadata → Publication in standard format

The system has been recently used for 3 university degree courses for biology sciences students and for computer engineering. Topics were: Perform models; Model analysis.

BlueBridge is a new VRE project in H2020 (INFRA-9-2015), with the involvement of D4Science. BlueBridge is planning to conduct at least 20 courses in the next 2 years. The project goal is

"To support capacity building in interdisciplinary research communities actively involved in increasing scientific knowledge about resource overexploitation, degraded environment

---

**Title of the Document / Number if required**

and ecosystem with the aim of providing a more solid ground for informed advice to competent authorities and to enlarge the spectrum of growth opportunities as addressed by the Blue Growth Societal Challenge.”

**Potential collaboration between EGI and D4Science:**

The D4Science environment is currently operates independently, but in parallel with EGI, sharing rather few components between each other. For example D4Science currently does not support the VOs that are in EGI (in Operations Portal and VOMS/Perun), Application codes and VMs from the EGI Applications Database, Authentication tokens from IGTF or EGI SSO, ARGO monitoring infrastructure, APEL accounting infrastructure, exposing experiment results through EGI storage sites. EGI and D4Science should build bridges between the two infrastructures to facilitate ‘openness’ and ‘repeatability’ of experiments and data analysis pipelines and to share cost of maintaining and innovating certain e-infrastructure components. Joint training modules and environments can be built on this shared e-infrastructure services. There are two priorities here:

1. As the first step D4Science is going to integrate the system with the EGI Federated Cloud in order to enable the use of compute and storage capacity by applications and pipelines of the D4Science system. The work require integration between the AAI and task and management systems of D4Science and EGI Federated Cloud.
2. Integrate the D4Science application configuration testing part with the AppDB GUI dissemination part in order to enable the testing of applications and VMs that are visible in AppDB and to facilitate the dissemination of D4Science applications through AppDB interface. The integration requires a common profile between the D4Science and AppDB registries.

## 5.6 SoBigData

A new Research Infrastructure project that will develop an e-infrastructure, platform with tools and models, user support and training for mining social data. The project includes 6 thematic clusters, each with own competences, methods and services (Text and social media mining; Social network analysis, Human mobility analytics, Web analytics, Visual analytics, Social data). The e-infrastructure currently used is a centralised data centre. The community does not use clouds and grids at the moment.

**Potential collaboration between EGI and SoBigData will be explored on:**

EGI should explore with the project potential need for a distributed e-infrastructure that could underpin the 6 thematic clusters of SoBigData collaborate in order for example to offer more scalable capacity and more flexible environments for applications, services and tools used in social mining. Depending on the interest EGI can offer/deploy reusable elements (technologies and services) from the EGI infrastructure and from the broader EGI community to build a distributed e-infrastructure for SoBigData. A shared e-infrastructure between EGI and SoBigData would be the basis for (1) the sharing of training services and modules, and (2) for the collaborative

---

Title of the Document / Number if required

development of custom modules for those wishing to develop services for social mining, or carry out social mining activities on the SoBigData RI.

## 6 Summary – Activity plan until February 2016

This section provides a summary of priority activities for the 1<sup>st</sup> project year. The list is numbered, but this does not define priorities among the items. Effort estimates are given for each item. This effort is needed from EGI-Engage T6.1 for the successful completion of the tasks. Additional effort will be required for some of the items (e.g. joint work with other tasks, or other projects for example to operate resources in the virtualised training infrastructure; to develop training modules on specific topics.)

### 1. Virtualised e-infrastructure:

- Pilot a virtualised e-infrastructure for training in the form of a new VO hosted on the EGI Federated Cloud, and complemented by a training access control mechanism. Deliver two tutorials based on this setup in July, then document the findings as a blueprint for future events. (1 PM)
- Work with Resource providers after July to identify suitable sites, services and business model for a virtualised training e-infrastructure that can operate 24/7 to serve EGI-related courses and self-paced training needs. (2 PM)

### 2. Training modules:

- Develop a reusable, customisable training module about the 'EGI Federated Cloud' solution by mid-July, use it during the two tutorials arranged in July, then make it available as reusable modules. (1 PM)
- Develop a joint training module with EUDAT by the end of 2015, based on the joint use case that has been identified and documented during the EGI-EUDAT meeting at the EGI Conference in Lisbon. (0.5 PM)
- Develop a training module by members of the 'long-tail of science' platform developer team about the use of the EGI long-tail platform. Target of the module will be the NGI user support teams, who can train and support long-tail users based on this knowledge. (2 PM)
- Develop a training module about security solutions and security interoperability solutions in the EGI and PRACE e-infrastructures. (0.5 PM)
- Develop a training module about Virtual Machine preparation and certification in the EGI Federated Cloud. (1 PM)
- Develop a training module about incident reporting and handling in the EGI Federated Cloud. (1 PM)
- In the second part of the first project year explore joint modules on (3 PM in total)
  - Open Data Processing solution from EGI (in collaboration with task JRA2.1).
  - GPGPU-computing in EGI (in collaboration with task JRA2.4 and NGI-BG).
  - Deploying clouds and federated clouds for scientific and educational purposes (in collaboration with Federated Cloud Task Force and Operations).
  - iRODS training (in collaboration with NGI-FR).

---

Title of the Document / Number if required

- Joint modules with the RIs that have partnership with EGI (e.g. Competence Centres, joint projects, MoUs).
3. Events:
- Deliver two tutorials about the EGI Federated Cloud in July and possibly at the EGI Community Forum in November. (1 PM)
  - Deliver training events with EGI's support at interested NGIs, using newly developed modules, resources and e-infrastructures as appropriate. (2 PM)
  - Continuously monitor the websites of partner RIs and projects, identify high-impact events to which EGI should contribute to. Arrange contributions from the local NGIs where possible. (1 PM)
4. Training Marketplace (0.5 PM):
- From September, in collaboration with the EDISON H2020 project, define and start the development of a new generation of the EGI Training Marketplace. The new generation should improve the sharing and integration of contributions from external projects for Data Scientists (the key beneficiaries of EDISON).
5. Partnerships (2 PM):
- Follow-up those partnerships that have been identified at the EGI Conference 2015 event. Monitor the broader context in which EGI training operates, recognise and establish partnerships with initiatives that can help EGI establish sustainable, value-added services for training.
6. Measuring success (1.5 PM):
- Collect the following metrics through the Training Marketplace:
    - Number of up-to-date training modules in the marketplace
    - Number of training events delivered (and thus registered in the training marketplace)
    - Number of downloads from the marketplace per item category and per individual item
    - Direct feedback on the registered items (in the form of comments and ratings)
  - Define and run surveys regularly among those who act as trainers within the community
  - Collect feedback from the attendees of EGI-related events through the trainers. Prepare a training survey template and promote this to EGI trainers to facilitate data collection and integration.

TOTAL effort needed from T6.1 during PY1: 20 PM



## Appendix I. Training module development steps

1. Who do you want to train?
  - a. What is their existing knowledge / skills that are relevant?
  - b. What extra knowledge, skills or behaviour do they want to acquire?
  - c. How many are there (e.g. demand/year)?
  - d. What resources can they contribute?
2. What is the syllabus?
  - a. New material (skills, methods, knowledge, judgement, behaviour) you plan to deliver.
  - b. This needs to be reviewed with the stakeholders identified in 1 to see if it is what they want.
  - c. But the stake holders are also their (future) employers, etc.
3. How will the content of the syllabus be developed and delivered so that people absorb the required increments to their knowledge and skills?
  - a. How can this be resourced from the point of view of developing the material and delivering the material?
  - b. How can the identified students (people engaging to learn) find (a) the prerequisites if they don't have them, (b) the time & engagement to learn, and (c) coping with the pace and duration? i.e. should there be identified stages?
  - c. How will the learners be supported, e.g. tele-tutoring and group discussions?
  - d. How will it be made concrete so progress is appreciated?, e.g. What practical exercises are there?
  - e. How are the preparation staff resourced to deliver?
  - f. How are the delivery staff resourced to support each replay of the course?
  - g. How are the technical support arrangements made, e.g. we need to book time on EGI machines to teach forward wave propagation modelling, about 20 nodes / student.
4. How do the courses deal with student feedback? Solicit it? Discussions during the course and a suitable time later?
  - a. Review of their progress by their organisations
  - b. Integration and distillation of the analysis
  - c. Identification of the areas needing action.
  - d. Revision of any of the previous stages.
  - e. This is key particularly the first rendition of the course must be considered a trial run, before it is considered prepared!