EGI-Engage

Training plan for the second period
(June 2016 – August 2017)

M6.5

Date	31 May 2016
Activity	SA2
Lead Partner	EGI.eu
Document Status	FINAL
Document Link	https://documents.egi.eu/document/2810

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 'SA2.1 Training' activity provides core training services and facilitation of training activities conducted by Competence Centres, NGIs, partner projects and partner infrastructures (including e-infrastructure and Research Infrastructures).

This milestone documents the training plan that is agreed with the broad EGI community and its partners for the second period of the EGI-Engage project (June 2016-August 2017). The document is a formal milestone towards the European Commission.

COPYRIGHT NOTICE

DOCUMENT LOG

Issue	Date	Comment	Author/Partner
v.1	3/May/2016	First draft	G. Sipos / EGI.eu
v.2	23/May/2016	All sections completed based on input from relevant parties.	G. Sipos / EGI.eu
v.3	1/Jun/2016	Updated section on EUDAT collaboration	G. Sipos / EGI.eu
FINAL	1/Jun/2016	Approved with no comment from EGI-Engage AMB	G. Sipos / EGI.eu

TERMINOLOGY

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

Contents

1 Introduction

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 main goal is "to empower researchers from all disciplines to collaborate and to carry out data and compute intensive science and innovation"[1], 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 following roles:

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 community → Internal training for EGI members.

3. **Bringing in new requirements** from users/communities participating in training → Capturing feedback about current solutions; Capturing requirements for new services (demand-pull)

These activities target researchers; IT support teams within research collaborations, research projects and industry; managers of research infrastructures; technical staff at service/resource provider institutes. Assembling relevant training content for them, delivering these efficiently at impactful events, monitoring the efficiency of training, capturing and evaluating feedback from training are the activities that the SA2.1 training task focuses on. SA2.1 has 37 person months effort for over 30 months, spread across 7 partners. The effort breakdown, together with the effort level that's still remaining for the second period is provided in Table 1 below:

---

[1] EGI Strategy: http://go.egi.eu/strategy

Partner	Effort for whole project (30 months)	Remaining effort for the second period	Scope of activity
EGI.eu	15 PM	13.08 PM	Activity coordination, EGI solutions
CESNET	1 PM	1 PM	AAI
CSIC (EGI.eu UCST position)	15 PM	12 PM	Cloud training
NIKHEF	2 PM	1.42 PM	Security
LIP	1 PM	1 PM	Security
STFC	1 PM	1 PM	Security, PRACE partnership
CERN	2 PM	2 PM	Incident handling

The next section provides a training activity plan for the second period of the project. The plan was prepared and agreed by the following members and partners of the EGI-Engage project:

Name, affiliation	Contribution	Role
Dave Kelsey, STFC	Security modules and events	EGI security training coordinator
Gergely Sipos, EGI	Overall structure, EGI content	SA2.1 coordinator, EGI User Community Support coordinator
Petr Holub, BBMRI-ERIC	BBMRI CC section	CC coordinator
Davor Davidovic, IRB	DARIAH CC section	CC coordinator
Ingemar Häggström, EISCAT	EISCAT_3D CC section	CC coordinator
Eric Yen, AS	Disaster Mitigation CC section	CC coordinator
Kimmo Mattila, CSC	ELIXIR CC section	CC coordinator
Daniele Bailo, INGV	EPOS CC section	CC coordinator
Jesus Marco, CSIC	LifeWatch CC section	CC coordinator
Alexandre Bonvin, UU	MoBrain CC section	CC coordinator
Peter Solagna, EGI	AARC project section	Project operational contact in EGI
Themis Athanassiadou, EGI	EDISON project section	Project operational contact in EGI
Yin Chen, EGI	ENVRIplus project section	Project operational contact in EGI
Jesus Marco, CSIC	Indigo-Datacloud section	Project training contact
Marjan Grootveld, DANS	EUDAT2020 section	Training coordinator in EUDAT2020

2   Training modules

A training module integrates training materials (e.g. slides), instructions (e.g. Wiki page) and links to required training services (e.g. training.egi.eu VO, fedcloud.egi.eu VO) to help specific target audience learn about a specific topic (e.g. how to use the OCCI interface of the Federated Cloud). A training module can be shared directly with trainees in the form of an online guide, or can target trainers who deliver face-to-face courses using the module as a guide.

Title	Description, Target audience	Priority for development	Way of delivery	Responsible person(s)
The EGI Federated Cloud for application developers	The module would provide guidance for application developers (PaaS, VRE, service) who want to integrate scientific software with the EGI Federated Cloud. The focus will be on the use of APIs for service discovery, compute and data management, user AA. The course will consist of two parts: 1. Standard-based clouds (OCCI) 2. OpenStack clouds (Nova)	High	• Online • Version for f2f delivery will be prepared only if there is request (e.g. by specific RIs)	Giuseppe la Rocca (EGI UCST)
Container based applications in the EGI FedCloud with Docker	User guide already exists on how to run individual containers inside Ubuntu VMs on the Federated Cloud. This new module would extend this with guide for application developers on how to create complex container-based applications, e.g. with Docker Swarm. The target audience is developers of scientific applications.	Medium (requires technology investigation)	• Online • Version for f2f delivery will be prepared only if there is request (e.g. by specific RIs)	Enol Fernandez (EGI UCST)
Creating data federations with the EGI Open Data Platform	The course will provide guidance for data providers on how to create data federations, with harmonised views, using geographically dispersed datasets. Such federations can simplify access and (re)use of data by distributed user communities.	High	• Online	Lukasz Dutka (CYFRONET, JRA2)
Hosting data-intensive services on EGI – the DataHub concept	The course would build on the Open Data Platform and extend it with instructions for scientific users and scientific application developers on how to bring 'computing to data', i.e. enable	High (depends on data federation course)	• Online	Matthew Viljoen (EGI, JRA2)

	and use virtualised applications on federated datasets that are sitting together with cloud compute services on EGI resources.			
GPGPU computing in EGI	The module provides guidance for application developers to integrate scientific applications with GPGPU resources of EGI. The course would consist of two parts: 1. GPGPUs in the cloud 2. GPGPUs in the grid	Medium	• Online	Themis Athanassiadou (EGI UCST)
Connecting scientific services with EGI resources using the EGI AAI Proxy	The course would target service operators in structured scientific communities and train them on how to integrate community-specific services with the authentication and authorisation 'layer' of EGI through the new proxy service.	Medium	• Online • Create version for F2F delivery in case of interest.	Christos Kanellopoulos (GRNET, JRA1)
EGI services for the long-tail of science	The module will consist of two parts and will • individual researchers and small research groups on the use of sciences that are integrated in the EGI Platform for the long-tail of science. • NGI user support teams on supporting users with this EGI platform.	High	• Online • To decide on version for F2F delivery based on feedback on online version.	Giuseppe la Rocca, Gergely Sipos(EGI UCST)
Platforms in the EGI Federated Cloud	Set of modules, each specialised on a platform (PaaS, SaaS, VRE, gateway) that provides high level abstractions and interfaces for application developers and/or scientists to access compute and storage services in the EGI Federated Cloud.	Medium	• Online • To decide on version for specific events (e.g. EGI Forums, RI conferences)	Each course created by the platform developers/providers. Giuseppe as supervisor (EGI UCST)
eInfrastructure user security awareness training	This module will act as an introduction for end users of eInfrastructures. This will be done jointly with other Infrastructures in the WISE working group on security training and is likely to involve pulling together existing material from various sources.	Medium	A set of slides tested at a F2F event, which could then be turned into an online course if time and effort permits	STFC (and LIP and possibly other SA2.1 members)
Security incident handling, methods and	This development will build on the earlier material produced in EGI-InSPIRE and EGI-Engage PY1. Some components are technical	High	Delivered in security training sessions at conferences and/or specialed security	Nikhef, CESNET and CERN

forensics	hands-on training and some will be role-play scenarios following Incident Handling procedures. The target audience includes system administrators of any resources or services within the whole EGI portfolio, managers/operators of virtualised services in the EGI Federated Cloud, managers of Science gateways and portals.		workshops related to such events. For example DI4R 2016, ISGC2017	
Security for Research Infrastructure and Research Community managers	A module to gather together the various security obligations in terms of policy, procedures and best practice. Audience is any person responsible for security or general managers in Research Infrastructures and Research Communities	Medium	A set of slides delivered at a F2F training session. We can build on the excellent work already done in this area funded by NSF in the USA in CTSC, probably via the WISE security training activity.	STFC (possibly with CERN and other SA2.1 members)

3 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. NGIs, CCs, the EGI Foundation are organising community events.

The below table is a snapshot of those events that have been recognised for the near future as good opportunities for EGI training to contribute to. The events are sorted by their date. The online version of this table is kept up-to-date at https://wiki.egi.eu/wiki/Community_events.

Type OR audience	Title	Date	Location	Website	Comment/Next action
Webinar	The EGI Open Data Platform - Towards Scientific Data Hubs	2016.05.17	Online	https://indico.egi.eu/indico/event/2969/	In collaboration with EGI-Engage JRA2.1
Science/ESFRI	EISCAT_3D User Meeting	2016.05.18-19	Uppsala, SE	http://www.space.irfu.se/workshops/EISCAT-3D_User2016/	Follow-up in EISCAT_3D CC. Talk and demo about data-compute portal from the CC
Scientific / ESFRI	ELIXIR Finland: Using clouds and VMs in bioinformatics training	2016.05.23-25	Helsinki, FI	https://csc.fi/web/training/-/cloud-vm-bioinformatics	EGI-Engage to contribute with a Federated Cloud - CHIPSTER tutorial
Webinar	WS-PGRADE gateway environment for compute intensive workflows on clouds	2016.05.26 (TBC)	Online	TBD	In collaboration with MTA SZTAKI
Webinar	Enabling access to EGI services with community-specific user identities: The new EGI IdP/SP Proxy service	2016.06 (TBD)	Online	TBD	In collaboration with EGI-Engage JRA1.1

Webinar	DIRAC4EGI system: Cataloguing and computing with data in EGI	2016.06.07	Online	TBD	In collaboration with Uni. of Barcelona
Scientific	DH BENELUX Conference 2016	2016.06.09-10.	Belval, Luxembourg	http://www.dhbenelux.org/	Regional event on Digital Humanities activities in Belgium, the Netherlands, and Luxembourg
e-Infrastructures	TNC 2016	2016.06.13-16.	Prague, CZ	https://tnc16.geant.org/	
Scientific	EUROMAR 2016	2016.07.03-07.	Aarhus, Demark	http://euromar2016.org/index.php	(INSTRUCT-related)
Scientific	Digital Humanities 2016	2016.07.12-16.	Kraków, Poland	http://dh2016.adho.org/	Follow-up in DARIAH CC. Workshop abstract was submitted, but rejected.
Scientific + ESFRI	European Conference on Computational Biology (ECCB)	2016.09.03-07.	Hague, NL	http://www.eccb2016.org/	Follow-up in the ELIXIR CC. Possibilities: Talk, poster, exhibition booth
Scientific	EURALEX Congresses for lexicographers	2016.09.06-10.	Tbilisi, Georgia	http://euralex2016.tsu.ge/	Follow-up in the DARIAH CC: Possible training workshop and paper in the proceedings.
RI & E-infra	Digital Infrastructures for Research (DI4R)	2016.09.28-30	Krakow, PL	http://www.digitalinfrastructures.eu/	Abstract Submission deadline: June 3. Follow-up by all EGI activities
RI / ESFRIs	ICRI-International Conference of Research Infrastructures	2016.10.03-05.	Cape Town, SA	http://ec.europa.eu/research/infrastructures	This seems relevant, however outside of Europe
ESFRI	DARIAH-EU annual event	2016.10.10-12	Gent, BE	http://dariah.eu	Follow-up in the DARIAH CC
Science/ RI	Digital Humanities Conference 2017	2017.08.08-11	Quebec, CA	NA.	Follow-up in DARIAH CC. Outside of Europe.

4 Support for Competence Centre training activities

4.1 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) is focused 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.

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

4.2 DARIAH CC

Work within the DARIAH-ERIC is carried out with a bottom-up approach, in Working Groups (WGs) that integrate community effort into few yearlong projects. 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"[2]. One of the key activities of the

---

[2] From the WG poster presented at the 4th DARIAH VCC in Ljubljana in April 2015.

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.

The CC is currently working on redefining its Wiki page into a service provider site that describes the services and demonstrator applications offered by the CC. In parallel with this work the CC will work on defining training modules (for Webinar or F2F delivery) to promote and educate researchers from the field about these services. Taking into account the heterogeneity of the DARIAH community and researchers with various backgrounds (humanists from numerous research fields, computer scientists, developers, service providers) the training effort will be divided into two directions. The first direction aims to provide a set of webinars and training materials for DARIAH application developers and service providers on how to use EGI cloud resources (e.g. using EGI infrastructure, create and start virtual appliances, container-based applications, etc.) and the second targeting end-users, scholar and scientists coming from Arts and Humanities on how to utilize the specific applications, tools and services developed by DARIAH-CC members (e.g. gLibrary, Parallel semantic search engine, DARIAH science gateway, etc.). The CC has identified a few events where workshop/tutorial type contributions will be submitted. (See in Events section.)

4.3 EISCAT_3D CC

In early 2016 the CC reached a pilot setup of the EISCAT_3D portal. The portal can be used for data queries (based on meta-data) and data download. Work is ongoing to integrate computing capabilities into the portal (e.g. visualise the data remotely instead of downloading it), and to lower the barriers of access (username-password instead of x509 certificate).

The portal will be presented and demonstrated to the EISCAT community in the second half of May in Uppsala, during the EISCAT_3D User Meeting (See event above). The feedback will complement the ongoing work towards stabilizing the portal into a service that can be opened for the EISCAT_3D community. Later in 2016 and in 2017 the CC will also give Webinar or F2F training for EISCAT researchers about the production portal. Try to co-locate F2F events with community conferences/workshops.

4.4 Disaster mitigation CC

The CC is currently focused on implementing disaster simulation cases. Training activities by the CC are envisaged in January 2017 the earliest. Possible venues for training can be APAN 43 (February 2017, New Delhi), where the DMCC is planning to hold a f2f meeting. Topic of the training would be the two simulation portals of tsunami and weather.

ASGC has been running sessions about earth science, meteorology and climate changes at International Symposium on Grid and Cloud for APGI annually for more than 10 years. Topics such

as case study, research method, application practices, trends, multidisciplinary innovation and regional collaborations are generally covered.

For the Disaster Mitigation CC (DMCC), the training requirement is based on the major DMCC objective to reliably support the high performance disaster impact analysis application environment. Three perspectives from the user, workflow, and resource provider should be included in the training plan. Promotion of the disaster simulation and analysis gateway is essential, the workflow, user interface and functionality will be revised according to the feedback from the user communities. Application development and support group have to be knowledgeable about the underlying distributed computing infrastructure, the application platform and the available EGI components such as the AAI, Cloud services, Data Management Services etc. For the resource provider, the operation technology and system efficiency are necessary skills for a reliable infrastructure.

Training delivered by DMCC will be primarily implemented as Webinars and f2f events co-located with community events.

- Training on tsunami simulation and weather simulation will be held at least once by end of 2015. The necessary data set, the simulation process, and the results access and analysis all will be covered.
- The 40th APAN meeting and the co-located TEIN meeting (http://www.apan.net/meetings/KualaLumpur2015/) will be considered for f2f training or at least demonstration of current simulation portals. DMCC will have a half-day F2F meeting at APAN40. DMCC will broaden the collaboration with the Earth Monitoring Working Group and related working groups of APAN.
- An Environmental Computing Workshop – possibly with training, is planned to be organized by the DMCC at ISGC2016. (A DMCC f2f meeting is planned at ISGC2016).

4.5 ELIXIR CC

At the time of writing the CC is nearly finished the setup of the first cloud compute sites as part of the ELIXIR Compute Platform. Once completed, the platform will serve as the underlying system for the demonstrator applications that will be ported to it by CC members (See M6.2 for further details on these applications). The CC is planning to contribute to ELIXIR community events with the results of these efforts. Presentations and demonstrations seem like the most suitable formats, but tutorials will be also considered.

In parallel with the CC activities other members of the EGI community – primarily the EGI User Community Support Team – is also looking for opportunities for contributing to ELIXIR events with tutorials. For example, they will contribute to an ELIXIR Training event in May 2016 in Finland with a Federated Cloud – CHIPSTER tutorial (See events section above for details).

4.6 EPOS CC

Members of the EPOS CC collect analyse and compare Earth Science community needs with EGI technical offerings based on specific use cases that are selected as drivers for the work. The CC identified, analysed and described three of such use cases in the recently published M6.4 milestone, and is currently working on their implementation: (1) Integration of EGI and EPOS AAIs, AAI, (2) Integration of the VERCE gateway with EGI Federated Cloud for running misfit analysis on data from EIDA / ORFEUS, (3) Create an environment for the development of value added services on top of the satellite data.

After an analysis and design phase, EPOS CC is planning to implement pilot implementations of use case (1) and (2). As a parallel activity, tools for training users about the usage of such a platform are need. To this respect, training material for use case (2) will be provided. Such material will introduce the platform basics and explain what waveform modelling is. It will also guide the user towards the registration process and installation of certificates and creation of proxy certificates. Also, it will train the user on fundamental operations as running a simulation, accessing the results in a grid / cloud environment, and running simulation using data uploaded by user.

4.7 LifeWatch CC

The goal of the LifeWatch EGI CC is to capture and address the requirements of Biodiversity and Ecosystems research communities. To achieve this the CC will (1) deploy cloud and GPGPU based e-Infrastructure services required to support data management, data processing and modelling for Ecological Observatories, (2) explore possibilities to increase the participation of citizens in data-intensive biodiversity research, (3) facilitate the adoption and exploitation of the EGI infrastructure by the LifeWatch user community.

During the first year of EGI-Engage the CC will be focussed on the evaluation and adaptation of EGI technologies and biodiversity applications. (Some of the preparatory work will happen outside of the CC, such as in the 'Accelerated Computing' task of EGI-Engage.) During this period the CC will benefit from the training courses offered and organised by EGI on fundamental e-infrastructure topics and EGI solutions. The CC will be active in training during the 2nd and 3rd project year, and the next issue of the EGI Training plan will cover specific activities.

4.8 MoBrain CC

The MoBrain CC already delivered training tutorials in the 1st project year, using services and tools that were developed/integrated with EGI by the CC. (For example INSTRUCT course in Utrecht, Cryo-EM tutorial in Taipei).

The CC is planning to contribute to the WestLife-MoBrain workshop that will be co-located with the iNEXT project meeting in the second half of October in Spain, close to Madrid. The

organisation of a MoBrain tutorial at the upcoming DI4R conference will be also explored, with the main goal to broaden uptake of MoBrain tools/services by researchers in Poland.

5 Joint activities with partner projects and partner infrastructures

5.1 AARC

Training by AARC in the next period will be focused on identity providers, providing them guidelines and best practices on managing and releasing user identities to facilitate access to federated e-infrastructure services. This topic is not directly relevant for EGI members (who are service providers and not identity providers).

EGI will be involved in the following technical activities with AARC in the next period:

- Token translation and attribute management cross-pilot: attributes managed in SAML attribute authorities to be translated into other standards, such as X.509, OpenIdConnect, CILogon. The topic can be relevant for service operators who need to translate user identity tokens and attributes between different formats.
- VO Entitlements and multi-VOs entitlements: how information about multiple VOs can be included in the same user's authentication assertion and be univocally interpreted, associating the sub-groups and roles to the right VO. Best practices to be suggested by AARC on this topic. These can be included also in EGI trainings that target user communities who manage their own attribute authorities.

5.2 EDISON

Within the EDISON project EGI.eu is involved in creating and supporting components of the 'EDISON Online Education Environment', and in the setup of a cloud for education. These activities both relate to EGI training:

The Education environment will provide a one-stop shop for educational and training resources to support both residential (university) and online education. The catalogue will make available on-line material and resources (from simple presentations to MOOCs or learning objects) using on the Data Science Taxonomy. Dynamic navigation will allow (1) the identification of gaps in the student or practitioner knowledge; (2) mapping existing courses or on-line materials to cover the identified gaps or the required specialisations towards the different target domains in Research, e-Infrastructures, the private sector and public organisations.

The EDISON education cloud will be based on the EGI training infrastructure and will offer a bookable and flexible infrastructure for demonstrations and hands-on exercises for Data Science classes. The cloud will be extended with Business Intelligence and Analytic tools (such as open source tools SpagoBI) to target wider audience of education and training programs. Guides and tutorials will be developed for the educator community on how to book, customise and use the

infrastructure for Data Science courses. Attention will be given to the overall user experience to meet the users' (students, graduates, as well as scholars and practitioners) expectation and way-of-working.

5.3 ENVRIplus

A training survey was conducted by EGI within the ENVRIplus project in May 2016, during the 2nd ENVRIplus week. The responses are currently under analysis, and in the next months will be used to:

- Prioritise training modules from EGI in terms of their relevance for ENVRIplus RIs
- Identify scientific applications that should be integrated with EGI, and should be promoted (with hands-on training) for ENVRIplus RIs
- Define a schedule of webinar and face-to-face training events to be conducted by EGI in the ENVRIplus training task. By looking at the short term: a f2f training at the next ENVRIplus week (Nov 2016, Prague), and a webinar event between now and that date is envisaged.

5.4 Indigo-Datacloud

RBI is leading T2.4, the training task in Indigo-Datacloud. This task is not active yet, but some initial ideas already emerged for training:

- Organising short courses for edX.org (Catania has some experience with this). (Only on topics that are not there yet.)
- Submit to RI4D a full session on "Training components for the Open Science Framework" where the project could present different levels, roles, basic background (in cloud and data management), etc.

The project will finalise its training plans in the next months and will involve EGI-Engage representatives in these discussions.

5.5 EUDAT2020

The following joint activities are planned in the area of training between EGI-Engage and EUDAT2020 project in the next 16 months:

- Harmonising existing e-infra training courses between EGI and EUDAT. This would help position their service offerings within the Open Science research workflows, and lead towards integrated service offerings from EGI and EUDAT.
- EUDAT has started working on API for B2Stage service. Based on this API the integrated EUDAT-EGI use case can be finalised (the prototype of the joint use case was demonstrated in Bari in 2015 November). B2Stage API releases for EGI testing are expected in the second half of 2016, with public release by the end of 2016. Based on the experiences with this API

the partners will create demos, prepare user guides and consultation for new communities. The exact dates and format of these will be defined in the second half of 2016.

6 Summary – Activity plan until Sept 2017

This section provides a summary of activities planned for the 2nd project period. The total funded effort that is available for these is 31.5 PMs.

1. Training modules (Section 2):
   o Prepare with high priority support from EGI-Engage:
     ▪ The EGI Federated Cloud for application developers (to broaden uptake of the Federated Cloud)
     ▪ Creating data federations with the EGI Open Data Platform (a new EGI service)
     ▪ Hosting data-intensive services on EGI – the DataHub concept (a future EGI service)
     ▪ EGI services for the long-tail of science (a new EGI service)
     ▪ Security incident handling, methods and forensics (for service providers)
   o Prepare with medium priority support from EGI-Engage:
     ▪ Container based applications in the EGI FedCloud with Docker (to respond to this popular topic)
     ▪ GPGPU computing in EGI (a new EGI service)
     ▪ Connecting scientific services with EGI resources using the EGI AAI Proxy (a new EGI service)
     ▪ Platforms in the EGI Federated Cloud (external contributions)
     ▪ eInfrastructure user security awareness training (jointly with other e-infrastructures)
     ▪ Security for Research Infrastructure and Research Community managers
2. Events (Section 3):
   o Continuous monitoring of upcoming events, submission and organization of tutorials to impactful events. Current list is provided in section 3 above.
3. Partnerships (Section 4 and 5):
   o Support for Competence Centres in conducting webinar and f2f training events about community specific e-infrastructure services. Support is given through generic EGI training services (training infrastructure, consultancy for trainers about VM preparation, training accounts, training registry, agenda setup, event announcements)
   o Work with AARC, EDISON, ENVRIplus, Indigo-Datacloud, EUDAT-OpenAire on joint training activities and modules.
4. Virtualised training infrastructure (Appendix 1):
   o Formalise the support provided by cloud sites for the training infrastructure. Operation Level Agreements to be setup to guarantee availability of resources for at least until the end of the project.
   o Work with the participating service providers on identifying funding sources and implementing operational model for beyond the project.

Appendix I. E-infrastructures available for training

During 2015 the EGI community has established a cloud based e-infrastructure for training, under the coordination of the EGI-Engage project. This training infrastructure is hosted as a dedicated resource pool, a Virtual Organisation (VO), on the EGI Federated Cloud infrastructure. The infrastructure provides resources and services for face-to-face events, online training courses (Webinars, MOOCs) or self-paced learning modules.

The training VO is integrated with the 'per-user proxy factory' solution of EGI to generate personal, but short-living proxy certificates in an easy way, for trainers and trainees. Such proxy certificates can identify students individually, and have limited lifetime - typically few hours or days, depending on the length of the training event.

The infrastructure can be extended with customised training environments on-demand. These environments can be implemented and deployed on the infrastructure in the form of Virtual Machine images. Reusable training exercises are available for those who wish to use the infrastructure for training.

The goals and possible usage modes of the EGI training infrastructure are:

1. Provide an infrastructure that demonstrates services of the EGI federated cloud. In this operational mode the infrastructure can accommodate courses that focus on the usage of the EGI cloud services themselves. Such courses typically target programmers or other technical members of scientific communities or projects.
2. Offer a baseline cloud infrastructure for training courses about scientific software and services. In this operational mode the representatives of the community (the trainers) deploy custom Virtual Machine images on the training infrastructure before the training, and these images offer the training environment for the students. Because of the cloud-based operational model the students can have dedicated training environments, and the community can benefit from the easy deployment, predictability and repeatability of courses.

The infrastructure currently includes resources from 5 cloud sites (CESNET, BIFI, UKIM, CETA-CIEMAT, INFN-Catania). Formalisation of participation with Operation Level Agreements (OLA) is currently ongoing between EGI.eu and the sites. The OLAs would guarantee availability of the sites in the infrastructure for at least until the end of EGI-Engage (Aug 2017).

Further information is available at https://wiki.egi.eu/wiki/Training_infrastructure.

Appendix II. Training module development steps

1. Who do you want to train, who is the target audience?
   a. What are their existing knowledge / skills that are relevant?
   b. What extra knowledge, skills or behaviour do they want to acquire?
   c. How many is 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 is the preparation staff resourced to deliver?
   f. How is the delivery staffs 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 the key particularly the first rendition of the course must be considered a trial run, before it is considered prepared!