

**EGI-Engage**

Training plan

DRAFT – TO BE FINALISED AFTER EGI CONFERNCE 2015 BASED ON YOUR FEEDBACK

SEND FEEDBACK TO gergely.sipos@egi.eu BEFORE MAY 27.

(Write input/updates with track changes into the document).

M6.1

|  |  |
| --- | --- |
| **Date** | 1 May 2015 |
| **Activity** | SA2 |
| **Lead Partner** | EGI.eu |
| **Document Status** | DRAFT |
| **Document Link** | <https://documents.egi.eu/document/XXX> |

Abstract

EGI-Engage in the T6.1 task (Training) of WP6 (Knowledge commons) has effort for training. This effort provides coordination to training activities across the EGI collaboration (beyond EGI-Engage!), and offers a framework to develop, integrate and deliver training content, events and infrastructures customised for various audiences. The task will connect training activities of the CCs, NGIs and external projects, and will strive to maximise their effectiveness and impact across e-Infrastructure communities. This milestone defines the plans for the first 12 months of EGI-Engage (March 2015-Feb 2016) to establish foundational services and customised services and modules in EGI and in collaboration with the global e-science community.

This document is a formal milestone delivered to the European Commission in May 2015, but will be continuously updated based on feedback from the EGI community and its partners.

**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**

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***Name*** | ***Partner/Activity*** | ***Date*** |
| **From:** | Gergely Sipos | EGI.eu-SZTAKI / SA2 | 28/05/2015 |
| **Moderated by:** |  |  |  |
| **Reviewed by** |  |  |  |
| **Approved by:** |  |  |  |

**DOCUMENT LOG**

|  |  |  |  |
| --- | --- | --- | --- |
| ***Issue*** | ***Date*** | ***Comment*** | ***Author/Partner*** |
| **v.1** | 22/05/2015 | First issue for external review | G. Sipos / EGI.eu-SZTAKI |
| **...** |  |  |  |
| **...** |  |  |  |
| **v.n** |  |  |  |

**TERMINOLOGY**

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

**Contents**

[1 Introduction – What kind of training should EGI-Engage enable? 6](#_Toc420065402)

[2 What training services are needed to achieve our goals? 7](#_Toc420065403)

[2.1 Foundational services 7](#_Toc420065404)

[2.2 Training modules 7](#_Toc420065405)

[2.2.1 What is a training module? 7](#_Toc420065406)

[2.2.2 Required modules 8](#_Toc420065407)

[2.3 Optional services 8](#_Toc420065408)

[3 Training plans and needs within EGI communities 9](#_Toc420065409)

[3.1 MoBrain CC 9](#_Toc420065410)

[3.2 BBMRI CC 9](#_Toc420065411)

[3.3 ELIXIR CC 10](#_Toc420065412)

[3.4 LSGC VRC 10](#_Toc420065413)

[3.5 … 10](#_Toc420065414)

[3.6 … 10](#_Toc420065415)

[3.7 … 10](#_Toc420065416)

[4 NGIs – Training status and plans 12](#_Toc420065417)

[5 Training collaborations to explore 13](#_Toc420065418)

[5.1 EDISON 13](#_Toc420065419)

[5.2 EUDAT2020 13](#_Toc420065420)

[5.3 Bio-Linux 13](#_Toc420065421)

[5.4 EOS Cloud 13](#_Toc420065422)

[5.5 JetStream 14](#_Toc420065423)

[5.6 D4Science 15](#_Toc420065424)

[5.7 SoBigData 16](#_Toc420065425)

[6 Summary - Status and next steps 17](#_Toc420065426)

[Appendix I. Training module development steps 18](#_Toc420065427)

[Appendix II. Typical issues during training – Solutions and recommendations 19](#_Toc420065428)

# Introduction – What kind of training should EGI-Engage enable?

* Training to support EGI’s Engagement activity with scientific communities
  + Training for CCs, NGIs, VRCs, international scientific communities (typically projects), long-tail of science, SME and industry (engagement).
* EGI-Engage provides (through T6.1)
  + Framework – infrastructure and logistical support for trainers
  + Core modules
  + (Evaluation?)
* CCs, VRCs, NGIs provide
  + Custom modules for domains, technologies, specific audiences
  + Training delivery
  + User support (post-training)
* Partners provide (e.g. universities, partner e-infrastructures)
  + External resources that EGI can build into its own service and modul portfolio

# What training services are needed to achieve our goals?

## Foundational services

* Training infrastructure – bookable and virtualised so can flexibly used for different types of courses.
* Pre-cooked VMs to simplify the setup of custom environments in the training infrastructure, specialising it for specific training purposes.
* Sample datasets that can be used for data-driven training topic, e.g. big data concepts, data analysis pipelines, etc.
* Training marketplace – a single place where existing training resources are advertised and made accessible for trainers and trainees.
* A Webinar programme that consist of a webinar system (Webex and Adobe Connect) and support staff at EGI.eu who sets up, records and shares courses delivered by external presenters.
* A ‘training module’ about each EGI-related topic that has value for the ERA and where we expect/predict uptake in the next years thus we need to train people about it:
  + Federated Cloud (IaaS core capabilities for application and platform developers)
  + High-level software tools – External contributions. For example
    - PaaS services compatible with the EGI IaaS cloud
    - Specialised services available from specific NGIs for problems in the HTC, AAI, Federated Data, … domains
  + Joint utilisation of EGI Federated Cloud and EUDAT services. For example
    - Staging data from EUDAT to EGI IaaS cloud for processing
    - Archiving data from EGI IaaS cloud to EUDAT
  + Operational tools and services for Research Infrastructures

## Training modules

### What is a training module?

Training modules fill foundational training services with content, or in other words the resason’d latre of foundational services is to enable effective training modules and their delivery for EGI. Each training module should consist of elements that simplify and maximise uptake within the main beneficiary communities considering the resources available for training package development within EGI. Possible elements of a training package can be:

* Online guides (e.g. in EGI Wiki)
* Reusable training materials (e.g. PPT slides, hands-on exercises, Webinar recordings)
* Reusable VM images ready for deployment in some EGI-related training infrastructure (e.g. in a training.fedcloud.egi VO; in specific NGI sites)
* Datasets prepared for training exercises
* Dedicated support team
* Dedicated trainers

### Required modules

* EGI Federated Cloud – required, with focus on …
* Operation tools – required, with focus on …
* DIRAC - optional

## Optional services

* Certification programme
  + E.g. Mozilla Open Badges
* E-learning environment
  + E.g. MOODLE instance
* Massive Open Online Courses (MOOCs)

# Training plans and needs within EGI communities

Several RIs have training programmes for their end-users for which eLearning tools are needed, or an e-Infrastructure is needed that can provide reference datasets for manipulation, tools, virtual labs etc.

## MoBrain CC

WeNMR is regularly running workshops in which we are using virtual machines pre-packed with the necessary data and software for the participants to run tutorials.

I guess this can be a standard cloud activity, although we also make sure that students can use those on their laptops even without networks connection. You never know how good the connectivity might be and hanging connections are a killer for training sessions.

Got once a grant from SURFSARA to provide students each with a cloud Linux system to run a particular over two months, part of a bachelor course I have been teaching.

And finally we have recordings of lectures and tutorials recorded in Taipei made available through YouTube (WeNMR channel).

Thinking of our university, we are currently using blackboard as e-teaching / communication environment.

## BBMRI CC

We have internally discussed this within BBMRI and it seems that we are fine with the common webinar + website 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.

That we can jointly perform with EGI is the development of training modules for biobankers, e.g. on deploying and operating clouds and services in those clouds. Priority topics are foreseen as

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

## ELIXIR CC

ELIXIR training is represented in the EGI CF by Brane Leskosek, focus is on eLearning. Brane, Eija and Donatella had a skype call and agreed on two types of contributions.

LMS is the e-learning system in ELIXIR, based on MOODLE environment and first modules to be developed on Galaxy and on CHIPSTER.

* Galaxy is a popular, open, web-based platform for data intensive bioinformatics research.
* CHIPSTER is a package of analysis tools in bioinformatics.

System and module setup have started. Authentication 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.

## LSGC VRC

* Would benefit from standard Federated Cloud training module
* Suggests EGI to include a DIRAC module in the portfolio because this makes the HTC platform usable. A VM-DIRAC element could be part of an EGI cloud module too.

## …

aaa

## …

aaa

## …

aaa

# NGIs – Training status and plans

# Training collaborations to explore

## EDISON

Aaa

## EUDAT2020

Aaa

## 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.

## 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 evens 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 though standard remote desktop tools.

Potential collaboration between EGI and JetStream 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.

## 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 of 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
* 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 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.

## 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. D4Sciecne 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 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 infrastructure 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 stroage 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.

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

# Summary - Status and next steps

|  |  |  |  |
| --- | --- | --- | --- |
| Service/modul/activity | Status | Next steps | Lead partner |
| IaaS-based training infrastructure | Under development | * FedCloud VO * Booking process for trainers * Proxy generation tool for trainers and trainees | EGI.eu |
| Pre-cooked VMs for training | Available in some communities and for topics (e.g. WeNMR) | * Stock-keeping existing VMs * Making existing VMs interoperable with EGI cloud training infrastructure (where appropriate) * Developing VMs where lacking and needed | Community topics: CC leaders  Generic topics: T6.1 coordinator |
| Sample datasets | ??? |  | JRA2.x (Federated Data) coordinator |
| Training marketplace | Available from EGI-InSPIRE (at STFC) | * Nothing at the moment. | STFC |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

1. Training module development steps
2. Who do you want to train?
   1. What is their existing knowledge / skills that are relevant?
   2. What extra knowledge, skills or behaviour do they want to acquire?
   3. How many are there (e.g. demand/year)?
   4. What resources can they contribute?
3. What is the syllabus?
   1. New material (skills, methods, knowledge, judgement, behaviour) you plan to deliver.
   2. This needs to be reviewed with the stakeholders identified in 1 to see if it is what they want.
   3. But the stake holders are also their (future) employers, etc.
4. How will the content of the syllabus be developed and delivered so that people absorb the required increments to their knowledge and skills?
   1. How can this be resourced from the point of view of developing the material and delivering the material?
   2. 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?
   3. How will the learners be supported, e.g. tele-tutoring and group discussions?
   4. How will it be made concrete so progress is appreciated?, e.g. What practical exercises are there?
   5. How are the preparation staff resourced to deliver?
   6. How are the delivery staff resourced to support each replay of the course?
   7. 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.
5. How do the courses deal with student feedback? Solicit it? Discussions during the course and a suitable time later?
   1. Review of their progress by their organisations
   2. Integration and distillation of the analysis
   3. Identification of the areas needing action.
   4. Revision of any of the previous stages.
   5. This is key particularly the first rendition of the course must be considered a trial run, before it is considered prepared!
6. Typical issues during training – Solutions and recommendations

* Strong linkage to specialised resources
* Moving from training environment to home environment
* Users break things in ways you never thought possible
* User get ahead of themselves