D15.4

Report on Available Training/Courses on the use of Data Processing Tools from the Existing RI's and the Priorities

WORK PACKAGE 15 – Training, e-Learning and Courses

LEADING BENEFICIARY: EGI FOUNDATION

Author(s):	Beneficiary/Institution
Giuseppe La Rocca, Gergely Sipos, Yin Chen	EGI.eu/EGI Foundation

Accepted by: XXX (WP 15 leader)

Deliverable type: [REPORT]

Dissemination level: PUBLIC

Deliverable due date: 31.10.2016/M18

Actual Date of Submission: 31.10.2016/M18

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654182

ABSTRACT

This document reports about relevant e-infrastructure courses that can underpin RI data processing tools, and about training events and webinars that are organized in the ENVRIplus context to facilitate the harmonized uptake of e-infrastructure services within the RIs.

The document is based on an extensive survey and intelligence gathering that was conducted by EGI during 2016 to build an understanding about the training priorities of RIs with respect to data storage and processing.

Project internal reviewer(s):

Project internal reviewer(s):	**Beneficiary/Institution**
Ari Asmi	University of Helsinki
Yuri Demchenko	University of Amsterdam (UvA)
Jacco Konijn	University of Amsterdam (UvA)

Document history:

Date	**Version**
30.9.2016	V1 - Draft for internal comments
07.10.2016	V2 – First complete draft, ready for internal review.
25.10.2016	V3 – Draft updated for external review
31.10.2016	Accepted by Jacco Konijn

DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the authors (Author names+email addresses).

TERMINOLOGY

A complete project glossary is provided online here:
https://envriplus.manageprojects.com/s/text-documents/LFCMXHHCwS5hh

PROJECT SUMMARY

ENVRIplus is a Horizon 2020 project bringing together Environmental and Earth System Research Infrastructures, projects and networks together with technical specialist partners to create a more coherent, interdisciplinary and interoperable cluster of Environmental Research Infrastructures across Europe. It is driven by three overarching goals: 1) promoting cross-fertilization between infrastructures, 2) implementing innovative concepts and devices across RIs, and 3) facilitating research and innovation in the field of environment for an increasing number of users outside the RIs.

ENVRIplus aligns its activities to a core strategic plan where sharing multi-disciplinary expertise will be most effective. The project aims to improve Earth observation monitoring systems and strategies, including actions to improve harmonization and innovation, and generate common solutions to many shared information technology and data related challenges. It also seeks to harmonize policies for access and provide strategies for knowledge transfer amongst RIs. ENVRIplus develops guidelines to enhance transdisciplinary use of data and data-products supported by applied use-cases involving RIs from different domains. The project coordinates actions to improve communication and cooperation, addressing Environmental RIs at all levels, from management to end-users, implementing RI-staff exchange programs, generating material for RI personnel, and proposing common strategic developments and actions for enhancing services to users and evaluating the socio-economic impacts.

ENVRIplus is expected to facilitate structuration and improve quality of services offered both within single RIs and at the pan-RI level. It promotes efficient and multi-disciplinary research offering new opportunities to users, new tools to RI managers and new communication strategies for environmental RI communities. The resulting solutions, services and other project outcomes are made available to all environmental RI initiatives, thus contributing to the development of a coherent European RI ecosystem.

TABLE OF CONTENTS

EXCUTIVE SUMMARY

This deliverable reports on the available training courses on the use of data processing tools and on the list of training courses and webinars that WP15 (specifically task 15.1 e-Infrastructure training part) has developed and delivered in the last year. The aim is to support those ENVRIplus research Infrastructures (RIs) who want to profit from e-Infrastructure technologies and build Virtual Research Environments (VREs) for (big) data management and processing.

To achieve these goals, we have been focusing on the following activities:

- Understand the ENVRIplus RIs training requirements and identify training priorities.
- Match training needs with existing e-infrastructure materials. Identify gaps and define priorities for filling those.
- Deliver training courses including:
    - Two face-to-face training courses to ENVRIplus communities in the conjunction of the ENVRIplus annual meetings.
    - Four online webinars in collaboration with the EGI-Engage project.
- Define plan for the development and delivery of training materials/events to respond to high-priority e-infrastructure needs.

This document reports our findings of ENVRIplus community's training requirements for data processing tools, and our efforts on planning and delivering suitable courses to the communities. The information on community training requirements is a supplement of Theme2 deliverable D5.1 - A Consistent Characterization of existing and planned RIs. Some requirements on training aspects have been reported in D5.1. This document provides detailed information that has been additionally collected in T15.1. The training activities reported in the second part of the document give supports to the tasks in WP9 – service deployment and validation. We closely work with WP9 to arrange training activities related to e-Infrastructure technology used in WP9 for use case integration testing and service deployments.

The information collected in this document will contribute to different ENVRIplus RIs for their community support activities. Combining training requirements and available courses from other environmental research communities, a RI can collaboratively organize training events and maximize the training impacts. The training plan provided in the second part of the report provides for ENVRIplus community an overview for future trainings. An ENVRIplus RI can select the related courses, and participate the events based on own interests.

1 INTRODUCTION

1.1 PURPOSE AND BACKGROUND

Handling of large volumes of data has been a challenge for many of ENVRIplus RIs. As an example, EISCAT 3D, a next generation incoherent scatter radar system, is expected to start with 2PB/year observation data and will need up to 3.5PB/year of permanent storage. Similar instruments come online in other ENVRIplus RIs as well, many still without a clear long term solution how the generated data will be stored, curated, archived, and shared with scientists. The problem will become even more acute as new RIs' facilities come on-stream in the next years.

The advent of "big data science", however, is not limited to environmental scientific domains. This new paradigm is emerging in any branch of science – from experimental, to theoretical, computational, humanities and social sciences. Data-intensive science consists of three basic activities: capture, curation, and analysis. Data comes in all scales and shapes, covering large

international experiments; cross-laboratory, single-laboratory, and individual observations. Making efficient use of large volumes of scientific data produced by this new paradigm is a critical issue and it introduces new challenges that have to be tackled such as:

- Access and preservation for re-use,
- Interoperability to allow cross-disciplinary exploration,
- Efficient computation,
- Intellectual property rights, etc.

Access and preservation for re-use

A primary aspect of scientific research is that its results need to be reproducible and therefore requires that all inputs and outcomes are made available to researchers. To support the reproducibility of data-intensive science it is important to provide open access to scientific publications, data repositories and associated software tools and create a link between the datasets and software tools adopted to generate research publications. This is a new challenge that we need to cope with.

Interoperability to allow cross-disciplinary exploration,

Many scientific domains are encountering similar technical problems when using large and heterogeneous datasets. Data may have different structures or may not be well structured at all and analytic tools to extract meaningful information from these large volumes of data are lagging. These technical problems are often more complex in interdisciplinary research when the huge amount of data to be processed cannot be moved around the network. For this reasons, novel solutions are needed to help and, in some cases, move the analysis where data is produced and made available.

Efficient computation

Since many years, the European Commission has promoted and invested into the development and the uptake of new advanced ICT infrastructures, so called e-Infrastructures, for research. These modern e-Infrastructures are rapidly changing the way we think about research and the way we do research empowering science by making possible massive interdisciplinary collaborations between researchers on a global scale. New IT technologies, such as cloud computing, can reduce infrastructure complexity and costs, deliver high-quality new services, help to significantly advance our understanding in the different scientific domains and aid scientists, researchers, policy makers and the general public in making decisions. To facilitate decision making, based on reliable scientific evidence, a new class of high-level tools and services, as well as new training modules are needed.

Intellectual property rights

The intellectual property (IP) has proven to be a key role to drive creativity and innovation, but the advent of the “big data science” introduced new challenges we have to cope with. Big data, no matter how massive it is, is just raw data that needs to be collected, stored and analysed before to produce results which need to be interpreted and communicated in a meaningful way. IP right issues usually arise during the collection, the storage, the analysis and the sharing of big data. In this scenario, e-Infrastructures can support seamless access, use, re-use, and trust of data.

Task 15.1 mainly focuses on technical training issues, so this deliverable will mainly touch upon the issues regarding e-infrastructure. In order to help the ENVRIplus RIs better understand available e-Infrastructure technology so as to benefit from using them to support their daily research activities, during the first 17 months of the project (M1-M17), WP15 Task 15.1 has put efforts in understanding the needs of ENVRIplus RIs on data processing, investigating which high-level training courses and open-source solutions match the specific needs of the ENVRIplus RIs, and preparing and delivering suitable training courses to them.

Task 15.1 (e-Infrastructure training part) is let by EGI Foundation. EGI Foundation operates one of the largest, collaborative e-Infrastructures in the World for the support of open science. It provides access to large scale computing, storage and data resources through a federation of national resource providers. Its aim is to support cutting-edge research, innovation and knowledge transfer in Europe. Today, this federation can count on more than 820,000 CPU cores which are federated across the world-wide network of e-infrastructures. EGI Foundation has also a stable cloud infrastructures with 22 cloud providers across Europe and there are 230 active projects with 48,000 of users accessing the federated infrastructure every day for scientific analysis.

EGI Foundation delivers IT solutions that enable domain specific research infrastructures. These solutions are currently adopted to enable the LHC experiment by providing distributed computing facilitate[1] to CERN and its partners institutes, support structural biology by providing different software and hardware solutions to the WeNMR community[2]. The existing high-level services[3] exposed to the end users by EGI Foundation form the EGI Service Catalogue:

- Compute services:
  - Cloud Compute: Run virtual machines on demand with complete control over computing resources.
  - Cloud Container Compute: Run Docker containers in a lightweight virtualised environment.
  - High-Throughput Compute: Execute thousands of computational tasks to analyse large datasets.
- Storage and Data:
  - Online Storage: Store, share and access your files and their metadata on a global scale.
  - Archive Storage: Back-up your data for the long term and future use in a secure environment.
  - Data Transfer: Transfer large sets of data from one place to another.
- Training:
  - FitSM training: Learn how to manage IT services with a pragmatic and lightweight standard.
  - Training Infrastructure: Dedicated computing and storage for training and education.

Other services such as the Open Data Platform (ODP) are under development, or in alpha/beta access mode.

As one of important European e-Infrastructures, EGI Foundation has long-term experiences to provide trainings for various scientific communities. EGI Foundation contributes to ENVRIplus

[1] http://wlcg.web.cern.ch/
[2] http://www.wenmr.eu/
[3] https://www.egi.eu/services/

training tasks by putting efforts to understand community's training requirements, designing suitable courses, and delivering the training in suitable ways. Within ENVRIplus, EGI Foundation also collaborates with other e-Infrastructures and technology providers, such as EUDAT, gCube, OneData, to introduce advanced e-Infrastructure technology to the community based on their own interests.

1.2 The approach used to collect inputs

The approach used by Task 15.1 to investigate which high-level training courses and open-source solutions match the specific needs of these ENVRIplus users' communities are as follows:

- Collection of training requirements: We have set up training questionnaires and conducted face-to-face short interviews to gather information about existing data analytic tools that are currently adopted within the ENVRIplus RIs communities, about training services already available for them, and about the need for training courses on data processing/analysis with e-Infrastructure. The responses helped the authors understand the level of maturity of the different communities with respect to data production and processing, and their needs for future training.
- Analysis of training requirements: All the responses and inputs collected through these questionnaires have been used to evaluate the level of maturity of the different RIs, understand their requirements in terms of training needs and identify priorities for delivering relevant training courses in ENVRIplus.

1.3 STRUCTURE OF THIS DOCUMENT

The document is organized as follows:

In Section 3, we report the approaches to collect training requirements and to identify the priorities. In Section 4 we provide a summary of the 7 advanced training and webinar events that were conducted already during the project (01/05/2015 – 30/09/2016). In Section 5 we provide a training plan for the next months.

2 COLLECTION AND ANALYSIS OF THE ENVRIplus TRAINING REQUIREMENTS

2.1 COLLECTION OF COMMUNITY TRAINING REQUIREMENTS

During the 2nd ENVRIplus week, EGI Foundation, member of WP15, invited participants to fill in a questionnaire[4]. The questionnaire was setup to gather information from ENVRIplus communities about existing data analytic tools, related training services and relevance of possible training courses on data processing/analysis with e-infrastructures. This questionnaire completes the view that EGI started to build with D15.1 on existing training activities, and on training requirements.

The questionnaire was organized in three different sections:

1. Information about the respondent and related Research Infrastructure (RI). Personal information will be stored and used by EGI Foundation for follow-up activities in the context of ENVRIplus WP15 work.
2. Information about software tools that are currently adopted within ENVRIplus RIs communities and information about training services already available about these software.
3. RI priorities on training topics that would help them establish relevant data processing and analysis environments on e-infrastructures.

Representatives from Euro-Argo, AnaEE, MBA-DASSH, Seadatanet, EISCAT_3D and several organizations from ICOS filled the questionnaire and provided inputs. With the aim to produce as much as possible accurate information, EGI foundation complemented the survey responses with information available from RI web sites, with training-related content from D5.1. Table 1 provides a summary of information sources that were used across the RIs for the intelligence gathering.

TABLE 1. INFORMATION SOURCES USED FOR THE INFORMATION GATHERING

RI name	Status of RI[5]	Responded to WP15 training survey?	Training-related information in D5.1	Training-related information on RI website (during June 2016)	Interviewed
ACTRIS	Entry		YES	YES	
AnaEE	Preparatory	YES	YES		
EISCAT-3D	Construction	YES	YES	YES	
ELIXIR	Operational				YES[6]
EMBRC	Constructional, Operational		YES		
EMSO	Operational, ERIC	YES	YES		
EPOS	Implementation				YES[7]

[4] Online https://www.surveymonkey.com/r/2G9RWYZ . Questions are also in the Appendix.
[5] According to ESFRI roadmap 2016: http://www.esfri.eu/roadmap-2016
[6] In ELIXIR Competence Centre of EGI
[7] In EPOS EGI Competence Centre

RI name	Status of RI[8]	Responded to WP15 training survey?	Training-related information in D5.1	Training-related information on RI website (during June 2016)	Interviewed
Euro-ARGO	Operational, ERIC	YES	YES	YES	
EuroGOOS	Operational		YES	YES	
FixO3	Implementation		YES	YES	
IAGOS	Operational		YES	YES	
ICOS	Operational, ERIC	YES	YES	YES	
INTERACT	Operational		YES	YES	
IS-ENES2	Integrated		YES	YES	
LTER	Operational		YES	YES	
SeaDataNet	Operational. Not yet an RI, but a project. Next priority is setup of ERIC.	YES	YES	YES	
SIOS	Interim		YES	YES	

The analysis of these inputs helped to better understand the ENVRIplus RIs training requirements and identify the following training priorities. The key findings about the RIs:

- **ACTRIS[9] (Aerosols, Clouds, and Trace gases Research Infrastructure)** addresses the scope of integrating state-of-the-art European ground-based stations for long-term observations of aerosols, clouds and short-lived gases.
  - ACTRIS data are available from the ACTRIS Data Portal[10] which provides access to the ACTRIS Data Base (different type of data). Three type of data repositories are available: near surface data (EBAS), aerosol profiles (EARLINET) and cloud profiles (CLOUDNET). Data can be accessed for free but for some data a registration is needed to know and understand who is using it. The ACTRIS community is looking for *best practices* and *solutions to discover data* (e.g. using EUDAT high-level tools), and improve their interoperability with other RIs. ACTRIS already started to work with other projects such as: ICOS, IAGOS and AeroCom (project outside EU). There is no a training coordinator in this RI and there are no training events scheduled in 2016.

- **AnaEE[11] (Analysis and Experimentation on Ecosystem)** is a RI that focuses on providing innovative and integrated experimental services for ecosystem research.
  - AnaEE is in its preparatory phase. There are no specific technological choices in play at a European level yet; however the emphasis is on achieving interoperability between data centres and institutions via a semantics-driven, web services approach, rather than enforcing specific software choices. AnaEE uses different software tools such as: R,

[8] According to ESFRI roadmap 2016: http://www.esfri.eu/roadmap-2016
[9] http://www.actris.net
[10] http://actris.nilu.no
[11] http://www.anaee.com/

Matlab, GIS (Qgis, Arcgis), Modelling platforms for ecosystems (RECORD, Vsoil, Coup Model, Lpj and Capsis). For the modelling platforms for ecosystems the RI organizes every year tutorial and training session. For the other software tools there are different level of support. Currently users access dedicated clusters offered by institutions to run this models. At this stage the RI is not mature enough to answer the questionnaire precisely. The RI may be interested in integrating applications into VRE (e.g. using WS-PGRADE, or use gCube), trainings for IT Service management (e.g. FitSM training) and trainings for computer system operators. There is no training coordinator contact for this RI and there are no training events scheduled in 2016.

- **EISCAT_3D**[12] is a RI that uses incoherent scatter radar to study the Earth's ionosphere, contributing to geospatial environmental research. Access to raw data is restricted to researchers operating in any of the countries that contribute to EISCAT_3D. The goal of this project is to facilitate a smooth and swift transition of EISCAT_3D from the Preparatory Phase to its implementation. EISCAT_3D uses DIRAC to access data and provides analysis software, based on Matlab, for reducing raw data into physical parameters. Software for visualization of low-level data (VIZU) is also available. The most popular software adopted within the RI is GUISDAP, a program package for Matlab.
  - The biggest problem for EISCAT_3D are: search and storage its data, AAI and the identification and citation of its datasets. To address these problems EISCAT_3D is interested in using e-Infrastructures for high-throughput, high-performance and cloud computing, GPGPU and access e-Infrastructures such as: EGI, EUDAT, PRACE and OpenAire. In particular, the project is collaborating with EUDAT to resolve the citation problems, while with EGI the project is working on the development of a portal tailored to serve the users' community needs[13]. EISCAT_3D usually runs regular summer schools and symposium on the use of their radar systems. The last training event has been organized in July[14]. No training contacts are available.

- **ELIXIR**[15] is a European infrastructure for biological information that unities Europe's leading life-science organizations in managing and safeguarding the massive amounts of data being generated every day by publicly funded research. This RI would like to improve the access to biological data and establish a closer collaboration with other RIs.
  - The ELIXIR's training strategy aims to facilitate the accessibility to Europe's bioinformatics resources, tools, data and compute services provided by ELXIR. The Training Programme is delivered in partnership with GOBLET[16]. To better coordinate the training activities ELIXIR has established a Training Coordinator's Group (TrCG) made up of training experts from each ELIXIR node. The TrCG meets regularly to coordinate national training activities and plan the activities to be implemented. There are several events about data analysis scheduled in the next months[17]. ELIXIR is working with EGI and EUDAT in the EGI-Engage and EUDAT2020 projects on building the 'ELIXIR Compute Platform'[18] a cloud-based storage and compute infrastructure to serve ELIXIR nodes.

- **EMBRC (European Marine Biological Resource Centre)**[19] is an RI that aims to become the major RI for marine biological research. It is now in its implementation phase and operations is planned to start in 2016-2017. The main purpose of EMBRC is to promote marine

[12] https://eiscat3d.se/
[13] EISCAT_3D Competence Centre in EGI: https://wiki.egi.eu/wiki/CC-EISCAT_3D
[14] https://eiscat3d.se/content/joint-eiscatnsf-incoherent-scatter-radar-school-2016
[15] https://www.elixir-europe.org/
[16] http://mygoblet.org/
[17] https://www.elixir-europe.org/events
[18] ELIXIR Competence Centre in EGI: https://wiki.egi.eu/wiki/CC-ELIXIR
[19] http://www.embrc.eu/

biological science by providing the facilities (labs), expertise and biological resources needed for carrying out biological research.

- EMBRC is interested in establish collaborations with the environment community, explore new workflows which make use of marine biological and ecological data, develop a new standards which can be used for other datasets and interact with other RIs. EMBRC will develop a training platform to foster the training of both staff and users of the infrastructure (it is centric around equipment, platforms, organisms and research techniques not on the e-Infrastructure). EMBRC will provide high quality facilities and services for marine biology training and education in Europe. EMBRC started in 2013 to develop the European Marine Training Portal[20] which is used to organize educational programs and courses related to marine sciences.
- The RI is planning to set up an e-Infrastructure to make easier the movement of data between labs. In Sept. they are organizing the workshop: "*Accessing the Sea & its Biodiversity for Science: What role for European Research Infrastructures?*" [21].

- **EMSO (the European multidisciplinary seafloor & water column observatory)**[22] is a RI in the field of environment sciences. Its main objective is to harmonize data curation and access, and improve the search of this datasets. The main objectives of this RI are to create a single infrastructure for a wide audience of scientific users to improve the study of the ocean environment; design and develop the **Data Management Platform (DMP)** to collect data coming different sensors (e.g. One Data Platform). EMSO is also investigating collaborations with EGI to use the distributed computing resources for implementing the DMP (e.g.: Hadoop running at RECAS-BARI cloud provider).

- **EPOS**[23] is a long-term plan for the integration of RIs for Solid Earth Science in Europe. Its main goal is to integrate the diverse European e-Infrastructures for Solid Earth Science and build new opportunity to monitor and understand the dynamic and complex solid-Earth System. EPOS needs advice to improve the interoperable AAI system and is interested in training courses on Open-Science and Open-Data.

- **EURO-ARGO**[24] aims at developing a capacity to procure, deploy and monitor 250 floats per year and ensure that all the data can be processed and delivered to users. Euro-Argo is interested to use cloud-based technologies to implement and deliver data subscription services. Gibor Obolenski[25] is the training contact for this RI. Euro-Argo uses different software tools for data processing and analysis such as: Matlab, Python and Java. Euro-Argo has also a web site dedicated to outreach and education[26]. Based on the response collected this RI is interested in running container-based applications in the cloud using Docker containers, access different e-Infrastructures such as: EGI and EUDAT and develop VRE (e.g.: WS-PGRADE or use gCube).

- **EuroGOOS (European Global Ocean Observing System)**[27] is an international not-for-profit organization. It promotes the use of oceanographic information and develops standards. This community does not use e-Infrastructure technology (maybe in the future) and for this reason are interested in courses on e-Infrastructure technology.

[20] http://www.marinetraining.eu/
[21] http://www.embrc.eu/postponed-accessing-sea-its-biodiversity-science-what-role-european-research-infrastructures
[22] http://www.emso-eu.org/
[23] https://www.epos-ip.org/
[24] http://www.euro-argo.eu/
[25] http://www.euro-argo.eu/News-Meetings/News/Freshly-hired-Project-Scientist-at-Euro-Argo!
[26] http://www.euroargo-edu.org/
[27] http://eurogoos.eu/

- **FixO3 (Fixed Open Ocean Observatory network)** is a research project that integrates oceanographic data gathered from a number of ocean observatories and provide open access to that data for academic researchers. FixO3 needs mechanisms for ensuring harmonization of datasets and improve the search of scientific data. For this issue, they are evaluating EUDAT high-level services.
  - Events: Training on acquisition, validation, quality control and access to biodiversity data. (June 2016)[28]

- **The In-service Aircraft for a Global Observing System (IAGOS)**[29] is RI which implements and operates a global observation system for atmospheric composition. IAGOS needs mechanisms to improve the discoverability of scientific data, metadata standardization, citation and DOI management. It expects services for citation, curation and provenance. Datasets in netCDF format can be shared. They prefer to use open-source software. They don't have an e-Infrastructure and they do not have a specific training plan.

- The **Integrated Carbon Observation system (ICOS)**[30] RI provides the long-term observations required to understand the present state and predict the future behaviour of the global carbon cycle and greenhouse gas emissions and concentrations.
  - The objectives of ICOS are to provide access to dataset to facilitate the analysis of gas emissions. ICOS needs tools and services in the fields of metadata curation, DOI and citation and provenance. At the moment ICOS does not have a common training plan as such. The Carbon Portal organizes occasional training events, e.g. on Alfresco DMS (the Document Management System used by ICOS RI). Representatives of ICOS have participated in training events organized by EUDAT, e.g. on PID usage and data storage technology. They are interested in using EUDAT and EGI services and in training courses that introduce state-of-the-art e-Infrastructure technologies. ICOS is also interested in developing web-services for discovering and accessing ICOS data products.

- The overall goal of **INTERACT** is to build capacity for identifying, understanding, predicting and responding to diverse environmental changes. The software and computational environments are located at the University of Uppsala. NORDGIS[31] is a geographic metadata information system at the moment holding information for nine sites stations. INTERACT is interested to collaborate with other Infrastructures and in training on e-Infrastructure technologies.

- **IS-ENES2** runs a federated data infrastructure based on few main data centres. Data is generated by climate modeling groups and post-processed according to the standards and agreements of the inter-comparison project (e.g. CMIP, CORDEX). As a next step, data is ingested at IS-ENES/ESGF data nodes, quality-controlled and then published to the IS-ENES/ESGF data infrastructure. Publication makes metadata available and searchable and data accessible via IS-ENES portals (as well as via APIs).
  - Data from IS-ENES is replicated to EUDAT for data curation purposes (long term archival). IS-ENES data is harvested by EUDAT metadata catalogue (B2Find). IS-ENES2 expects to obtain advices for data catalogues to compare their model data with other data. ESGF infrastructure is based on Grid/Cloud and HPC. From time to time they organize workshops and training courses. They are potentially interested in training courses that introduce state-of-the-art e-Infrastructure technology.

[28] http://www.fixo3.eu/events/training-on-acquisition-validation-quality-control-and-access-to-biodiversity-data-2/
[29] http://www.iagos.org/
[30] https://www.icos-ri.eu/
[31] http://www.nordgis.org/sites

- **Long-Term Ecosystem Research (LTER)** aims at providing information on ecosystem (functions and processes) of the whole eco-system via a discovery portal. This information is very diverse in its technical formats. The purpose of the RI is to focus on harmonized data products. The *metadata* on the available datasets (including time-series) are provided via the DEIMS platform centrally (based on Drupal). DEIMS[32] provides a central repository of metadata on: a) Research sites (representing the Location on different Levels for the data acquisition), b) data sets (including Service based data Provision), c) persons (including institutions and Networks). Metadata are free without restrictions.
  - For *data storage* LTER uses a wide range of solutions (file based: csv, netCDF, Excel; databases: Postgres, ORACLE, MySQL; GIS: spatial databases, shapefiles, grids etc.). Data are free in principal if collected in European scientific projects but local restrictions could apply.
  - Problems to resolve:
    - Migration of DEIMS portal to Drupal 7;
    - Integration of data repository into the workflow (e.g. B2SHARE);
    - Improve datasets harmonization.

- **SeaDataNet** implements and efficient distributed Marine Data Management Infrastructure to manage large datasets coming from observations of seas and oceans. SeaDataNet needs to improve expertise on observation networks and data management. SeaDataNet requires an AAI to manage users' authentication.
  - Serge Scory[33] is the training contact for this RI. SeaDataNet uses different software tools for data visualization and interpolation such as: Ocean Data View[34] and Diva[35]. Euro-Argo has also a web site dedicated to outreach and education[36]. Training courses organized by SeaDataNet are available here http://www.seadatanet.org/Events/Training-courses.
  - Some of these SW are already deployed at CINETA. A proposal has been submitted to collaborate with DKRZ and EUDAT. Interested to collaborate also with EGI. Based on the response collected this RI is interested in accessing different RIs (e.g. EUDAT, EGI, etc.) and integrate data, tools and applications on VRE.

- **SIOS (Svalbard Integrated Earth Observing System)** is an Earth Observation System built on existing infrastructures to better understand the on-going and future climate changes in the Arctic. Currently SIOS created a distributed data management system called SIOS Knowledge Centre.
  - A core element of SIOS is the **Knowledge Centre** (KC)[37] in Longyearbyen. The Centre will enable open access and interaction between observation, modelling and process research. It will also be a facilitator of strategic processes, a service point to user communities, a platform for data handling and utilization, and a facilitator and arena for training. The SIOS Knowledge Centre will initiate and coordinate education and training of scientists and research technicians in polar research methods through dedicated field and lab courses. SIOS will offer specialized courses related to the infrastructure being made available from SIOS partners in Svalbard.
  - Training activity will be centred on the following categories:

[32] http://data.lter-europe.net/DEIMS
[33] serge.scory@naturalsciences.be
[34] https://odv.awi.de/
[35] http://modb.oce.ulg.ac.be/mediawiki/index.php/DIVA
[36] http://www.euroargo-edu.org/
[37] http://www.sios-svalbard.org/servlet/Satellite?blobcol=urldata&blobheader=application/pdf&blobheadername1=Content-Disposition:&blobheadervalue1=+attachment;+filename="D8.3SIOSKnowledgeCentreImplementationPlan–functionsandservicesdefinition-conceptpaper.pdf"&blobkey=id&blobtable=MungoBlobs&blobwhere=1274505790460&ssbinary=true

- Interdisciplinary field courses and training activities.
- Interdisciplinary courses and training activity linking observations and modelling.
- Interdisciplinary collaborative courses and training activities utilizing infrastructure and data management systems.
- Courses and training activities on how to use satellite data and their combination with in-situ data.

- The SIOS Training Programme aims at providing the research and other user communities with the skills and opportunities to make best use of the infrastructure. It will provide SIOS with specific training to user communities enabling them to perform state-of-the-art research using SIOS services.

2.2 ANALYSIS OF TRAINING PRIORITIES

The following topics emerged as high priority items from the survey:

- Creating data federations;
- Hosting data-intensive services in EGI;
- E-infrastructure services for the long-tail of science;
- Security incident handling, methods and forensics.

2.2.1 EGI's response to training needs

To define priorities for e-Infrastructure training, we matched the RI training needs (based on Section 3.1) with the EGI e-Infrastructure training topics (based on EGI service catalogue from Section 2.1). The training priorities reported in this table are based on the analysis of requirements received from the different ENVRIplus RIs as described in the previous sections:

TABLE 2. MATCHED RI TRAINING NEEDS WITH EGI E-INFRASTRUCTURE TOPIC

E-infrastructure training topic	Description, Target audience	Relevant for	Priority for development (from ENVRIplus point of view)	Way of delivery
EGI Overview (focus on HTC and cloud)	Introduction with EGI with focus on its the 'oldest' service: High-throughput computing and storage in the grid.	ACTRIS, EISCAT_3D, EMBRC, EMSO, Euro-Argo, EuroGOOS, IAGOS (low), ICOS, INTERACT, IS-ENES2, SeaDataNet, SIOS	Content already exists	Can be delivered as Webinar or F2F (short, max half day)
Introduction to EGI Federated Cloud	Introduction to the federated cloud system, with its standard-based and OpenStack-based realms.	ACTRIS, EISCAT_3D, EMBRC, EMSO, Euro-Argo, EuroGOOS, IAGOS (low), ICOS, INTERACT, IS-ENES2, SeaDataNet, SIOS	Content already exists	Can be delivered as Webinar or F2F (at least 1.5 hour, max half day)
Joint usage/interoperability of multiple European e-Infrastructures (e.g. EUDAT, EGI, PRACE, OpenAire)	Positioning the major e-infrastructures of Europe, and guiding communities about the joint usage of relevant services.	EISCAT, ACTRIS, EISCAT-3D, EPOS, Euro-Argo, FixO3, IAGOS, ICOS, LTER, SeaDataNet	Medium	
Integrating applications, data and online tools into community portals (Virtual Research Environments)	Introducing the concept of Virtual Research Environments, and guiding users on the best approaches and technical frameworks to setting up such VREs.	AnaEE, Euro-Argo, ICOS, SeaDataNet, EISCAT_3D (DIRAC)	Low/Medium	
Training for computer system operators	Introduction to IT server management.	Euro-Argo, AnaEE, SeaDataNet, EISCAT-3D	Low	
IT service Management	IT service management, based on the lightweight FitSM practice.	AnaEE , Euro-Argo	Low	

E-infrastructure training topic	Description, Target audience	Relevant for	Priority for development (from ENVRIplus point of view)	Way of delivery
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)	SIOS (webinar), EMSO (Hadoop)	Medium	• Online • Version for f2f delivery will be prepared only if there is request (e.g. by specific RIs)
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.	Euro-Argo (high), SeaDataNet	Medium (requires technology investigation)	• Online • Version for f2f delivery will be prepared only if there is request (e.g. by specific RIs)
Creating data federations with the EGI Open Data Platform	The course will provide guidance for data providers on how to create data federations, with harmonized views, using geographically dispersed datasets. Such federations can simplify access and (re)use of data by distributed user communities.	EMSO, ICOS, EMSO, SeaDataNet	High	Online
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 and use virtualized applications on federated datasets that are sitting together with cloud compute services on EGI resources.		High (depends on data federation course)	Online
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	EISCAT_3D (high), Euro-Argo (low),	Medium	Online

E-infrastructure training topic	Description, Target audience	Relevant for	Priority for development (from ENVRIplus point of view)	Way of delivery
Connecting scientific services with EGI resources using the EGI CheckIn service	The course would target service operators in structured scientific communities and train them on how to integrate community-specific services with the authentication and authorization 'layer' of EGI through the new proxy service.		Medium	• Online • Create version for F2F delivery in case of interest.
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.
Platforms in the EGI Federated Cloud	Set of modules, each specialized 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)
e-Infrastructure user security awareness training	This module will act as an introduction for end users of e-Infrastructures. 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
Security incident handling, methods and forensics	This development will build on the earlier material produced in EGI-InSPIRE and EGI-Engage PY1. Some components are technical 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 virtualized services in the EGI Federated Cloud, managers of Science gateways and portals.	AnaEE, EISCAT_3D (low), Euro-Argo (low), SeaDataNet (low)	High	Delivered in security training sessions at conferences and/or specialed security workshops related to such events. For example DI4R 2016, ISGC2017

E-infrastructure training topic	Description, Target audience	Relevant for	Priority for development (from ENVRIplus point of view)	Way of delivery
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	AnaEE, EISCAT_3D, Euro-Argo (low), SeaDataNet (low)	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.
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	AnaEE, EISCAT_3D, Euro-Argo (low), SeaDataNet (low)	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.

2.2.2 External initiatives with relevant activities

From a technical point of view, there are several networks/initiatives that contribute to exchange best practices, training contents for supporting the European researcher during his/her day-by-day work. These initiatives are extremely important to engage with new researchers and optimize the use of European RIs which usually play a key role to promote cutting-edge research, innovation and technology transfer. In this sub-section we want to highlight some external projects/initiatives that perform/plan activities that are relevant to respond to the RI's training needs.

The first project which is worth to mention, for its effort to support cutting-edge research and favor the uptake of Open Access vision, is FOSTER (Facilitate Open Science Training for European Research)[38], an EU-funded project which aimed to produce an European-wide training programme to help researchers, postgraduate students, librarians and other stakeholders to incorporate Open Access approaches into their existing research methodologies. The project pursued these objectives through the identification of already existing contents that can be reused in the context of the training activities, the creation of a specific portal to support e-Learning courses, the dissemination of training materials, and the delivery of face-to-face training meeting. The activities carried out by the project are relevant for supporting the development and the uptake of Open Science paradigm, which is also what EGI is trying to

[38] https://www.fosteropenscience.eu

address, and perfectly in line with what the High Level Expert Group on Scientific Data submitted its final report to the Commission in late 2010. The main conclusion of the report is that there is a need for a "collaborative data infrastructure" for science in Europe and globally.

The key role of RIs, and the importance to create new professional skills to help modern researchers to cope with challenges coming from the "Data Intensive Science", is recognized not only by EGI, but also by several projects/initiatives such as: RITrain (Research Infrastructure Training Programme) [39] and CORBEL (Coordinated Research Infrastructures Building Enduring Life-science Services)[40]. Both projects are involved to develop flagship training programmes to help future leaders/operators of RIs to better understand the RI landscape in Europe which is complex and rapidly growing. Beside to these initiatives, we can also mention the activities carried out within the Education for Data Intensive Science to Open New science frontiers (EDISON)[41] project which focus on the creation of the data scientist profession, the training modules to target the professionalization of data supporters developed by Data Archiving and Networking Service (DANS)[42] and the UK Data Service collections[43] which cover a range of topics relating to data reuse and management. We can also cite the Digital Curation Centre (DCC)[44] , a specialized center which provides tailored support for RIs and offers training modules to help researchers and data custodians to manage and share data effectively, and the ESIP Federation[45] which aims to share the community's knowledge with scientists interested to become data managers.

We are interested in contacting with these projects/initiatives, learning from their experiences, and checking out relevant courses suitable for ENVRIplus community.

3 E-INFRASTRUCTURE TRAINING IN ENVRIPLUS

Having understood the training requirements of ENVRIplus research community, members of the e-infrastructure training task have prepared and organised a number of training events. This section provides a summary of 7 advanced knowledge transfer events organized and held during the project lifetime (18/11/2015 – 07/06/2016).

3.1 TRAINING EVENTS SO FAR

3.1.1 The 1st ENVRI week

The 1st ENVRI week was held in Prague – Czech Republic on the 18 November 2015. The aim of this event was to introduce EGI and the Federated Cloud service to the ENVRIplus RIs. Tutors from Hungary (EGI.eu) and the Czech Republic (CESNET) delivered the event. The event consisted of 2x 90 minutes long sessions over 2 days. During the first session, a general overview of the e-Infrastructure solutions from EGI was provided along with some real community examples which have already profited from EGI solutions. The second session was devoted to the technical details of the EGI Federated Cloud, with further examples of how communities are using it, or working on the integration of it into their own RI setups.

[39] http://ritrain.eu/
[40] http://www.corbel-project.eu
[41] http://edison-project.eu/
[42] https://dans.knaw.nl
[43] https://www.ukdataservice.ac.uk/use-data/advice
[44] http://www.dcc.ac.uk
[45] http://commons.esipfed.org

The programme of the event is available at https://documents.egi.eu/document/2650 and includes all the presentations. The event was widely announced on the ENVRIplus and EGI official channels.

3.1.2 A training workshop in the 2nd ENVRI week

During the 2nd ENVRI week, a training workshop was organized in Zandvoort – The Netherlands (05-13 May 2016). The overall goal of this interactive workshop was to provide a first line of support, know-how and expertise, to analyse the ENVRIplus use cases requirements for using e-Infrastructures, provide deployment solutions and work plans for supporting the use cases.

The half-day workshop was intended for:

- ENVRIplus use case providers.
- e-Infrastructure providers.
- Members of the agile groups involved in the implementation of the ENVRI+ use cases.
- Research Infrastructures interested to get familiar with e-Infrastructures technologies to address emerging research challenges.

The programme of the event is available at https://indico.egi.eu/indico/event/2966/ and includes all the presentations. The event was widely announced on the ENVRIplus and EGI officially channels. During the event two different use cases have been discussed.

- Thierry Carval from IFREMER presented the Euro-Argo use case. The project wants to use the EGI Federated Cloud Infrastructure resources to implement a data subscription of EuroArgo datasets. For this use case EGI Foundation provided support to use Hadoop technology to search for datasets based on user's criteria (e.g., time, spatial, update period of delivery, etc.).
- Abraham Nieva from the Cardiff University introduced the MBA-DASSH use case. DASSH supports the publishing and discovery of data about marine species and habitats. The main issue is try to integrate this system with other RIs and communities and make it interoperable. For this use case EUDAT services have been presented as possible solution to support the integration problem.

On the average, the workshop was attended by 12 participants (about the 23% were women) plus one remotely (Bartosz Kryza from Cyfronet).

3.1.3 The Webinar on the use of Hadoop technology

The webinar on the use of Hadoop technology on the EGI Federated Cloud Infrastructure was held online, in the virtual room of a Adobe Connect web-meeting platform, on the 17 March 2016. The webinar was organised in two 30-min lectures presented by Tamas Kiss from the University of Westminster – UK and Carlos Blanco from the University of Cantabria – Spain.

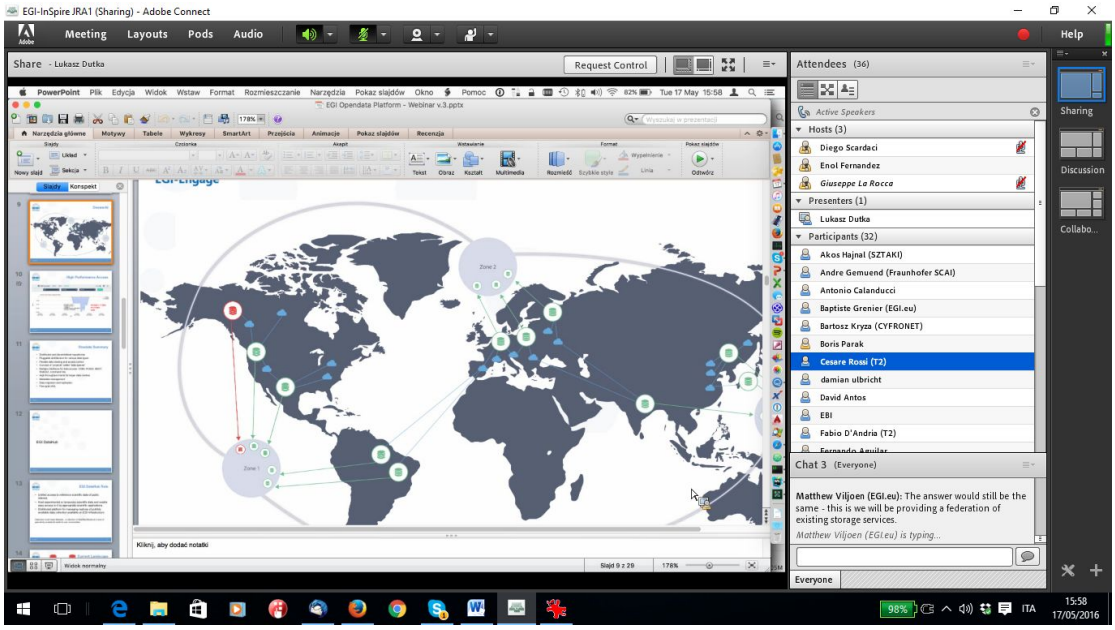
The aim of this webinar was to present an overview about the Hadoop technology, provide guidelines on how to use cloud computing resources to setup an Hadoop cluster and integrate Hadoop with workflows and the WS-PGRADE portal. The programme of the webinar is available at https://indico.egi.eu/indico/event/2931/ together with all the slides presented and the recording. On the average, about participants attended the webinar. The webinar was jointly organized by the EGI-Engage project.

3.1.4 The Webinar on the EGI Open Data Platform (ODP)

The webinar on the EGI Open Data Platform – Towards Scientific Data Hubs was organized online on the 17 May 2016. The goal of this webinar was to present to all research communities,

technology providers and RIs the status of the new data platform that EGI is developing in the context of the EGI-Engage project. Lukasz Dutka, from Cyfronet, presented the platform to store and discover research data, publish with open or controlled access, access and reuse data with the EGI computing services. Based on the training requirements collected by EGI Foundation, this platform can help to resolve some of the technological problems raised by several ENVRIplus RIs.

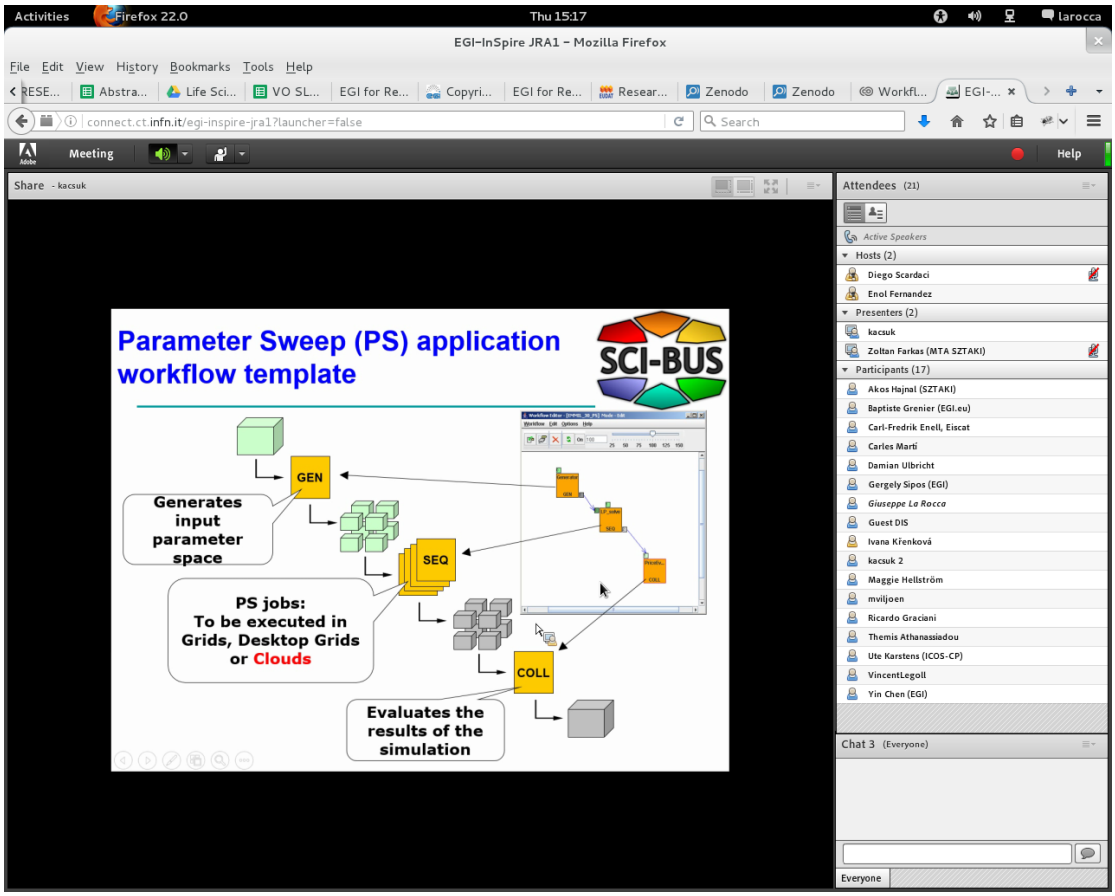
The programme of the webinar is available at https://indico.egi.eu/indico/event/2969/ together with all the slides presented and the recording. The webinar was jointly organized by the EGI-Engage project. The webinar was attended by 36 participants (20% were women). The following figure shows a screenshot taken during the webinar:



3.1.5 The Webinar on workflow applications on EGI with WS-PGRADE

The webinar on how to create workflow applications on the EGI e-Infrastructure with the WS-PGRADE portal was organized on the 26 May 2016. Its main goal was to provide to technology providers, researchers communities and application developers with the necessary common knowledge to create scientific workflows on the EGI e-Infrastructure using the high-level functionalities exposed to end-users by the WS-PGRADE portal.

The programme of the webinar is available at https://indico.egi.eu/indico/event/2977/ together with all the slides presented and the recording. The webinar, organized by the EGI-Engage project, was tutored by prof. Pèter Kacsuk and Zoltàn Farkas from the MTA SZTAKI LPDS. The webinar was organized in two parts: the first was a general overview and introduction of the WS-PGRADE portal, while the second was a live demo about how to run a general-purpose application on the cloud-based resources of the EGI Federated Cloud Infrastructure. The webinar was attended by 19 participants (26% were women). The following figure shows a screenshot taken during the webinar:

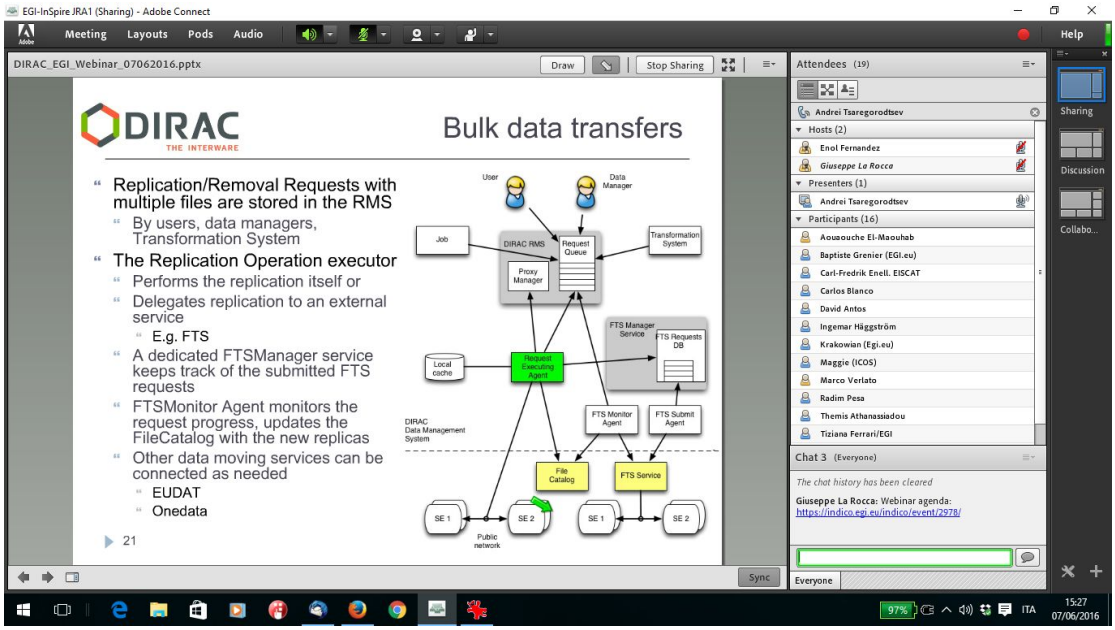


The high-level functionalities exposed by the WS-PGRADE portal matches the requirements and the needs of many ENVRIplus RIs.

3.1.6 DIRAC system

The webinar on how to use DIRAC to create workflow applications for EGI communities was organized online on the 07 June 2016. The webinar chaired by Andrei Tsaregorodtsev from CPPM-IN2P3-CNRS, Marseille, aimed to introduce DIRAC project, a general purpose distributed computing framework which provides all the necessary components to build ad-hoc grid- and cloud-based infrastructures interconnecting computing resources of different types, allowing interoperability and simplifying interfaces. During the first part of the webinar, the DIRAC computing infrastructure, as well as the DIRAC File Catalogue (DFC), the central component of the DIRAC Data Management system, were introduced. The DFC allows to define a single logical name space for all the data managed by DIRAC. This is an important feature which is already adopted by some RIs such as EISCAT_3D. The overview about the framework was also accompanied by a live demo where Andrei Tsaregorodtsev demonstrated how using EGI computing resources and storage facilities using CLI and web portal.

The programme of the webinar is available at https://indico.egi.eu/indico/event/2978/ together with all the slides presented and the recording. Demo materials to try out off-line can be found at https://github.com/DIRACGrid/DIRAC/wiki/Quick-DIRAC-Tutorial. The webinar was attended by 19 participants (31% were women). The following figure shows a screenshot taken during the webinar:



3.2 TRAINING MATERIALS AND USER GUIDES

Various resources are available online the most relevant EGI e-infrastructure technologies. These can be used by ENVRIplus RIs:

- User guide with examples about the EGI Federated Cloud: https://wiki.egi.eu/wiki/Federated_Cloud_user_support
- Container compute: https://wiki.egi.eu/wiki/Federated_Cloud_Containers
- HTC & Online storage (with gLite): https://agenda.ct.infn.it/event/232/
- Open Data Platform: http://www.digitalinfrastructures.eu/content/egi-datahub-and-open-data-platform-0

4 FUTURE PLANS FOR TRAINING

EGI.eu, member of the ENVRIplus project consortium suggests and will work on very specific activities to meet the identified needs of the ENVRIplus RIs. Based on the identified priorities this section provides plan for the development and delivery of relevant training services for ENVRIplus communities:

1. Creating data federations with the EGI Open Data Platform: the module will provide guidance for data providers on how to create data federations, with harmonised views, using geographically dispersed datasets. Such federations can simplify the integration, access and (re)use of data by distributed user communities. The initial version of the training course was piloted for[46] the Digital Infrastructure for Research (DI4R) conference (28-30 September 2016, Krakow). Based on the experiences of this pilot course a more mature version will be created and delivered at the next ENVRIplus Week in Prage in November 2016.
2. Data-intensive applications in the EGI cloud: A new training module is needed to support application developers integrating applications with the EGI cloud through APIs. the module would provide guidance for application developers (PaaS, VRE, Science Gateways) who want to integrate scientific application with the EGI Federated Cloud Infrastructure. The focus will be on the use of standard-based clouds APIs (OCCI) for service discovery, compute and storage management. The initial version of the training

[46] http://www.digitalinfrastructures.eu/content/egi-datahub-and-open-data-platform-0

course was delivered[47] during the Digital Infrastructure for Research (DI4R) conference (28-30 September 2016, Krakow). An online guide will be prepared by the end of 2016 based on this material, and will be promoted to ENVRIplus communities. If requested, a f2f training can be held based on this content at future ENVRIplus Weeks, or RI-specific conferences/symposiums.

3. The 'EGI easy access platform' for the long-tail of science is under finalisation and will be opened in 'beta' mode later in 2016. The service will provide a cloud and grid resource pool for long-tail researchers, and easy-to-use interfaces to sign-up for usage and to conduct data-intensive simulations based on pre-defined applications (e.g. R), or by porting own code/software. The platform will include user guides that will demonstrate usage for newcomers.
4. Security incident handling, methods and forensics: this will be a training module about security policies and procedures to handle security incidents. The target audience for this training course includes IT service administrators, managers/operators of virtualised services in the ENVRIplus RIs, managers of Science Gateways and portals. The initial version of the training course was delivered[48] during the Digital Infrastructure for Research (DI4R) conference (28-30 September 2016, Krakow). Discussions recently started how to turn this into a 'Security training service' that EGI could deliver for RIs.

Besides the above high-priority topics EGI will also work on the following training modules, mostly from effort in the EGI-Engage project:

5. GPGPU-computing in the EGI e-Infrastructure: the module will provides guidance for application developers to integrate scientific applications with GPGPU resources of EGI.
6. Containers based applications in the EGI Federated Cloud infrastructure with Docker: A new online guide was developed recently[49], providing guidelines for application developers on how to create complex container-based applications, e.g. with Docker and Docker Swarm.

However, there are initiatives and projects outside of ENVRIplus, with relevant activities for e-infrastructure training. Partnership and collaboration will be sought for during the remaining time of ENVRIplus with them, and content/services will be brought onboard to ENVRIplus. This activity will run at the level of the WP15 management, given that scope of these external contributions is broader than e-infrastructure training.

5 CONCLUSIONS

The activities carried out by Task 15.1 second part (led by EGI Foundation) during the project lifetime are in line with what stated in the DoA. All the knowledge dissemination events have been focused on the development of training events and materials to address the needs of the main stakeholders in the Environmental RI community. These events have also been organized in collaboration with other projects (e.g., EGI-Engage), and other e-Infrastructure technology providers (e.g., EUDAT, OneData, gCube), in order to exploit synergies and maximize impact.

The topics covered in these training events include instructions to:

- Develop and deliver of a face-to-face event to demonstrate how to access and use cloud-based resources of the EGI Federated Cloud Infrastructure. This training event

[47] http://www.digitalinfrastructures.eu/content/egi-federated-cloud-developers
[48] http://www.digitalinfrastructures.eu/content/federated-aai-meets-reality-security-incident-handling-role-play-di4r
[49] https://wiki.egi.eu/wiki/Federated_Cloud_Containers

was intended for IT operators of RIs and those who need to build and operate IT infrastructures to support environmental sciences data, applications and scientific data.
- Integrate scientific applications into general purpose distributed environment and web-based portals.
- Develop and deliver a webinar to introduce the Open Data Platform, the new solution to store and discover research data, publish with open or controlled access, access and reuse data with the EGI computing services.

In total, more than 86 unique people attended the courses, with a good percentage of women (about 25% on average).

5.1 IMPACT ON PROJECT

This deliverable continued the investigations that were started in the project about RIs' readiness and preferences for adopting harmonized approaches. The work strongly built on WP5, and particularly D5.1 and deepened the consortium's understanding on training priorities, using surveys and face-to-face interviews.

5.2 IMPACT ON STAKEHOLDERS

The deliverable identified priorities for e-infrastructure training topics to facilitate the harmonized adoption of e-infrastructure services for data-intensive science within the ENVRIplus RIs. The document defines a work-plan for developing training modules and deliver training events that meet the high-priority needs. The expected impact of this work will be stronger uptake of e-infrastructure services within those RIs that are ready for large-scale applications of IT resources.

6 APPENDIX – E-infrastructure training survey

EGI.EU training session – e-Infrastructure survey

ENVRIplus week, Prague // November, 2015

Name:

……

Affiliated Research Infrastructure (RI) and your role:

……

Email:

……

Question	Low	Medium	High	Comment (explain your answer)
What is the priority for your RI to engage with e-infrastructures?	☐	☐	☐	
How relevant do you see EGI overall for your RI?	☐	☐	☐	
What is the level of relevance of EGI's 'High throughput data analysis' solution to your RI?	☐	☐	☐	
What is the level of relevance of EGI's 'Federated Cloud' solution to your RI?	☐	☐	☐	
What is the level of relevance of EGI's 'Federated Operations' solution to your RI?	☐	☐	☐	
What is the level of relevance of EGI's 'Community-driven Innovation' solution to your RI?	☐	☐	☐	
What did you find most useful in this training session?				

What did you find <u>least useful</u> in this training session?	
About which e-infrastructure topic(s) would you like to learn more about in the future?	

7 APPENDIX – 'Building e-infrastructure environments' training survey

EGI Technical Support for ENVRIplus Use Cases WS - survey

2nd ENVRIplus week, Zandvoort, The Netherlands // May, 2016

Part 1 – General Information

Name of the respondent ……

Your Organisation ……

E-mail Address (for possible follow-up ……

Training topics	Not Relevant	Somehow Relevant	Relevant	Very Relevant	Comment (explain your answer)

Which RI community are you representing ? ……

Which is your role within the represented community ?

……..

Part 2 – Training in Research Infrastructures

Does your RI has a training contact ? If yes, please provide name and contact information (for possible follow-up by WP15).

……..

Which are the most popular software tools in your RI community for data processing/analysis ?

……..

How do members of your community learn about these software tools ? (e.g. name relevant training courses, online learning sites, summer schools, etc.)

……..

Does your community develop training materials, organize training events about these tools ? If yes, please specify which training materials and training events are available

……..

Would your community be interested in extending these popular data processing tools with e-Infrastructure capabilities, and organizing training about them ? If yes, which software tools should we focus on ?

……..

Part 3 – Relevance of e-Infrastructure training topics

Training topics	Not Relevant	Somehow relevant	Relevant	Very Relevant	Comment
Using e-Infrastructures for high-throughput, high-performance and cloud computing					
Creating and running container based applications in the cloud (e.g. Docker)					
Using GPGPU from e-Infrastructures for scientific applications					
Joint usage of multiple European e-Infrastructures (e.g. EUDAT, EGI, PRACE, OpenAire)					
Integrating applications, data and online tools into community portals (Virtual Research Environments)					
Establishing community grids or community clouds from geographically dispersed computers					
Training for computer system operators (e.g. security, storage management, user support, etc.)					
IT service management (with topics such as capacity management, customer relationship management, incident management, service availability, etc.)					

In this section we ask you to priorities e-Infrastructure training topics in terms of their relevance to your RI community. Indicate your assessment by placing a cross in the appropriate box for each offered topic. The marks have the following meaning:

1. **Not relevant**: I think my community would not be interested in such training.
2. **Somehow relevant**: A few people in my community could be interested, but overall this is not a priority topic for us.
3. **Relevant**: My community will need such training, but we are not yet ready to receive courses on this topic.
4. **Very relevant**: My community needs such training and we are ready to receive i