WP6 (SA2) Knowledge Commons

Work package number	WP 6 (SA2)		**Lead Part.**	EGI.eu	**Start**	M1	**End**	M30
Work package title	Knowledge Commons							
Participant number	1	2	3	6	7	8		
Short name of participant	EGI.eu	OEAW	VLIZ	CESNET	FRAN HOFER	GWDG		
Person/months per participant:	30	3	6	23	4.5	9.5		
Participant number	9	10	11	12	13	14		
Short name of participant	CSIC	CSC	CNRS	INRA	GRNET	SRCE		
Person/months per participant:	88	16	15	7	8.5	18.5		
Participant number	15	16	17	18	21	22		
Short name of participant	MTA SZTAKI	INFN	CIRM MP	IRCCS FBF	SURFsara	CYFRONET		
Person/months per participant:	15.5	22.5	8	3	26.5	4.5		
Participant number	23	24	26	29	30	31		
Short name of participant	LIP	ICETA	SNIC	STFC	BBMRI ERIC	EMBL		
Person/months per participant:	18	3	15	4.5	13	5		
Participant number	33	36	37	38	39	40		
Short name of participant	EISCAT	GNUBI LA	UI (OSG)	AS	ASTI	ITB		
Person/months per participant:	14	4	5.5	48	2	4		
Participant number	41	42	43					
Short name of participant	KISTI	UPM	NSTD A					
Person/months per participant:	1	3	2					

Objectives

WP6 (SA2) coordinates the provisioning of services for scientific communities and supports the co-design, co-development and dissemination of new community-specific e-Infrastructure services. These services will support novice and semi-experienced user communities as well as members of the long-tail of sciences to become active and self-sufficient users of e-Infrastructure solutions. In addition, some of the services will support mature e-Infrastructure communities in taking their production experiments on distributed computing and data infrastructures to the next level. The work package includes dedicated tasks for eight Competence Centres (CCs) to support high-impact Research Infrastructures/communities by joint development of customised services for these based on core EGI capabilities, by user engagement and training. Two of the CCs (BBMRI and EISCAT_3D) are supported by both EGI and EUDAT within the Horizon2020 programme. The work package also contributes to fostering open science, and nurturing smaller or less structured research groups by providing training and direct technical support to them. Specific objectives of WP6 are: 1) Identify and support communities and users from EGI and its partners; 2) Facilitate the integration of scientific applications with EGI's e-Infrastructure services; 3) Co-design and co-develop services for sustainable, structured scientific communities; 4) Promote and support the uptake of new services within scientific communities; 5) Provide a training framework, foundational training services and domain specific training events for scientific communities, EGI members and partners.

Description of work (broken down into tasks), lead partner and role of participants

Task SA2.1 – Training (Lead: EGI.eu, M1 – M30)

Estimated task effort: 37PM

This task provides coordination to training activities across EGI, and a framework to develop, integrate and deliver training content and events 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. A plan of join training activities with EUDAT will be defined at PM12 and PM24. The foundational services to be provided by the task are: 1) Operating an up to date section in the EGI website about training events resources; 2) Develop generic training content for customised courses prepared by CCs, NGIs and scientific communities; 3) Contribute to the setup and delivery of tutorials at high-impact events organised by CCs, the NGIs, scientific communities and partner projects.

Task SA2.2 - Technical User Support (Lead: CSIC, M1 – M30)

Estimated task effort: 48PM

This task operates a distributed user support team across multiple NGIs to provide technical support to the integration of domain-specific applications with EGI platforms. The support will include, for example, integration of scientific code with EGI cloud and HTC platforms, workflow development, development of community gateways or resource allocation. The effort will be focused on new communities coming from scientific disciplines that do not have dedicated EGI Competence Centres and lack sufficient support within the NGIs. The task will result in community-specific systems underpinned by resources and generic e-Infrastructure services from the NGIs and new communities.

Task SA2.3 - ELIXIR (Lead: CSC, M6 – M24)

Estimated task effort: 36.5PM

ELIXIR is a pan-European RI supported by 17 European governments united in a vision to build a sustainable European infrastructure for biological information, supporting life science research and its translation to medicine, agriculture, bioindustries and society. Members of the ELIXIR CC will select key life science use case workflows between partners that are representative of many in the community and will have high impact for scientist end-users and use these as drivers to develop demonstrators that will assess the use of EGI cloud resources for tool and/or data services recognised within the ELIXIR community. The CC will also create best practices materials for life science service providers using the experience gained in those tasks and organise events to facilitate discussions between e-Infrastructure and life science communities. The goal is that by the end of the two-year effort ELIXIR service providers feel motivated to use Big Data tools coming from e-Infrastructures, and that e-Infrastructure providers have established processes to support particular life science data analyses. OSG, as an associate partner will organise collaborative discussions co-located with existing events to discuss life science drivers for virtual organisations and research groups.

Task SA2.4 - BBMRI (Lead: SURFsara, M1 – M24)

Estimated task effort: 29PM

Thousands of biobanks in Europe have been collecting data, samples and images of millions of individuals in different stages of their lives, during disease and after recovery. Biobanking is currently evolving from local repositories to a pan-European RI the BBMRI-ERIC. The BBMRI CC facilitates the implementation of big data storage in combination with data analysis and data federation by integrating technologies from community projects, EGI and other e-Infrastructures. The CC will capture requirements and provide technology demonstrators to:

- Increase biobank interoperability and data discovery in BBMRI community by providing a secure and standard way to share biobank high-throughput data.
- Provide biobanking community with a federated infrastructure for big data storage and intensive data analysis.
- Facilitate the efficient use of bio-resources by supporting visibility and sharing, while also respecting the protection level required by owners of the data and samples.
- Facilitate the efficient use of economic resources in BBMRI by providing a common infrastructure for storage and processing of big data.

OSG, as an associated partner will provide interoperability mechanisms for the virtual organisations of the BBMRI competence centre to use OSG resources, and will organise a joint information sharing workshop to discuss interoperability and new technologies that may affect international resource sharing.

Task SA2.5 - MoBrain (Lead: SURFsara/BCBR, M1 – M30)

Estimated task effort: 59.5PM

The MoBrain CC will lower barriers for scientists to access modern e-Science solutions to simulate life science processes from micro to macro-scales. The CC will build on grid- and cloud-based infrastructures and on the existing expertise available within WeNMR/INSTRUCT and NeuGrid4You. The CC will integrate molecular structural biology and medical imaging services, e.g. Scipion[21], and data and will kick-start the development of a larger, integrated, global science virtual research environment for life and brain scientists worldwide. The task will deliver a MoBrain integrated infrastructure based on a variety of e-Infrastructure solutions, including grid, cloud and accelerated computing systems accessible through cloud interfaces and a virtual research environment to serve translational research from molecule to brain.

Task SA2.6 – DARIAH (Lead partner: SRCE/RBI, M1 – M30)

Estimated task effort: 62.5PM

The DARIAH CC aims to widen the usage of the e-Infrastructures for Arts and Humanities research. The CC will develop and provide a workflow-based science gateway based on the generic-purpose WS-PGRADE and gLibrary technologies, adapted and tailored to the needs of users coming from the field of Arts and Humanities. The gateway will provide access and compute services for data residing in distributed grid and cloud storages. The gateway will be validated and enriched with the 'Multi-Source Distributed Real-Time Search and Information Retrieval' application (SIR). The CC will engage with Arts and Humanities communities to attract more applications and users to the gateway.

Task SA2.7 - LifeWatch (Lead partner: CSIC, M1 – M30)

Estimated task effort: 60PM

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.

Task SA2.8 - EISCAT_3D (Leader: EISCAT, M1 – M30)

Estimated task effort: 28PM

The design of the next generation incoherent scatter radar system, EISCAT_3D introduces significant challenges in handling large-scale experimental data which will be massively generated at great speeds and volumes. The CC will build an e-Infrastructure to meet the requirements of the EISCAT_3D data system, will support the EISCAT science community in their acquisition, curation, access to and processing of the data, and will train data scientists who can explore new approaches to solve problems via new data-centric way of conceptualising, organising and carrying out research activities.

Task SA2.9 - EPOS (Leader: INFN/INGV, M16 – M30)

Estimated task effort: 30.5PM

The task will operate a CC to drive the future design of the use of grid and cloud for the integrated solid Earth Sciences research as part of the European Plate Observing System (EPOS). The CC will (1) identify and validate authentication and authorisation services, (2) will test cloud resources and usage models, (3) provide knowledge transfer services between e-Infrastructure and EPOS communities.

---

[21] Scipion (http://scipionwiki.cnb.csic.es/). Scipion is a workflow-oriented software integrator with desktop-based as well as web interfaces. It currently is capable to handle most of the 14 major software suites in the field. In the context of this project, we envision to deploy Scipion in the EGI Federated Cloud, handling data- and computation-intensive jobs by MPI, while being accessed by the Instruct EM Access sites through its web interface.

Task SA2.10 – Disaster Mitigation (Leader: AS, M1 – M30)

Estimated task effort: 60PM

The objective of the this CC is to make available customised IT services to support the climate and disaster mitigation researchers to gain a deeper understanding of the most serious natural disasters that affect Asia (e.g. earthquakes, tsunamis, typhoons) and to mitigate multi-hazards via data-intensive, e-Science techniques and collaborations. The task strongly builds on experts from the Asia-Pacific region who will create virtual research environments with embedded services and simulations that enable the sharing of disaster-related data, tools, applications and knowledge among field-workers, scientists, and e-Infrastructure experts, shortening the time they can respond to natural disasters. The CC will have two associated partners, the Leibniz Supercomputing Centre (LRZ, Germany) and University of St. Andrews (UStA, UK) to connect the activities with related projects and initiatives in Europe.

Deliverables (brief description and month of delivery)

D6.1 - Assisted pattern recognition tools integrated with EGI for citizen science (OTHER) (M9)

Tools integrated with the most suitable types of resources from EGI for the LifeWatch community.

D6.2 - Data repository for DARIAH (OTHER) (M11)

A digital repository based on gLibrary framework, the DARIAH community's requirements, and the EGI Federated cloud and grid infrastructures.

D6.3 - Production portal for EISCAT_3D (OTHER) (M12)

Improved version of the current pilot portal of EISCAT_3D using all the existing EISCAT data in the archive.

D6.4 - Fully integrated MoBrain web portal (OTHER) (M12)

An entry portal by the MoBrain Competence Centre that integrates and/or gather information form the various web portals relevant to the WeNMR, INSTRUCT and N4U research communities.

D6.5 - Final version of Multi-Source Distributed Real-Time Search and Information Retrieval application (OTHER) (M12)

Real-time search and data retrieval platform developed in the DARIAH Competence Centre based on the SIR software and EGI services.

D6.6 - Data flow handler and basic R tools to integrate and process data from Ecological Observatories on EGI (DEM) (M12)

Tools for Ecological Observatories by the LifeWatch Competence Centre of EGI.

D6.7 - Implementation and evaluation of AMBER and/or GROMACS (R) (M13)

Evaluation of the deployment of AMBER and/or GROMACS on a GPGPU testbed of EGI by the MoBrain Competence Centre.

D6.8 - Analysis of requirements on biobank and study workflows (R) (M14)

This document provides a summary and an analysis of the requirements captured by the BBMRI-ERIC Competence Centre for workflows from the biobank community.

D6.9 - web portals for tsunami wave propagation simulations and for WRF-based weather simulation (OTHER) (M14)

Simulation portals developed by the Disaster Mitigation Competence Centre

D6.10 - Infrastructure tests and best usage practices for life science service providers (R) (M15)

This report will summarise tests and best usage practices of the EGI Federated Cloud services for life science service providers.

D6.11 - Security toolset release for BBMRI-ERIC (OTHER) (M15)

The final version of the toolset and security libraries released by the BBMRI-ERIC Competence Centre. This includes implementation and testing of the AAI in the project.

D6.12 - GPGPU-enabled web portal(s) for MoBrain (OTHER) (M16)

Development of the GPGPU-enabled web portal(s) by the MoBrain Competence Centre.

D6.13 - EPOS EGI pilot incorporating data retrieval, transfer and linkage (DEM) (M18)

Demonstrator developed based on EGI services by the EPOS Competence Centre for data retrieval, transfer

---

and linkage.

D6.14 - Scipion cloud deployment for MoBrain (OTHER) (M21)

Deployment of the Scipion workflow-oriented software integrator tool with the EGI Federated Cloud.

D6.15 - Demonstrator for ELIXIR workflows implemented in the EGI Federated cloud (DEM) (M24)

Demonstrators in the form of Virtual Machine Images in the EGI Federated Cloud.

D6.16 – Evaluated cloud environment and demonstrator of analysis workflow for biobank studies (DEM) (M24)

A demonstrator of a biobank research workspace running in the EGI cloud, together with an evaluation of its suitability in a typical NGS use case.

D6.17 - Production level gateway for Arts and Humanities (OTHER) (M24)

A WS-PGRADE based science gateway customised for the Arts and Humanities according to the requirements captured by the DARIAH Competence Centre.

D6.18 - Report on the installed LifeWatch applications and their usage record (R) (M24)

A report to summaries the software applications that have been integrated and installed on EGI by the LifeWatch Competence Centre and to provide statistics about their use.

D6.19 - EPOS EGI pilot integrating computational resources and services (DEM) (M24)

Demonstrator developed based on EGI computational services by the EPOS Competence Centre.

D6.20 - Application of the simulation portals for scientific scenario in disaster mitigation (DEM) (M24)

Demonstrators built by the Disaster Mitigation Competence Centre within the simulation portals.

D6.21 - Second version of enhanced portal for EISCAT_3D (OTHER) (M28)

Improved version of the EISCAT portal with reanalysis features implemented.

D6.22 - Design larger scale multi-hazards simulation attempting to reduce the uncertainty of climate change assessment (DEM) (M29)

Demonstration of multi-scale hazards simulation by the disaster mitigation competence centre.

D6.23 - Report on evaluation of EPOS EGI pilot studies (R) (M30)

The report will provide an evaluation of the pilots and demonstrators that have been developed by the EPOS Competence Centre during the project.

The following table gives an **overview of all work packages** of the EGI-Engage project.

Table 4 (3.1b) List of Work Packages

WP	Title	Type of activity	Lead Participant	Lead Participant Number	Person Months	Start Month	End Month
WP1 (NA1)	**Project Management**	MGNT	EGI.eu	1	96 PM	PM1	PM30
WP2 (NA2)	**Strategy Policy and Communications**	OTHER	EGI.eu	1	199 PM	PM1	PM30
WP3 (JRA1)	**e-Infrastructure Commons**	RTD	INFN	16	153 PM	PM1	PM30
WP4 (JRA2)	**Platforms for the Data Commons**	RTD	EGI.eu	1	156 PM	PM1	PM30
WP5 (SA1)	**Operations**	RTD	EGI.eu	1	106 PM	PM1	PM30
WP6 (SA2)	**Knowledge Commons**	RTD	EGI.eu	1	451 PM	PM1	PM30
TOTAL					1161 PM		