D10.6 Requirements and gap analysis report v2

Lead Partner:	INFN
Version:	1
Status:	Under EC review
Dissemination Level:	Public
Document Link:	https://documents.egi.eu/document/3666

Deliverable Abstract

The Technical Coordination Work Package (WP10) in the EOSC-hub project plays an important role aimed at supporting both participants (e.g. Thematic Services, Competence Centres and Business pilots) and external user communities (not directly involved in the project but engaged through other channels, e.g. the EOSC Portal) in the process to integrate their services into the Hub. In the first report WP10 has defined and is currently operating procedures to elicit, assess and track the technical requirements of such communities and, consequently, provide adequate assistance to the user communities to adopt the service of the Hub.

In order to better structure and consolidate this support WP10 extended such activities by launching the Early Adopter Programme for research communities interested in exploring the latest state-of-art technologies and services offered by the European Open Science Cloud.

In this respect, details are provided about the outcome of the above mentioned technical support activities, together with the related requirements and gap analysis extracted from the needs of the different communities involved.

EOSC-HUB RECEIVES FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT NO. 777536.

COPYRIGHT NOTICE

DELIVERY SLIP

Date	***Name***	***Partner/Activity***	***Date***
From:	Giacinto Donvito Alessandro Costantini Diego Scardaci	INFN/WP10 INFN/WP10 EGI Foundation/WP10	
Moderated by:	Małgorzata Krakowian	EGI Foundation/WP1	
Reviewed by:	Miguel Caballer Anabela Oliveira	UPV/WP6 LNEC/WP7	14/09/2020 17/09/2020
Approved by:	AMB		

DOCUMENT LOG

Issue	***Date***	***Comment***	***Author***
1	14/06/2020	ToC ready	Alessandro Costantini
2	27/8/2020	Document ready for review	Alessandro Costantini Giacinto Donvito Diego Scardaci
3	14/09/2020	Document reviewed	Miguel Caballer
4	16/09/2020	Suggestion from reviewers addressed by the authors	Alessandro Costantini Giacinto Donvito
5	13/10/2020	Final version	Alessandro Costantini Giacinto Donvito Diego Scardaci

TERMINOLOGY

https://wiki.eosc-hub.eu/display/EOSC/EOSC-hub+Glossary

Contents

Executive summary

Among the roles and activities provided by the Technical Coordination work package (WP10) in the EOSC-hub project there is the support to the internal project participants, expressed in the form of thematic services, competence centres and business models, to integrate their services into the Hub.

The same approach has been adopted also for the support to external communities, not directly involved in the project but engaged through other channels, e.g. the EOSC Portal.

As from the D10.5, WP10 defined and operated a procedure to elicit, assess and track the technical requirements of such communities and, consequently, provide adequate assistance to the user communities to adopt the service of the Hub. This procedure, named SOCRM-04 (of the SOCRM process of the EOSC SMS) contains all the steps, and the related interfaces towards processes and project activities, that have been put in place to provide technical support.

In the present deliverable D10.6, the technical support activities described in the D10.5 have been consolidated and for each use case, the detailed information about identified user stories, use cases, technical requirements and related status are given and updated in the Community Requirements DB.

In D10.6, particular emphasis has been given to the outcomes of the EOSC-hub Early Adopter Programme (EAP). The EAP, both 1st and 2nd calls, is intended as an opportunity for research communities interested in exploring the latest state-of-art technologies and services offered by the EOSC. In particular, the EAP is aimed at allowing research communities to scale up the in-house infrastructure and to access a richer set of resources by providing expertise, resources and the related support, to enable active usage of the EOSC and foster a culture of co-operation between researchers and EOSC providers.

1 Introduction

The present document contains the update of the D10.5 Requirements and gap analysis report v1.

This second report describes, in section 2, the updates of the technical support activities carried out during the project activities for the Thematic Services and Competence Centers.

Sections 3 and 4 describe, in detail, the technical support activities carried out to support the applications selected during the 1st and 2nd call of the EOSC-hub Early Adopter Programme (EAP). The EAP is intended for research communities interested in exploring the latest state-of-art technologies and services offered by the European Open Science Cloud (EOSC).

The technical support activities are presented with information about identified user stories, use cases and technical requirements with related status.

2 Technical support activities – update

The present section describes the support activities carried out within Competence Centers (CC) and Thematic Services (TS) providing user stories, use cases, technical requirements and related status as reported in the EOSC-hub Community Requirements DB[1].

No.	Community	Type	Link
1	CLARIN	TS	https://wiki.eosc-hub.eu/display/EOSC/CLARIN
2	DARIAH	TS	https://wiki.eosc-hub.eu/display/EOSC/T7.8+DARIAH
3	Marine	CC	https://wiki.eosc-hub.eu/pages/viewpage.action?pageId=29851902
4	Fusion	CC	https://wiki.eosc-hub.eu/pages/viewpage.action?pageId=33064292
5	EPOS-ORFEUS	CC	https://wiki.eosc-hub.eu/pages/viewpage.action?pageId=33064352
6	EO Pillar	TS	https://wiki.eosc-hub.eu/display/EOSC/EO+Pillar
7	WeNMR	TS	https://wiki.eosc-hub.eu/display/EOSC/WeNMR
8	OPENCoastS	TS	https://wiki.eosc-hub.eu/display/EOSC/OPENCoastS
9	ECAS/ENES	TS	https://wiki.eosc-hub.eu/pages/viewpage.action?pageId=23234555
10	Radio astronomy	CC	https://wiki.eosc-hub.eu/pages/viewpage.action?pageId=33064389
11	EISCAT_3D	CC	https://wiki.eosc-hub.eu/pages/viewpage.action?pageId=29851949

[1] Community requirements DB: https://wiki.eosc-hub.eu/display/EOSC/Community+requirements+DB

12	ELIXIR	CC	https://wiki.eosc-hub.eu/pages/viewpage.action?pageId=26418954
16	DODAS	TS	https://wiki.eosc-hub.eu/display/EOSC/DODAS

2.1 CLARIN

2.1.1 User Stories

No.	User stories
US1	A researcher wants to find relevant resources by the available metadata, using keywords or other search dimensions (facets) such as date, location, language, format, etc. to use in their work. Many of such resources are available through the CLARIN infrastructure.
US2	A researcher wants to be able to find (software) tools that can be used to process the data that they have found. For instance, they want to find a tokenizer for the Dutch language.
US3	A repository manager wants to make a repository and its resources findable for researchers. There may be various forms of resources which may have anywhere from no metadata to well-defined elaborate metadata based on specific schema.
US4	A community manager wants to make some language technology tools findable for researchers. The tools have minimal metadata.
US5	A user wants to be able to discover & access the content associated with a (virtual) collection. The collection can be discovered via a search engine or other means.
US6	A researcher wants to manage a group of resources (not limited to a single existing collection or site) that are relevant for her in a way that they are easily findable, accessible, and citable.
US7	A community manager wants to group related resources from their repository in citable collections.

US8	A researcher wants to know what tools can be used to process a given resource. The resource could have been found through an EOSC compatible repository or discovery service or it may have been produced by the researcher. The resource itself may also be a virtual collection. The researcher would like to have an overview quickly showing a selection of tools that are relevant and useful.
US9	A researcher or software engineer has developed a tool for processing resources. They want to make this tool available, findable and accessible to as many researchers and users as possible. They prefer if they can make the tool available and maintain it themselves without having to ask help from a middle layer
US10	A user of an EOSC-hub compatible data discovery tool wants to be able to find and access virtual collections from within the service they are using.
US11	A linguist using one of the EOSC-hub compatible discovery or repository services wants to be able to see what linguistic tools they can use to process a given data object, without leaving the environment of the service that they are using.
US12	A user of the VLO (CLARIN or other instance) wants to create and exploit semantic annotation added to VLO metadata. Semantic annotations are meaningful texts that describe or comment on the tagged resource.
US13	CLARIN centers need to backup & archive their data for persistency.

2.1.2 Use Cases

Step	Description of action	Dependency on 3rd party services (EOSC-hub or other)

UC1	The VLO is a search tool that can be used to find relevant data by searching through metadata ● Searching can be done various facets, e.g. date, location, genre, collection ● The primary users are researchers working in a specific domain for which the set of facets is optimised. ● Community data managers can use the VLO to make their data available to a larger audience B2FIND is a general metadata catalogue for non-specific researchers an easier to use, but less specific tool to find resources ● B2FIND will be used for CLARIN collection level metadata and not for individual resource metadata	VLO, B2FIND (US1-4)
UC2	The VCR is a repository of collection metadata ● In the VCR virtual collections are defined that contain links to resources stored in other repositories ● Researchers can use the VCR to group data into virtual collections of their choosing and make these collections findable and citable ● Researchers can seamlessly access virtual collection content	VCR, B2ACCESS (US5-7)
UC3	The LR switchboard can provide a set of relevant tools for specific data types. ● It is not meant as a stand-alone service, but as a feature that can be integrated with other services. ● Users can use this feature to find relevant tools for processing data objects directly from within the data discovery or hosting service. ● Tool and service providers can make their tools available via the LR switchboard by specifying its features (tool metadata).	LR Switchboard (US8-9)

UC4	Reverse integration of VCR with B2SHARE (B2DROP/B2STAGE). ● Users can access and use the VCR from within relevant EOSC services to download or copy the data to a new destination.	VCR, B2SHARE, B2DROP (US10)
UC5	Reverse integration of LR Switchboard with B2FIND/B2SHARE/B2DROP ● The LR Switchboard can be accessed by users from within relevant EOSC services to show what tools can be applied on a given resource	LR Switchboard, B2SHARE, B2DROP (US11)
UC6	Integration of B2NOTE with the VLO. ● B2NOTE can be used from the VLO as an added discovery mechanism using semantic annotations in addition to metadata. Users can add semantic annotations and search for them.	B2NOTE, VLO (US12)
UC7	B2SAFE/B2STAGE use for safe data replication. ● This use already occurs, but CLARIN needs clear costs and SLA statements. The B2SAFE HTTP JSON REST API is used to create and manage persistent identifiers (PIDs) associated with a file or iRODS collection.	B2STAGE, B2SAFE, B2SAFE HTTP API (US13)

2.1.3 Requirements

Requirement ID	EOSC-hub service	GAP (Yes/No) + description	Requirement description	Source Use Case

RQ1	VLO, B2FIND	Yes, not all CLARIN metadata is harvested (scaling and mapping problem on the B2FIND side)	Make VLO metadata also findable in B2FIND. ● At least the available collection metadata should be harvested. ● B2FIND to harvest CLARIN collection level metadata only. ● This is an adapted technical req (orig. harvest collection metadata through a managed VLO OAI-endpoint.)	UC1
RQ2	VCR, B2ACCESS	Yes.	AAI integration between CLARIN and EOSC-hub. Different levels of integration are conceivable: ● (a) CLARIN thematic services part of EOSChub ID federation ● (b) CLARIN users able to access EOSChub services. CLARIN can indicate which extra users (IdPs) will be enabled using EOSC services in this way. Note that this (SAML federation interoperability) is not a technological but a legal & administrative problem.	UC2

RQ3	VCR, B2FIND, B2SHARE	Yes. Link from VCR to EOSC-hub services	Make metadata resources available in relevant EOSC-hub services, e.g. B2SHARE, B2FIND, etc., accessible from the VCR and in suitable format for building virtual collections	UC2
RQ4	VCR, B2SHARE, B2FIND, EGI DataHub, …	Yes. Link from EOSC-hub services to VCR	Make the VCR available when using EOSC-hub repository and discovery services, in a way that this data can be downloaded e.g. used to make new collections	UC4
RQ5	LR Switchboard, B2DROP, B2SHARE, B2FIND, ...	Yes. Invoking the LR Switchboard from within EOSC-hub services	Adding a capability to EOSC-hub services to invoke the LR switchboard for a data item. This requires information about the data items to be sent to the LR switchboard via a call to its API	UC3, UC5
RQ6	B2NOTE	Yes B2NOTE not integrated with CLARIN VLO nor connected to CLARIN SPF	Integrate B2NOTE in the VLO, add B2NOTE to CLARIN SPF (note any legal requirements)	UC6
RQ7	B2SAFE/B2STAGE	Yes No single costs and condition listing available	Provide information on all relevant aspects including costs and conditions	UC7

2.2 DARIAH

2.2.1 User Stories

No.	User stories
US1	An authenticated user needs to transfer data files from local to remote storage and *vice versa.* The user may also want to transfer files between different storage services which may require different protocols. Within the current EGI DARIAH Gateway, this is currently performed using the Data Avenue service (https://data-avenue.eu/) developed under the SCI-BUS Project (supported by the FP7 Capacities Programme under contract n°RI-283481). Using the B2 storage services will allow users to use different types of storage at different stages of the data lifecycle.
US2	The EGI DARIAH Service manager would like to make use of better supported data transfer technologies, in order both to improve user quality of experience and ensure long term support for the data transfer protocol.
US3	The EGI DARIAH service manager wants to be able to make use of cloud infrastructures to support an increased number of users. These cloud infrastructures could be running docker images or create hadoop clusters.
US4	Users would like to be able to do large scale data analysis on existing accessible data sets through the DARIAH EGI Gateway.
US4	Users would like to be able to use docker containers-based applications on the EGI federated cloud.
US5	The EGI DARIAH service manager would like to allow data distribution across multiple sites for easy data storage, discovery and access.

2.2.2 Use Cases

Step	Description of action	Dependency on 3rd party services (EOSC-hub or other)
UC1	Replace Data Avenue with B2STAGE	None

2.2.3 Requirements

Requirement ID	EOSC-hub service	GAP (Yes/No) + description	Requirement description	Source Use Case
RQ1	B2STAGE	POTENTIAL: Need to check whether EGI DARIAH AAIcertificates are generated/accepted.	POTENTIAL: B2STAGE can be improved to support file transfer among different architectures (not just B2 service)	
RQ2	B2SHARE	No	Integrate with DARIAH Gateway	UC2
RQ3	B2DROP	No	Integrate with DARIAH Gateway	
RQ4	B2ACCESS	UNKNOWN	Integrate DARIAH Gateway AAI (to investigate as other available EOSC AAI mechanisms can be adopted)	
RQ5	B2FIND	NO: Existing DARIAH metadata already harvested	Make data discoverable through metadata searches	
RQ6	B2SAFE	POTENTIAL: May need to allow harvestable metadata from B2SAFE instances	Integrate with DARIAH Gateway	

RQ7	PaaS Orchestrator	UNKNOWN - Plan is to integrate with WS-Pgrade	DARIAH Gateway is not a deployable software. So, a central instance is used.	
RQ8	EGI FedCloud/ FedCloud Containers/ Nova-docker/OneDock	YES: Can users run arbitrary containers on fedcloud resources/ can DARIAH user containers be made available for general usage	Extend the gateway functionalities to support docker-based applications and tools in the cloud environment (EGI FedCloud) by using Indigo components (OneDock, OpenStack Nova Docker)	
RQ9	FutureGateway	UNKNOWN	The science gateway will internally deploy a copy of FutureGateway and the other components will integrate with its APIs.	
RQ10	B2FIND/B2SHARE	UNKNOWN: Can Semantic Search Engine. harvest from B2FIND B2SHARE can already be harvested	Make data stored in EUDAT repositories findable through the Semantic Search Engine.	

RQ11	EGI FedCloud/ OneData/B2 SHARE/ B2STAGE	NO	The VLE will be integrated with the EGI compute and storage resource infrastructure and integrated as a new user-oriented service of the EGI DARIAH CC Gateway. The current client-server model will ported to the EGI FedCloud infrastructure, while the underlying data management will be integrated with the INDIGO OneData solution and EUDAT B2Share and B2Stage service to transfer data to EGI FedCloud for batched processing of editing procedures.	

2.3 Marine

2.3.1 User Stories

No.	User stories
	Argo user stories
US1	A data provider should be able to link its data production instruments into the 'back-end' of the Marine CC setup and become a data provider for the CC users.
US2	A scientist should be able to browse the connected data source networks (e.g. Argo, EMSO, SeaDataNet, etc.) and define preferences for the data records he/she is interested in. The system should make matching records visible in his/her personal access folder.
US3	A user should be able to access his/her personal data access folder via a Jupyter system and perform data analytics on the data.

No.	User stories
	SeaDataNet user stories
US4	As a user, I want to be able to use my legacy/desktop software to process and analyze data stored on the cloud.
US5	As a data provider, I want to only have to update erroneous files during import to only transfer the data required once.
US6	As a user, I want to be able to access my requested data through cloud computing tools within my cloud environment.
US7	As a user, I want to be able to find relevant datasets available within the cloud environment in (semi) real-time.

2.3.2 Use Cases

Step	Description of action	Dependency on 3rd party services (EOSC-hub or other)
	Argo use cases	
UC1	Data discovery and subsetting-subscription service on Argo observations.	
UC2	DIVA data-interpolating variational analysis on Argo floats oxygen data, running on a Jupyter notebook.	
UC3	Data scientist manages his workspace within JupyterHub : save and share notebooks, run codes on the datasets pushed by Research Infrastructures on EOSC (such as Argo) and his individual datasets.	
	SeaDataNet use cases	

UC4	Cloud migration for legacy applications: Many software applications used by the Marine Community were developed during multiple projects spanning many years. These applications have specific requirements regarding File system operations. The most common assumption is that files are available on local storage. To simplify the processes of migrating these (mostly desktop) applications to the cloud we would use EGI DataHub as a file access layer providing seamless access to files distributed in the cloud environments.	
UC5	Reducing redundant data transfer: Processing datasets from partners require that the files first be transferred to a staging area and then processed. However, these operations are not always successful. In the case of quality control, certain datasets may be rejected and have to be revised before being submitted again. This means that some complete datasets are transferred multiple times before being accepted. Using the direct access provided by EGI DataHub to these files we can process only the required amount of data.	
UC6	Virtual space for user data storage and delivery: After a user searches for specific data he sends a request for a subset of datasets, a process is started to collect the datasets from many partners. This collection is an asynchronous process. The process can take weeks to collect all the requested files from the partners. This process is dependent on the resources available at the partners. We intend to use the features of Space and privileges management provided by EGI DataHub to streamline these processes. We would provide the end users access to his requested files through a shared space. Ideally, such a space can be used to make his requested files available for further processing in a cloud environment.	

UC7	Interface for distributed search using metadata queries: In order to find specific files in a distributed environment, we use proprietary search indexes. These indexes are inflexible and only allow querying of predetermined fields. This increases the time required to process and index all available datasets. With the current search interface, we process changes daily. However, the use of datasets within workflows would benefit from up to date information on the available datasets. To extend the discovery capabilities of a cloud application we can leverage the advanced metadata querying functionalities of the EGI DataHub platform.	

2.3.3 Requirements

2.3.3.1 Argo use cases

EOSC-hub services	Amount of requested resources	Time period
B2SAFE	100GB for Argo data	From 2018
B2DROP	EOSC hub data scientist user default account on B2DROP	
B2ACCESS	100 users should be able to access the services	From 2020
JupyterHub	EOSC hub data scientist default account on Jupyter Hub	From 2018 with Cineca for DIVA analysis

Host the data subscription web GUI with its Cassandra and Elasticsearch databases	Ongoing request for capacity with IN2P3 Alternative possibilities : CSC or Cineca	From mid 2019

2.3.3.2 SeaDataNet use cases

EOSC-hub services	Amount of requested resources	Time period
EGI DataHub	The 4 SeaDataNet use cases can be run on a testbed consisting of 3 sites: 2 as data providers, 1 as cloud compute provider. The sites should scale to the following level: The average number of files requested and processed is 500. Each file has a size of tens of KB but occasionally some larger files that average 500 MB are processed. A request is usually for ease of transfer and is usually between 50 to 100 MB.	From 2018

2.4 Fusion

2.4.1 User Stories

No.	User stories

US1	The Fusion CC wish to demonstrate making use of EOSC computational and storage resources for running containerised modelling applications (primarily HPC and HTC). This requirement derives from the fact that local resources are not scaled for peak demand and we wish to use the infrastructure provided by EOSC (and public cloud providers) as a scalable, non vendor specific resource. At a high level, this is an opportunistic use case where we wish to make use of any spare resources at sites, and thus going through an ordering process would be non optimal since the user would not know local resources are exhausted until they submitted their job. It may be that some sort of framework agreement would be needed between the community and the sites to allow this opportunistic use beyond the small number of cores already presented through EGI. Different parts of the workflows may involve different computational requirements, from simple single core machines to many core/multi-node. In the first case we would request resources on a single site, but only instantiate the number of machines required for a specific element of the workflow. It is desirable to develop this further to allow the instantiation of machines for different parts of the workflow to meet the requirements of each step. We are also interested in using both using traditional workflows and workflows within the ITER Integrated Modelling and Analysis Suite (IMAS) which is anticipated to become the standard framework for both modelling and analysis work in the future. In most cases the steps of the workflow communicate through files, with each stage producing its own unique file which acts as an input to the next stage. While running at a single site, it would be possible to request storage at that site of appropriate size. However, in the desirable case of different stages of the workflow running at different sites, either a storage system accessible to many sites will be required for these intermediate files, or they would need to be transferred between sites. Final output data (and possibly intermediate data) should be accessible to the end user.

US2	As a site data manager I am looking at how we can improve user and computational access to experimental data, in addition to being driven to allow more open access to users beyond the fusion community by national or/and international funding bodies. However, in common with most science disciplines my site places an embargo on experimental data to allow researchers to publish. In addition, some data will not be made public where it has no scientific value (engineering tests for example), or where work is done on behalf of industry. Significant analysis work is performed on the MARCONI/Fusion supercomputer based at CINECA and for data sets which will be accessible it would be beneficial to my users if data could be hosted there. In addition, we want to offload public data to partner sites for hosting and access to the wider science community and general public, so that data used by the fusion community is kept on site and only accessed by fusion users. This, combined with the restricted roadmap for tape technologies is pushing me towards replication as a means of bit preservation in the longer term. The community already has a data access mechanism (UDA) which it uses and must be usable at each site where the community will access data. General access will be via HTTP. Thus as a site manager I would ideally like to put a full copy of my data on a trusted site which will prevent unauthorised access to the data (although not the metadata) during the embargo period but will make it accessible following that. That site should be able to provide me with data download statistics on an annual basis as part of my reporting to senior management and fundholders. Additionally, I would like that data to be copied from the trusted site to CINECA so it can be used optimally for analysis on the MARCONI computer and I would like a third copy to ensure there are four copies on my data availability for high availability (one local and three offsite copies).

2.4.2 Use Cases

Step	Description of action	Dependency on 3rd party services (EOSC-hub or other)
UC1	On submission of a containerised workflow, sufficient resources are provisioned on a remote site(s) to allow execution of that workflow. The users home credentials (or the services services credentials) must be accepted on the remote site(s)	Orchestrator/Kubernetes EGI Fed Cloud/Other cloud Suitable AAI mechanism (note the fusion community does not have a centralised IdP or AAI system, but each site provides its own authentication based on username/password) PROMINENCE cloud execution service

UC2	During workflow execution, intermediate files should be written to a location which will be accessible to later stages of the workflow. The final results should also be accessible at the users home institute	EGI FedCloud/B2DROP integration based on either user or service credentials
UC3	Data hosted off-site shall be accessible to code using the Fusion Unified Data Access (UDA) middleware so that the same code can be used to access data regardless of locality.	B2SAFE with three way replication - preferred sites would be CINECA (for the MARCONI link), STFC (geographically close) and PSNC (another fusion site running this service). Will require suitable AAI mechanism and integration with UDA.
UC4	Any attempt to access data hosted off site should determine whether the data is 'open' or 'embargoed'. In both cases, users should authenticate themselves to allow traceability of who has accessed the data. In cases where the data is embargoed, the hosting site should deny access to unauthorised users.	B2SAFE and embargo periods/OneData
UC5	Data placed at an offsite location should be replicated to CINECA and at least one other partner site within the fusion community	B2SAFE/EGI DataHub

2.4.3 Requirements

Requirement number	Requirement title	Source Use Case
RQ1	Storage requests for WP8.2 Fusion CC	UC5
RQ2	Provide a homogenised AAI for single sign on access to all services	UC4

RQ3	Support Opportunistic usage of cloud computing	UC1

2.5 EPOS-ORFEUS

2.5.1 User Stories

No.	User stories
US1	As a **provider** of an EIDA data centre I want to provide users with an authentication and authorisation service in order to enable them to securely access restricted and embargoed data.
US2	As a seismological **researcher** I want to search for datasets offered by EIDA and stage them on the available cloud infrastructure offered by EOSC providers.
US3	As a seismological **researcher** I want to analyse my data in a Jupyter environment, pre-populated with my preferred libraries and with access to my pre-staged datasets. I want to store results in my personal workspace/storage area and eventually share them with my colleagues.
US4	As an EIDA **data manager** I want to define my data management (DM) policies and share them with my colleagues at EIDA data centres. I want to enable them to understand, adjust and apply DM policies at their data centres.

2.5.2 Use Cases

Step	Description of action	Dependency on 3rd party services (EOSC-hub or other)
UC1	A researcher requests to access the services of the EPOS-ORFEUS CC. He is redirected to the CC Authentication Service (relying on B2ACCESS in the background) where he can log in at his home institution or create a local account if needed. He receives a token. Depending on his profile he might be authorised to use the services of the CC. Profiles include information about the groups he belongs to (e.g. read permission of particular restricted data).	B2ACCESS

UC2	A researcher (authenticated and authorised) searches for datasets of her interest, by using an API. She selects one or more staging nodes from the available ones received, by interrogating the API. Finally she initiates the data movement by calling a dedicated method of the same API.	B2STAGE, B2SAFE, EIDA WFCatalog (with Dublin Core extension)
UC3	A seismologist (authenticated and authorised) wants to perform an analysis on a datasets previously selected and staged. He logs in the Jupyter environment close to the staged datasets. He selects and launches a kernel containing his preferred seismological libraries. When the correspondent Jupyter notebook is up and running the datasets are available in a local directory and he can perform his analysis. He might choose to pause his work and save it for later. Finally he can download results on his PC, move them to his personal cloud storage folder or make them available on a local folder.	EGI Notebook, B2DROP
UC4	A data centre acquires and stores waveform data by connecting to servers or devices. A network operator indicates data publicity policies to the data centre. At a next phase, a check of the expected data is performed as well as computation and ingestion of waveform data quality metrics (e.g. percentage availability). Meanwhile, manually data maintenance (e.g. gap filling) and replication is being implemented. Concerning replication, the data are transferred from the data archive to external resources by using B2SAFE. Finally, data requests via services are being traced and summed up to statistics regularly.	B2SAFE-DPM
UC5	A seismologist wants to analyse data that is available (previously staged) at different distributed compute centres. After accessing one of the available Jupyter environments, he writes and tests his analysis code. When he is satisfied with the results he might decide to run such an analysis code on a selected number of compute centres.	EGI Notebook

UC6	Despite that EPOS works with and provides access to open data, some seismic networks related to temporary experiments need to keep data embargoed for a short period of time. The management of the Access Control List (ACL) is in charge of the Network Operator or project PI. Originally, the PI contacted the data centre to include/exclude users from the ACL. EPOS would like to give permissions to PIs, so that they can manage the ACL by themselves using B2ACCESS groups from the B2ACCESS GUI. This way, data centres can configure their systems to grant access to groups (not individuals) and the PIs manage group members in a decentralised way.	B2ACCESS

2.5.3 Requirements

Requirement number	Requirement title	Source Use Case
RQ1	iRODS instance accessible from the Jupyter environment and federated with local B2SAFE/iRODS instances	UC2, UC3
RQ2	Customisable and permanent kernels in Jupyter (EGI Notebook)	UC3
RQ3	Personal data folder with staged data available for mounting in the Jupyter notebook	UC3
RQ4	Operating SeedLink slarchive and rsync for data acquisition, while ArcLink and FDSNWS for data exposal. WFCatalog for quality metrics collection and distribution, as well as B2SAFE for data replication and Webreqlog for statistics sum-ups	UC4
RQ5	Execution of distributed Jupyter notebooks	UC5
RQ6	A centralised catalogue of policies. It should collect descriptions of data management policies and make them available (via API and metadata)	UC4

2.6 EO Pillar

2.6.1 User Stories

No.	User stories
US1	As a researcher interested in EO data, I want to be able to access and analyse those data with VMs or using HPC-like facility
US2	As a researcher interested in EO data, I want to be able to explore EO-data produced in EOSC-hub in an interactive way
US3	As a provider of the Sentinel Playground, I want to be able to integrate OGC compliant WS from EOSC-hub into the service
US4	As a provider of EO Pillar TS, I want to be able to deploy the service components into VMs hosted at the EO-data providers
US5	As a non-expert user, I want to do simple analysis on maps (e.g. vegetation growth in the last year) by zooming and panning to inspect areas of interest.
US6	As a GIS-experienced user, I want to import EO-data into a GIS platform (e.g. QGIS)
US7	As a expert user, I want to perform my own analysis using python/R by calling remote OGC WS facilities
US8	As a provider of EO Pillar TS, I want to offer my services to existing ESA SSO users
US9	As a GeoHazards expert, I want to interactively find relevant data for a given area of interest and time period, and invoke processing from a predefined set of functions over that area
US10	As a GeoHazards expert, I want to download via OGC APIs the data of interest

US11	As a GeoHazards expert, I want to periodically process data from areas of interest and have the results available

2.6.2 Use Cases

Step	Description of action	Dependency on 3rd party services (EOSC-hub or other)
UC1	Discovery of available EO data	Browsable/Searchable EO Data catalogue with pointers to the providers
UC2	Execution of analysis of EO data	EGI Cloud Compute/EGI HTC
UC3	Provide EO data to the compute environment with POSIX interface	EGI Cloud Compute/EGI DataHub
UC4	Discovery of OGC WS compliant services	EOSC catalogue
UC5	Use OGC WS services with federated identities	EOSC-hub AAI
UC6	Login into the services with ESA SSO identities	ESA SSO

2.6.3 Requirements

Requirement ID	EOSC-hub service	GAP (Yes/No) + description	Requirement description	Source Use Case
RQ1	Catalogue	Yes	Browsable catalogue of EO data that allows to discover the providers hosting the relevant data for the user	UC1

RQ2	EGI Cloud Compute	No	Create VMs on providers hosting the data	UC2
RQ3	EGI Cloud Compute / EGI DataHub	No: DataHub can provide that functionality	Provide data with a POSIX interface on VMs	UC3
RQ4	Catalogue service	Yes: there is no catalogue of OGC WS endpoints in the EOSC-hub	Discovery of OGC WS compliant services	UC4
RQ5	EOSC-hub AAI	No: OpenID Connect allows non-browser access to services	Use OGC WS services with EOSC-hub identities	UC5
RQ6	EOSC-hub AAI	Yes: ESOC-hub AAI doesn't support ESA SSO	EOSC-hub AAI should accept ESA SSO ids	UC6
RQ7	EOSC-hub Accounting	Yes	Tracking the resource usage and monitoring the data via EOSC-hub accounting tool	

2.7 WeNMR

2.7.1 User Stories

No.	User stories

US1	As a portal administrator, I want to allow my users to transparently use DIRAC4EGI to run their workloads.
US2	As a user, I want to be able to access my data stored in B2DROP and Onedata (and possibly some other repository like Dropbox) directly from the web portal (avoiding local transfer steps), both for job submission and for uploading results.
US3	As a user, I want to be able to access all my portals using the same credentials, including authentication through the EGI CheckIn when desired.

2.7.2 Use Cases

Step	Description of action	Dependency on 3rd party services (EOSC-hub or other)
UC1	Portal's administrator integrates into the portal job submission by users.	DIRAC4EGI
UC2	A portal user uses Virtual Folder service to access his data stored on remote services	West-Life Virtual Folder VM
UC3	West-Life Virtual Folder exposes through webdav multiple backend data storage services	B2DROP EGI DataHub
UC4	A user is able to authenticate to the portals using its own IdP supported by WeNMR	EGI CheckIn

2.7.3 Requirements

Requirement ID	EOSC-hub service	GAP (Yes/No) + description	Requirement description	Source Use Case
RQ1	DIRAC4EGI	No	DIRAC4EGI usage should be integrated into the portals	UC1

RQ2	West-Life Virtual Folder VM	Yes: The West-Life Virtual Folder VM and service is not provided/offered by EOSC-Hub. Additionally, West-Life project will come to an end at the end of 2018.	West-Life Virtual Folder service should be operated by EOSC-hub	UC2
RQ3	B2DROP	No	B2DROP should be integrated within West-Life Virtual Folder	UC3
RQ4	EGI DataHub	Yes: Onedata isn't integrated yet within West-Life Virtual Folder	Onedata should be integrated within West-Life Virtual Folder	UC3
RQ5	EGI CheckIn	Yes: EOSC-hub and EGI-CheckIn AAI doesn't support the IdP(s) used by WeNMR	AAI should accept the IdP(s) used by WeNMR	UC4

2.8 OPENCoastS

2.8.1 User Stories

No.	User stories
US1	As a service user, I want to log into the OPENCoastS web portal through my home institution credentials
US2	As a service user, I want to download data from the simulation of the model
US3	As a service owner, I want to preserve high-quality/premium forecasts to be offered for re-analysis
US4	As a service owner, I want to publish the catalogue of high-quality/premium forecasts to be offered to the service users

US5	As a service user, I want to be able to search the forecast catalogue based on a given set of characteristics
US6	As a service owner, I want to promote the simulation data as Open Data
US7	As a service owner, I estimate 40TB per month of storage consumption
US8	As a service user, I want to run my simulations up to 72 hours
US9	As a service owner, I want to run jobs in HTC, Grid and Cloud environments
US10	As a service owner, I want to be able to deploy the service in an automated way using the Cloud
US11	As a service owner, I want to obtain monitoring and accounting information of my running service

2.8.2 Use Cases

Step	Description of action	Dependency on 3rd party services (EOSC-hub or other)
UC1	User logs in the OPENCoastS service using eduGAIN	EGI Check-in
UC2	OPENCoastS service manages the authorization/attribute provision	OPENCoastS platform
UC3	OPENCoastS service obtains a x509 certificate for Grid submission	WaTTS
UC4	User constructs a broadcast simulation	OPENCoastS platform
UC5	User submits jobs with the broadcast simulation	DIRAC4EGI
UC6	OPENCoastS service may support docker container execution	udocker
UC7	User obtains outputs from the simulation for the next (at most) 72h	DIRAC4EGI

UC8	OPENCoastS service automatically performs quality checks to identify high-quality/premium forecasts	OPENCoastS platform
UC9	OPENCoastS service stores high-quality/premium forecasts	Data preservation service EGI DataHub
UC10	User searches the catalogue of high-quality/premium forecasts for re-analysis	Data discovery Metadata and provenance service
UC11	OPENCoastS service is deployed automatically	Ansible
UC12	OPENCoastS service is deployed in the Cloud as a long-running service	PaaS orchestrator
UC13	OPENCoastS service is monitored	ARGO
UC14	User resource consumption (compute, data) is tracked and accessible	Accounting

2.8.3 Requirements

Requirement ID	EOSC-hub service	GAP (Yes/No) + description	Requirement description	Source Use Case
RQ1	EGI Check-in	No	eduGAIN support for EOSC-hub AAI	UC1
RQ2	OPENCoastS	No	Attribute provision for OPENCoastS	UC2
RQ3	OPENCoastS	No	Forecast simulation composition	UC4
RQ4	WaTTS/MasterPortal	No	OpenID Connect token translation to x509 certificate for Grid submission	UC3

RQ5	DIRAC4EGI	No	Multi-site job submission	UC5
RQ6	udocker	No	User-space Docker container execution	UC6
RQ7	DIRAC4EGI	No?	Job output management	UC7
RQ8	B2SAFE	No	OPENCoastS need to store high-quality/premium forecast archives in permanent storage	UC9
RQ9	EGI DataHub B2FIND B2NOTE	No	OPENCoastS need to handle metadata for user access to historical data	UC10
RQ10	B2HANDLE B2DROP B2SHARE	No	OPENCoastS need to expose historial catalogue as Open Data	UC10
RQ11	<NOT_AVAILABLE>	Yes: EOSC-hub does not provide Ansible consulting	Automated deployment using Ansible of OPENCoastS service	UC11
RQ12	PaaS Orchestrator	No	OPENCoastS service deployment in the Cloud	UC12
RQ13	ARGO	No	Monitoring as a service	UC13
RQ14	Accounting	No	Tracking compute and storage consumption	UC14

2.9 ECAS/ENES

2.9.1 User Stories

2.9.2 User Stories

No.	User stories
US1	As a user I want to be able to perform detailed analysis on large volumes of data in parallel using scalable cloud resources in order to achieve more rapid results than sequential processing and avoiding downloading large quantities of data to local storage.
US2	As a user I want the results of my analysis available to me anywhere and be able to share it with colleagues before publishing in order to discuss and confirm the outcomes.
US3	As a user I want to ensure my input data is accessible regardless of physical location (for example, by making use of persistent identifiers), since then I do not need to implement my own code to deal with these changes.
US4	As provider of the Climate gateway I want to empower researchers from academia to interact with datasets stored in the Climate Catalogue, and bring their own applications to analyse this data on remote cloud servers
US5	As a data producer I would like scientists to be able to reference the source data used for downstream analysis and get accreditation in any subsequent publications.
US6	As a data manager I want any analysis to generate provenance metadata in order to understand what analysis has been performed to allow both confirmation of results and increase confidence in the scientific methods and analysis.
US7	As a decision maker I want to have confidence in the scientific results on which I rely to make policy decisions.
US8	As a research infrastructure provider I would like to (*link up the community AAI portal/ensure my users only need to use a single EOSC portal to interact with ECAS*) so that my users can still use the portal they are familiar with to access resources outside of the community

US9	As a user, I want to access ECAS from my familiar community portal or the workflow I am used to without having to make use of additional services that I first would have to learn about in order to make use of ECAS.
US10	As an infrastructure manager I want to reduce the effort of maintaining client side code support
US11	A user would like to run a climate data analysis experiment across CMIP51 or CMIP62 data. The targeted model output data come from multiple modelling groups across the globe and are therefore hosted at different ENES data sites across Europe. For a specific target experiment, as a preliminary step, the user runs a distributed search on the ENES data nodes to discover the required input files, which will result in a list of input dataset PIDs. The user then assembles a processing job specification and submits it to the ENES data analytics service. The service will arrange for the data to be available at the processing site; data locality will be exploited by default, but data movement could be also needed. Once the data are available, a data analytics workflow runs on the service instance. After the analysis has completed, the user receives an e-mail notification with a link to a shared folder at a publication service folder from which she can access and download the processing results. The challenge is to do the server-side and parallel computation on the distributed data by making transparent the access to the data (including data movement), the allocation and use of the computational resources, the analytics experiment, the provenance tracking, and the overall experiment orchestration.
US12	At a later point, another user discovers the data from the previous analysis on the publication service. In order to be certain that the data can be used for the purpose at hand, she would like to evaluate its generation history, including the details of the processing that was done and the specifics of the input data used. She retrieves data and accompanying metadata from the publication service and uses the PIDs of the processed data to discover additional information on the processing and the full list of PIDs of all input datasets. She can then make a judgement based on the assembled information whether the data are fit for her purpose without having to contact the user who originally requested the data processing.
US13	As a user, I want to be able to make selected results of my data analysis or the analysis script I developed available to others. These recipients may be my immediate colleagues but also a wider range of external third parties. The workflow to make these data available should be largely hassle-free for me.

2.9.3 Use Cases

Step	Description of action	Dependency on 3rd party services (EOSC-hub or other)
UC1	User needs to discover the location of all required input data	ESGF Metadata Service/B2FIND
UC2	Input data must have a PID associated with it.	Community solutions assigning PIDs, possibly via B2HANDLE
UC3	ENES Data Analytics Service must be able to transfer data from its current location to the processing site based on PID (Low priority - It depends on how data input integration ultimately looks like and what can be done with limited effort.)	gridFTP/other?
UC4	Output data must be moved to a site where users can share it for others so they can access it via a link provided by the ECAS system.	B2DROP
UC5	Users will need to register to use the ECAS service	Appropriate EOSC-AAI Solution
UC6	Data must be movable between the output storage in UC4 to a data publication service, where it must be given appropriate metadata and a PID	B2SHARE
UC7	Output data shall have appropriate and sufficient metadata and provenance information associated to enable other users to have trust in the data.	ECAS, B2HANDLE profiles (possibly their usage by B2DROP)
UC8	A link between the output data and the sources must be maintained, in addition to provenance information related to the processing steps.	ECAS, B2HANDLE profiles (possibly their usage by B2DROP)

UC9	Input data must be accessible to the computation regardless of location.	B2HANDLE usage by communities and the DataHub. Support for B2HANDLE PID profiles by DataHub.
UC10	Published output data must be assigned a PID	B2SHARE, DataHub
UC11	The provenance information must be accessible for published output data	B2SHARE & DataHub usage of B2HANDLE profiles
UC12	Users will select individual files or entire directories from their ECAS workspace and then select to publish them. The ECAS workspace will inquire a destination location for the files in the user's B2DROP workspace. The publishing workflow for the users will start from the ECAS workspace but end with a view on the publishing repository (B2DROP) showing the newly published files as confirmation.	B2DROP

2.9.4 Requirements

Requirement ID	EOSC-hub service	GAP (Yes/No) + description	Requirement description	Source Use Case
RQ1	EOSC-hub AAI	ESGF AAI not integrated to any AAI services	Integration of ESGF AAI to one of EOSC AAI services	UC5

RQ2	B2DROP	Can be a central service; no need for local installation. User has no interface to the B2DROP filesystem; currently the user logs in to jupyter with username and password. Files automatically moved to B2DROP without user intervention. GAP: Need to integrate AAI to B2DROP. For training purposes, consider using a proxy user for training purposes.	Need to be able to write directly to B2DROP (via mount point inaccessible to users), or have the workflow copy data in using NextCloud OpenCloudMesh API. Will require separate instances for training and production	UC4

RQ2.1	B2DROP	Publishing files from an ECAS workspace to B2DROP will not require the user to log in to B2DROP separately. Aside from selecting files to publish and a destination folder, the user should also not be asked for additional information (e.g. metadata).	B2DROP must be able to understand and accept IAM security tokens provided by ECAS. Possibly additional detail questions to clarify wrt session management (transparent authentication, selecting destination folder, initiating and confirming transfer as one seamless workflow).	UC12
RQ3	B2DROP	GAP - UNSURE - If data is moved using OpenCloudMesh, the security needs to be considered. NextCloud website recommends using SSL since user information is passed in plain text. Need to check how B2DROP is configured.	B2DROP must run with SSL enabled	UC4

RQ4	B2SHARE	GAP - NO (if RQ2 is satisfied), YES (otherwise) Enable users to push files to B2SHARE. If RQ2 works there is no gap to deal with as the bridge exists. Unless RQ2 works, then beed to integrate AAI to B2SHARE	B2DROP/B2SHARE Bridge required	UC6
RQ5	EGI Datahub	GAP - UNCLEAR Data publishing and data ingest. Allows contacting multiple communities.	See the GAP column	

RQ6	B2HANDLE	GAP - UNCLEAR Both input data and published derived data must be assigned a PID. For third-party users to access provenance information, B2SHARE and possibly also B2DROP need to support recording of minimal provenance information, possibly organized via B2HANDLE profiles.	See the GAP column	UC7, UC8, UC11

2.10 Radio Astronomy

2.10.1 User Stories

No.	User stories
US1	As an Observatory, we want to offer Single Sign On to our users and manage community access through a central federated collaboration management service, such as COmanage, to improve user experience and consolidate user administration for services.

US2	As a scientific user, I want to perform LOFAR data analysis on archived data-products using available scalable compute infrastructure, allowing for long-term storage and inspection of results. It must be possible to automate initiation and monitoring of processing workflows including data staging and storage. I want to use portable software deployment such that time spent on porting applications is minimized, an integrated workflow management framework, and user workspace to store data products that are to be evaluated or further processed.
US3	As a scientific user, I want to enter science-grade data products in a science data repository that supports the FAIR principles to ensure long-term data preservation and attribution of effort. This will further improve sharing of data with colleagues and access to data from other science domains. It should be possible to access data in the science data repository using direct links to individual data objects via an anonymously accessible public URL such that other services, e.g. those provided by the Virtual Observatory, can be built to provide access to the data.

2.10.2 Use Cases

Step	Description of action	Dependency on 3rd party services (EOSC-hub or other)
UC1	Observatory registers service with federation	Federated AAI (Comanage, e.g. EGI Check-In)
UC2	User registers for LOFAR collaboration(s)	Federated AAI (Comanage, e.g. EGI Check-In)
UC3	Observatory manages collaboration(s)	Federated AAI (Comanage, e.g. EGI Check-In)
UC4	Collaboration administration provisions non-web service(s)	Federated AAI (Comanage, e.g. EGI Check-In)
UC5	User authenticates to federation	Federated AAI (Comanage, e.g. EGI Check-In)
UC6	Observatory releases data analysis software in repository	-

UC7	User enters data analysis software in repository	-
UC8	User selects data for retrieval/processing	-
UC9	User requests/stages data-products	-
UC10	User retrieves data-products	dCache WEBDAV/Macaroons
UC11	User initiates data analysis workflow	HTC processing cluster supporting Singularity, CWL workflows, CVMFS (application distribution)
UC12	User inspects output data	dCache WEBDAV/Macaroons
UC13	User registers output data	B2SHARE, PID service (EPIC)
UC14	Observatory registers archived data	B2FIND (option), PID service (EPIC)

2.10.3 Requirements

Requirement number	Requirement title	Source Use Case
RQ1	EOSC-hub to support integration of LOFAR services in EGI Check-In federated AAI	UC1
RQ2	EOSC-hub to support customization of the COmanage Community Organisation for the LOFAR.	UC3
RQ3	EGI Check-In to allow any IdP in EduGAIN, plus social identity providers Google, Orcid, etc., to be allowed for SSO access to federated LOFAR services.	UC2
RQ4	EOSC-hub to provide a LOFAR scoped Persistent ID generation service.	UC13, UC14

RQ5	EOSC-hub to provide a Science Data Repository (B2Share) that supports LOFAR data sizes and a radio astronomy oriented metadata model.	UC13
RQ6	B2Share to provide direct data access URL's	UC13
RQ7	EOSC-hub to provide a B2FIND service for harvested registration of archive data using LOFAR PID's and a LOFAR metadata model (option)	UC14
RQ8	EOSC-hub to support CVMFS mounting on processing clusters	UC6, UC11
RQ9	EOSC-hub to support Singularity containerized applications on processing clusters.	UC11
RQ10	EOSC-hub to support CWL based workflow management framework(s)	UC11
RQ11	EOSC-hub infrastructure partners for LOFAR to support required dCache features (WEBDAV, Macaroons, User workspace, staging API/GFAL)	UC9, UC10, UC12
RQ12	EOSC-hub infrastructure partners for LOFAR to integrate storage and compute systems in accordance with requirements for LOFAR processing workflows	UC11

2.11 EISCAT 3D

2.11.1 User Stories

No.	User stories
US1	Any researcher should be able to access the portal and browse metadata. The portal grants/denies access to data and processing based on affiliation. (Metadata should be available for all researchers. The real data for authorised users only.)
US2	Authorised researchers should be able to select the EISCAT_3D data they are interested in for download or for analysis.
US3	Authorised researchers should be able to browse reference applications in the portal, select an application for use, feed their data in (from US2), visualise or download the analysis result.

2.11.2 Use Cases

Step	Description of action	Dependency on 3rd party services (EOSC-hub or other)
UC1	Authentication: 1. User enters the portal. 2. Portal directs to SSO 3. Portal grants authorisation	1. Portal is running with metadata and file catalogue 2. EGICheckin <-> B2Access, EISCAT idp, other social IDPs
UC2	Data access: 1. User selects data 2. Portal accepts authorisation 3. Portal directs to data store	1. Dirac file catalogue 2. dirac → sso 3. Dirac
UC3	Computation: 1. Data selected 2. Software selected/uploaded 3. Data staging at HPC 4. Processing of data 5. Results presented	1. Dirac 2. Dirac → HPC 3. E.g. B2Stage 4. HPC, VMs and/or containers 5. Visualisation, B2drop, B2share

2.11.3 Requirements

Requirement ID	EOSC-hub service	GAP (Yes/No) + description	Requirement description	Source Use Case
RQ1	EOSC-hub AAI	Yes: Dirac	Dirac interface to EGI Check-in	UC1
RQ2	Data staging	Yes: File system	Temporary file storage close to computing	UC3

RQ3	Cloud Compute	Yes: HPC	Run VMs and/or containers for portal and data processing	UC3
RQ4	PID/DOI service	Yes: Registry	Registration of digital objects: data collections (experiment, low → high levels of data)	UC2,3
RQ5	EOSC-hub Check-in	Yes: Check-in	EISCAT-3D community includes institutes in Japan and China. Currently, Checkin has been integrated into the EISCAT data portal based on DIRAC4EGI technology. However, google and facebook cannot be used in China thus users from there cannot yet login via Check-in. Other social logins are to be implemented, see GGUS ticket: https://ggus.eu/index.php?mode=ticket_info&ticket_id=139172&come_from=submit	UC1

2.12 ELIXIR

2.12.1 User Stories

No.	User stories

US1	ELIXIR wants to establish a federation of cloud sites, each providing storage and compute capacity for researchers. The federated clouds should be connected to a data replication service (Reference Data Set Distribution Service with the ELIXIR terminology - RDSDS) that enables ELIXIR to stage 'ELIXIR Core Data Resources' to the cloud sites on-demand. As a result, the cloud sites become data hosting nodes which are equipped with CPUs/GPUs and are suited for large-scale data analysis and analytics. Centrally provided and curated datasets can ensure high-quality research in any of the partner states/regions. Researchers can go to their 'local' ELIXIR cloud provider, choose an already pre-staged ELIXIR dataset or request the staging of an ELIXIR dataset, choose an application of their choice (from a VM catalogue or container catalogue), maybe upload some additional data and then perform data analysis/analytics. Different conditions of access may apply at the different cloud sites, but it is expected that the cloud compute resources would be free at the point of use for national/local researchers, while pay-for-use or other special conditions apply for foreigners. The replication of community assets to national cloud providers maximises the utilisation of national funding and lowers the total cost of access for researchers. The services in the setup should recognise users via their ELIXIR identity, therefore ELIXIR AAI (Life science AAI) should be integrated with the RDSDS as well as with the national clouds.
US2	The cloud federation can be also equipped with a 'container replication and orchestration service' that enables application providers to deploy containerised community/reference applications to any of the federated cloud sites, and users to instantiate and use the applications on those sites. Having a centrally managed or self-managed container orchestration service will allow users who do not have access to cloud resources of their own to deploy containerised workflows co-located with existing datasets in cloud locations. The services in the setup should recognise users via their ELIXIR identity, therefore ELIXIR AAI (Life science AAI) should be integrated with the RDSDS as well as with the national clouds.

2.12.2 Use Cases

Step	Description of action	Dependency on 3rd party services (EOSC-hub or other)

UC1	Joining the cloud federation with a cloud site (cloud provider perspective): 1. AAI integration between ELIXIR and EOSC 2. Connect ELIXIR cloud compute or data storage location with EOSC 3. Policy compliance with policies (ELIXIR, EGI cloud, and EOSC-hub policies)	ELIXIR AAI ↔ EOSC AAI ELIXIR Cloud Services ↔ EOSC Compute / Storage Catalogue Policy compliance between ELIXIR and EOSC
UC2	Making reference/core datasets available for replication to the federated cloud providers (data provider perspective): 1. Data provider publishes reference dataset to RDSDS using ELIXIR AAI 2. Based on a defined dataset replication policy RDSDS triggers one or more data transfer/synchronisation process via a central EOSC FTS service 3. EOSC FTS manages the transfer process between source and location and notifies RDSDS on completion 4. RDSDS notifies Data Provider that the data has been synchronised and/or any errors	Assumption: User is allowed to perform this action EOSC-hub centrally provided data distribution service (a new requirement to EOSC-hub) EOSC-hub centrally provided application distribution and orchestration service (a new requirement to EOSC-hub) ELIXIR cloud federation policies, protocols and interfaces.
UC3	Requesting the replication of a reference/core dataset to my local cloud (researcher perspective): 3 User searches and finds for a dataset with the RDSDS catalogue 4 User initiates a transfer of a dataset to their local cloud resource using the EOSC central FTS 5 EOSC FtS notifies the completion of the data transfer and/or errors	Same as above.

UC4	Making virtualised, reference/core applications available for replication and orchestration on the federated cloud providers (data provider perspective): 1. User searches the EOSC Service Catalogue to identify a container orchestration service 2. The user uses their ELIXIR credentials to instantiate a container orchestration service 3. User deploys a containerised application or workload to the container orchestration service	Assumption: User is allowed to perform this action ELIXIR AAI ↔ EOSC AAI Kubernetes as a Service

5.1.1 Requirements

Requirement number	Requirement title	Source Use Case
RQ1	EOSC-hub to provide an FTS data transfer service	UC1, 2, 3
RQ2	EOSC-hub to provide Kubernetes as a service	UC1, 4, 5

5.2 DODAS

5.2.1 User Stories

No.	User stories
US1	As a DODAS user I want a service that simplifies the process of provisioning, creating, managing and accessing a pool of heterogeneous computing resources, including private and public clouds.
US2	As a DODAS administrator, I want to be able to monitor DODAS services (DODAS Core services)

US3	As a DODAS administrator I want to be able to monitor processes and services running on the dodas clusters (check status resources consumption etc.)
US4	As a DODAS end user (e.g. physicist) I want to transparently access my remote data
US5	As a DODAS end user (e.g. physicist) I want to temporary cache input and output data from dodas cluster
US6	As a community adopting DODAS, I want to be able to globally share libraries and software (e.g. runtime environment) more in general.
US7	As user of DODAS I want to be able to move my sandboxes (input/output) through dropbox like solution
US8	As a DODAS user, I want to be able to access all my portals using the same credentials, including authentication through the EGI Check-In when desired.

5.2.2 Use Cases

Step	Description of action	Dependency on 3rd party services (EOSC-hub or other)
UC1	The DODAS admin requests the automated deployment of a DODAS cluster	TOSCA template to be submitted to the PaaS Orchestrator
UC2	The DODAS admin checks the status of the DODAS core services	Integration with monitoring service already available in EOSC-hub
UC3	DODAS Users access remote stored data from the DODAS cluster	Cluster configuration shall support EGI Datahub/XRootD integration. XRootD is maintained externally to EOSC-hub
UC4	User jobs running on DODAS clusters store temporary data on local cluster	

UC5	User jobs running on DODAS clusters access software from specific file system path	
UC6	User jobs require to move input and output sandbox without the needs of moving data through storages (manually)	
UC7	DODAS admin and users authenticate against a single AAI system	INDIGO-IAM
UC8	Resource providers authenticate and authorize DODAS requests with INDIGO-IAM tokens	Integration through AAI EOSC-hub solutions
UC9	Accounting data are gathered from the resource providers	

5.2.3 Requirements

Requirement ID	EOSC-hub service	GAP (Yes/No) + description	Requirement description	Source Use Case
RQ1	INDIGO PaaS	No	Automated deployment of the DODAS cluster on top of heterogeneous cloud environments through TOSCA orchestration	UC1
RQ2	Monitoring	Yes: check if EOSC-hub provides a monitoring service for collecting monitoring data	Monitoring information gathering	UC2

RQ3	OneData/XRootD	No	User data remote access	UC3
RQ4	OneData/XRootD	No	Store temporary data locally	UC4
RQ5	CVMFS (client and stratum 0)	No	Usage of specific global file system for software distribution	UC5
RQ6	OneData	No	Job input and output sandbox data movement	UC6
RQ7	INDIGO-IAM	No	Provide authN/authZ for DODAS users	UC7
RQ8	ESACO, AAI EOSC-hub	ESACO is being successfully used by the providers of the enabling facility	Resource providers integration with INDIGO-IAM	UC8
RQ9	APEL	Yes: how to extract accounting data from heterogeneous providers (including public clouds)	Accounting data gathering	UC9

6 Early Adopters selected during the 1st call

The EOSC-hub Early Adopter Programme (EAP)[2] is intended for research communities interested in exploring the latest state-of-art technologies and services offered by the European Open Science Cloud (EOSC). In particular, the EAP is aimed at allowing research communities to scale up the in-house infrastructure and to access a richer set of resources. EOSC-hub will provide expertise and resources to enable active usage of the EOSC and foster a culture of co-operation between researchers and EOSC providers.

This section is giving information about identified user stories, use cases and technical requirements with the related status of the applications selected during the 1st call of the EOSC-hub Early Adopter Programme (EAP).

6.1 STARS4ALL

STARS4ALL is a foundation that gives support to a community concerned with the light-pollution problem. The community derives its name from the STARS4ALL project funded by the European Union H2020 Programme (688135) that was created to create awareness among citizens about the light pollution problem. For this purpose, it deploys a platform to give support to some light pollution initiatives. These initiatives include a photometer network (http://tess.stars4all.eu) to continuously monitor the light pollution, using photometers to measure the sky brightness. In this context, it deploys a platform (http://tess.dashboards.stars4all.eu) to show the measurements in some dashboards. Besides the photometer network, STARS4ALL gives support to citizen science initiatives like Cities At Night . All data is published openly in the STARS4ALL Zenodo community (https://zenodo.org/communities/stars4all). The project was completed in 2018 but STARS4ALL continues the work through the STARS4ALL foundation created for this purpose.

STARS4ALL made a proposal for the first call for Early Adopter Pilots with as major objectives to make the current STARS4ALL infrastructure more robust and obtain better and wider visibility for its activities and data. Alongside these improvements to the STARS4ALL technical infrastructure and data management practices on the basis of EOSChub services will also be implemented.

The user-stories/use-cases are derived from that original STARS4ALL EAP application and further insights and discussions with the project PI, but is a newer reworked version of the community requirement database STARS4ALL entry

6.1.1 User stories

Infrastructure provider perspective.

- **US1**. The motivation for the EAP is that the number of photometers has been continuously growing and is expected to do so in the future. There is a need to increase the capacity and robustness of the STARS4ALL technical infrastructure.

[2] https://eosc-hub.eu/eosc-early-adopter-programme

- There is also a wish to improve data management possibilities by introducing Research Objects that can bundle (references to) different types of information. Currently the information (data, code, presentations, videos) generated by STARS4ALL initiatives is scattered over the STARS4ALL Zenodo community and other repositories, thus being difficult to access and discover. Potentially Research Objects offer a mechanism to bundle all this information.

End user perspective

- **US2**. Each photometer, besides the measurements, has its own metadata associated (sensor type, location, calibration parameters, etc …). As a user, I want to create a persistent id for each photometer so I can access all information that pertains to this photometer
- **US3** As a user I want to describe and bundle the information related with my research
- **US4** As a user I want to deposit the monthly datasets generated by my photometer in a data repository as B2SHARE where other users/researchers can access freely to them, as well as they can cite correctly.
- **US5** As a user I want to have reliable access to STARS4ALL observation data
- **US6** As a user I want to use Jupyter Notebooks and analyse my data
- **US7** As a user I want to discover STARS4ALL research objects via relevant discovery services such as B2FIND. The research objects should be actionable e.g. references to the research objects associated data can be resolved for instance by clicking on a link in a UI.
- **US8** As a user I want to discover and reference the data deposited in B2SHARE from a visual platform like GEOSS to increase the visibility of my location.

6.1.2 Use cases analysis

The need for a more robust higher capacity infrastructure can be handled by duplicating the current STARS4ALL infrastructure components (sensor network, database and repository systems) using a series of VMs.

The wish for global unique identification of the photometers including the associating of relevant metadata can be handled by current PID infrastructure such as provided by B2HANDLE.

Research Objects can be partly accommodated by specifying a suitable research object metadata schema that provides qualified links to different associated data. Note that to achieve a full implementation, the community metadata schema should be highly flexible. B2SHARE allows specific community metadata schemas but is not infinitely flexible e.g. it does not support qualified links to other resources nor deeper structures beyond a list.

Metadata in B2SHARE can be automatically harvested by B2FIND. The B2FIND UI provides actionability of URI links to associated data in identifiers. A special community metadata mapping has to be provided.

Integrating STARS4ALL metadata in GEOSS for discovery is possible by offering suitable metadata for harvesting by GEOS DAB via OAI PMH. To simplify matters this should be a (filtered version) of the metadata deposited in B2SHARE

Jupyter Notebooks provided by the EGI Notebook service can access B2SHARE deposited data via standard APIs. These notebooks can be used for researching purposes (to analyze my data) or for educational purposes (to teach students interesting concepts such as timezones, moon phases, etc.)

6.1.3 Handle technical requirements

The current available EOSChub services will enable most of the STARS4ALL technical requirements.

The duplication of current STARS4ALL infrastructure will be handled by 4 Virtual Machines with the following characteristics

- Dashboards: 1x(1xvCPU, 8GB RAM, 20GB storage)
- API: 1x(1xvCPU, 4GB, 10GB)
- Databases: 2x (1xvCPU, 4GB RAM, 50GB storage)

A EGI Notebook environment has been provided and currently suffices the needs.

The standard B2HANDLE service will be sufficient to manage issuing PIDs for photometers and handling the associated metadata.

6.2 EMSO ERIC Data Management Platform

European Multidisciplinary Seafloor and water-column Observatory (EMSO)[3] is a large-scale European distributed Research Infrastructure (RI) for ocean observation. The RI aims to explore the oceans and to explain the critical role they play in the broader Earth systems, focussing on climate change, risks of biodiversity loss and natural hazards. EMSO's observatories are platforms equipped with multiple sensors to measure biogeochemical and physical parameters such as ocean temperature, dissolved oxygen concentration, and ocean current speed and direction. A fundamental IT component of the EMSO cyber-infrastructure is represented by the Data Management Platform (DMP) developed within the EMSODEV[4] project.

In a nutshell, the project aims at transitioning the DMP to pre-production first and then to full production. The transition of the DMP to pre-production and production will enable data and services to be harmonized to bring accessibility and, when consistent and relevant, it will be enhanced through enriched metadata.

The prototype DMP has been deployed in the EGI FedCloud and its current technology readiness is at level 8. Its transition to level 9 is part of the goals of this project. The transition of the DMP to pre-production and production will enable data and services to be harmonized to bring accessibility and, when consistent and relevant, it will be enhanced through enriched metadata. The existing basic and non-homogeneous EMSO data web services will be homogenized and standardized (e.g., using OGC standards) across EMSO nodes and made interoperable with the subdomain according to FAIR principles, which has the potential to foster interoperability between EOSC services.

[3] http://emso.eu/
[4] https://www.emsodev.eu/

6.2.1 User stories

- **US1**: As a user, I want a service that simplifies the ingest, process and archive of data of the regionally distributed EMSO nodes.
- **US2**: As a user, I want to use federated authentication mechanisms to access the EMSO-ERIC Data Management Platform, including authentication through the EGI Check-In when desired.
- **US3**: As a user, I want to use services and tools for data anonymization such as OpenAIRE Amnesia.
- **US4**: As an administrator, I want to be able to monitor and account for the use of the EMSO-ERIC resources.

6.2.2 Use cases analysis

To support the use cases described above, the following technical solutions have been identified:

- EGI Cloud Compute service,
- EGI Cloud Container Compute service,
- EGI Online Storage service,
- OpenAIRE Amnesia service,
- EGI AAI Check-In service,
- EOSC-hub accounting and monitoring services

6.2.3 Handle technical requirements

To support this application, the following technical requirements have been identified:

- **RQ1**: A dedicated VO[5] has been registered in the EGI Operations Portal for supporting accounting and monitoring issues.
- **RQ2**: A Service Level Agreement (SLA)[6] has been agreed between the customer (EMSO-ERIC) and the two cloud providers of the EGI Federation. With this agreement the following computing and storage resources have been provisioned to support the transition of the EMSO Data Management Platform in pre-production and production:
    - CESGA (Spain): 192 vCPU cores, 512GB of RAM and 600GB HDD of block storage.
    - RECAS-BARI (Italy): 300 vCPU cores, 1.2TB of RAM and 10TB of block storage.
- **RQ3**: To integrate federated authentication mechanisms in the EMSO-ERIC Data Management Platform a new COU group has been created in the development instance of the EGI AAI Check-In service[7]. This set-up will be used for testing the integration of the DMP with the EGI AAI Check-In service.
- **RQ4**: To enable data anonymisation of sensitive data in the EMSO-ERIC Data Management Platform, the open-source OpenAIRE Amnesia[8] will be adopted.

[5] https://operations-portal.egi.eu/vo/view/voname/vo.emso-eric.eu
[6] https://documents.egi.eu/document/3539
[7] https://ggus.eu/index.php?mode=ticket_info&ticket_id=147853
[8] https://amnesia.openaire.eu/

6.3 Mapping the sensitivity of mitigation scenarios to societal choices

6.3.1 User stories

The project aims to perform modelling studies to explore how future energy systems can evolve and to quantify the tradeoffs, co-benefits and interlinkages between different aspects of the global energy systems in the context of international climate change policy and sustainable development.

Such analyses utilize the so-called Integrated Assessment Models (IAMs), which are models of the energy, environment and economic systems in order to quantify key variables of interest in these scenarios, such as emissions pathways consistent with international climate policy goals, tradeoffs of climate mitigation with land use and food security, among others.

This project will provide a proof-of-principle platform aimed at performing large scale (10-15k model runs) analyses.

6.3.2 Use cases analysis

The IAM MESSAGEix-GLOBIOM (considered by the applicants at TRL9) will run sequentially on the selected resources where each job is independent from the other in a parametric fashion. The parametrized simulations are submitted by making use of a batch system manually deployed by the applicants.

An exploratory activity has been performed by applicants for running the full software stack in a containerized environment (using docker) on larger compute systems (e.g., HTC). Even if this activity is likely TRL5, it is envisioned as a key software infrastructure product to be promoted to TRL9 during this project. Starting from such an assumption, a Mesos/Marathon cluster can be instantiated on the provided cloud resources and the parametrized simulations can run in it as independent containers.

The output carried out from the simulations is stored in a distributed environment where it can be accessed for post-processing analysis.

6.3.3 Handle technical requirements

The MESSAGEix-GLOBIOM integrated assessment model (IAM) relies, via GAMS, on the commercial CPLEX solver, for which applicants have a current academic license. As part of the pilot, applicants have resolved issues related to housing these solvers on EOSC infrastructure. Currently, the license is valid as long as the work is carried out through IIASA. The longer-term solution will be to move towards supporting additional, free/open-source solvers besides CPLEX.

To support this proposal, the following requirements have been addressed

- RQ1: Deploy virtual machines on EGI Cloud Compute: INFN-Bari, vo.iiasa.ac.at: 200 vCPUs cores, 800GB of RAM.
- RQ2: Enable federated identity management using one of the available AAI solutions provided by EOSC-hub.

- RQ3: Setup ONEDATA for handle 6TB of distributed storage
- RQ4: Setup a Database (PostgreSQL) managed by the applicants
- RQ5: Containerization of the application and setup on the VMs a Mesos/Marathon cluster as scheduler for the parametric jobs

A first effort has been made to understand the different components (authentication, registration, VPN, tooling, etc.) needed to access the cloud resources.

In addition, an intensive activity is still ongoing to introduce changes to the model code that are necessary to automate job creation, execution and reporting.

In this respect, work has been performed by running a few jobs implementing the above-mentioned changes and automation and a GITHUB repo has been created to track activities and speed-up communications.

The recent COVID-19 pandemic has heavily influenced the activities foreseen in the work plan for the present EAP application.

All the activities will be gradually resumed as soon as the key-personnel will be operational again, but a remodulation of the related roadmap shall be taken into consideration.

6.4 Towards an e-infrastructure for plant phenotyping

6.4.1 Towards an e-infrastructure for plant phenotyping

As an IT architect, I contribute to provide an european e-infrastructure for high throughput plant phenotyping data management. Such an objective will not be possible without the support and services offered by EGI Foundation.

The open-source Phenotyping Hybrid Information System PHIS (Neveu et al. 2019 New Phytologist, 221: 588–601) has been proposed to organize these data and make them accessible and reusable to a larger scientific community.

Three use cases have been proposed to explore which EGI services are the most appropriate to support an european plant phenotyping e-infrastructure.

6.4.2 Use cases analysis

US1. The PHIS information system and the Galaxy environment will be deployed on **EGI virtual machines.** The storage layer is based on the existing **FranceGrilles iRODS** infrastructure. An authentication layer based on the **EGI check-in service** and a computing layer provided with the **EGI Notebooks service** will be added (see Figure 1 for details).

First pilot

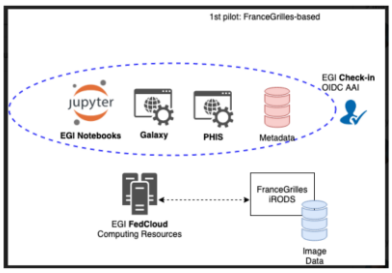

Figure 1: First pilot

U2. Compared to the previous pilot **the storage layer is based on the B2SAFE service** supported by the EGI infrastructure (see Figure 2 for details).

Second pilot

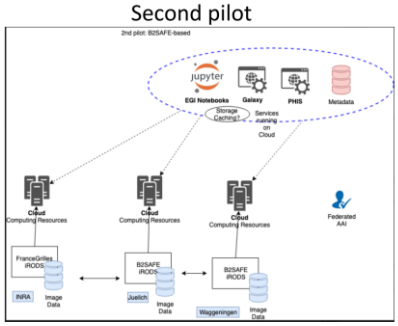

Figure 2: Second pilot

US3. Compared to the previous pilot the **storage layer is based on the Data Hub service** supported by the EGI infrastructure (see Figure 3 for details).

Third pilot

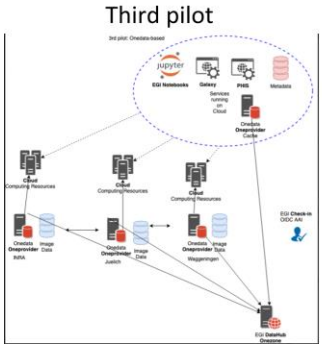

Figure 3: Third pilot

6.4.3 Handle technical requirements

US1:

- **RQ1** Deploy virtual machines: CESNET-MCC (or other sites if more performant), vo.emphasisproject.eu (EGI cloud compute):
    - 1VM; 4CPUs with 32GB RAM;
    - 80GB of storage for the system
    - + 100GB of additional storage (Mongodb)
- **RQ2** Install PHIS information system on the VMs - PHIS team
- **RQ3** Connect iRODS data with PHIS IS - PHIS team
- **RQ4** Deploy Jupyter Notebook: community-deployment for notebooks for 4 concurrent users (2 vCPUs cores, 4GB of RAM and 40GB of storage per notebook) - EGI
- **RQ5** Data available in Notebooks - PHIS team
- **RQ6** Deploy Galaxy environnements - Galaxy.eu
- **RQ7** Provide persistent identifier to the data - B2HANDLE EUDAT/GRNET
- **RQ8** Federated authentication should be integrated within PHIS IS. - Check-in EGI/GRNET

US2:

- **RQ9** Deploy virtual machines: CESNET-MCC (or other sites if more performant), including support with B2HANDLE vo.emphasisproject.eu: 1VM; 4CPUs with 32GB RAM; 80GB of storage for the system + 100GB of additional storage (Mongodb) - EGI cloud compute
- **RQ10** Install PHIS information system on the VMs - PHIS team
- **RQ11** Provide 10TB of storage in B2SAFE - B2SAFE – EUDAT/CINES
- **RQ12** Connect PHIS IS with B2SAFE - B2CONNECT – EUDAT/Juelich + PHIS team
- **RQ13** Deploy Jupyter Notebook: community-deployment for notebooks for 4 concurrent users (2 vCPUs cores, 4GB of RAM and 40GB of storage per notebook) - EGI
- **RQ14** Provide persistent identifier to the data - B2HANDLE EUDAT/GRNET
- **RQ15** Data available in Notebooks - PHIS team
- **RQ16** Deploy Galaxy environnements - EGI

US3:

- **RQ17** Provide virtual machines: IN2P3-IRES vo.emphasisproject.eu - 1VM for PHIS IS; 4CPUs with 32GB RAM; 80GB of storage for the system + 100GB of additional storage (Mongodb) - Oneprovider VM with 8 vCPU, 32GB RAM with SSD - EGI cloud compute
- **RQ18** Install PHIS information system on the VMs - PHIS team
- **RQ19** Provide 10TB of storage in EGI DataHub - EGI
- **RQ20** Support to Connect PHIS IS EGI DataHub - EGI

- **RQ21** Provide persistent identifier to the data - B2HANDLE EUDAT/GRNET
- **RQ22** Deploy Jupyter Notebook : community-deployment for notebooks for 4 concurrent users (2 vCPUs cores, 4GB of RAM and 40GB of storage per notebook) - EGI
- **RQ23** Data available in Notebooks - PHIS team
- **RQ24** Deploy Galaxy environments - EGI

6.5 Big Data Analytics for agricultural monitoring using Copernicus Sentinels and EU open data sets

The key aspect in the early adopter demonstrator is to show how federated EOSC resources can facilitate a range of Sentinel data applications across agricultural user domains. This extends the Copernicus DIAS concept, aimed at business users, to scientific and public users, by ensuring interoperability between EOSC resource providers and exposing the Copernicus high resolution Sentinel data archive with standardised processing services through tested standard interfaces.

6.5.1 User stories

- **US1** As a farmer, I want direct access to Copernicus Sentinel imagery that covers my parcels, so that I can check how they compare in crop development within and across sets
- **US2** In order to make a precise analysis as a user, I want the extracted imagery to be correctly georeferenced and available for the full history of acquisitions
- **US3** As an inspector in the Paying Agency, I want to generate consistent statistical time series for the whole region of interest, which may contain up to 1 million parcels, in a reasonable time frame (several days)
- **US4** As a scientist I want to be able to analyse time series from Sentinel-1 and Sentinel-2 for 3 types of summer crops in this area to understand how the signals relate to crop phenology trends
- **US5** As a service company, I want to develop a method to inventorize all wheat parcels in a production area, to help in the planning of the deployment of combines for the harvesting operations
- **US6** As a developer, I need to integrate imagery as GeoTIFFs and time series profiles as JSON formatted responses in my farm advisory service app.

6.5.2 Use cases analysis

- **UC1** Access to full Copernicus Sentinel archive. *Dependency*: One or more DIAS instances need to be federated
- **UC2** Marshall scalable compute resources to facilitate parallel analysis tasks
- **UC3** Provide transparent accounting across federated resource instances
- **UC4** Single sign on and allow transparent resource marshalling across federated resources
- **UC5** Provide (temporary) storage for caching essential data sets relevant for data analytics *Dependency*: Open data access platforms (e.g. national OGC compliant data servers)
- **UC6** Support interactive data analysis with advanced geospatial data visualization and libraries of advanced method (including ML)

- **UC7** Provide service endpoints (e.g. RESTful) for client-side access to both raw data selections and analysis results

6.5.3 Handle technical requirements

- **RQ1** EOSC-hub to provide access to the full Copernicus archive (UC1)
- **RQ2** EOSC-hub to provide single sign on to marshall compute resources (UC2, UC4)
- **RQ3** EOSC-hub to provide accounting services, across federated resources (UC3)
- **RQ4** EOSC-hub to provide scalable data transfer, storage for temporary data assets (UC5)
- **RQ5** EOSC-hub to provide multi-user data analytics platform (e.g. JupyterHub) and relevant data processing libraries (e.g. GDAL, python modules) (UC6)
- **RQ6** EOSC-hub to develop fast data access mechanism to block storage (e.g. smart caching, optimized storage formats) (UC2, UC5, UC7)

7 Early Adopters selected during the 2nd call

To continue the description of the supported applications within EAP, this section is giving information about identified user stories, use cases and technical requirements with the related status of the applications selected during the 2nd call of the Programme.

7.1 Towards a global federated framework for open science cloud

The ultimate goal of the project is to empower the scientific researchers from Africa and China to interact with EOSC services and data - publish data sets and data analysis services to EOSC Marketplace and access them from outside of Europe.

7.1.1 User stories

- **US1**: As a user, I want to use federated authentication mechanisms based on EGI AAI Check-In service to access the federated cloud infrastructure composed by EGI and CNIC CAS resources.
- **US2**: As a user, I want to use the OPENcoastS and DMCC+ services to analyse the high resolution (8 m) satellite data exposed by the CASEarth service for the simulation of tsunami, hurricane, typhoon, floods and extreme weather. The high resolution satellite datasets will be to predict the trajectory of typhoons.
- **US3**: As a user, I want to use sensor and satellite data to perform disaster assessments.
- **US4**: As a user, I want to use genomics datasets for analysing genetic make up of diseases.

7.1.2 Use cases analysis

To support the use cases described above, the following technical solutions have been identified:

- EGI Cloud Compute service,
- EGI Online Storage service,
- OPENCoastS service,
- DMCC service,
- AGROS service.

7.1.3 Handle technical requirements

To support this application, the following technical requirements have been identified:

- **RQ1**: The Chinese Academy of Science (CAS) IdP joined the eduGAIN federation. Thanks to this integration users can access the cloud resources of the CNIC infrastructure using the EGI AAI Check-In service. The integration of the CNIC infrastructure with the EGI Federation is still in progress.
- **RQ2**: To analyse the high resolution (8 m) satellite data the OPENcoastS has been extended in order to perform storm surges in the coast of Taiwan.

7.2 EOSC DevOps framework and virtual infrastructure for ENVRI-FAIR common FAIR data services

7.2.1 User stories

The project goal is to deploy a DevOps environment, with necessary capacity of Cloud Infrastructures and services for testing ENVRI-FAIR development. The project aims to automate the testing/integration of the FAIR data services developed by the teams in ENVRI-FAIR.
Three use cases have been identified:

- **Automated Cloud execution for data workflow**: In VREs, scientific workflow (including workflow logic, services, and data objects) need to be executed in a Cloud environment dynamically at runtime. The VRE needs to automate those steps: virtual infrastructure (networked VM) provisioning, software deployment (Dockers or RESTful services), workflow execution, runtime monitoring, and provenance.
- **Continuously testing and integration for ENVRI services:** ENVRI Knowledge Base ontology/demonstrator are described as RDF files in the git repositories, each update will trigger an automated testing/integration activity via the DevOps pipeline.
- **Notebook based environment for FAIR data access and processing:** ENVRI users (users of RIs/VRE) will perform their scientific research using data, software service and models. They often perform such kinds of activities using Jupyter notebook.

7.2.2 Use cases analysis

- **Automated Cloud execution for data workflow**
    - EGI Cloud Compute
    - https://marketplace.eosc-portal.eu/services/egi-cloud-compute
        - Access via IaaS API in order to test the VM provisioning, software deployment, Workflow Execution

- **Continuously testing and integration for ENVRI service**
    - Jelastic PaaS
      https://marketplace.eosc-portal.eu/services/jelastic-platform-as-a-service
      The PaaS will be installed on EGI Federated Cloud resources. The number of users accessing the service will be from 10 to 20.

- **Notebook based environment for FAIR data access and processing**
    - EGI Notebooks
      https://marketplace.eosc-portal.eu/services/egi-notebooks
        - possible GPU requirements
        - max 10 users in parallel
    - EGI DataHub
        - access to storage for sharing data between notebook users.

7.2.3 Handle technical requirements

- **Automated Cloud execution for data workflow**

    - **RQ1**: Create the VO vo.envri-fair.eu to grant users access to resources

- **RQ2**: Deploy 4VMs at INFN-CATANIA
  - 4 cores and 8G memory and 100GB storage
- **RQ3**: Deploy 1 VM at CESGA
  - 12 cores,16GB RAM and 1.5 TB storage.

- **Continuously testing and integration for ENVRI service**

  - **RQ4**: Deploy 2 VMs for Jelastic Installation (1 for infrastructure services, 1 for user services)
    - 8 vCPU cores, 24 GB RAM, 1TB storage
    - 12 vCPU cores, 24GB RAM, 1.5 TB storage
    - Multiple Public IPs associated to the VMs
    - dedicated domain with delegation (j.fedcloud.eu)
  - **RQ5**: Jelastic installation performed by Jelastic engineers

- **Notebook based environment for FAIR data access and processing**

  - **RQ6**: Deploy 4 VMs at INFN-CATANIA to extends the Kubernetes cluster hosting the EGI Notebook instance
    - 8 vCPUs, 16 GB RAM and 120GB storage each
  - **RQ7**: Deploy an instance of the EGI Notebooks dedicated to ENVRI over EGI Cloud computing service
  - **RQ8**: Provide storage on a EGI DataHub OneProvider available from the EGI Notebooks at CESGA
    - 1 VM, 8vCPU, 32GB RAM and 50 GB storage
    - 10 TB via NFS

7.3 AGINFRA+: virtual research environments to support agriculture and food research communities

7.3.1 User stories

The primary objective of AGINFRA+ project is to exploit cloud-based Virtual Research Environments (VREs) for the (a) Agro-climatic and economic modelling, the (b) Food safety risk assessment, and the (c) Food security research communities. The AGINFRA+ VREs rely on the operation of a DataMiner facility to provide big-data analytics features to researchers. Consequently:

- The DataMiner facility is required to be deployed in an Infrastructure as a Service (IaaS) cloud management framework.

The DataMiner facility shall be monitored in order to provide an accurate tracking of the analytics tasks being performed.

7.3.2 Use cases analysis

- The IaaS platform provides the means to expose the DataMiner as a long-running service.
  - The EGI Cloud Compute service has been the selected EOSC service in order to provide Cloud capabilities.

- The DataMiner resource consumption needs to be tracked.
    - The EOSC Accounting service is integrated with EGI Cloud Compute in order to facilitate compute and data usage.
- The DataMiner operation needs to be monitored.
    - The EOSC Monitoring service provides the means to check analytics tasks performance through the DataMiner Application Programming Interface (API)

7.3.3 Handle technical requirements

- RQ1: DataMiner cluster deployment in the EGI Cloud Compute service
    - Task complete using IFCA-LCG2 resource centre: 4 VMs with 16 vCPUs and 32G RAM, 250GB local disk space.
- RQ2: Tracking compute consumption
    - ~70K hours in period from Apr 15th to July 15th (source: EGI Accounting Portal)
- RQ3: Get performance metrics: number of analytics tasks
    - May: 36 analytics tasks (15K hours, 241 GBytes)
    - June: 3 analytics tasks (25K hours, 170 GBytes)
    - July: 1 analytics task (14K hours, 65 GBytes)
- RQ4: Get performance metrics: availability/uptime (percentage)
    - Task in progress
- RQ5: Display of performance metrics in Nagios
    - Task in progress

7.4 OpenBioMaps data management service for biological sciences and biodiversity conservation

7.4.1 User stories

The objective of OpenBioMaps is to create a service with EOSC that allows multiple users to run tasks that are above the level of a PC through the same interface. In fact, we would like to develop a “service in service” - specifically for projects that collect nature conservation and biodiversity data.

To serve these diverse tasks we need a fully configurable VM which let us deploy our service interface (API) which will be available in the OpenBioMaps Network and provide computation capacity access to the involved projects.

According to our recent experiences in our PC based local computational cluster, the number of processors is the most important in these ecological analyzes. A “typical” analysis is now running at an acceptable rate on 16 threads. The parallel computing requirements of image analysis can be much higher, and GPU usage can be interesting there. Some analyzes, for example, genetic analyzes or larger spatial analyzes require a lot of memory.

7.4.2 Use cases analysis

The OpenBiomaps system is deployed as a single VM on EGI Cloud Compute. A single VM will attach two pre-created volumes for persistence with a total size of 2TB. Services are launched using docker compose.

It will be evaluated the usage of the following services:

- B2FIND
- EGI DataHub
- B2DROP

7.4.3 Handle technical requirements

- **RQ1**: Deploy virtual machines on EGI Cloud Compute: @IFCA-LCG2, vo.lifewatch.eu: 1VM; 48CPUs with 96GB RAM; 2TB of storage for the system.
- **RQ2**: Install docker and docker compose on the VM.
- **RQ3**: Launch and configure OpenBiomaps services using docker compose.
- **RQ4**: Create automated recipes and TOSCA template to deploy the infrastructure through the Infrastructure Manager.

7.5 VESPA-Cloud

VESPA (Virtual European Solar and Planetary Access) is a mature project, with 50 VESPA providers distributing open access datasets throughout the world (EU, Japan, USA). In October 2019, the current number of data products available within the VESPA network reaches 18.3 millions (among which 5 millions products from the ESA/PSA, Planetary Science Archive).

The VESPA team is supported by the Europlanet-RI-2024 project (started on Feb 1st 2020 for 48 months, H2020 grant agreement No 871149).

Each VESPA provider (institutes, scientific teams...) is hosting and maintaining a server (physical or virtualized) with the same software distribution (DaCHS, Data Centre Helper Suite), which implements the interoperability layers (from IVOA, International Virtual Observatory Alliance, and VESPA) and following FAIR principles. Each server hosts a table of standardized metadata with URLs to data files or data services. Data files can be hosted by the VESPA provider team, or in an external archive (e.g., ESA/PSA - Planetary Science Archive).

The VESPA architecture relies on the assumption that data provider’s servers are up and running continuously. The VESPA network is distributed but not redundant. For small teams with little or no IT support available locally, the services are down regularly. Thus a more stable and manageable platform for hosting those services is needed. The EOSC-hub “cloud container compute” service would solve this problem.

The usage of the EOSC infrastructure to host VESPA provider's servers (through a controlled deployment environment with git-managed containers) is proposed.

The open-source DaCHS framework is developed for Debian distribution. A docker containerization will be used to facilitate the framework deployment on other Linux environments.

7.5.1 User stories

Identifier	Description
Overarching goal	
US0	I am a small data provider willing to join VESPA, but I don't have any IT support team to implement and maintain a VESPA server. I want to share my data products so that the community can find and use my datasets. Thanks to VESPA-Cloud, I can set up a VESPA data service, focussing on the science interface, and let the VESPA team manage the server.
As VESPA-Cloud admin	
US1	As a VESPA-Cloud administrator I want to create a pre-configured science-enabled VM to be instantiated on demand for VESPA providers and dynamically configured for them.
US2	As a VESPA-Cloud administrator I want to manage authentication and authorisation to e-infrastructures (EGI and EUDAT) resources using eduTEAMS as my community AAI.
US3	As a VESPA-Cloud administrator I want to manage authentication and authorisation to the VESPA-Cloud server instance via SSH or HTTPS using eduTEAMS as my community AAI.
US4	As a VESPA-Cloud administrator I want to monitor the status of my deployments and be warned in case of problems.
US5	As a VESPA-Cloud administrator I want to collect accounting information about access to the scientific data.
As a VESPA provider	
US6	As a VESPA provider I can configure my VESPA metadata ingestion scripts on the VESPA gitlab server and push it to my VESPA-cloud instance.
US7	As a VESPA provider, I can upload my data via EUDAT/B2SAFE (iRODS) or object storage.
US8	As a VESPA provider I want to access a pre-configured Virtual Machine and access it over SSH to process my data accessed via EUDAT/B2SAFE (iRODS) or object storage.
US9	As a VESPA provider I can order a VESPA-Cloud service through the EOSC Marketplace.

US10	As a VESPA provider I want to have my service registered and harvested by B2FIND and IVOA registry to make it discoverable.
As a VESPA end-user	
US11	As a VESPA end-user I want to discover VESPA-Cloud data products through VESPA client and include them in my astronomy pipeline.

7.5.2 Use cases analysis

The table below contains the various use cases that will have to be implemented and document the services they rely on.

Identifier	**Description of action**	**Dependency on 3rd party services (EOSC-hub or other)**
UC1	VESPA-Cloud administrators pre-configure a VM image with the VO framework.	CloudCompute, Check-in, eduTEAMS
UC2	VESPA-Cloud administrators setup auto-configuration of VM on deploy for a specific VESPA provider.	CloudCompute, Check-in, eduTEAMS
UC3	VESPA-Cloud administrators manage access to the VM instance via SSH or HTTPS using eduTEAMS	CloudCompute, Check-in, eduTEAMS
UC4	VESPA-Cloud administrators access and receive monitoring notifications for the VESPA-Cloud applications and VM instances.	CloudCompute, Check-in, eduTEAMS, ARGO
UC5	VESPA-Cloud administrators access accounting information about the scientific data in the VESPA-Cloud VMs.	CloudCompute, Check-in, eduTEAMS
UC6	VESPA providers order VESPA-Cloud service through the EOSC Marketplace.	CloudCompute, Check-in, eduTEAMS, EOSC Marketplace, SOMBO
UC7	VESPA Providers access a VM via SSH with the VO framework installed	CloudCompute, Check-in, eduTEAMS
UC8	VESPA Providers access their data in the VM via VESPA Gitlab, B2SAFE or object storage.	CloudCompute, Check-in, eduTEAMS, B2ACCESS, B2SAFE
UC9	VESPA end-users discover VESPA-Cloud data products through VESPA client.	B2FIND, IVOA

7.5.3 Handle technical requirements

- **RQ1**: Integrate eduTEAMS as community AAI and EGI Check-in as e-infrastructure AAI
- **RQ2:** Integrate eduTEAMS as community AAI and EUDAT B2ACCESS as e-infrastructure AAI
- **RQ3:** Interact with EGI ComputeCloud resources via authorisation managed in eduTEAMS
- **RQ4:** Use eduTEAMS to manage access to SSH on EGI ComputeCloud VM using SSH keys
- **RQ5:** Monitor the VESPA Cloud applications using ARGO
- **RQ6:** Collect accounting information about the scientific data in the VESPACloud VMs
- **RQ7:** Allow to order VESPA-Cloud instances via the EOSC Marketplace
- **RQ8:** Access user data from the VM via VESPA gitlab
- **RQ9:** Access user data from the VM via B2SAFE
- **RQ10:** Access user data from the VM via object storage
- **RQ11:** Make VESPA-Cloud produced data discoverable via B2FIND and IVOA.

While VESPA-Cloud progressed on various points (access to most of the resources was tested by the VESPA-Cloud team, SLA for those resources are already in place, cross provider integration of AAI already took place,...), the overall implementation is still being discussed and refined, and to accommodate technical integration there could be some variations in how things will be eventually implemented.

7.6 Open AiiDA lab platform for cloud computing in materials science

The AiiDA lab brings the AiiDA workflow manager for computational science (www.aiida.net) to the cloud. While domain experts can install AiiDA on their own hardware, the AiiDA lab web platform gives novice users access to their personal pre-configured AiiDA environment in the cloud. AiiDA is a workflow manager for computational science with a strong focus on provenance, performance and extensibility. When executing a workflow, AiiDA records the provenance − calculations performed, codes used and data generated − in a directed acyclic graph tailored to provide full reproducibility of any given result. The AiiDA engine relies on a message queue in order to support high-throughput use cases of up to 50k calculations per hour, and the relational database backend enables performant queries on graphs of tens of millions of nodes. AiiDA (TRL 7-8) is used in production for high-throughput calculations.

The pilot will further develop AiiDA lab to:

- Provide an open AiiDA lab for researchers in Europe capable of supporting ~100 concurrent users
- Support of the order of ~1000 users in the system
- Test scalable kubernetes set-up so resources can be adjusted to the load as required.

7.6.1 User stories

US1 As a computational scientist I want to login into the AiiDA lab using my institutional credentials.

US2 As a computational scientist, I need to have a personal space to store data that persist between logins.

US3 As a materials scientist I would like to use the AiiDA lab to run and manage materials science workflows on remote compute resources.

US4 As a computational scientist I would like to use the AiiDA lab to participate in AiiDA tutorials so that I do not need to set up the software on my local machine.

7.6.2 Use cases analysis

UC1 Access via institutional credentials to the service (Dependency on EOSC-hub AAI)

UC2 Provide an online version of AiiDA lab completely accessible via browser that does not require installation of software

UC3 Access to a persistent storage pool for users personal space (Dependency on EGI Online Storage)

UC4 Access to compute resources to run users workload (EGI Cloud Compute)

UC5 Provide a kubernetes deployment to manage access to compute and storage resources for the AiiDA lab platform (Dependency on EGI Cloud Container Compute)

UC6 Provide scalable kubernetes setup that can adapt to workload during trainings (Dependency on EC3)

7.6.3 Handle technical requirements

RQ1 EOSC-hub to provide a kubernetes managed pool of resources to deploy AiiDA lab (UC5, UC6)

RQ2 EOSC-hub to provide a single sign-on services that supports institutional credentials (UC1)

RQ3 EOSC-hub to provide a scalable kubernetes cluster to accommodate varying load during training events (UC6)

RQ4 EOSC-hub to provide enough resources to host planned training events (UC2, UC3, UC4)

7.7 Supporting FAIR data discoverability in clinical research: providing a global metadata repository (MDR) of clinical study object

7.7.1 User stories

In order to fully assess and review the evidence generated in clinical research, it is necessary to have access not only to the published results but also to the source individual participant data and related study documents (e.g. study protocol, statistical analysis plan, case report form).

As data and document sharing becomes more and more common, however, the researcher is faced with a bewildering mosaic of possible source locations and access modalities. There is an urgent need to develop a central resource that can catalogue all the diverse data and documents associated with a clinical study, and then make that information searchable by using a central web portal.

In this pilot the following user stories are of concern:

- As a researcher I want to find discover additional sources of a clinical study through a centralized service
- As a researcher I want to find additional documents of a clinical study through a centralized service
- As a general user I want to discover the MDR portal and its sources in a general discovery service

7.7.2 Use cases analysis

The use cases of concern here are captured in the activities that are taken to enable the various services and their integrations.

Provide access to published results, source individual participant data and related study documents
The MDR portal will provide a user with information and links to additional source material of a clinical study.

This case's activities include:
- Revise web-portal
- Investigate a new mechanism of ECRIN metadata 'injection' and upgrading on OneData platform

Develop central catalogue resource
This case's activities include:
- Extend the MDR demonstrator to run in production in the EOSC environment
- Developing ElasticSearch-based APIs

Make publications searchable using web portal
This case's activities include:
- Make MDR demonstrator part of the EOSC catalogue
- Collect data on user action and users feedbacks
- Enable harvesting by EUDAT B2FIND

7.7.3 Handle technical requirements

The ECRIN MDR system is deployed as a single virtual machine for the production instance. In addition, a development instance will be available for development purposes. The instances will be connected to a database server that contains the gathered information.

It will use the following services:

- EGI OneData
- INFN ElasticSearch + hardware
- B2FIND

The success of the pilot will be partly due to the establishment of a generic metadata schema that can be used in the MDR and for exchange with EUDAT's B2FIND service.

- **RQ1**: Deploy VMs for revision of current web-portal
- **RQ2**: Establish integration of portal with ElasticSearch server

- **RQ3**: Enable harvesting of a single MDR instance by B2FIND instance

7.8 Integration of toxicology and risk assessment services into the EOSC marketplace

OpenRiskNet is providing an e-infrastructure to the communities involved in safety assessment, including toxicology and especially predictive toxicology, systems and structural biology, bioinformatics and its subtopics toxicogenomics, cheminformatics, biophysics and computer science specifically targeting the EU's chemical manufacturing industries, e.g. pharmaceutical companies, chemical and agrochemical industries and cosmetic industries, and the corresponding regulatory agencies.

During the lifetime of OpenRiskNet (2016-2019), 45 services were integrated, that can be grouped into seven categories: 1) Toxicology, Chemical Properties and Bioassay Databases, 2) Omics Databases, 3) Knowledge Bases and Data Mining, 4) Ontology Services, 5) Processing and Analysis, 6) Predictive Toxicology and 7) Workflows, Visualisation and Reporting. The service integration also included the development of workflows to support the case study work by automating complex tasks only achievable by the combination of multiple services. Additional services will be integrated over time to complete the portfolio to allow full risk assessment of chemical compounds and nanomaterials. This is specifically supported by the infrastructure project NanoCommons developing, and harmonizing such services for the nanosafety starting community represented by the projects of the EU NanoSafety cluster.

7.8.1 User stories

The general OpenRiskNet infrastructure and specific services are already listed on the ESOC marketplace. However, a better integration, harmonisation and interoperability with other EOSC services are anticipated. This is started by the early adopter pilot concentrating on the four user stories:

- **US1**: As a user, I want a public instance of the OpenRiskNet/NanoCommons virtual environment (VE) to test and get first experience with the data and modelling services provided by OpenRiskNet/NanoCommons, especially also in connection to other EOSC service, before deploying a VE internally or ask for additional EOSC resources.
- **US2**: As a user, I want to use federated authentication mechanisms to access the public VE, including authentication through the EGI Check-In when desired.
- **US3**: As a toxicology data and modelling service provider, I want to have my service listed in the EOSC marketplace based on information I provided to OpenRiskNet/NanoCommons.
- **US4**: As a toxicology data service provider, I want to find solutions to index, link and make the data available via general EOSC services like B2FIND and OpenAIRE improving their FAIRness.

7.8.2 Use cases analysis

EOSC services best supporting the user stories have been identified and service requests have been submitted. This are:

- **US1:** cloud resources provided by INFN-BARI as previously agreed (200 CPU cores, 360 GB of RAM, and ~1TB of disk space)
- **US2:** EGI Check-in as an single-sign-on solution
- **US3:** interactions with EOSC teams related to the information collection and representation of the EOSC marketplace
- **US4:** interactions with teams from B2FIND and OpenAIRE to discuss how to handle data managed by specialised solutions provided and used standard data management systems by the toxicology community.

7.8.3 Handle technical requirements

OpenRiskNet is already operating a public VE to promote its services and make them available for testing, which is hosted at the Johannes Gutenberg Universität Mainz and will be available until mid 2021. For long-term sustainability and better integration into the EOSC infrastructure, transfer to EOSC resources is planned and the early adopter pilot is performing analysis of the steps needed to do so. Unfortunately, due to a security breach at the German research computing infrastructure, which also had a huge impact on the system in Mainz, the VE was down for over two months. This prohibited many of the tasks planned for the early adopter pilot. However, we were able to already organise the interactions with the EOSC service providers and are now able to continue with the work, hopefully making up for the lost time. Specific requirements, which have already been addressed are:

- **RQ1**: OpenRiskNet already used GitHub and LinkedIn as social account providers. Compatibility of the system with EGI Check-in as additional e-infrastructure AAI has been verified and the service access request has been approved.
- **RQ2**: The service access request for cloud resources has also been approached. The migration to the EOSC system should also include the update of the used containerisation, orchestration and monitoring systems. Needed updates as well as replacements of specific tools have been evaluated and phrased as requirements for the new deployment, including e.g. the benefits and disadvantages of using the extensions of OpenShift.

8 Update on the procedure used to provide technical support

As from D10.5 - the procedure to provide technical support (SOCRM-04) has been defined as part of the Service Order and Customer Relationship Management (SOCRM) process of the EOSC SMS. Aim of this procedure is creating a well-defined communication channel between the technical support team and first line of support working with requests from user communities received through the EOSC Portal, identified by the Stakeholder Engagement activity (T3.2) or by other channels.

The procedure has been deeply tested to support different use cases and different applications as described in Sections 3, 4 and 5 and to provide material for technical support and an homogeneous communication channel.

The procedure is based on the following key concepts and professional figures resumed in the following sections.

8.1 Technical support

Each community has assigned a unique WP10 contact. The main role of this responsible person, named shepherd, is to drive the integration of the service in the Hub by bridging users and resource or service providers by providing a proper use case analysis, according to the specific needs of the customers, and harmonizing the coordination of the technical activities.

8.2 Community Requirements DB

The primary source of information is the Community Requirements DB[1]. Each entry in the database points to a document containing an analysis of the technical requirements of a use case. The structure of this document has three main sections corresponding to the three core components of the agile process as reported in Section 4 and 5: user stories, use cases and technical requirements.

Whenever a new use case or application comes in, the shepherd initially holds an interview with the user community's main contacts in order to build a shared understanding that will yield to the elaboration of the user stories. In subsequent iterations, the shepherd will break down the stories into use cases and use cases into technical requirements. The whole process will be transparent for the customers as the shepherd works directly in the Community Requirements DB.

8.3 EOSC Portal Catalogue and Marketplace

The EOSC Portal Catalogue and Marketplace[9] is an integrated platform that allows easy access to lots of services and resources for various research domains along with integrated data analytics

[9] https://marketplace.eosc-portal.eu/

tools. Browse by scientific domain, service category or provider the EOSC Portal represents a single point of access able to provide

- Information about the EOSC governance and players, the projects contributing to its realisation, funding opportunities for EOSC stakeholders, relevant European and national policies, important documents, and recent developments.
- A gateway to the multitude of services and resources offered by the providers to the researchers.