



EOSC-hub

D10.2 EOSC-hub Technical Roadmap v2

Lead Partner:	INFN
Version:	1
Status:	FINAL
Dissemination Level:	Public
Document Link:	https://documents.egi.eu/document/3705

Deliverable Abstract

This deliverable introduces the second version of the Technical Roadmap for the EOSC-hub services. It describes the capabilities, the features and the plan for the technical evolution of the services within the EOSC hub service portfolios, with a focus on the activities needed to improve the interoperability among the services and to enable service composition in the wider EOSC environment. Plans depicted in this deliverable will be implemented in EOSC-hub follow-up projects, notably those funded under the INFRAEOSC-03 (EOSC Future) and INFRAOESC-07 calls (EGI-ACE, DICE, etc.).



COPYRIGHT NOTICE



This work by Parties of the EOSC-hub Consortium is licensed under a Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>). The EOSC-hub project is co-funded by the European Union Horizon 2020 programme under grant number 777536.

DELIVERY SLIP

<i>Date</i>	<i>Name</i>	<i>Partner/Activity</i>	<i>Date</i>
From:	Giacinto Donvito Alessandro Costantini	INFN/WP10 INFN/WP10	11/03/2021
Moderated by:	Sjomara Specht	EGI Foundation/WP1	
Reviewed by:	Matthew Viljoen Anabela Oliveira	EGI Foundation INEC	
Approved by:	AMB		

DOCUMENT LOG

<i>Issue</i>	<i>Date</i>	<i>Comment</i>	<i>Author</i>
v.0.1	14/06/2020	ToC ready	Alessandro Costantini Giacinto Donvito Diego Scardaci
v.0.2	10/02/2021	Document ready for review	Alessandro Costantini Giacinto Donvito Diego Scardaci
v.0.3	17/02/2021	Document reviewed	Matthew Viljoen Anabela Oliveira
v.0.4	04/03/2021	Suggestion from reviewers addressed by the authors	Alessandro Costantini Diego Scardaci
v.0.5	06/03/2021	Document reviewed	Matthew Viljoen Anabela Oliveira
v.1	11/03/2021	Final version	Alessandro Costantini Giacinto Donvito Diego Scardaci

TERMINOLOGY

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

EOSC	European Open Science Cloud
AAI	Authentication and Authorisation Infrastructure
AARC initiative	Authentication and Authorisation for Research and Collaboration
SAML	Security Assertion Markup Language
OIDC	OpenID Connect
CODATA	COMmittee on DATA
RDA	Research Data Alliance
TCOM	Technology Committee
OAIS	Open Archival Information System
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
BDII	Berkeley Database Information Index
gRPC	gRPC Remote Procedure Calls
AMS	Argo Messaging Service
AWS	Amazon Web Services
SQA	Software Quality assurance
EIF	European Interoperability Framework
SRIFTI	Security Incident Response Trust Framework for Federated Identity
DPA-CoCo	Data Protection Code of Conduct
WISE-AuP	WISE Baseline AUP

Contents

1	Introduction.....	6
2	Service integration and composability	7
2.1	Technical Specifications and Interoperability guidelines: state of the art	7
2.1.1	Common services	7
2.1.2	Federation services	12
2.2	Participation to standardization bodies.....	18
3	Resource access and integration in the EOSC Portal & Marketplace.....	19
3.1	Discoverability	19
3.2	Access.....	23
3.3	Marketplace projects.....	24
4	EOSC-hub services roadmap - Evolution	26
4.1	Plans for service evolution in the EOSC-hub technical areas (TCOM)	26
4.2	Thematic Services roadmap.....	38
5	Integration activities	43
5.1	Thematic services	43
5.2	Competence Centers	45
5.3	EOSC Early Adopter Programme	46
5.4	Business pilots	47

Executive summary

In this deliverable, the final version of the Technical Roadmap for the EOSC-hub services is presented and described together with the capabilities, the features and the technical evolution of the services achieved during the project for the services within the EOSC-hub portfolios. With respect to the first version, the focus here is on the activities put in place to improve the interoperability among the services and to enable service composition in the wider EOSC environment. Considering the EOSC-hub project is approaching its end, the implementation of the plans described in this document is expected to happen in the follow-up projects, notably those funded under the INFRAEOSC-03 (EOSC Future) and INFRAEOSC-07 calls (EGI-ACE, DICE, etc.).

It has to be pointed out here that the Technical Roadmap, as it has been defined in the D10.1 “EOSC-hub Technical Roadmap v1”¹, has primarily been shaped taking into account the requirements collected from user communities within (WP7 thematic services, WP8 competence centers, WP9 business pilots) and outside of the projects (e.g. communities that requested services and technical support in the EOSC Portal and Marketplace or that have been identified in the context of the stakeholder engagement programme). Moreover, the effective collaboration between the technical support team (T10.3) and the product teams working on service development (WP5, WP6 and WP7) played an important role by allowing the inclusion of the users’ requests into the roadmap and the release of new versions of the services satisfying such needs.

The technical solutions carried out during the project lifetime, and defined in the Technical Roadmap, have been able to lower the barriers to integrate and compose services, enabling the combined use of multiple services to effectively deal with common issues of scientific applications, and foster the sharing of services between adjacent communities. The adoption of well-known standards and protocols, that simplifies and reduces the costs to combine services, the inclusion of end-to-end integration activities and the increase of the set of services that can be jointly exploited, open the possibility to combine different solutions able to satisfy common and relevant user requirements. These solutions are now available in the EOSC service portfolio, ready to be used by a large set of communities.

Due to the need to properly promote the evolution of the EOSC-hub technical solutions depicted in the roadmap, the EOSC-hub access channels also needed to evolve. In particular, the EOSC Service Catalogue and Marketplace, as the main EOSC access channels, have been enriched with a series of features to simplify the discovery, usage and exploitation of the EOSC resources, and facilitate and promote the combined usage of multiple resources, assisting the users in identifying and composing compatible or integrated resources.

As a result of all these analyses, an improved version of the roadmap for the most relevant EOSC-hub services has been depicted, including those for the thematic services and their integration with federation and common services.

¹ <https://www.eosc-hub.eu/deliverable/d101-eosc-hub-technical-roadmap>

1 Introduction

As the first version of this deliverable, the document focuses on the principal features of the EOSC-hub services and their planned evolution and on the large effort that has been done to enhance service interoperability and foster their composition. Follow up of these activities have been planned in new EOSC projects started (or close to start) in 2021, notably those funded under the INFRAEOSC-03 (EOSC Future) and INFRAOESC-07 calls (EGI-ACE, DICE, etc.).

The document is organised as follows:

- Section 2 summarises the state of the art of the technical specifications and interoperability guidelines developed during the project activities.
- Section 3 contains the description of the activities to enrich the EOSC Service Catalogue and Marketplace, the main EOSC access channels, and a series of features to improve discoverability, accessibility and facilitating the combination of additional services.
- Section 4 presents the roadmap evolution for the most relevant EOSC-hub services including the integration activities between Thematic and federated/common ones.
- Section 5 describes the service integrations/compositions achieved by the project.

2 Service integration and composability

EOSC-hub adopted a twofold approach in order to make service integration and composability in the EOSC a reality.

The EOSC-hub technology committee developed a set of EOSC technical specification and interoperability guidelines to identify standards, well-known interfaces and best practices to facilitate the service integration in EOSC. These guidelines cover both common/horizontal services (part of the future EOSC Exchange) and federation/core services (part of the future EOSC Core) and can be considered the basis on which building the EOSC Interoperability Framework.

In addition, technical experts from the project participated in the activities of a series of standardisation bodies like the Research Data Alliance (RDA), the Storage Networking Industry Association (SNIA), the HEP Software Foundation (HSF), etc. to contribute to the enhancement of the standards taking into account the EOSC community requirements².

This section describes these activities presenting the most relevant work done in the two aforementioned areas.

2.1 Technical Specifications and Interoperability guidelines: state of the art

This section describes the most relevant technical specifications and interoperability guidelines developed by EOSC-hub. They are classified according to the EOSC-hub technical areas they belong to.

The technical specifications are publicly available in the EOSC-hub website³.

2.1.1 Common services

2.1.1.1 Cloud Compute and container orchestration⁴

Cloud Compute - IaaS VM Management

This specification covers those services that provide on-demand API-based access to computing resources as Virtual Machines that can run user-defined arbitrary software (including operating systems and applications). Services in this category also allow management of block storage that can be associated to the VMs and network management to provide connectivity between VMs and external networks. This kind of computing resources allow for full customisation of the software to execute and the hardware environment to use (e.g., number of cores, RAM, disk) and hence potentially allows users to deploy and execute any kind of application.

Cloud Compute - IaaS Container Management

² <https://wiki.eosc-hub.eu/display/EOSC/Community+requirements+DB>

³ <https://www.eosc-hub.eu/technical-documentation#overlay-context=>

⁴ <https://marketplace.eosc-portal.eu/categories/compute>

The IaaS Container Management services provide on-demand API-based management of container-based applications. These services support the (Automated) Orchestration of container-based applications which manage the deployment of a complete lifecycle of the containers that compose an application into a set of computing resources. Similarly, to the IaaS VM Management, any container-based application could be executed with complete control on the resources to use.

Cloud Compute - IaaS Orchestration

The deployment of applications on IaaS services requires the coordinated management of different kinds of resources. The IaaS Orchestration services cover tools that automate the deployment of applications by executing a set of tasks that interact with the cloud services to start Virtual Machines, create storage devices and objects, configure networking and any other kind of related services to install and run the application. These tools normally use some sort of domain specific language or script that defines your application deployment process, TOSCA is the main standard in this area.

2.1.1.2 HTC/HPC³

HTC/HPC Compute - Multitenant job submission

Most scientific challenges require running computationally demanding tasks. Typically, these computing challenges can be tackled by gathering several computing resources that concurrently run the tasks. In some cases, the computational problem can be addressed by multiple and loosely coupled tasks that can run over different data blocks or different parameter sets, and in some cases, the problem requires gathering several computing elements together to solve every single task in a closely coupled parallelism. The former is addressed through High-Throughput Computing (HTC) execution approaches and the latter by the High-Performance Computing (HPC) ones. In the HTC/HPC Compute TCOM we address services for running a large set of independent tasks and to jointly use several computing resources to run a parallel job.

In this specification, we expose a macro-feature for Multitenant job submission, which relates to the capability of submitting HPC/HTC jobs with predefined constraints (both at resources and software) without a previously deployed virtual infrastructure. This service should be able to run a bunch of batch job on HTC/HPC and cloud compute resources, interfacing with storage solutions and seamlessly integrated with the authentication mechanisms. The main difference between HPC and HTC jobs is the requirement of multiprocessing (OpenMP or MPI for example). Some sites support both types of jobs, using different queues and specifications in the batch job to differentiate and provision the most appropriate resources.

HTC/HPC Compute - Multitenant Containerised job submission

The containerised job submission is an important feature to run computationally demanding tasks. In the last years, the use of containerised jobs has boosted due to the enormous convenience of containers for application delivery. Application dependencies are embedded into the containers reducing the effort and side-effects of the installation of software. However, popular container technologies such as Docker use daemon processes that run only for privileged users, which is not

acceptable by many data centre policies. There are solutions for running jobs on containers that run-in user space (such as u-docker). This approach reduces the capabilities of a containerised job to those of the user running the job, which makes it suitable for HPC, HTC and Cloud Compute infrastructures.

This macro-feature is complementary to the Multitenant Job Submission⁵ and it should be considered as an extension.

HTC/HPC Compute - HPC/HTC clusters on demand

Scientific portals expose convenient interfaces that typically run partially customisable jobs on computing infrastructures. Scientific portals normally have a higher computational demand on their back-ends than conventional portals, so they have to be provisioned with enough resources to deal with the potentially unexpected workload peaks. However, users in scientific portals could also be prepared to accept longer delays on retrieving the results, as they are more used to queuing systems.

In this regard, we identify the need for provisioning self-managed elastic clusters supporting mainstream job managers such as PBS, SLURM and especially scheduling systems based on Kubernetes resource orchestrators. Opposite to the multitenant job management service, this service will explicitly deploy a single-tenant cluster backend to be used by the user community managed by the user who deployed it.

2.1.1.3 Metadata Management and Data Discovery⁶

Metadata Cataloguing and Indexing

Metadata Cataloguing and Indexing comprise the entire metadata ingestion workflow, i.e.

1. Metadata harvesting from community repositories.
2. Metadata mapping on common schemas including curation and validation and
3. Uploading and indexing of metadata records in the metadata catalogue, to enable Data Discovery and Access, see related macro feature⁷

Data Discovery and Access

Data Discovery and Access comprises the ability for end-users to search for data resources and access the referenced data. This functionality requires and is based on the existence of an indexed metadata catalogue (see macro feature Metadata Cataloguing and Indexing).

Annotation Service

The Annotation Service enables end-users to extend descriptions of datasets or parts of datasets with user-defined content, without modifying the underlying dataset, e.g., adding comments, free text keywords or semantic tags (keywords from ontologies). These annotations can be used to

⁵ https://docs.google.com/document/d/166AcIBzyk5GrwKPliKdWzzCu_LYjmS77R-zBuF0b4k/edit#

⁶ <https://marketplace.eosc-portal.eu/categories/data-management>

⁷ <https://wiki.eoschub.eu/display/EOSCDOC/Metadata+Management+and+Data+Discovery?src=contextnavpagetreemode>

search and aggregate datasets or parts of datasets into user- defined datasets, either localised in a unique data repository or throughout a heterogeneous and distributed set of data repositories.

2.1.1.4 Data Platforms for Processing⁸

Onedata is a distributed *eventually consistent* virtual filesystem, providing a unified namespace for user data and supporting several backend storage solutions including POSIX, Ceph, GlusterFS, S3, Swift and WebDAV. Onedata is the basis for the EGI DataHub service, providing large-scale reference data sets, easily accessible on worker nodes, including virtual machines and containers.

Onedata has already been integrated with several EOSC-hub services in the previous reporting period including B2ACCESS - via the OpenID protocol, B2STAGE - through implementation of WebDAV storage driver, B2HANDLE - by adding support for automatic PID minting when publishing open data sets via Onedata, B2FIND - through the OAI-PMH protocol.

2.1.1.5 PaaS Solutions³

The PaaS (Platform as a Service) solution adopted in this project allows the users to deploy virtualised computing infrastructures with complex topologies (such as clusters of virtual machines or applications packaged as Docker containers) using standardized interfaces based on REST APIs and adopting the TOSCA (Topology and Orchestration Specification for Cloud Applications) templating language for the description of Cloud-based applications. The PaaS layer features advanced federation and scheduling capabilities ensuring the transparent access to the different IaaS back-ends including on-premises Cloud Management Frameworks such as OpenStack and OpenNebula, public Cloud providers such as Amazon Web Services and Microsoft Azure and, finally, Container Orchestration Platforms such as Apache Mesos and Kubernetes. The selection of the best cloud provider to fulfil the user request is performed considering criteria like the user's SLAs, the services availability and the data location.

2.1.1.6 Workflow management, user interfaces and Data analytics⁷

Marketplace

Marketplace is a dedicated platform where services are presented to the users and made available to get access to. Is a place where the Service Organisations can define and present to the users dedicated service offers, users can issue an order for those offers and handle different phases of the ordering process. Along with SPMT supports Service Management, and along with Service Order Management Back Office it provides Service Order Management in EOSC-hub.

⁸ <https://marketplace.eosc-portal.eu/categories/processing-analysis>

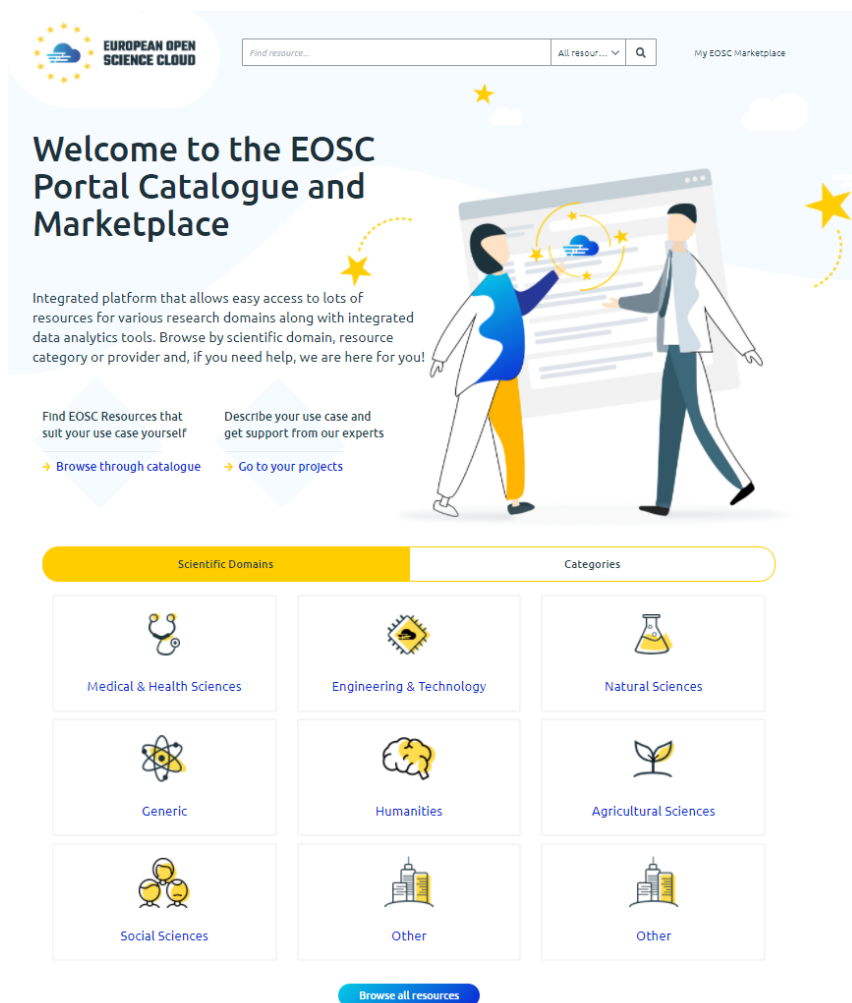


Fig. 1. The EOSC Portal Service Catalogue & Marketplace

Machine Learning/Deep Learning data analytics services

The data analytics service provides with the required¹ tools and mechanisms to build, encapsulate and execute Artificial Intelligence, Machine Learning and Deep Learning applications across different platforms, covering the whole development life-cycle. This cycle comprises the phases of model creation, training, test and evaluation, as well as model publication, serving (as a service) and sharing. It includes both CLI libraries as well as an SDK for developers to ease building and composing application architectures to be deployed on the Cloud. It should provide links to the existing storage and data management solutions, in order to provide efficient data access to the required data sets. Standards and best practices should be encouraged for the whole process, in order to ensure interoperability across the served applications and models. A model exchange should be desirable, in order to publish and share, with common metadata schemas, the built applications.

2.1.2 Federation services

2.1.2.1 Accounting

The EOSC Accounting service collects, stores, aggregates, and displays usage information of HTC compute, storage space, cloud VM and data set resources. This usage data is collected from the Resource Centres of the EOSC infrastructure.

Accounting information is gathered from distributed sensors into a central Accounting Repository where it is processed to generate summaries that are made available through an Accounting Portal. Depending on the use case the data may go via intermediate repositories that collate accounting data for particular regions, infrastructures or communities.

The Accounting Repository has a database backend and needs to ensure the exchange of accounting information with peer e-Infrastructures. The Accounting Portal receives and stores the resource centre, user, and user groups (e.g., Virtual Organisation/VO) level aggregated summaries generated by the Accounting Repository and provides views via a web portal. For example, by grouping resource centres in a country on specific time intervals a customized view can be generated and displayed. The databases are organized into a resource record database (e.g., CPU, storage, dataset, etc), a User record database, and a topology database.

The main features of the EOSC Accounting can be grouped by two target groups.

Main features offered to the user are:

- Aggregated views of their usage wherever that usage occurred.
- Views that allow usage to be checked against allocation.

Features for resource providers:

- Provider-centric views of resource usage by users.
- Views that allow comparisons to be made between resource providers within and between regions and communities.

The prototype accounting system for EOSC delivered by EOSC-hub has been implemented following these specifications.

- Management of incident or disruption of Hub services
- Interface for communicating with other service providers ticketing systems
- First level support for EOSC integrated services as a service
- Interface with a Known Errors Database and with a Change Management Database

The helpdesk for the EOSC Portal¹⁰ delivered by EOSC-hub has been implemented following these specifications.

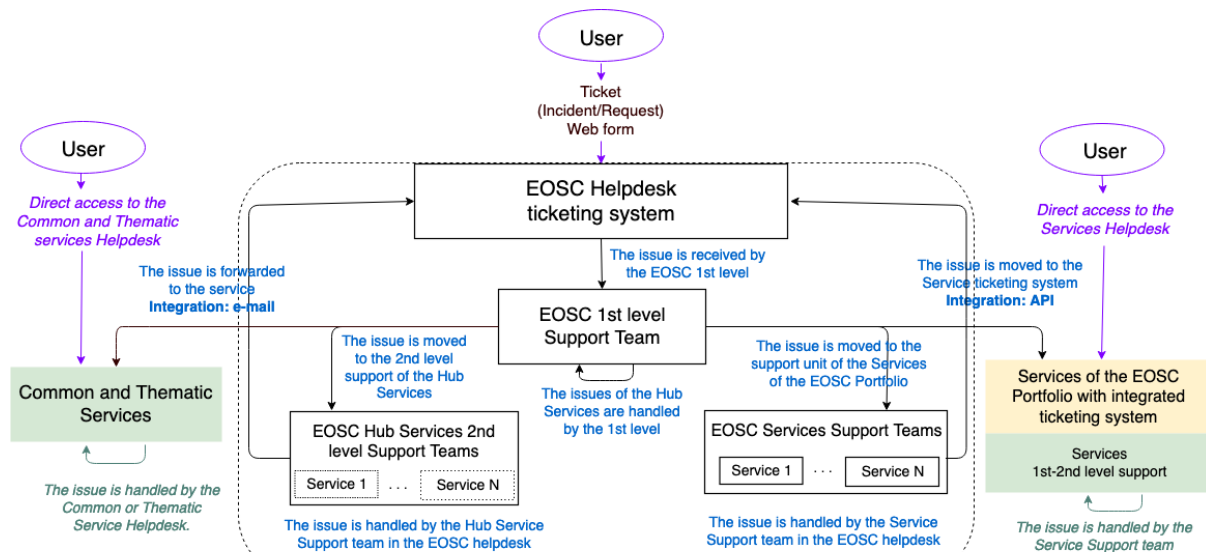


Fig. 3. High-level architecture of the EOSC Helpdesk

2.1.2.3 Monitoring

Monitoring is the key service needed to gain insights into an infrastructure. It needs to be continuous and on-demand to quickly detect, correlate, and analyse data for a fast reaction to anomalous behaviour. The challenge of this type of monitoring is how to quickly identify and correlate problems before they affect end-users and ultimately the productivity of the organization they belong to. Management teams can monitor the availability and reliability of the services from a high-level view down to individual system metrics and monitor the conformance of multiple SLAs. The key functional requirements are:

- Monitoring of services
- Reporting availability and reliability,
- Visualization of the services status,
- Provide dashboard interfaces,
- Sending real-time alerts.

¹⁰ <https://helpdesk.eosc-portal.eu/>

The dashboard design enables easy access and visualisation of data for end-users. APIs also allows third parties to gather monitoring data from the system through them.

The key requirements of a monitoring system are:

- Support for multiple entry points (***different types of systems can work together***)
- Interoperable
- High availability of the different components of the system
- Loosely coupled: support APIs in the full stack so that components are independent in their development cycles
- Support for Multiple Tenants, Configurations, Metrics and profiles to add flexibility and ease of customisation.

The monitoring tool for the EOSC Portal¹¹ delivered by EOSC-hub has been implemented following these specifications.

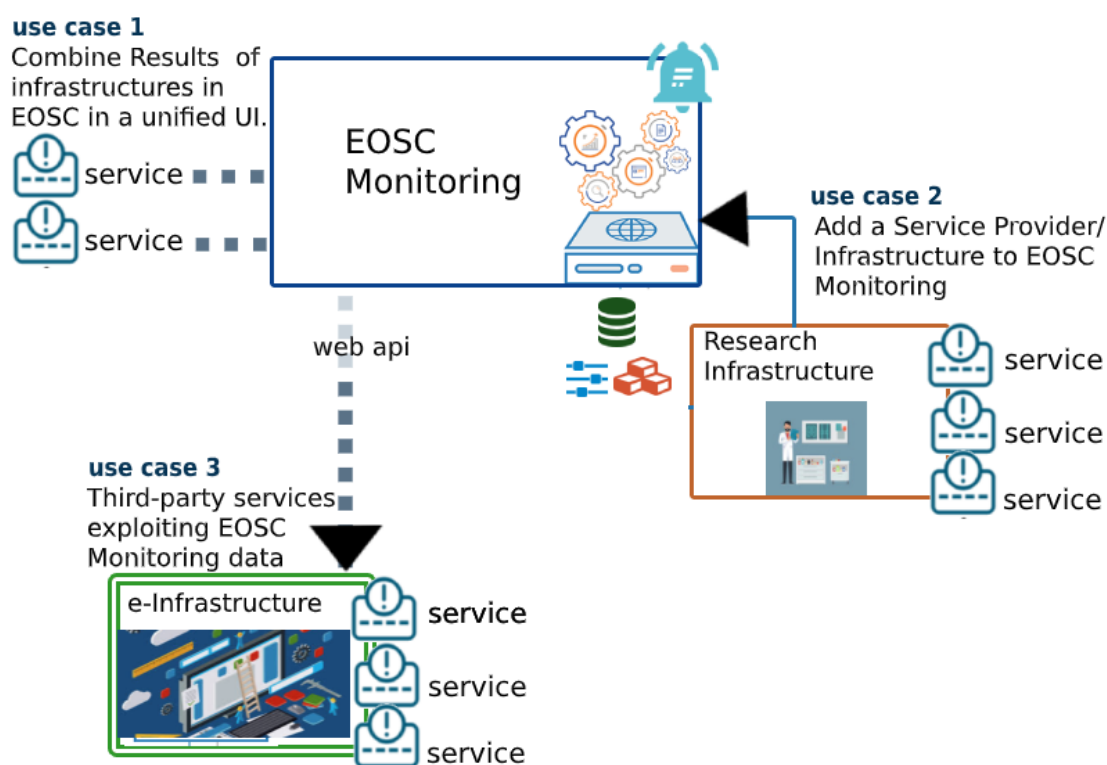


Fig. 4. Integration scenarios in the EOSC Monitoring

2.1.2.4 Security

Federation security is established by having policy and practices for infrastructure service providers, for end users and their communities, and between infrastructures (including any national CSIRT¹²).

¹¹ <https://argo.eosc-portal.eu/>

¹² CSIRT = Computer Security Incident Response Team

These must cover establishing a baseline of service provisioning adequate to the task at hand, while protecting users and administrators from each other, and protecting everyone from unauthorised outsiders. Broadly, these measures are aimed at preventing incidents; it is also necessary to have measures to specify how an incident is dealt with when it happens. Using sufficiently secure communications, information is communicated between participants to try to contain the incident, while gathering information about the attacker, if applicable. Between the proactive mitigation and the reactive handling are measures to deal with potential breaches, such as known software vulnerabilities which may not yet be exploited, or managing the risks associated with granting users elevated privileges such as when they deploy virtual machines in a cloud.

In the EOSC Technical Specification, we identified three standards as key to interoperation: SIRTFI, SNCTFI, and the WISE Trust Framework for Security Collaboration among Infrastructures. Further security measures, usually at the level of participant organisations include cybersecurity awareness training for employees, incident detection, ransomware detection, endpoint protection and pentesting (penetration testing), VPNs, detection measures from SIEM and intrusion detection to threat hunting.

D10.1 briefly noted the need for software security, citing the need for assurance in the marketplace in particular, using automated checks (of container images). The deliverable also noted the link to AAI (specifically for the assurance of user authentication and traceability, and authorisation).

2.1.2.5 Software Quality Assurance

The software quality assurance (SQA) is the process responsible for the overall supervision of both software development lifecycle ensuring that the required quality level is achieved together with service quality. The SQA encompasses not only all software development processes starting from the definition of requirements, coding, release, testing and integration but also covers the service production delivery.

This technical area covers ways to deliver quality software and services for EOSC consumption and favours the adoption of automated solutions over the traditional manual-based validation mechanisms. The automation allows not only to speed up the development tasks but as well improves the reliability of the developments “ensuring the fast execution of defined tests at each change in the codebase” and keeping them aligned with the initial user requirements and design “Fast feedback received at any development stage - faster release of quality software”.

The EOSC Technical Specification for the software development follows well establish practices and standards adopted by the open-source community while for the EOSC Service Quality it's based on Maturity Levels, via the operational definition of the Technology Readiness Level (TRL) indicators: TRL, 7, 8 and 9¹³.

¹³ EOSC-hub Service Maturity Classification:
<https://wiki.eosc-hub.eu/display/EOSC/Service+Maturity+Classification>

2.1.2.6 AAI services

Access to resources in EOSC play a crucial role. It is therefore essential to have an Authentication and Authorisation Infrastructure (AAI) as well as AAI building blocks in place that can address the variety of requirements and the distributed nature of EOSC.

Work to this end was initiated by the Authentication and Authorization for Research Collaboration (AARC) project. AARC, whilst acknowledging the existence of production AAIs, recognised that distributed research collaborations need mechanisms to control access to their resources and their users in a federated manner.

The results of the AARC project, that is the Blueprint Architecture and policy recommendations [AARC-Community] were the starting point for the work carried out in the EOSC-Hub project. During the EOSC-Hub project further work was dedicated to expanding aspects related to the Community AAI as well as the Infrastructure Proxy. It became clear that the infrastructure proxy would be a federated proxy not a single one; this approach offers the benefits to integrate, consolidate, and connect e-infrastructures AAIs.

With the aim to enable seamless access to research data and services in EOSC-Hub and in EOSC in general, the AAI Technical Specifications have been produced to collect protocols, APIs and standards that have to be followed by AAI architects to ensure interoperability. The set of guidelines focuses on both technical and policy aspects. The AAI Technical Specifications have identified standards, such as SAML, OIDC and X.509 as well as technical and policy guidelines. Example of these guidelines at technical level are, REFEDS R&S, VO Membership and Group Information (AARC-G002), AARC guideline on affiliation information (AARC-G025), AARC Guideline on resources a user is allowed to access (AARC-G027) and REFEDS Assurance Framework.

On the policy side the AAI technical specifications require adherence to GÉANT Data Protection Code of Conduct (DPA-CoCo)¹⁴, Sirtfi¹⁵ and WISE-AuP¹⁶.

The EOSC Portal AAI¹⁷ delivered by EOSC-hub has been implemented following these specifications.

¹⁴ <https://wiki.geant.org/display/eduGAIN/Data+Protection+Code+of+Conduct+Cookbook>

¹⁵ <https://aarc-project.eu/policies/sirtfi/>

¹⁶ <https://wise-community.org/wise-baseline-aup/>

¹⁷ <http://aai.eosc-portal.eu/>

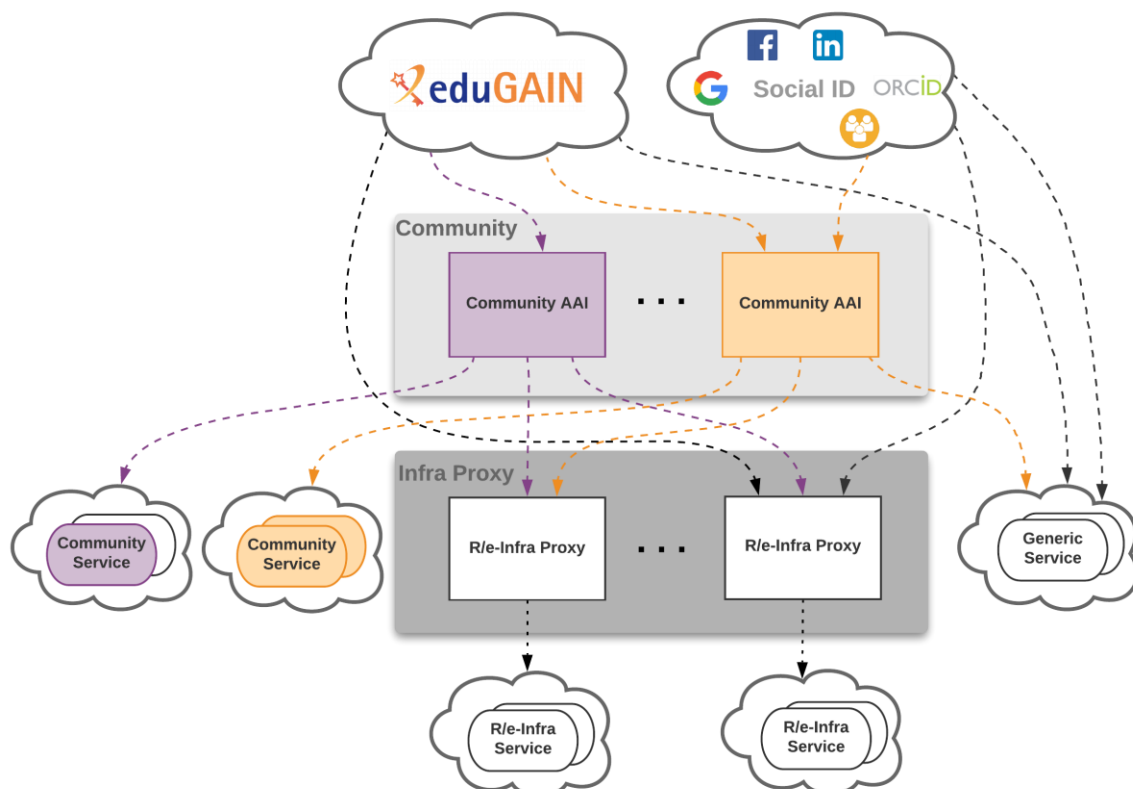


Fig. 5. High-level view of AAI architecture for access to EOSC resources

2.2 Participation to standardization bodies

2.2.1 Research Data Alliance (RDA)

EOSC-hub and RDA actively collaborate in several areas. Interactions with RDA includes both requirements gathering from the RDA members that use EOSC-hub services and the active participation of members of the project in different working groups (WGs) and interested groups (IGs) like, for example, the “Research Data Repository Interoperability” WG, the “Metadata” IGs and WGs, the “Sensitive Data in the Open Science” WG, the “Data Usage Metrics” WG and the “RDA-COVID19” WG. In particular, the last WG has delivered in June 2020 the RDA COVID-19 Recommendations and Guidelines for Data Sharing¹⁸.

It is worth mentioning that since 2018 EOSC-hub is actively collaborating with the CODATA19 International Science Council in the organisation of the CODATA-RDA School of Research Data Science, aimed at providing the principles and practice of Open Science and research data management and curation, and skills such as the use of a range of data platforms and infrastructures, large scale analysis, statistics, visualisation and modelling techniques, software development and annotation, just to mention some of them.

¹⁸<https://www.rd-alliance.org/group/rda-covid19-rda-covid19-omics-rda-covid19-epidemiology-rda-covid19-clinical-rda-covid19-1>

¹⁹<https://codata.org/about-codata/our-mission/>

3 Resource access and integration in the EOSC Portal & Marketplace

This section describes the latest achievements of the project to provide the EOSC-hub with access channels that, from one side, simplify the discovery, usage and exploitation of the EOSC resources and, from the other side, facilitate and promote the combined usage of multiple resources assisting the users in identifying and composing compatible or integrated resources.

3.1 Discoverability

During this reporting period we enhanced the Marketplace resource discoverability by implementing the tool that allows users to compare resources and services. A user can choose three resources and compare them taking into consideration the most relevant information about them. This feature will be refactored and improved in the future in the EOSC Enhance project, more fields will be added to the comparison engine.

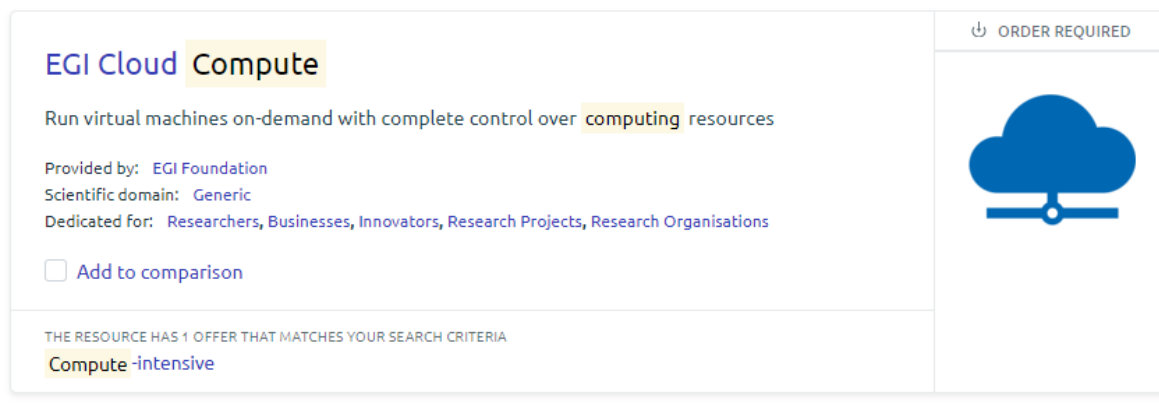


Fig. 6 "Add to comparison" button on the resource view.

Home > Resources > Comparison




Comparison CLEAR ALL			
	 EGI High-Throughput Compute With High-Throughput Compute you can run computational jobs at scale on the...	 EGI Cloud Container Compute Cloud Container Compute (in Beta phase) gives you the ability to deploy and...	 EGI Cloud Compute Cloud Compute gives you the ability to deploy and scale virtual machines...
Providers	EGI Foundation	EGI Foundation	EGI Foundation
Resource order type	Order required	Order required	Order required
Scientific domain	Generic	Generic	Generic
Dedicated for	Researchers, Businesses, Research Projects, Research Organisations	Researchers, Businesses, Innovators, Research Projects, Research Organisations	Researchers, Businesses, Innovators, Research Projects, Research Organisations
Resource life cycle status	Production	Beta	Production
Geographical availabilities	World	European Union	World
Languages	English	English	English
Tags	high-throughput large datasets	cloud compute container iaas	pre-configured virtual appliances virtual machines on-demand

Fig. 7 Resources comparison tool

Another enhancement that has been developed in the Marketplace is the enrichment of the search tool with the possibility to search by the resource offer. In the current implementation the engine takes into consideration: resource name, resource description and resource offer name.

Fig. 8 Resource and resource offers search tool.

The last but not least important enhancement made in collaboration with EOSC Enhance project is a new design of the resource presentation page. The new layout includes the most necessary information about the resource and is designed in a way that helps users find resources most relevant to their work. The implementation was done based on results of the survey conducted as a scope of the EOSC Enhance project. The main advantages for users:

- a new visual template for the resource detail site in EOSC Portal helps users quickly find the most relevant information about the resource they need, because the most important information is prioritized in the new view.
- new design helps users order proper resources.

Fig. 9 New layout of the resource presentation page

ABOUT
DETAILS
REVIEWS (0)

Cloud Compute gives you the ability to deploy and scale virtual machines on-demand. It offers guaranteed computational resources in a secure and isolated environment with standard API access, without the overhead of managing physical servers. Cloud Compute offers the possibility to select pre-configured virtual appliances (e.g. CPU, memory, disk, operating system or software) from a catalogue replicated across all EGI cloud providers. With Cloud Compute you can: Execute compute- and data-intensive workloads (both batch and interactive). Host long-running services (e.g. web servers, databases or applications servers). Create disposable testing and development environments on virtual machines and scale your infrastructure needs. Select virtual machine configurations (CPU, memory, disk) and application environments to fit your requirements. Manage your Cloud Compute resources in a flexible way with integrated monitoring and accounting capabilities.

Resource offers

ORDER REQUIRED

General purpose

Base preformance instance type. Features: Accessible in opportunistic or reserved ways, CPU cores could be overcommitted. Ideal for: Web services, Micro-services, Development...

[Show more](#)

PARAMETERS	
Number of CPU Cores	1 - 8
Amount of RAM per CPU core	1 - 4 GB
Local disk	10 - 40 GB
Number of VM instances	1 - 50
Number of days	1 - 730

[Select an offer](#)

ORDER REQUIRED

Compute-intensive

Optimised instance for computing tasks. Features: High performance CPU cores, Until 64 CPU cores, Real CPU cors (non-overcommitted), Low latency network, Reserved instances. Ideal for: Batch...

[Show more](#)

PARAMETERS	
Number of CPU Cores	8 - 64
Amount of RAM per CPU core	2 - 8 GB
Local disk	10 - 40 GB
Number of VM instances	1 - 50
Number of days	1 - 730

[Select an offer](#)

ORDER REQUIRED

High-memory

Optimised instances for tasks that require more memory relative to virtual CPUs. Features: High amount od RAM per CPU core, Up to 240 GB of RAM in total, Reserved instances. Ideal for: Running...

ORDER REQUIRED

GPU

GPU-enabled instances. Features: 1 or 2 GPU cores, 9 CPU cores for each GPU core, large memory. Ideal for: Graphics and general purpose GPU compute applications.

PARAMETERS	
Number of CPU Cores	1 - 8
Amount of RAM per CPU core	1 - 4 GB
Local disk	10 - 40 GB
Number of VM instances	1 - 50
Number of days	1 - 730

SCIENTIFIC CATEGORISATION

- Generic
- Generic

CATEGORISATION


- Compute
- Virtual Machine Management

TARGET USERS

- Researchers
- Businesses
- Innovators
- Research Projects
- Research Organisations

RESOURCE AVAILABILITY AND LANGUAGES

- English



Highcharts.com © Natural Earth

[More about EGI Cloud Compute](#) →

Fig. 10 New layout of the resource presentation page, tab “About”

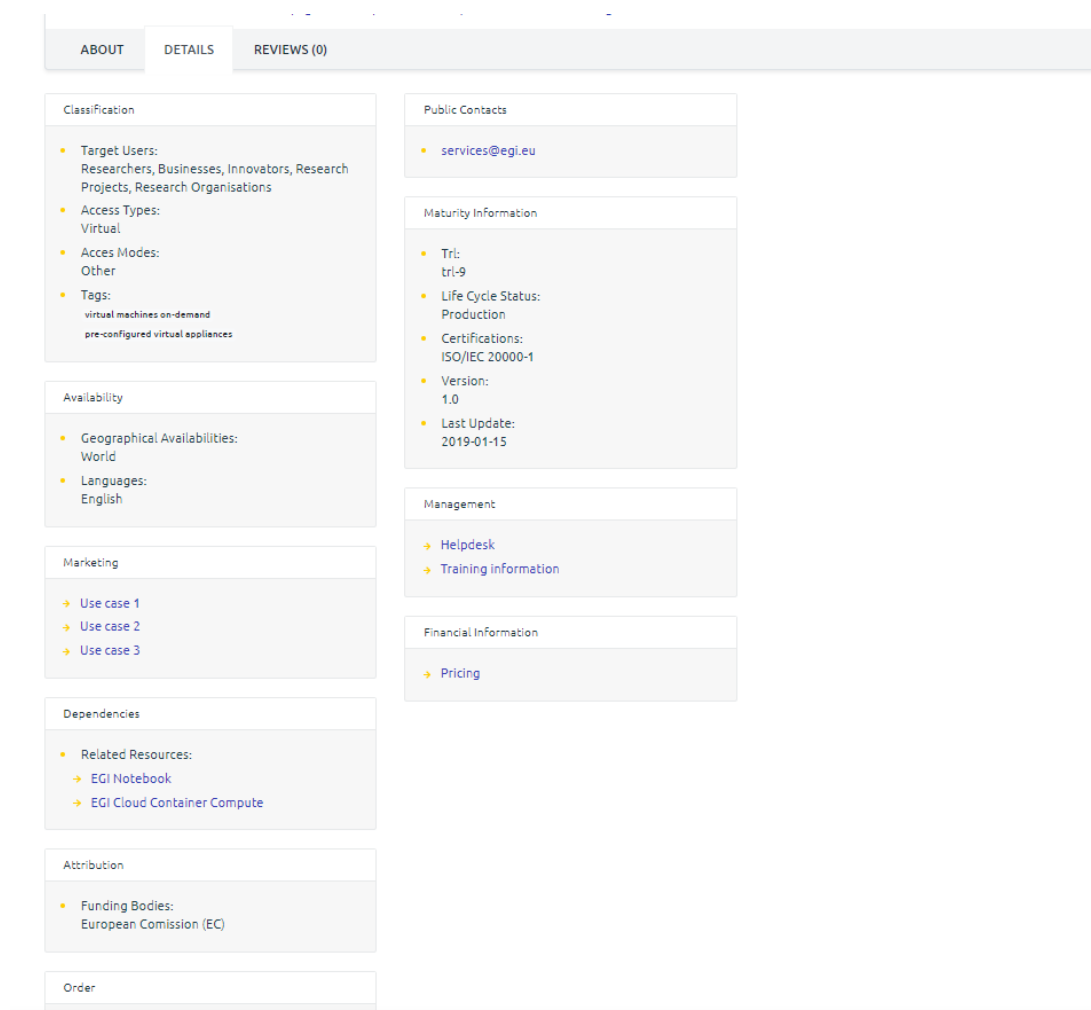


Fig. 11 New layout of the resource presentation page, tab “Details”

3.2 Access

In order to offer a better user experience and properly guide the user, it became clear that the user should be able to undertake different paths depending on the access policies of the selected resources. Since Deliverable 10.1, in collaboration with the EOSC Enhance project, we adopted and mapped the new types of resource access:

1. **Request/Order required** - Resource requires an ordering procedure.
2. **Open access** - No ordering procedure necessary to access the resource but requires user authentication.
3. **Fully open access** - No ordering procedure necessary to access the resource and no user authentication required.

Service type is shown on the main view of the resource page and described in detail after clicking “Access the resource” button.

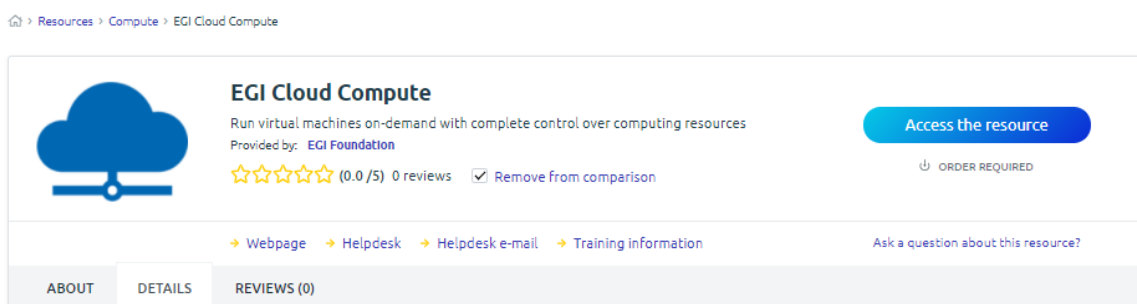


Fig. 12 “Access the resource” button

3.3 Marketplace projects

With Marketplace Projects users can organise their resources and resource orders into logical blocks to reflect a common scientific purpose and gain support for the created Marketplace Project.

Since Deliverable 10.1 we enhanced the Marketplace projects in various dimensions. As the resources available in the Marketplace are available for all European researchers, a user can use projects to precise his/her customer typology. Much effort was devoted to improving user-facing features and integrational aspects. First of all, we mapped a Project to a unique hierarchical order which helped us to improve the support for the operations behind the order management system. Now, after a project is created, the user can start a dialogue with an EOSC Hub expert to ask for guidance and get access to the technical support as it is needed. Marketplace Projects are the place to follow the status of orders, get all relevant information about the resource access when the access is granted and contact the EOSC support if necessary. In the end we improved the user interface to be more friendly and easy to use. The current version of the Marketplace project is visible in the picture below.

🏠 > My projects > Test

MY PROJECTS

Test

+ CREATE NEW PROJECT

Test

Created at 25.11.2020 — Single user — Cyfronet

EDIT

DUPLICATE

DELETE

RESOURCES

PROJECT DETAILS

CONTACT WITH PROJECT SUPPORT

PROJECT DETAILS

Project name

Test

Email

a@a.pl

CUSTOMER DETAILS

Customer Typology

Single user

Organization

Cyfronet

Origin country

Poland

Webpage

http://www.cyfronet.pl

Scientific Domains

Aerospace Engineering

USAGE

Reason to request access to the EOSC resources

Test reason

Fig. 13 New layout and functions of the Marketplace projects.

4 EOSC-hub services roadmap - Evolution

4.1 Plans for service evolution in the EOSC-hub technical areas (TCOM)

This section describes the planned evolutions for the most relevant EOSC-hub services grouped per technical areas. It focuses on the presentation of the roadmap prepared by the technical teams developing the services, as a result of the analysis of the use cases supported by the project. Moreover, the reported plans also consider the technical requirements for integration among existing EOSC-hub services. This approach will help to ensure that, in the future, combined usage of more services of the EOSC portfolio will be much easier.

4.1.1 TCOM area: Data Platforms for Processing

During the framework of EOSC-hub project this area has focused on integrating the various data processing services within the community such as EGI DataHub²⁰, B2SAFE, B2FIND or Rucio. In particular EGI-DataHub has been integrated with several services including B2SAFE, B2ACCESS, B2HANDLE and B2FIND allowing data access and discovery across the communities using EGI and EUDAT data services. EGI-DataHub is built on Onedata²¹, a distributed *eventually consistent* virtual filesystem, providing a unified namespace for user data and supporting several backend storage solutions including POSIX, Ceph, XRootD, GlusterFS, S3, Swift and WebDAV. Furthermore, complete integration with Jupyter Notebooks has been added, allowing to directly access data stored in EGI-DataHub or EUDAT from the notebook code, as well as storing entire notebooks and related data within EGI-DataHub.

The future plans for the evolution of the data platforms will be oriented on supporting long-term archiving and of preservation use cases from the community, backed with elastic QoS mechanisms allowing easy control of the data replication, storage quality and preservation requirements. EGI DataHub will be extended with support for OAIS compliant standards such as BagIt, PREMIS and METS, allowing ingestion and curation of legacy data collections, storage in formats ensuring long-term preservation and easy access to archived data on demand, for Cloud processing, verification or reproducibility.

4.1.2 TCOM area: Metadata Management and Data Discovery

This EOSC-hub area focuses on three major use cases:

Metadata Cataloguing and Indexing

Metadata Cataloguing and Indexing comprises the entire metadata ingestion workflow, i.e., from harvesting of metadata from community data repositories, over metadata mapping on a common schema including curation and validation until uploading and indexing of metadata records in the

²⁰ <https://datahub.egi.eu>

²¹ <https://onedata.org>

metadata catalogue, to enable Data Discovery and Access, (see functionality described in the next subsection). The technical implementation usually comprises several modules as metadata harvesting endpoints on data provider site, metadata ingestion comprising normalisation, homogenisation and mapping of the specific community standards on service provider site. Beside this technical service this task requires huge effort in community outreach to satisfy the various and specific schemas and vocabularies. Especially for an interdisciplinary metadata catalogue it is important that standardized protocols as OAI-PMH are used and that metadata must be available and harvestable in a known metadata schema and format.

This functionality has reached a good level of maturity and it is extensively supported by technologies such as EUDAT-B2FIND as the central metadata catalogue of EOSC-hub.

Data Discovery and Access

Data Discovery and Access comprises the ability for end-users to search for data resources and access the referenced data. This functionality requires and is based on the existence of an indexed metadata catalogue, (see above section “Metadata Cataloguing and Indexing”).

The technical implementation of a data discovery and access service enabling searching for and identifying digital data should comprise graphical and command line user interfaces. To allow end users a simple and comfortable way to search research data, a discovery portal with an intuitive graphical user interface with faceted search and filtering options was provided. An additionally offered Command Line Interface allows machine readability and users to embed data discovery in a data processing workflow. Typically, this is implemented by a RESTful Search API with functionalities to identify referenced data collections by persistent identifiers and a search indexer relying on a comprehensive metadata catalogue (see macro feature ‘MD cataloguing and indexing’).

This functionality has reached a good level of maturity and there are, on an interdisciplinary level, many services which provide such discovery portals. To name just two of them there is the ‘Google Dataset Search’ (<https://toolbox.google.com/datasetsearch>) which allows users to find records collected from schema.org endpoints using a simple keyword search and EUDAT-B2FIND (<http://b2find.eudat.eu/>) as a cross-domain discovery service based on metadata steadily harvested from research data collections from EUDAT data centres and other repositories offering faceted browsing.

Data Annotation

The Annotation Service enables end-users to extend descriptions of datasets or parts of datasets with user-defined content, without modifying the underlying dataset, e.g., adding comments, free text keywords or semantic tags (keywords from ontologies). These annotations can be used to search and aggregate datasets or parts of datasets into user-defined datasets, either localised in a unique data repository or throughout a heterogeneous and distributed set of data repositories.

4.1.3 TCOM area: HTC/HPC Compute

The support of High-Throughput (HTC) and High-Performance Computing (HPC) workloads have focused on three major use cases:

Multitenant Job Submission

Multitenant job submission relates to the capability of submitting HPC/HTC jobs with predefined constraints (both at resources and software) without a previously deployed virtual infrastructure. This service should be able to run a bunch of batch jobs on HTC/HPC and cloud compute resources, interfacing with storage solutions and seamlessly integrated with the authentication mechanisms. The main difference between HPC and HTC jobs is the requirement of multiprocessing (OpenMP or MPI for example). Some sites support both types of jobs, using different queues and specifications in the batch job to differentiate and provision the rightmost resources. This functionality has reached a good level of maturity and it is extensively supported by technologies such as DIRAC4EGI WMS²², INDIGO-DC orchestrator²³ and others.

Multitenant Containerised Job Submission

In the last years, the use of containerised jobs has boosted due to the enormous convenience of containers for application delivery. Application dependencies are embedded into the containers reducing the effort and side-effects of the installation of software. However, popular container technologies such as Docker use daemon processes that run on privileged users, which is not acceptable by many data center policies. There are solutions for running jobs on containers that run on the user space. This approach reduces the capabilities of a containerised job to those of the user running the job, which makes it suitable for HPC, HTC and Cloud Compute infrastructures.

In this focus, three areas have been explored:

- The execution of containerised workloads natively supported by the sites. This may require rebuilding container images to exploit native hardware capabilities, such as low-latency networks or specific accelerator devices. It will require Synchronization and distribution of container images, standardized metadata annotation on the containerization technologies supported, standardized customized base image naming, among other features.
- The execution of regular jobs as containers. Resource providers may offer different execution environments through containers to run regular batch jobs. This will increase isolation, facilitate the support of multiple versions of software dependencies and facilitate administration at resource level. In this case the user will not provide a container but will select one of the container images that fit the dependencies of the execution. This will require a standardized way to specify container images in the multi-site environment so users can discover, select and test the most suitable container images.
- An efficient support of user-space container technologies. The use of jobs embedded in containers even in batch systems which have not any of the classical container engines is feasible by using containers in the user space such as uDocker. However, container image distribution is an issue as it is inefficient to pack it with every job instance. Mechanisms for container image distribution inside the resources could mitigate this issue.

²² <https://dirac.readthedocs.io/en/latest/>

²³ <https://indigo-dc.gitbook.io/indigo-paas-orchestrator/>

HTC / HPC Clusters on demand

Scientific portals expose convenient interfaces that typically run partially customisable jobs on computing infrastructures. Scientific portals normally have a higher computational demand on their back-ends than conventional portals, so they have to be provisioned with enough resources to deal with the potentially unexpected workload peaks. However, users in scientific portals could also allow longer delays on retrieving the results, as they may not have the same urgency in getting the results as an industrial application has.

In this regard, we identify the need for provisioning self-managed elastic clusters supporting mainstream job managers such as PBS²⁴, SLURM²⁵ and especially scheduling systems based on Kubernetes resource orchestrators. Opposite to the multitenant job management service, this service will explicitly deploy a single-tenant cluster backend to be used by the user community managed by the user who deployed it.

The support of elastic, customizable and fully automated HTC/HPC clusters supporting batch systems as SLURM or TORQUE²⁶, popular workflow environments such as Galaxy and container management platforms such as Kubernetes or Apache Mesos could facilitate enormously the provisioning of back-ends. Main public cloud providers already offer such types of services.

This feature should be supported by infrastructure-as-code specifications (such as OASIS TOSCA) that could be easily customized and could be supported by different cloud orchestration tools.

4.1.4 TCOM area: Cloud Compute (including containerisation and orchestration)

The EOSC-hub services in the area of Cloud Compute are divided in three categories, each with its own roadmap:

IaaS VM Management

This is a service relying on a set of providers deploying the OpenStack cloud management framework. The providers update the underlying OpenStack as part of their regular operations, bringing new features that can be used for the EOSC-hub communities. The federation is delivered via a set of additional components, their roadmap is described below:

- cloud-information-provider: the information system will finalise transition to AMS²⁷ as transport instead of the current BDII-based²⁸. Any future evolution of the GlueSchema²⁹ will be implemented into the component. New renderings of the information (mainly json) will be implemented to facilitate the integration with new clients.
- cloud accounting: the accounting extractors will be extended to report usage information of public IP addresses and block storage.

²⁴ <http://www.pbspro.org/>

²⁵ <https://slurm.schedmd.com/queue.html>

²⁶ https://en.wikipedia.org/wiki/TORQUE_Resource_Manager

²⁷ <https://confluence.egi.eu/display/EOSC/ARGO+Messaging+Service+-+AMS>

²⁸ <https://en.wikipedia.org/wiki/BDII>

²⁹ <https://www.ogf.org/documents/GFD.147.pdf>

- cloudkeeper: the image synchronization tool will upgrade to v2, with a complete revision of the gRPC³⁰ specification for communication between components and capable of running scoped to single VOs.

IaaS Container Management

The EGI Cloud Container Compute service is delivering container management in EOSC-hub. Initially this service was relying on the provisioning of docker-enabled VMs in the infrastructure, but now the service is adopting EC3 as an orchestrator to create elastic kubernetes clusters. This brings the industry standard tool for container management (kubernetes) to EOSC-hub. The service is planned to further evolve in the future as follows:

- Facilitate the configuration of the clusters, with automatic detection of provider parameters (e.g. VM image to use, network configuration, authorisation)
- Improve the reliability and elasticity management of the kubernetes deployments
- Update deployment recipes to follow kubernetes releases (every 3 months)

IaaS Orchestration

The Orchestrator layer allows to coordinate the provisioning of virtualized compute and storage resources on Cloud Management Frameworks, both private and public (like OpenStack, OpenNebula, AWS, etc.), and the deployment of dockerized long-running services and batch jobs on Apache Mesos clusters. It receives the deployment requests, expressed through templates written in TOSCA (Simple Profile in YAML version 1.0), and orchestrates the deployments on the best available cloud sites. In order to select the best site, the Orchestrator implements a complex workflow: it gathers information about the SLAs signed by the providers with the user, the monitoring data about the availability of the compute and storage services and the location of the data requested by the user (if any). Hybrid deployments spanning multiple sites are supported. Using the Orchestration layer and the TOSCA templates, the end users can exploit computational resources without any knowledge about the IaaS details. The service is planned to further evolve in the future as follows:

- Add direct support to EGI CheckIn (already supported as IdP of INDIGO IAM).
- Add support to launch jobs on Kubernetes clusters.
- Obtain information from the SOMBO (Service Order Management Back Office) component to obtain information about negotiated SLAs.
- Add modifications to upstream cloud-info-provider to get all the information needed by the Orchestrator.

4.1.5 TCOM area: Software Release and SQA

The SQA roadmap follows the current best practices criteria used for opensource projects. While the software criteria are focused on software development, the services criteria are focused on deployment and delivery of EOSC services.

³⁰ <https://grpc.io/>

The set of criteria described in Technical specification for the SQA do not follow any exclusive models, although some of the criteria are very similar to some of the models. The criteria used considers the complexity of the EOSC ecosystem and the fact that software and services currently available in EOSC marketplace were developed, deployed and delivered without any guidelines or central coordination and may not be open source.

Taking this into account the roadmap for both the Quality Assurance process for Software and Services features a set of quality criteria based on best practices meant to be verified in an automated way and are agnostic concerning the technologies or services used. Note that, that many of those services do not have tools for automated verification or measurement of the characteristics or metrics proposed.

The evolution of this area will have the following objectives in mind:

1. Define well known criteria and disseminate best practices of Software development or service operation across EOSC service providers.
2. Establish criteria which can be automatically verified.
3. Always decouple the criteria from the tools and technologies used for validation providing abstraction allowing developers, service managers and/or infrastructure operators to choose which tools and technologies to use for each purpose.

Together with these objectives, another set of wider objectives will be pursued:

4. Establish a well-established set of technical guidelines and policy for services in EOSC involving service providers in the process definition.
5. Provide within EOSC an SQA as a Service (SQaaS) which would allow automatic verification of the most important criteria.
6. Establish a reward mechanism for services complying the criteria (e.g., through badges or other means).

4.1.6 TCOM area: Federation Tools

Details of the main achievements and technical roadmap of the EOSC Hub Federation Tools are available in the deliverable D5.5 “Second report on maintenance and integration of federation and collaboration services”³¹.

It is worth to mention that the Federation Tools area is also working on further developing the technical architecture and interoperability guidelines of tools like AAI, accounting, monitoring and helpdesk already mentioned in section 2. The guidelines are being enhanced according to the requirements collected by research communities and other relevant EOSC stakeholders through dedicated surveys.

³¹<https://www.eosc-hub.eu/deliverable/d55-second-report-maintenance-and-integration-federation-and-collaboration-services>

4.1.7 TCOM area: Workflow management, user interfaces and Data analytics

User interfaces

- Marketplace - described in the section 3, the planned roadmap is:
 - New features introduced in scope of EOSC-hub:
 - Introduce service offers for ordering (technical and quality parameters, conditions to use the service)
 - Non-free offers support (e.g., HelixNebula Vouchers)
 - Comparison between services
 - Research project grouping (collect all services needed for research project and manage them for defined research team)
 - Compatible Marketplaces - Possibility to create white-label marketplaces sharing the same services and related processes
 - Implementation of reference API for Marketplace orders handling:
 - Implementation of the order handling integration API
 - Implementation of an API for provider offering/offering parameters integration
 - Further work on the White label solution:
 - Synchronization between different service catalogues
 - Increase of the customizability degree by the application configuration (instead of customizability in the source code)
 - Further roadmap
 - In the scope of EOSC Enhance project (plans for the near future):
 - Tours implementations as a way to demonstrate a new feature in action, present solutions to timely issues or showcase resources in entirely new ways
 - A/B Testing as the process of testing multiple new designs of a webpage against the original design of that page with the goal of determining which design generates more conversions
 - Recommendation for the users (recommendations for new EOSC Resources based on similar uses, recommendations for new EOSC Resources based on EOSC Resources a user already searched or ordered)
 - Further extension planned in the context of EOSC-Future and EOSC Enhance
 - FutureGateway
 - The FutureGateway is a complete framework aiming at supporting the creation of Science Gateways. It includes many components for installation and management. It provides a set of REST APIs to address final user interfaces, intended to address distributed computing resources using three logical entities named: *Infrastructures*, *Applications* and *Tasks*. The Task element consists of application instances, running on top of a given distributed infrastructure. FutureGateway provides services to install and maintain the system and encourages its customisation in order to best fit the needs of the adopter. Roadmap:

-
- Future activities will be much more focused on reproducibility and reusability use cases and hence on the interaction of FG-powered science gateways with FAIR principles compliant repositories. It will improve data reproducibility in science. Software solutions will allow the user to keep track of computational studies from the beginning to the final outcome and enable sharing of the data for others to reproduce the results.
 - Add Science Software on Demand (SSoD) to EOSC Marketplace. SSoD is built on top of the FutureGateway framework. It demonstrates capabilities of building a dedicated science gateway for research communities.

 - Thematic portal services
 - The following services have already been added to EOSC Marketplace:
 - DARIAH Science Gateway³² provides various web-based applications and services for the Digital Humanities researchers, institutes and communities- roadmap described in 4.2.6
 - GEOSS portal³³ - the main entry point for discovering and accessing GEOSS data. GEOSS Platform interconnects more than 170 data systems globally, providing discoverability of more than 400M datasets. Roadmap described in 4.2.4
 - EISCAT_3D portal³⁴ - user portal for EISCAT data access and analysis, which provides services for data cataloguing, discovery and pre-defined analysis,
 - Lifewatch user interface - roadmap described in 4.2.7
 - OPENCoastS Portal³⁵ - builds on-demand circulation forecast systems for user-selected sections of the North Atlantic coast and maintains them running operationally for the timeframe defined by the user.

³² <https://marketplace.eosc-portal.eu/services/dariah-science-gateway>

³³ <https://marketplace.eosc-portal.eu/services/geoss-web-portal>

³⁴ <https://marketplace.eosc-portal.eu/services/eiscat-data-access-portal>

³⁵ <https://marketplace.eosc-portal.eu/services/opencoasts-portal>

Data analytics services

The following services has already been added to the EOSC marketplace:

- ENES Climate Analytics Service (ECAS)³⁶ - Geo big data analytics services and their roadmap are described in 4.2.3
- Datacubes (<https://marketplace.eosc-portal.eu/services/rasdaman-eo-datacube>) - data analytics service with a multi-sensor, -scale and -purpose datacube approach. Satellite datacubes ready for spatio-temporal analysis and visualization.
- Geohazards Exploitation Platform (<https://marketplace.eosc-portal.eu/services/gep-eo-services-for-earthquake-response-and-landslides-analysis>) is focused on the integration of Ground Segment capabilities and ICT technologies to maximise the exploitation of EO data. The Geohazards TEP (GEP) is an enhancement of the precursor platforms (G-POD, SSEP), and is designed to support the Geohazard Supersites (GSNL) and the Geohazards community via the CEOS WG Disasters. One of the main points of the common roadmap in regard to EOSC-hub was integration with EOSC Hub AAI to allow seamless access to the platform services to a growing number of research communities.

4.1.8 TCOM area: Security

Starting from the previous roadmap (D10.1), the roadmap for TCOM Security is based on (a) updated requirements from the communities, (b) the European Interoperability Framework (EIF), (c) best current practices in the IT security industry.

Common to all User Stories is the (sometimes implicit) requirement for interoperation in security. Indeed, security (in the general sense) and privacy are together one of the core EIF principles; the EIF identifies them as a requirement to ensure that end users can trust services in the digital single market. While the TCOM Security area has a narrower focus of operational security, the interoperability of security is an essential component of establishing and maintaining trustworthy EOSC services. Moreover, as from the discussion below, it may veer into more general security topics, which may help to minimise the risk of security incidents. Since the importance of interoperability of security is recognised by the EIF, it would make sense to look at it from the legal, organisational, technical and semantic perspectives (there will be some overlap between the topics.)

Legal.

Operational security could be affected by the use of external resources (i.e., resources hosted by external organisations, such as commercial cloud providers, or social media IdPs) - the agreement with them is between the user as a customer and the resource provider, unless there is a more comprehensive framework, such as when an NREN provides a framework for educational/research use. The risk is that the external organisation will not participate in the resolution of security incidents. The roadmap should include NREN's facilitation of the use of cloud resources, considering EOSC Secretariat Sustainability WG legal/policy work.

³⁶ <https://marketplace.eosc-portal.eu/services/enes-climate-analytics-service>

Some data and software are licensed only to specific communities, so community membership management needs to be sufficiently robust. Data security measures in general may be also necessary to address GDPR. Citizen science projects should be supported (e.g., STARS4ALL). Communities may not have the legal expertise, and in any case, the infrastructure should assess the legal risks in order to protect itself. As a part of the onboarding of each new community, there should be an assessment of their legal issues wrt GDPR and licensing.

Organisational.

For incident handling, there is already an established baseline in SIRTFI. A core (infrastructure) security team should organise regular security service challenges and track how well each organisation responds. Similarly, it may make sense to also provide training either at the organisational or infrastructure level (or both). Most “breaches” by insiders are unintentional, not malicious - the security industry is quite keen on “insider threat management”, but the primary mitigation is usually training (e.g., anti-phishing). As a part of the security roadmap, there should be regular re-evaluations of security training needs for administrators and users.

More generally, security risks can be mitigated by organisational best practices (as required by SIRTFI) on service maintenance and patching. Data centres should consider implementing Data Transfer Zones³⁷ in order to facilitate fast low latency data transfers between them.

Organisational attributes must be published in a way that allows interoperable authorisation based on organisational membership (see also semantics, below.) Organisations should understand the assurance, accuracy and timeliness of their attributes, and should be willing to publish this³⁸. There should be a means of updating the set of attributes from time to time, in response to changing practices or requirements.

Technical.

Organisational and community membership attributes need to be maintained in usable forms (for authorisation, account, and auditing), i.e., they must be persistent, unique, and meaningfully named. The need for meaningful naming arises from the need to inspect log files in an incident, and to ease the process of authorisation management based on these attributes. End user identifiers should be unique and persistent and have meaningful names unless there is good reason not to. Security contacts should be available (cf SIRTFI), with appropriate contact information. It is necessary to implement methods to maintain the integrity of the metadata - both for the publisher to assert correctness (and origin authentication if necessary) and for the consumer to verify integrity. The immediate activity for the roadmap is to encourage all IdPs and SPs to be SIRTFI compliant (and R&S), while the future action is to do the same with future participants.

It may be necessary to consider not just core users of IaaS but also at the PaaS and SaaS levels: suppose User A obtains IaaS resources from EOSC and deploys services for the community. EOSC must know who A is, in order to manage authorisation and to intervene if there is a problem with the deployment. Now A grants access to the services to Users B and C who are both members of the community. When B and C access the services, should EOSC know who they are, or is it sufficient to

³⁷ <https://fasterdata.es.net/science-dmz/>

³⁸ <https://wiki.refeds.org/display/ASS>

rely on A to keep track of it³⁹? If B, say, violates the terms of the EOSC IaaS AUP, should EOSC contact (a) User A, (b) User B, or (c) the community's security contact? If instead User B's violation affects only User C, should the resolution stay wholly within the community or does EOSC need to be involved? In general, the answer depends on the type of service. Conversely, there may be cases where User A cannot share the identity of User B with EOSC.

In adapting community AAls (e.g., ELIXIR, CLARIN), it is necessary to have SNCTFI compliance. It may be necessary to be able to automatically produce security reports - or alerts? - accounting - for communities, site administrators, and for users, their jobs and delegated tasks/data transfers. A delegated credential or authorisation token needs to be traceable to its issuer and to the end user to whom it was issued. Moreover, delegation may happen with limited user intervention, such as in distributed workflows (ECAS/ENES), or renewals for long-running tasks, so sufficient controls must be in place to ensure traceability and limit abuse.

Security of container images and legacy software needs to be addressed - it may be necessary to sandbox certain containers and applications. Access controls should protect software with restricted licensing - i.e., licensed to a specific organisation or community.

Non-web access is required, at least for some communities, to support access to web services. For the foreseeable future, it will also be necessary to support generation of X.509 credentials for users and automated clients.

Semantic.

It is expected that the language of communication remains English, except where an incident happens and is handled entirely within a single (non-English) language region (consequently, no roadmap item is needed, as the situation is unchanged from current practices.)

Security policies should be human readable (to allow users to understand the policies), lawyer readable (for legal use), and machine readable (for service discovery/selection and automation/orchestration). It is necessary to do additional work on the machine-readable versions.

If finer grained organisational authorisation (i.e., based on organisational roles, as opposed to just "membership") is needed (ELIXIR have expressed a requirement for "researcher" status), more work is needed to make these semantically consistent and interoperable. There have been several attempts to date, but none has been successful.

4.1.9 TCOM area 12: AAI

The TCOM AAI area focused its activities in supporting the day-to-day work carried out by WP5. In The TCOM AAI worked at a higher level to ensure that AARC principles would be adopted and that technical specifications focused on interoperability would be produced.

As highlighted in D10.6, common to many users' stories there is the need for service providers and research collaborations to have an easy way to enable single sign on as well as to integrate their

³⁹ The service AUP in the AARC PDK says (item 7) "You may control access to your Service for administrative, operational and security purposes and shall inform the affected users where appropriate". The PDK also contains a Security Incident Response Procedure for participants.

existing workflows with their AAI. In many cases research communities have already AAI services in place and the discussion revolves on how to evolve these systems to consume services and offer services from/to EOSC. There is certainly awareness that AARC compliant AAIs are the way to go to achieve interoperability with research and e-infrastructures as well as EOSC.

Further evolution and refinement of the EOSC-Hub AAI is expected; there is consensus among different groups, that the EOSC AAI will be implemented following the architectural and policy recommendations defined by the AARC project. It has been by now demonstrated that this model enables interoperability across different SP-IdP-Proxy services, each of which acts as a bridge between the community proxies (these are the Community AAIs) managing the researchers' identity and the generic services offered by Research and e-Infrastructures (termed R/e-Infrastructures or Infrastructures).

Beside the technical and policy aspects that are being addressed not only in EOSC-Hub but also in the EOSC Architecture Working Group as well as in the AARC community, it would be important to focus on the user experience as a whole. GDPR requires that users be informed on where their data goes; this is certainly important, but it may pose challenges from an implementation perspective in a multi-layered and distributed environment such as EOSC. We should strive to ensure that users can easily access their services with a limited number of clicks and the necessary level of information about GDPR related matters.

Also related to the user experience, further investigation is needed to address:

- Multiple user registrations: Users are asked to register with different AAI services as they access resources protected by different infrastructure proxies. There are already activities in progress to align user attributes and of AUPs which should continue to enable seamless access across different domains.
- Multiple IdP discovery steps: The EOSC-hub AAI is based on the AARC BPA “community-first” approach; this may result in the need for users to go through multiple IdP discovery steps: for instance, to select their Community AAI and then to select their Home Organisation. This process is not yet completely streamlined. The AARC community is working on a guideline, “IdP hinting” protocol proposed in AARC-G049) to simplify this process by making the selection process transparent.

Lastly effort is needed to keep promoting the [AARC Interoperability Framework](#) more widely particularly where different protocols are being used and now that more research collaborations are engaging.

4.2 Thematic Services roadmap

4.2.1 CLARIN - Component Metadata Infrastructure

Virtual Language Observatory (VLO)

The main aspects of the current roadmap for the VLO we would like to highlight are (1) improved monitoring of issues related to input data, interoperability or accessibility; (2) closer integration with the LRS at the user interface level; (3) implementing compliance with the latest VCR API; (4) support for filtering by temporal coverage of the indexed resources; (5) retrieval and display improvements through enhanced metadata mapping on basis of common metadata standards ('CMDI core components' initiative).

Language Resource Switchboard (LRS)

The LRS roadmap is centred on improving user experience and integration capabilities; the next main planned steps are: (1) extend input file format detection capabilities (2) add support for different input modalities, as file pairs, collections or batches of files (3) facilitate the integration at the user interface level, in data repositories and discovery tools as the VLO (4) improve the management and monitoring of the tools catalogue, feeding the LRS with more information about the current status and limitations of the tools offered, improving the relevance of the results and better informing the user about potential limitations of the tools on display.

Virtual Collection Registry (VCR)

The VCR will focus on three areas in the current roadmap: (1) improve and simplify the user workflow to create and edit virtual collections, (2) further improve compatibility with existing standards, such as assigning DOIs and exporting DOI compatible metadata and (3) further improve the potential for integration in third party environments (such as data catalogues, portals, etc). In order to improve this potential, we have discussions with communities interested in integrating the VCR. Based on these discussions we have defined common scenarios from which we can extract requirements, such as collaborative collection management.

4.2.2 DODAS

DODAS is a Platform as a Service whose aim is to guarantee deployment of complex and intricate setup on "any cloud provider" with almost zero effort. As such it implements the paradigm of Infrastructure as code: driven by a templating engine to specify high-level requirements. DODAS allows instantiating on-demand container-based clusters to execute software applications.

DODAS completely automates the process of provisioning by creating, managing, and accessing a pool of heterogeneous computing and storage resources. As a consequence, it drastically reduces the learning curve as well as the operational cost of managing community-specific services running on distributed clouds. DODAS Thematic Service supports automated cluster creation both managing big data platforms and more traditional batch systems as a service. At the time of writing the scientific community integrating DODAS in the computing model are: CMS, AMS and FERMI. Other initiatives are also ongoing within the context of WLCG and Virgo.

From the technical perspectives the plan for the evolution of the DODAS services is based on two main assets:

- hybrid resources federation
- storage management and data management.

Regarding the hybrid resources federation, the current vision is on one hand to keep integrating the INDIGO PaaS Orchestrator features natively providing resources federation capabilities. On the other hand, the plan foreseen to integrate application-level solutions to integrate hybrid resources including HPC together Cloud/HTC. It is worth to mention that one of the key aspects of all of this, is Identity Management. In this respect any deeper integration/federation available in the EOSC-hub portfolios and beyond will be considered and evaluated.

For what concerns Storage and Data Management one of the main objectives is to increase the level of compatibility between DODAS and available Services, in the EOSC-hub portfolios and beyond. An example here could be the integration with B2Share.

Other features and capabilities foreseen in the plan are those oriented toward a more comprehensive support to the heterogeneous data handling and processing. This in turn translates into the need of supporting data integration solutions (for collecting data coming from external storages and repositories), automated meta data handling as well as automated solutions for data validation. The ultimate objective is to provide an effective support for the schema on read data processing model.

4.2.3 ECAS

The ENES Climate Analytics Service (ECAS) enables scientific end users to perform data analysis experiments on large volumes of climate data, by exploiting a server-side, PID-enabled, and parallel approach. The task is aimed at providing a paradigm shift for the ENES community with a strong focus on data intensive analysis and server-side approaches as opposed to the current ones that are mostly client-based, sequential and with limited (sometime missing) end-to-end analytics workflow and provenance capabilities.

The integration and training activities of ECAS will continue until the end of the project. A pre-operational service was made available by M18, while the planned integration activities were completed before M22. Additional activities started during the final year of the project as part of the amendment A3, to further strengthen the integration with EOSC-Hub services. After a first integration of the ECAS components with one of the EOSC-HUB AAI providers (i.e., IAM), both at the level of Ophidia and JupyterHub, the target is now the integration with another EOSC-HUB AAI service (i.e., EGI Check-in).

ECAS will integrate B2SHARE most likely via the existing integration between B2DROP and B2SHARE. This requires the definition of the necessary metadata and a process to acquire it in the ECAS-Lab environments. Within ENES, the definition of the core metadata scheme has not yet been finalized, so the integration between B2SHARE and ECAS will be able to take place during the project period of EOSC-hub. This also affects the integration with B2HANDLE, as the transfer of result data to B2SHARE is not yet possible. At present, it is being investigated to what extent the standard PID profile used by B2HANDLE meets the requirements of ECAS.

The integration of a subset of ECAS components with the EGI FedCloud through the EC3 LToS service, for the dynamic deployment of the ECAS service, has been completed. An Ansible role and a RADL file have been provided to support automatic deployment of ECAS with the Infrastructure Manager (IM). The service has also been made available through the Marketplace (EGI AppDB) as a self-contained VMI. Extensions to the VMI and the Ansible role are planned in order to integrate a higher number of ECAS components in the deployment.

Several training events have been carried out during the whole project and additional training courses (also as virtual and online events) are planned for the future. Updated presentations and hands-on materials are under preparation to support end users in the exploitation of the ECAS service.

Inter-thematic-service collaboration (even beyond the end of the project) will be supported over the next months to discuss new integrated scenarios relying on multiple thematic services. In particular, based on preliminary interactions with OPENCoastS and the EGI team, joint ECAS & OPENCoastS use cases as well as training events could be planned/organised, pending evaluation of the use cases for possible common approaches.

4.2.4 GEOSS

EO DAB

This service is deployed in a production environment and is stable. Expected improvements will mainly focus on supporting additional services/APIs for discovering and accessing data made available by GEOSS providers.

VLab

The service is deployed in a pre-production environment and is regularly updated. The next developments will include: enhancing the support for multi cloud environment (e.g., Copernicus DIAS platforms), accessing to additional EO data/products natively stored in supported cloud platforms, and implementing a model chaining functionality. The VLab framework is adopted in other H2020 projects (e.g., ERA-PLANET) where such new developments will be implemented.

4.2.5 OPENCoastS

The OPENCoastS (On-demand Operational Coastal Circulation Forecast Service) service builds on-demand circulation forecast systems for user-selected coastal systems and keeps them running operationally for the time frame defined by the user. This daily service generates forecasts of relevant variables (water levels, 2D or 3D velocities, wave parameters, 3D salinity and temperature) over the spatial region of interest for periods of 48 hours, based on numerical simulations of all relevant physical processes. Currently, the service is deployed at two computing sites (NCG-INGRID-PT and IFCA).

As of June 2020, OPENCoastS has evolved to:

- Integrate EOSC core services such as UDocker and DIRAC4EGI
- Integrate new World and European forcing services such as CMEMs
- Develop integration with several EUDAT services and showcase them in simple OPENCoastS deployments

In the scope of the 2nd EAP call, OPENCoastS is supporting a new service for typhoon forecast and the following integration activities are planned:

- Integration with WRF atmospheric predictions
- Integration with Sentinel and Chinese remote sensing images

After the end of the project, the following actions are planned:

- Full integration with a data management system: EUDAT was tested and reviewed the API implementation in OPENCoastS, but the required resources were too vast. This issue will be pursued in following projects
- Extension to water quality predictions – the need to comply with several European Directives (such as the Water Framework Directive or the Bathing Water Directive) has prompted OPENCoastS users to request its extension to water quality variables. This extension is part of the EGI-ACE project.
- Extension to hindcast runs – while forecasts have a large community of users and OPENCoastS has already been used for establishing a model in the EAP Taiwan service, full operationalization of hindcast runs is especially important for the whole coastal community and can benefit greatly from the core services and infrastructure resources in the EOSC. This extension is considered in the EGI-ACE proposal.
- Integration with other EOSC services to provide added-value applications. An example is ECAS, where the requirements analysis to integrate its post-processing tool in OPENCoastS workflow was initiated in EOSC-hub.
- Allow to receive a broader users' community from OpenID Connect provided from EGI Check-in proxy service.
- Deploy the scripts environment in any EGI High-Throughput compute cluster using udocker to execute a container in user space without requiring root privileges.

4.2.6 DARIAH

DARIAH Science Gateway

The DARIAH Science Gateway is a mature service deployed on the EGI FedCloud infrastructure and is regularly maintained. The gateway provides several specific services and tools targeting the scholars and research from the domain of digital arts and humanities. There are no plans for technical improvements or further integration with other EOSC services. The roadmap foresees the extension of the gateway with other services and tools coming from digital arts and humanities, if required. In that case, if required, the gateway might be technically improved and integrated with

new or extend the usage of the existing EOSC services. The integration/extension actions will be decided upon for each new service/tool, individually.

Invenio-as-a-Service portal

The Invenio-as-a-Service portal is an operational service exploiting the EGI FedCloud infrastructure for running both, the Portal and the Invenio-repository instance deployment. No further technical improvements or extensions are foreseen.

DARIAH Repository

The DARIAH Repository is a production research data repository for the Arts & Humanities communities. It hosts a growing number of collections for a multitude of scholars, projects, and institutions. It is planned to massively extend the search capabilities of the repository through an integration with the search functions of the CLARIN ERIC. Furthermore, it is planned (as part of the SSHOC project) to integrate the CLARIN Switchboard, which is part of the respective Thematic Service offer in EOSC, into the DARIAH Repository. Depending on progress of the German national research data infrastructure funding, the DARIAH repository will also play a central role in a national German infrastructure for the humanities and will be developed well beyond the lifetime of the EOSC-hub project, but still offered through its Marketplace.

DARIAH Community Notebook

DARIAH Community Notebook is a new service planned to be integrated in the DARIAH Thematic Service. The DARIAH Community Notebook is a service based on the EGI Notebook service and will serve various DARIAH user, scholars, researchers and projects coming from the domain of digital arts and humanities domain, who are in need for an interactive Python-based platform to conduct their data preparation, data analysis and visualization. The service is planned to be integrated with the EGI Notebook service, EGI AAI and EGI DataHub. From the functionality, the service aims to extend the features for the users by allowing the notebooks and data to be shared among users.

4.2.7 LIFEWATCH

Plant Classification App

This service is deployed in a production environment and has been maintained stable for a year, supporting the requests from users via the web portal and the app API. There are no plans to extend this service more than improving the image classification, which is transparent for the final user.

Remote Monitoring and Smart Sensing

Although at the architectural level the service will remain in a similar way, it is still developing new features for the user, especially new functionalities for data retrieval. This includes some changes in the different code repositories that the service uses, which can be updated automatically.

Glacier Lagoons of Sierra Nevada

From the project "Glacier Lagoons of Sierra Nevada", no technical improvements are expected to be implemented. However, content updates on the lagoons and related events are (and will be) made. Additionally, it is important to note that users are contributing almost daily.

5 Integration activities

EOSC-hub is actively working on service integration and composability with a twofold approach. The project is defining an overall framework to facilitate service integration and composability leveraging the definition of interoperability guidelines as described in section 2. EOSC services conforming with these guidelines will offer well-established and documented interfaces for usage and integration, based on well-known standards or APIs, facilitating the exploitation of EOSC services from user communities willing to create new scientific services and the combined usage of EOSC services.

In parallel, the project is integrating flagship use cases from ESFRI, relevant scientific communities, SMEs and industries.

5.1 Thematic services

30 thematic services have already been integrated from several large research communities (CLARIN, CMS/DODAS, ECAS/ENES, GEOSS, OPENCoastS, WeNMR, EO Pillar, DARIAH, LifeWatch) and are now available to the European researchers through the EOSC Portal. Other additional cases will be fully integrated by the end of the project.

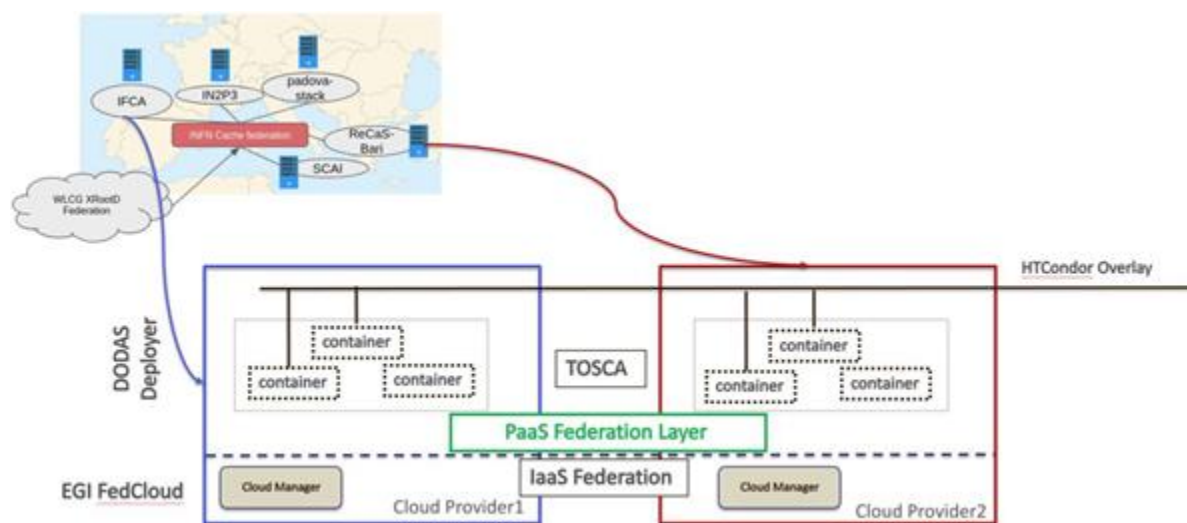


Figure 14. DODAS integration with the EGI Federated Cloud.

Before being published in the portal, these research-oriented services have been integrated with several EOSC-hub services (delivered by EGI, EUDAT and INDIGO) to implement essential features like the AAI and the monitoring and to re-use services supporting the whole data lifecycle from creation to processing, analysis, preservation, access and reuse. Examples of these services are cloud orchestrators, workflow engines, workload managers, data catalogues and repositories, etc.

Overall, the research-oriented services delivered by EOSC-hub **successfully achieved 40 integrations with 19 different services**. Some integrations covered multiple services, for example the integration of the EGI Workload Manager with WeNMR enabled the usage of this service in the

7 biological tools offered by the WeNMR suite. The integration process is continuing facilitated by the interoperability guidelines being delivered by the Technology Committee of the project.

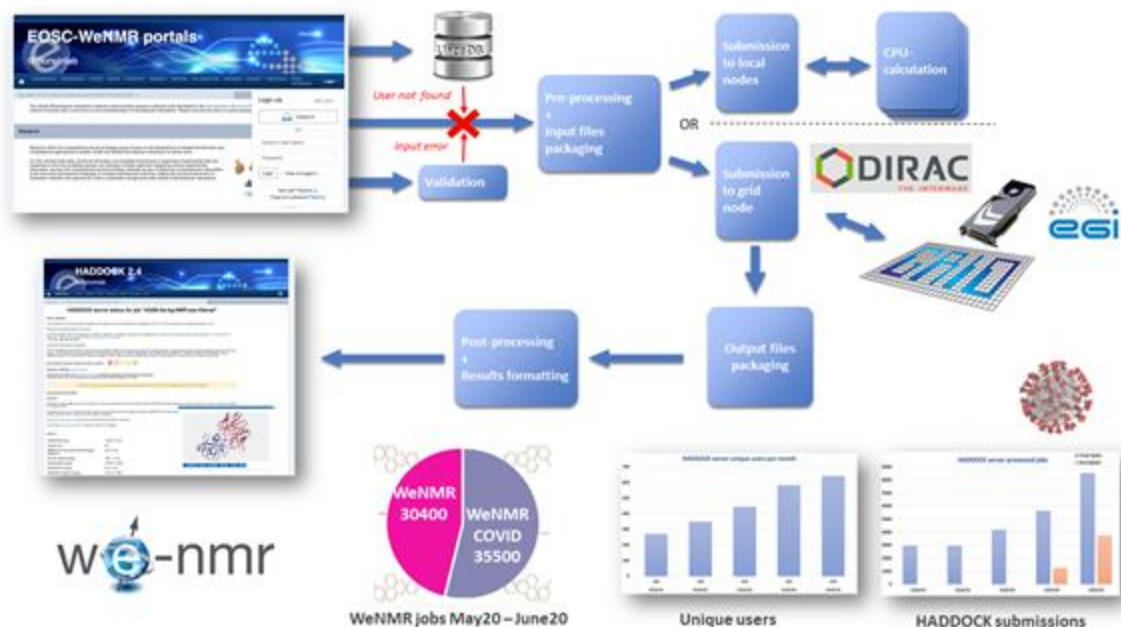


Figure 15. The WeNMR portal and HADDOCK, with its background machinery integrating various EOSC services. The bar and pie plots at the bottom report usage statistic, also highlighting the fraction of COVID-related submissions.

The publication of the services in the EOSC Portal fostered their uptake as demonstrated by key virtual access metrics. For example:

- The number of average monthly visits to metadata search portal for CLARIN service: 647 with a baseline value of 425, accounting for a +52% increase.
- The number of newly registered users per period in the WeNMR portal: 3844 with a baseline of 1750, an increase of 120% with a corresponding increase of the CPU hours consumed in the EGI infrastructure (from an average of 15M CPU/hours in an 8-month period to 18.5 CPU/hours, an increase of the +23%).
- One of the thematic services, OPENCoastS, was opened for cross-border access at European level and can now count 114 cross-border deployments.

The WeNMR portal has also seen a huge increase of registrations over the last months with many users indicating they intend to use the HADDOCK WeNMR service for COVID-19 projects. For this purpose, together with EGI/EOSC experts, the team is looking both into expanding the processing capacity of the HADDOCK portals on top of the EGI infrastructure and providing customized solutions to support researchers.

5.2 Competence Centers

The project includes eight Competence Centers (CCs) that work on establishing infrastructures to support users cope with the data deluge, with the challenges of various compute intensive data analysis scenarios. Each CC operates as a project on its own, with a small consortium composed of representative institutes from the Research Infrastructures, experts of relevant e-infrastructure services, and software/technology developers. CCs expect to bring scalable setups for ELIXIR, Fusion (ITER), Argo, SeaDataNet, EISCAT_3D, EPOS-ORFEUS, LOFAR and SKA, ICOS, eLTER and Disaster Mitigation communities. The overall objective of the CCs is to co-design and co-develop services for these communities by mobilising generic services from the EOSC Hub service portfolio.

All the 8 CCs successfully piloted EOSC-hub services and **19 services from the EOSC-hub** catalogue successfully passed the CC assessment and **were integrated with the community services**. Integration was completed for other 6 services that are being assessed by the communities while other 15 services are planned to be integrated by the end of the project.

Piloting of the technologies will continue until the end of the project and the number of EOSC-hub services that will be adopted by the CCs is expected to increase.

While most of the CCs aim to reach prototype and pilot setup with their community-specific services, some aim for production service level and making those **services available for access via the EOSC Portal**. In particular:

- The **Fusion CC** reached an initial version of their PROMINENCE service for the EOSC Portal launch event (Nov 2018), registered PROMINENCE in the EOSC Portal and now updating it with additional delivery options.
- The **Marine CC** aimed to reach a mature-enough ARGO data platform for EOSC by month 18. The service is in the EOSC Portal Marketplace⁴⁰ since November 2019.
- The **EISCAT_3D CC** published the Data Access Portal in the 2020.
- The **Disaster Mitigation CC** reached a production version of their tsunami simulation portal and applied for registration in the EOSC Portal. It is in the onboarding pipeline at the moment.

⁴⁰ <https://marketplace.eosc-portal.eu/services/european-marine-science-openaire-community-gateway>

EUROPEAN OPEN SCIENCE CLOUD

Find service... All services My EOSC Marketplace

Services > Compute > PROMINENCE

PROMINENCE
 Enabling HTC & HPC applications opportunistically across private, academic and public clouds
 Provided by: UK Atomic Energy Authority
 Research area: Interdisciplinary
 Dedicated for: Researchers, Research organisations, Business, Research group
 (0.0 / 5) 0 reviews Add to comparison

Access the service

ABOUT **REVIEWS (0)**

PROMINENCE is a platform which allows users to exploit idle cloud resources for running scientific workloads with a simple batch system style interface. Key features include:

- Jobs can be submitted from anywhere using any OS and any language.
- All jobs are run in containers to ensure they will run reliably anywhere and are reproducible.
- Multi-node OpenMPI jobs can be run in addition to HTC jobs.
- On-premises resources can be utilised with the ability to burst onto external clouds at times of peak demand.
- Goes beyond bursting onto a single external cloud with hierarchical cloud bursting. For example, burst onto national research clouds, to EGI FedCloud and finally to public clouds.
- All infrastructure provisioning is handled completely automatically and is totally transparent to the user.
- Clouds are selected automatically based on job requirements and preferences, and any failures are handled automatically.
- Output data can be accessed from cloud-based object storage.

PLACES AND LANGUAGES

- Europe
- English

SERVICE AVAILABILITY

Map showing service availability across Europe.

Figure 16. The PROMINENCE service from the Fusion competence center in the EOSC Portal.

5.3 EOSC Early Adopter Programme

EOSC-hub has also launched an EOSC Early Adopter Programme (EAP) for research communities interested in exploring the latest state-of-art technologies and services offered by the European Open Science Cloud (EOSC). As a result, 13 research projects were selected with 75 planned integrations. The EAP research projects are expected to publish their services in the EOSC Portal.

Baseline Implementation of EMSO ERIC Data Services

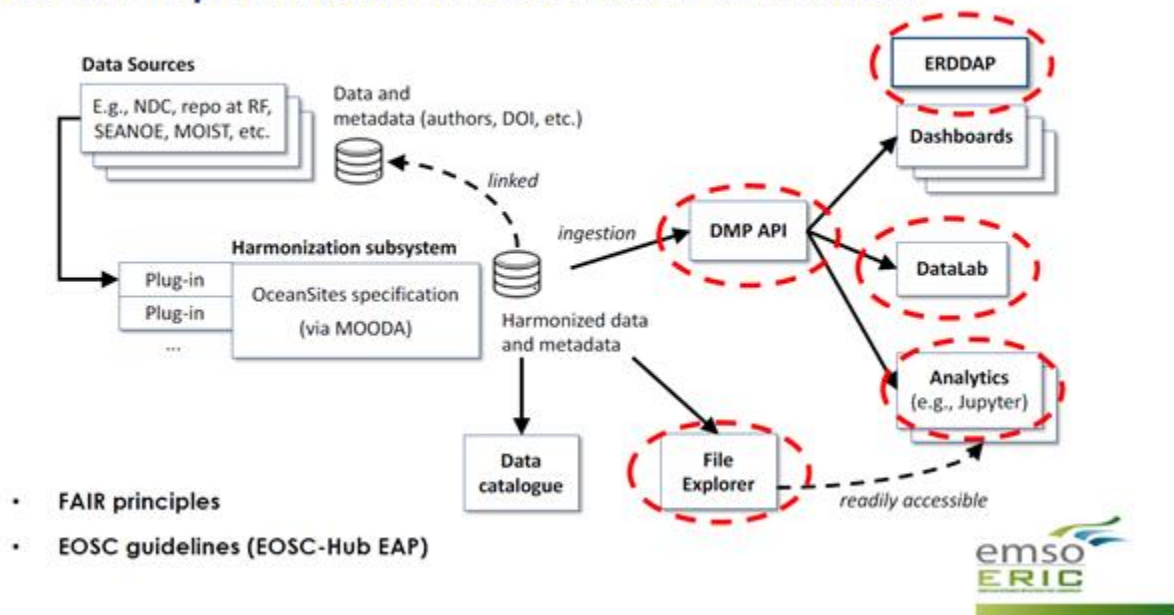


Figure 17. Architecture of the EMSO ERIC data management services deployed in the EGI Federated Cloud.

5.4 Business pilots

Finally, EOSC-hub also worked with SMEs and industries with the EOSC Digital Innovation Hub (EOSC DIH) that supported 11 pilots that achieved 21 integrations with EOSC-hub services.

5.4.1 Pilots

Business Pilot 1: CyberHAB (Water body management sector) focusing on the management of harmful algae blooms, exploiting Data Cloud Services (DCS) to support the key processes required (data processing, modelling, integration of images).

Service integrated: EGI Cloud Compute, INDIGO IAM, EGI DataHub, PaaS Orchestrator.

Business Pilot 2: Sports Smart Video Analysis (Sports sector) that has developed a mobile-friendly cloud platform, provided as a SaaS, for data-driven video analysis and automatic processing of videos of athletes' training sessions.

Service integrated: EGI Cloud Compute.

Business Pilot 3: ACTION Seaport (Local coastal authorities) ACTION Seaport is an advanced mobile-friendly platform providing accurate environmental and operational performance, capable of serving simultaneously multiple Port Authorities (as well as coastguards and other maritime authorities worldwide) in decision support to improve safety.

Service integrated: EGI Cloud Compute.

Business Pilot 4: Bot Mitigation Engine (Cybersecurity sector) that created a solution, called Guardonic, for online service providers in the business sector to prevent online services from botnet attacks such as web scraping, online fraud, digital ad fraud and spam. It behaves as a filter between global networks and a client's online services independent of where they are running.

Service integrated: PSNC Openstack cloud-based infrastructure.

Business Pilot 5: Space Weather Data Services for the future DRACO Observatory (Climate sector) developed a cloud-based framework for handling of the DRACO observatory data.

Service integrated: EGI Cloud Container Compute, HPC computing, intensive workloads, high throughput, large databases and application server.

Business Pilot 6: Furniture Enterprise Analytics - DataFurn (Furniture industry sector) developed a furniture analytics Platform-as-a-Service that collects, analyses and visualises online content (from social media and blogs to online portals), detects useful product-related content, extracts relevant furniture product-service topics/features, monitors brand influence and customer interactions and early predicts furniture trends for the upcoming seasons (e.g., regarding colours or textiles).

Service integrated: EGI Cloud Compute.

Business Pilot 7: Kampal - Artificial Intelligence for rare disease diagnosis: Assessing the probability of development of further diseases in Gaucher disease patients (finished).

Service integrated: EGI Cloud Compute.

Business Pilot 8: BI Insight - Business Intelligence, Artificial Intelligence and Big Data technologies

- Access the knowledge contained in artefacts: presentations, text documents, sheets and others.

Service integrated: EGI Cloud Compute, to be integrated: DEEP ML/DL services.

Business Pilot 9: BBC R&D - video coding and compression - Transforming video content through compression and large-scale processing.

Service integrated: PSNC HPC infrastructure.

Business Pilot 10: DCP - dynamic resource allocation and accounting in a digital marketplace -

Automating resource allocation and multi-metric accounting in a federated digital marketplace.

Service integrated: EGI Cloud Compute.

Business Pilot 11: NetService - Blockchain for university certificates Enabling public institutions to issue valid official documents in a digital form on the blockchain.

Service integrated: EGI Check-in, EGI Cloud Compute.