



EOSC-hub

D10.1 EOSC-hub Technical Roadmap v1

Lead Partner:	INFN
Version:	1
Status:	FINAL
Dissemination Level:	Public
Document Link:	https://documents.egi.eu/document/3502

Deliverable Abstract

This deliverable introduces the first 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.



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	INFN/WP10	4/10/2019
Moderated by:	Małgorzata Krakowian	EGI/WP1	
Reviewed by:	Catalin Condurache Bartosz Wilk Bartosz Kryza	EGI ACK Cyfronet ACK Cyfronet	
Approved by:	AMB		

DOCUMENT LOG

<i>Issue</i>	<i>Date</i>	<i>Comment</i>	<i>Author</i>
V.0.1	28/06/2018	Initial structure	Giacinto Donvito (INFN), Diego Scardaci (EGI F.), Stefano Nicotri (INFN), Mark Van De Sanden (SurfSARA), Lukas Dutka (Cyfronet), Giuseppe Fiameni (CINECA), Heinrich Widmann (DRKZ), Ignacio Blanquer (UPV), Enol Fernandez (EGU.eu), Joao Pina (LIP), Marica Antonacci (INFN), Marcin Plociennik (PSNC), Jens Jensen (STFC), Michal Prochazka (CESNET), Licia Florio (GEANT)
V.0.2	28/07/2018	Completed content	Diego Scardaci (EGI F.) Giacinto Donvito (INFN)
V.0.3	3/09/2018	Revision and corrections	Diego Scardaci (EGI F.) Giacinto Donvito (INFN)
V.0.4	10/09/2018	Revision and corrections	Diego Scardaci (EGI F.) Giacinto Donvito (INFN)
FINAL	15/10/2018	Final revision	Diego Scardaci (EGI F.)

			Giacinto Donvito (INFN)
--	--	--	-------------------------

TERMINOLOGY

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

Contents

1	Introduction.....	7
2	How the roadmap of each service is defined	8
2.1	Live process.....	8
2.2	From customers' requirements to service roadmap.....	9
3	Enhancing services to foster/facilitate integration and composability.....	10
3.1	Promoting standards and interfaces.....	10
3.2	Service integration and composability.....	19
4	Enhancing the EOSC-hub to facilitate service access and integration	25
4.1	Discoverability.....	25
4.2	Access.....	29
4.3	Marketplace projects.....	30
5	EOSC-hub services roadmap	31
5.1	Individual plans for service evolution	31
5.2	Planned integrations among EOSC-hub services.....	38
5.3	Long term evolution of the EOSC-hub services.....	39
5.4	Thematic Services roadmap.....	40
6	Collaboration with external projects.....	49
6.1	Collaboration with OpenAIRE Advance.....	49
6.2	Collaboration with GN4-2/GN4-3	50
6.3	Collaboration with other EOSC projects	50
7	References	56

Executive summary

This deliverable introduces the first 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.

The Technical Roadmap has primarily been shaped taking into account the requirements collected from user communities within (thematic services, competence centers, 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). The effective collaboration between the technical support team (T10.3) and the product teams working on service development (WP5, WP6 and WP7) allowed the inclusion of the users' requests into the roadmap and the release of new versions of the services satisfying such needs.

One of the main drivers behind the definition of the Technical Roadmap has been the ambition 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. To achieve this aim, well-known standards and protocols have been adopted by EOSC-hub services, and experts from the project are contributing to the definition and evolution of standards within standardisation bodies and collaboration like SNIA¹, RDA², OGF³, etc. Supporting standards simplifies and reduces the costs to combine services for the needs of the user communities. Furthermore, to increase the set of services that can be jointly exploited, the technical roadmap has also included worthy end-to-end integration activities that have been identified after the use cases analysis. Some of them have already been successfully accomplished and new combined solutions satisfying common and relevant user requirements are now available in the EOSC service portfolio, ready to be used by a large set of communities.

EOSC-hub access channels also need to be properly evolved to enable the combined usage of services and to promote their adoption between adjacent communities. For this aim, the roadmap includes activities to enrich the EOSC Service Catalogue and Marketplace, as the main EOSC access channels, with a series of features to improve discoverability, accessibility and facilitating the combination of more services. In this area, the EOSC Marketplace 'Project' feature deserves a special mention since it allows users to create a virtual lab where they can order, access, monitor and combine all the services of their interests.

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

¹ <https://www.snia.org/>

² <https://www.rd-alliance.org/>

³ <https://www.ogf.org/ogf/>

The EOSC-hub technical roadmap has been developed also taking into account collaborations with other projects working on the EOSC establishment and, in general, on tools for the Open Science that are being onboarded in the EOSC service portfolio. This document ends with a description of such collaborations and how they contributed to and influenced the roadmap.

1 Introduction

The main points which this deliverable focuses on are the principal features of the EOSC-hub services and their planned evolution within the timeline of the project. A large effort will be put into the enhancement of interoperability among services and their composition.

The document is organised as follows:

- Section 2 summarises the process to gather user requirements that drive the definition of the roadmap;
- Section 3 describes the already ongoing activity to improve service interoperability and composability;
- Section 4 contains the description of the activities to enrich the EOSC Service Catalogue and Marketplace, the main EOSC access channels, with a series of features to improve discoverability, accessibility and facilitating the combination of more services;
- Section 5 presents the roadmap for the most relevant EOSC-hub services including the integration activities between Thematic and federated/common ones;
- Section 6 depicts collaboration with external projects and their contributions to the definition of the EOSC-hub technical roadmap.

2 How the roadmap of each service is defined

The EOSC-hub technical roadmap is continuously revised and updated according to the requirements collected from user communities. This section summarises the requirement gathering process developed by the project, a detailed description is reported in the D10.5 *Requirements and gap analysis report v1*⁴.

2.1 Live process

The process of gathering requirements in EOSC-hub has been defined in the early stages of the project, and it is heavily inspired by the agile methodology for software development. It follows a top-down approach where high-level customer needs are progressively refined in order to identify the low-level technical requirements that emerge from them. Consequently, new developments targeted to the common, federation and collaboration services in the Hub may arise from the customer requirements, driving the technical evolution of those services' offerings over time.

The process starts with the definition of high-level user stories, then, from those, the use cases are defined and linked to the technical requirements that derive from them:

- *User stories* are informal descriptions in natural language of one or more features of a software system or service. Thus they appear as the highest level of abstraction of a requirement, usually taken as it was formerly formulated by the customer through the individual interviews.
- *Use cases* are the next step in the requirement gathering process, which identifies the single actions to be performed for each user story, where interactions between an actor and a system are implied. Since a system is involved, the dependencies with the services in the Hub are highlighted at this stage.
- Lastly, the *technical requirements* are the smaller units of development or integration work required to enable the use cases. They represent the outcome of the continual refinement process, and thus, the technical needs of the customer are acceptably scouted and delimited as part of their definition.

Technical requirements are assigned to the maintainers of each specific service. Further refinement might be done at this stage, as discussions flow among service maintainers, technical support and the customers themselves. In this regard, the technical support team oversees the progress of the implementation of the requirements, acting as a bridge between developers and customers. As the new individual features are implemented, the product roadmap is elaborated, mapping features to releases, with the ultimate objective of managing and aligning customer expectations in the short and long term.

This process supports the seamless integration of any new customer requirements in the Hub. It also supports the adaptive planning and fast response to changes in customer needs, as an outcome of the technical support teams' work.

⁴ <https://documents.egi.eu/document/3463>

2.2 From customers' requirements to service roadmap

The aforementioned requirement gathering process is supported by a set of tools allowing to track the requirements throughout their lifetime and enhancing the communication among the involved stakeholders, i.e. customers, technical support teams and developers.

The primary source of information is the Community Requirements Database⁵ that contains the documentation resultant from the requirement gathering process that outlines the outcomes from the technical assessment of each community under analysis. Once the technical requirements for each use case have been identified, they are monitored through the issue tracking system or activity dashboard. Each of these requirements is mapped to an issue created under the technical coordination queue and managed by the technical support teams. Once created, the issue is eventually assigned to the relevant development or integration teams and linked in the Community Requirements Database for future reference.

Issues in the activity dashboard are the primary source of discussion among the stakeholders. As a result, the requirements are able to evolve in the right direction under the supervision of the technical support teams and, at some point, some of them may result in requests for changes in the targeted services. When this happens, the delivery and deployment of the new features are tracked in a separate queue, specific to change management processing, within the same activity dashboard.

⁵ <https://wiki.eosc-hub.eu/display/EOSC/Community+requirements+DB>

3 Enhancing services to foster/facilitate integration and composability

Enabling the combined usage of different services and fostering their interoperability is a key element of the process of establishing a valuable and effective EOSC bringing advanced digital innovation for science. The availability of clear interoperability guidelines and integrated service offers would lower the barriers for user communities to simultaneously use more federation and common services (e.g. AAI, accounting, Cloud PaaS, workflow engine). Consequently, researchers may focus their work on actual scientific problems, profiting from advanced tools for exploiting IT capabilities and leveraging a powerful and easy-to-use infrastructure. Interoperability would also facilitate the creation of new workflows made of services developed by adjacent research communities creating opportunity for new collaborations and high-impact projects.

This section describes how the project is working to increase the interoperability between services in the EOSC portfolios and simplify the composition and integration of multiple services. A twofold approach has been followed:

- from one side, standards and widely used protocols are promoted and adopted wherever possible, since they are considered essential enablers to simplify service integration and composition. A list of EOSC endorsed standards is being defined in collaboration with other relevant EOSC stakeholders⁶;
- from the other side, specific end-to-end integration activities have been accomplished to create offers of integrated services. The integration activities have been identified and prioritised taking into account technical analysis and, in particular, the user requirements as they are gathered using the process described in section 2.

The main objective of this work is to contribute to the establishment of an EOSC where end users can deploy high-level community services that consume other EOSC services in an easier way, putting the ground for a creation of a dynamic, on-the-fly composition mechanism. This approach would simplify and reduce the costs of the integration of scientific services to EOSC and its federating core (EOSC shared resources plus a compliance framework)⁷ facilitating the growth of the EOSC portfolio and, in general, of its uptake.

3.1 Promoting standards and interfaces

Since the very beginning of the project, EOSC-hub has worked to foster open and well defined standards whenever possible. This has been accomplished by actively participating to standard bodies and boards, and by promoting the adoption of standards within its portfolios. Services that initially populated EOSC-hub portfolios were selected during the preparation of the proposal, taking

⁶ D10.4 EOSC Hub Technical Architecture and standards roadmap v2:

<https://documents.egi.eu/document/3495>

⁷ D2.3 Briefing Paper - EOSC Federating Core Governance and Sustainability

<https://documents.egi.eu/secure/ShowDocument?docid=3479>

into account their adherence to well-known standards. Furthermore, dedicated effort has been allocated to the definition of technical specifications and interoperability guidelines for Access Enabling and Federation services, that will compose the EOSC Federating Core, and Common services. EOSC-hub is intended to promote the identified standards in the context of this activity as EOSC endorsed standards and foster their adoption to the wider EOSC environment. All these actions have been undertaken with the final aim to foster the service composability in EOSC.

This section recaps the main activities in this area.

3.1.1 Participation to standardization bodies

Being aware of the importance of standard protocols and interfaces, partners of the EOSC-hub project are not only adopting standards in their services, but often they are also contributing to relevant standards bodies and working groups to develop enhancements aimed at addressing gaps in current solutions. This is a common effort across all the technical areas of the projects which leads to effective collaborations with different standardization bodies such as the Research Data Alliance (RDA)⁸, the Open Grid Forum (OGF)⁹, the Storage Networking Industry Association (SNIA)¹⁰, the HEP Software Foundation (HSF)¹¹, the Authentication and Authorisation for Research Collaborations (AARC) project¹², the Interoperable Global Trust Federation (IGTF)¹³, and others.

As examples, collaborations and related activities with some of these entities are described in this section.

3.1.1.1 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 and the “Data Usage Metrics” WG.

It is worth mentioning that EOSC-hub has a key role in the RDA working group on “Data Management Plans (DMP) Common Standards” that works to implement machine-actionable DMPs. The larger goal of this WG is to improve the experience of the involved people by exchanging information across research tools and systems and embedding DMPs in existing workflows. As a result, parts of the DMPs can be automatically generated and shared with other collaborators or funders. To achieve this goal, the WG is working on delivering: (1) a broader understanding of the variety of research data management planning processes, (2) a research data management infrastructure, (3)

⁸ <https://rd-alliance.org>

⁹ <https://www.ogf.org>

¹⁰ <https://www.snia.org>

¹¹ <https://hepsoftwarefoundation.org>

¹² <https://aarc-project.eu>

¹³ <https://www.igtf.net>

a common data model for machine actionable DMPs. Work on DMP is performed in collaboration with the OpenAIRE Advance project¹⁴.

3.1.1.2 Storage Networking Industry Association (SNIA)

The SNIA is a non-profit global organization dedicated to developing standards and education programs to advance storage and information technology. Activities carried out with SNIA are running since the INDIGO-DataCloud project¹⁵, where the official reference implementation of the SNIA Cloud Data Management Interface (CDMI), an ISO standard, has been adopted to develop a set of functionalities aimed at Improving QoS capabilities of storage resources for better support of high-level storage requirements, such as flexible allocation of disk or tape storage space and support for data life cycle. The collaboration with SNIA is successfully continuing in EOSC-hub.

3.1.1.3 HEP Software Foundation (HSF)

The HEP Software Foundation facilitates cooperation and common efforts in High Energy Physics software and computing at international level. To realize the physics programs of the planned and/or upgraded HEP experiments over the next 10 years, the HEP community will have to address a number of challenges in the area of software and computing. It is expected that the computing models will need to evolve and a significant “software upgrade” is required.

The LHC experiments and HSF have been appointed by the WLCG project to reach out to other HEP experiments around the world and start up a collaboration to define a community strategy and a roadmap for software and computing R&D in HEP for the 2020s.

EOSC-hub members involved in HSF are working on the definition of such common strategy with the aim to extend envisaged solutions to other science sectors and make significant progress on the reduction of the fragmentation of the IT facilities and digital tools.

3.1.1.4 Authentication and Authorisation for Research Collaborations (AARC)

In order to build a federated environment of e-infrastructures within EOSC, a reliable and secure A&A Infrastructure, able to support today's federated access requirements, has to be adopted.

In this context, the EOSC-hub AAI team is contributing to the definition of the EOSC infrastructure implementation roadmap by enabling seamless access to a system of research data and services provided across nations and disciplines. The EOSC-hub AAI is built on existing AAI solutions from EGI Federation, EUDAT CDI, and INDIGO-DataCloud that have successfully delivered a portfolio of operational services in this field over the last few years. The EOSC-hub AAI has been and continues to be shaped according to the recommendations delivered by the AARC project and other relevant alignments that facilitate the use of eduGAIN and federated access for research and thematic services. At the same time, the adoption of standards and open technologies, including SAML 2.0, OpenID Connect and OAuth 2.0, facilitates interoperability and integration with the existing AAI of other e-Infrastructures and research communities.

¹⁴ <https://www.openaire.eu/openaire-advance-project>

¹⁵ <https://www.indigo-datacloud.eu/>

3.1.2 Standards and interfaces currently adopted in EOSC-hub services

This section lists the standard interfaces and protocols (including *de jure* and *de facto* standards) currently adopted by a relevant set of services in the EOSC-hub portfolios. Services are grouped per technical areas.

3.1.2.1 Data discovery and access

3.1.2.1.1 B2FIND

B2FIND (<http://b2find.eudat.eu>) is an interdisciplinary discovery portal for research data that are stored within EOSC-hub and beyond. Therefore, metadata collected from heterogeneous sources are indexed in a comprehensive joint metadata catalogue and made searchable via an openly accessible web interface. B2FIND provides transparent access to the scientific data objects through the given references and identifiers in the metadata, thus supporting (at least) the first two pillars of FAIR¹⁶ data principles.

B2FIND offers Guidelines for Data Providers, including research data management recommendations, references to FAIR data principles and technical requirements concerning harvesting methods as well as advice for aggregation levels and metadata quality in general:

- <http://b2find.eudat.eu/guidelines/introduction.html>
- <http://b2find.eudat.eu/guidelines/providing.html>
- <http://b2find.eudat.eu/guidelines/harvesting.html>

B2FIND offers a training module in GitHub:

- <https://github.com/EUDAT-Training/B2FIND-Training>

All B2FIND code is openly accessible and reusable in GitHub:

- <https://github.com/EUDAT-B2FIND>

3.1.2.1.2 EGI DataHub

EGI DataHub (<https://datahub.egi.eu>) is a service aimed at provisioning large reference open data sets, based on the Onedata distributed virtual file system platform, available to end users over standard POSIX interface. For service management and integration, Onedata provides comprehensive REST API for each of its constituting services:

- **Onezone** - <https://onedata.org/#/home/api/latest/onezone>
- **Oneprovider** - <https://onedata.org/#/home/api/latest/oneprovider>
- **Onepanel** - <https://onedata.org/#/home/api/latest/onepanel>

For data access Onedata offers 3 main options:

1. **POSIX** - available using the Oneclient command line tool, which creates a mount point with a virtual filesystem containing data accessible to a given user,
2. **CDMI** - standard HTTP data access and management interface (<https://www.snia.org/cdmi>),

¹⁶ <https://libereurope.eu/wp-content/uploads/2017/12/LIBER-FAIR-Data.pdf>

3. **Web GUI** - easy to use web graphical interface enabling basic uploading and downloading of files.

3.1.2.1.3 Identity and Access Management (IAM) Service

The Identity and Access Management Service (<https://www.indigo-datacloud.eu/identity-and-access-management>) is an integrated AAI solution which provides a layer where identities, enrolment, group membership and other attributes and authorisation policies on distributed resources can be managed in a homogeneous way.

IAM can be integrated with relying services via standard OpenID Connect/OAuth interfaces and heterogeneous authentication mechanisms (SAML, X.509 certificates, plain username/password).

IAM offers a RESTful provisioning and management API based on the [SCIM](#) standard as well as the ability to be natively integrated to Grid services via a [VOMS](#) compatibility layer.

3.1.2.1.4 B2STAGE

B2STAGE (<https://www.eudat.eu/b2stage>) is a suite of services aimed at transferring data to and from EUDAT data nodes and exposes two protocols for staging data: GridFTP and HTTP API.

GridFTP (via the EUDAT Data Storage Interface) is a service aimed at large data transfers, and to the movement of a large number of files between HPC centers and EUDAT in order to store and process them and, possibly, move the results back.

The HTTP APIs service is a set of RESTful endpoints that allow access to both B2SAFE data and metadata and it is targeted at small to medium datasets. This service offers programmatic access to data and allows for smooth integration of such data into other applications and data services.

3.1.2.1.5 B2DROP

B2DROP (<https://www.eudat.eu/services/b2drop>) is a Sync&Share service offering researchers an easy way for collaborative working on documents and synchronisation of data across multiple devices. Beside the common functionalities of sync and share services, B2DROP is connected to other services, such as EUDAT B2SHARE or CLARIN Switchboard, and offers a one-click file transfer to them.

It offers a web frontend for interactive use, and WebDAV API for clients and connected services.

3.1.2.2 Federated Compute

3.1.2.2.1 EGI Cloud Compute

EGI Cloud Compute (<https://www.egi.eu/services/cloud-compute/>) provides users with a distributed computing service to deploy and scale virtual machines on-demand. It gives access to API-controlled computational resources in secure and isolated environments, freeing the final user from the hassle of managing physical servers.

Some features of EGI Cloud Compute are:

- GUI access: AppDB VMOps <https://dashboard.appdb.egi.eu/vmops>
- API/CLI access:

-
- Discovery: AppDB is API (REST and GraphQL) accessible via https://wiki.egi.eu/wiki/Federated_Cloud_Discovery#AppDB
 - IaaS Federated Access Tools, see https://wiki.egi.eu/wiki/Federated_Cloud_IaaS_Orchestration
 - Direct IaaS access, via several APIs, depending on the provider: https://wiki.egi.eu/wiki/Federated_Cloud_APIs_and_SDKs

3.1.2.2.2 Cloud Container Compute

Cloud Container Compute allows to deploy and scale Docker containers on-demand using Kubernetes technology¹⁷. The service uses Kubernetes clusters on EGI Cloud Compute resources that can be scaled and upgraded without the overhead of installing, managing and operating the nodes.

Native Kubernetes API with OpenID Connect authentication is used, see at

<https://kubernetes.io/docs/reference/access-authn-authz/authentication/#authentication-strategies>

Kubernetes API: <https://kubernetes.io/docs/concepts/overview/kubernetes-api/>.

3.1.2.2.3 EGI Workload Management

EGI Workload Management allows users to manage and distribute their computing tasks in an efficient way, maximising the usage of computational resources.

The Workload Manager service is based on DIRAC technology and is suitable for users that need to exploit distributed resources in a transparent way. The service has a user-friendly interface and also allows easy extensions for the needs of specific applications via APIs.

DIRAC documentation is available at <https://dirac.readthedocs.io/en/latest/index.html> .

3.1.2.2.4 EGI Online storage

EGI Online storage is a service that allows to store data in a reliable and high-quality environment and share it across distributed teams. Data can be accessed through different standard protocols and can be replicated across different providers to increase fault-tolerance.

Online Storage is offered via different APIs depending on the available providers and user needs:

- Block Storage is offered via OCCl/OpenStack APIs: <http://occi-wg.org/>, <https://api.openstack.org/>
- Object Storage is offered via Swift: <https://api.openstack.org/>
- File-based Storage is offered via SRM: <https://sdm.lbl.gov/srm-wg/doc/SRM.v2.2.html>, WebDAV and GridFTP.

¹⁷ <https://kubernetes.io/>

3.1.2.2.5 EGI High-Throughput Compute

EGI High-Throughput Compute allows running computational jobs at scale on the EGI infrastructure. It allows to analyse large datasets and execute thousands of parallel computing tasks on a distributed network of computing centres, accessible via a standard interface.

The service is offered via three different APIs, depending on the providers and users preferences:

- CREAM: <https://wiki.italiangrid.it/twiki/bin/view/CREAM/WebHome>
- ARC: <http://www.nordugrid.org/arc/>
- QCG: <http://www.qoscosgrid.org/qcg-now/en/>

3.1.2.3 Advanced IaaS

3.1.2.3.1 uDocker

uDocker (<https://github.com/indigo-dc/udocker>) allows the execution of applications and services within virtualized environments similar to Linux containers. It enables the execution of Docker containers without requiring any privileges for both installation and execution, making it especially suitable to execute applications in batch systems or other environments where the end user does not have system administrator privileges. It is being used for high throughput computing, grid computing, GPU computing and high-performance computing.

uDocker is meant to be used from the command line and offers a Docker-like command line interface.

3.1.2.3.2 CVMFS

The CernVM File System (CernVM-FS or CVMFS) is a read-only file system designed to deliver scientific software onto virtual machines and physical worker nodes in a fast, scalable, and reliable way.

CernVM-FS is a file system with a single source of data. This single source, the Stratum-0 repository is maintained on a dedicated release manager machine or publisher.

CernVM-FS documentation is available at <https://cvmfs.readthedocs.io/en/stable/>

3.1.2.4 Processing and Orchestration

3.1.2.4.1 TOSCA for Heat

TOSCA for Heat is a tool that translates TOSCA templates to Heat Orchestration Template (HOT) format. For detailed description of the service see D6.1 and <https://github.com/indigo-dc/heat-translator>. A Command Line Interface (CLI) is used.

3.1.2.4.2 Infrastructure Manager

Infrastructure Manager (IM) is a tool that orchestrates the deployment of complex and customized virtual infrastructures on multiple Cloud providers.

IM offers a web-based GUI¹⁸, an XML-RPC API¹⁹, a REST API²⁰ and a command-line application²¹.

3.1.2.4.3 PaaS Orchestrator

The PaaS Orchestrator is a service that allows to:

1. coordinate the provisioning of cloud resources and the deployment of virtual infrastructures on heterogeneous cloud environments like private clouds (OpenStack, OpenNebula) and public clouds (Amazon Web Services, Microsoft Azure);
2. manage the deployment of dockerized long-running services or the execution of dockerized batch-like jobs on top of Apache Mesos clusters.

The PaaS orchestrator exposes REST API endpoints documented at <https://indigo-dc.github.io/orchestrator/restdocs/>; request/response data is transferred in the compact and easy-to-use JSON data-interchange format.

The Orchestrator supports the TOSCA standard for describing the topology of the virtual infrastructures and the services to be deployed. The deployment requests submitted by the user to the orchestration tools must adhere to the TOSCA template syntax defined by the TOSCA's YAML Simple Profile that specifies a rendering of TOSCA providing a more accessible syntax as well as a more concise and incremental expressiveness of the TOSCA DSL (Domain Specific Language).

The adoption of the TOSCA standard ensures the portability of the deployment topology description across different cloud providers and the support of the cloud bursting use-case.

3.1.2.4.4 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.

FutureGateway foresees the following software components:

- **fgSetup** collects scripts and procedures to install and maintain the FutureGateway.
- **fgAPIServer** is a Python-based implementation of the FutureGateway APIs.
- **fgAPIServerDaemon** is a daemon process that addresses physical distributed resources. This component uses a set of sub-component named **ExecutorInterfaces** that can be developed in order to address any kind of distributed environment. At the moment, the available executor interfaces are: '*ToscaIDC*' and '*Grid and Cloud Engine*'.

¹⁸ <https://github.com/grycap/im-web>

¹⁹ <https://imdocs.readthedocs.io/en/latest/xmlrpc.html>

²⁰ <https://imdocs.readthedocs.io/en/latest/REST.html>

²¹ <https://github.com/grycap/im-client>

3.1.2.5 Metadata Management

3.1.2.5.1 B2HANDLE

B2HANDLE (<https://www.eudat.eu/services/userdoc/b2handle>) is a service to assign persistent identifiers (PIDs) to digital objects, making them referenceable across middleware services. B2HANDLE relies on the Handle System to achieve this.

B2HANDLE offers multiple service interfaces, including the native (Java) Handle API and the native Handle REST API for PID management (create, read, update, delete), the B2HANDLE search service at each service provider and at the B2HANDLE Central PID Catalogue. To facilitate easy programmatic access to these interfaces, including search, B2HANDLE provides the `b2handle/pyhandle` Python libraries. The `pyhandle` library also supports interaction with the B2HANDLE raw data database (for administrative purposes only) and supports the generic Handle Batch format to accommodate asynchronous bulk operations.

3.1.2.5.2 B2SAFE

B2SAFE is a long-term preservation and policy based data service. It allows community repositories to implement data management policies on their research data that is distributed across multiple administrative domains.

B2SAFE is accessible through various interfaces.

- IRODS legacy: the interfaces offered by the iRODS component. They are CLI and API, java, python and C (<https://irods.org>). Moreover a WebDAV interface is available as a separate component (<https://github.com/UtrechtUniversity/davrods>).
- HTTP API: it is a RESTful interface which exposes functions to upload and download data (see 3.1.2.1 for B2STAGE).
- GridFTP: a bridge between iRODS and the GridFTP service is available as explained in the paragraph about B2STAGE. That allows uploading and downloading data relying on the high-performance transfer features of GridFTP.
- Data Policy Manager Web UI: the user web interface that allows to define data policies and store them in a DB, from where they can be distributed to multiple B2SAFE instances.
- Data Policy Manager REST API (BaseX API): the HTTP API of the BaseX XML DB component, which stores the data policies XML documents.

3.1.2.5.3 B2SHARE

B2SHARE (<https://eudat.eu/services/userdoc/b2share-usage>) is a data storage and sharing service for research communities and individual researchers. It allows discovery and publication of research datasets by providing detailed descriptions in the form of standardized metadata.

B2SHARE is accessible through a web interface and a REST API that allows a user to create, modify and manage records. The service is integrated with B2DROP to allow direct uploads, B2HANDLE to mint new handles, B2NOTE for additional annotation of files in records and B2ACCESS for authentication.

For metadata harvesting, an OAI-PMH endpoint is available that supports multiple metadata prefixes for compatibility with B2FIND, OpenAIRE RCD and other metadata catalogues.

3.1.2.5.4 B2NOTE

B2NOTE (<https://eudat.eu/catalogue/B2NOTE>) is a data annotation service integrated with data repositories/data publication services. It allows the users to add extra information without modifying the underlying data record. Annotations are structured using the W3C Web Annotation data model, serialized in JSON-LD and stored in a document database (MongoDB). These annotations can then be used to organize and retrieve datasets based on the user's needs.

B2NOTE offers two different interfaces:

- a User Interface, available as a widget for the integration within the User Interface of partner services
- a RESTful API to initialize the annotations and retrieve stored annotations.

The User Interface has been designed to offer functionalities for the creation, management and usage of annotations despite the reduced size. This interface should be extended to offer more convenient functionalities for users.

The RESTful API is currently limited to a small subset of functionalities but should be extended depending on the various integration requirements.

3.2 Service integration and composability

This section describes the end-to-end integration activities that have been identified, planned and, in some cases, already successfully accomplished to create new combined solutions that satisfy the most common and relevant user requirements.

Service integrations have been identified in several ways. They may be proposed through a number of different channels, either internal to the project (e.g. through the various engagement activities) or external (e.g. requests from the EOSC portal or from related projects, including other EOSC projects). Once received, the requests are prioritised, and precedence will be given to those that can be adopted in various contexts and re-used by more communities.

Some of the services in the current EOSC-hub portfolios offer basic features (e.g. federation and common services) and may be considered as *building blocks* useful to create more complex services. User communities can pick-up two or more of these services and combine them to create or enhance scientific tools. The project provides the technical support and starts any needed integration activities to enable the combined usage. When two services are integrated, their combination can be adopted also by other cases. The Competence Centers (Work Package 8) are typical examples of user communities testing different combinations of basic services to check how they can fulfil their needs.

Service composition and integration represents a fundamental way to enhance and enrich the solution portfolio offered by EOSC-hub and has thus been recognised as beneficial by many research communities. This is the reason why EOSC-hub provides a qualified and strong support to it.

3.2.1 EOSC-hub interoperable services, current status

The current status of integration activities in the area of federation and common services (WP6) is described in detail in deliverable 6.2 “First report on the maintenance and integration of common services”²², a brief summary of these activities, split in four domains (Data Discovery and Access, Federated Compute, Processing and Orchestration, and Data and Metadata Management), is provided here.

3.2.1.1 Data Discovery and Access

EOSC-hub is working on providing a common data discovery and access layer which supports the FAIR data principles. This layer is formed from numerous services including B2FIND, EGI Datahub, INDIGO IAM, B2STAGE and B2DROP. To achieve this objective, these services were further developed and several integration activities were undertaken including exposing data in EGI DataHub via B2FIND, integration of EGI DataHub and B2SAFE, B2STAGE and IAM to enable seamless data access and staging across services and integration of B2DROP with B2SAFE to retrieve and store small datasets.

Completed and planned integration activities for each of these services are listed below. Planned integration activities will be implemented in the second reporting period taking into account the users’ requests.

3.2.1.1.1 B2FIND

Completed integration activities:

- B2FIND is integrated with B2SHARE, incrementally harvesting on a daily basis for records in B2SHARE.
- B2FIND is integrated with EGI DataHub via its OAI-PMH endpoint exposed by EGI DataHub, which exposes the metadata of all published open data sets in EGI DataHub.

Planned integration activities:

- B2FIND can be integrated with B2SAFE as shown in Herbadrop Use Case.

3.2.1.1.2 EGI DataHub

Completed integration activities:

- EGI DataHub is integrated with B2FIND through its OAI-PMH endpoint, which exposes the metadata of published open data sets in EGI DataHub. Moreover, EGI DataHub allows easy minting of data handles (including DOI and PID), which enables the assignment of permanent identifiers to the published data sets, which can then be referenced. This feature allows to implement dataset discoverability in the OpenAIRE Community Dashboard.
- EGI DataHub is integrated with B2HANDLE, allowing users to automatically mint a PID handle while publishing an open data set.
- EGI DataHub is indirectly integrated with B2ACCESS, as it already supports login via EGI Check-in (which is integrated with B2ACCESS).

²² <https://documents.egi.eu/document/3480>

-
- EGI DataHub is integrated with B2STAGE via the WebDAV protocol, enabling data transfers between the EGI and EUDAT users.
 - EGI DataHub is integrated with the INDIGO IAM.

3.2.1.1.3 INDIGO IAM

Completed integration activities:

- IAM has been integrated via SAML and OpenID Connect interfaces with the INDIGO PaaS services, Onedata (the technology underlying the EGI DataHub), StoRM WebDAV and other services.
- IAM has been successfully integrated with the EGI Check-in service in order to enable federated authentication, leveraging both the OpenID Connect and SAML interfaces. This integration has been demonstrated using the IAM instance dedicated to the DODAS thematic service.
- IAM is integrated with EGI DataHub.

3.2.1.1.4 B2STAGE

Completed integration activities:

- B2STAGE is integrated with B2SAFE by supporting all authentication protocols (native, GSI and PAM). B2STAGE is able to expose data stored into B2SAFE by both referring to PIDs and data paths.
- B2STAGE is integrated with B2ACCESS by implementing the full OAuth2 authorisation protocol and managing both access and refresh tokens provided by B2ACCESS.

Planned integration activities:

- B2STAGE can be integrated with B2SHARE by using B2ACCESS as common authentication layer. With this integration, users will be able to retrieve data from B2STAGE and share into B2SHARE.

3.2.1.1.5 B2DROP

Completed integration activities:

- B2DROP is integrated with B2ACCESS and B2SHARE.

3.2.1.2 Federated Compute

The Federated Compute activities focused on integration activities of the cornerstone services, Cloud Compute, Cloud Container Compute and High Throughput Compute, designed to improve service deployment and management.

3.2.1.2.1 EGI Cloud Compute

Completed integration activities:

- Cloud Compute is integrated with EGI Check-in.

Planned integration activities:

- It can be integrated with B2DROP and EGI DataHub.

3.2.1.2.2 EGI Cloud Container Compute

Completed integration activities:

- EGI Cloud Container Compute is integrated with EGI Check-in, Cloud Compute.

Planned integration activities:

- It can be integrated with B2DROP and EGI DataHub.

3.2.1.2.3 EGI Workload Management

Completed integration activities:

- EGI Workload Management is integrated with High Throughput Compute, Online Storage.

Planned integration activities:

- It can be integrated with EGI Check-in and Cloud Compute.

3.2.1.2.4 EGI Online Storage

Completed integration activities:

- EGI Online Storage is integrated with Cloud Compute and High-Throughput Compute.

3.2.1.2.5 EGI High Throughput Compute

Completed integration activities:

- EGI High Throughput Compute is integrated with Online Storage.

3.2.1.2.6 Advanced IaaS: uDocker

Planned integration activities:

- uDocker can be integrated with EGI High-Throughput Compute, EGI Workload Management and EGI Cloud Compute.

3.2.1.2.7 CVMFS

Completed integration activities:

- CVMFS is integrated with High-Throughput Compute.

Planned integration activities:

- It can be integrated with the Cloud Container and Cloud Compute.

3.2.1.3 *Processing and Orchestration*

In the Processing and Orchestration domain, the project is working on the integration of orchestration services with the federated compute services. The developments have mainly been based on the use of TOSCA templates to describe deployments and services such as Infrastructure

Manager (IM), INDIGO PaaS and the FutureGateway to enable the deployment of complex environments or workflows. Additional activities were undertaken to improve integration with the Configuration Management Database (CMDB) and Service Level Agreement Manager (SLAM)

3.2.1.3.1 Heat-translator

Planned integration activities:

- Heat-translator can be used by TOSCA-aware services such as the IM and PaaS Orchestrator in order to deploy resource stacks in OpenStack Heat.

3.2.1.3.2 Infrastructure Manager (IM)

Completed integration activities:

- IM is integrated with the PaaS Orchestrator, EC3, EGI VMOps Dashboard and EGI FedCloud.

3.2.1.3.3 PaaS Orchestrator

Completed integration activities:

- The PaaS Orchestrator is integrated with INDIGO IAM, Infrastructure Manager, CMDB and SLAM, and Onedata.

3.2.1.3.4 FutureGateway

Completed integration activities:

- The FutureGateway framework is integrated with the INDIGO PaaS Orchestrator.

Planned integration activities:

- It can be integrated with EGI Check-in and INDIGO IAM.

3.2.1.4 *Data and Metadata Management*

This domain focuses on common repository service and policy driven data management/stewardship. The activities have included the further development and integration of services such as B2HANDLE, B2SAFE, B2SHARE and B2NOTE.

Working on improving the authentication workflow with B2ACCESS and the interoperability among data services.

3.2.1.4.1 B2HANDLE

Completed integration activities:

- B2HANDLE is already integrated with B2SAFE and B2SHARE.

Planned integration activities:

- B2HANDLE is now working on improving the integration with B2SHARE, supporting PID metadata profiles, and initial integration with EGI DataHub, Online Storage and the Federated Data Manager.

- B2HANDLE supports integration with EOSC-conformant AAI services wherever possible and/or useful.

3.2.1.4.2 B2SAFE

Completed integration activities:

- B2SAFE is integrated with B2ACCESS and B2STAGE

Planned integration activities:

- It can be integrated with EGI DataHub, B2DROP, B2SHARE, B2ACCESS and B2STAGE.

3.2.1.4.3 B2SHARE

Completed integration activities:

- For monitoring purposes, a Nagios plugin that interfaces with the EUDAT Argo monitoring service has been developed and released.

Planned integration activities:

- Design documents have been written to describe the requirements and necessary changes to B2SHARE in order to improve the interfacing to OpenAIRE Community Dashboard, EGI DataHub and Online Storage and B2NOTE.

3.2.1.4.4 B2NOTE

Completed integration activities:

- B2NOTE is integrated with B2SHARE.

Planned integration activities:

- B2NOTE can be integrated with B2FIND and with OpenAire data services such as Zenodo and the Research Community Dashboard.

Integration with B2SHARE is planned to be improved.

4 Enhancing the EOSC-hub to facilitate service access and integration

This section describes the on-going work of the project to provide the EOSC-hub with access channels that, from one side, simplify the discovery, usage and exploitation of the EOSC services and, from the other side, facilitate and promote the combined usage of multiple services assisting the users in identifying and composing compatible or integrated services. This is being achieved by enriching the EOSC Marketplace with a series of features to discover, access and combine services that are described below.

4.1 Discoverability

The main challenge is to create a platform allowing the user to find a suitable solution in the EOSC Catalogue, even if the researcher is not aware of what is available. Services need to be defined and described in a way which is understandable to the user, so that it can be easily assessed whether a given service might be useful. The discoverability in the platform (e.g. identifying the most suitable services) can be guaranteed and progressively improved fostering and constantly refining the mechanism related to the interactions with service providers and users. These are described below.

4.1.1 Service search and filtering tools

Implementing powerful service search and filtering tools requires:

- an adequate metadata schema to describe services and the implementation of filters and tags based on these metadata;
- thoughtful revision of the service information in collaboration with the service providers.

The screenshot shows a search interface for the EOSC Marketplace. The search term is 'cloud'. The results are filtered by 'Researchers' and 'Research group'. The search results are sorted by 'Best match' and show 10 items per page. The results list three services:

- EGI Cloud compute**: Run virtual machines on-demand with complete control over computing resources. Provided by: EGI Federation. Research area: Interdisciplinary. Dedicated for: Researchers, Research organisations.
- EGI Cloud container compute BETA**: Run Docker containers in a lightweight virtualised environment. Provided by: EGI Federation. Research area: Interdisciplinary. Dedicated for: Researchers, Research organisations.
- Elastic Cloud Compute Cluster (EC3)**: Deploy a virtual cluster on top of IaaS infrastructures with a few clicks. Provided by: GryCAP-I3M-UPV. Research area: Interdisciplinary. Dedicated for: Researchers, Research organisations, Business.

The left sidebar contains a 'CATEGORIES' list with counts: All Services (65), Compute (12), Data management (13), Networking (1), Processing & Analysis (26), Security & Operations (3), Sharing & Discovery (14), Storage (8), and Training & Support (5). Below this is a 'FILTERS' section with expandable categories: Research Area, Providers, Dedicated for (Business, Providers, Researchers, Research group, Research organisations), Related Infrastructures and platforms, Rating, and Provider location. A 'Filter' button is located at the bottom of the sidebar.

Fig. 1. Service list provided by the EOSC Marketplace as result of a search based on the word 'cloud'.

4.1.2 Identifying similar services

Finding (through search and filtering capabilities) and testing a service might bring further interest to a user. After a customer has identified a suitable working service, he/she might be tempted to find another similar one to further explore the EOSC and make comparisons. This use case can be supported with a tag mechanism allowing to identify similar services. This kind of tag mechanism already exists in the EOSC Marketplace. Tags can be triggered in two ways:

- inserting a tag into the search bar;
- looking up the tags connected with a service in its description page and clicking on one of them. The EOSC Marketplace will show a list of services associated with the selected tag.

Alien and Invasive Species Virtual Research Environment (VRE) is a web-based, comprehensive, collaborative working environment supporting decision makers and scientists in predicting the spread of an invasive species (possibly alien) in a new environment. The VRE hosts examples of suitable habitat maps produced for today and 2050 in new areas for more than 11,000 species and provides models and workflows to combine environmental data with species observations in their habitats to predict their future spread.

Features:

- A service providing state-of-the art Data Analytics algorithms suitable for alien and invasive species spread prediction;
- A service for seamless discovery and access to species occurrence data and taxa names from several providers providers (eg. BrazilianFlora, OBIS, WoRMS, GBIF, ITIS, CatalogueOfLife);
- A complete spatial data infrastructure for discovering, accessing and publishing spatial datasets according to OGC standards;
- A catalogue service for discovering, accessing and publishing any research object (multi-part objects with actionable parts) and promoting its re-use;
- A social networking service supporting communication among VRE members by sharing and commenting posts;

PLACES AND LANGUAGES

- Europe
- English

RELATED INFRASTRUCTURES AND PLATFORMS

- EGI
- D4Science
- GBIF
- OBIS
- SeaDataCloud
- gCube System

SUPPORT

- [Webpage](#)
- [Manual](#)
- [Helpdesk](#)
- [Tutorial](#)

DOCUMENTS

- [Corporate SLA](#)
- [Terms of use](#)
- [Access policies](#)

SERVICE RESTRICTIONS

- No restrictions

SERVICE PHASE

- Production (min. TRL 8)

Tags: [applications](#) [data analytics](#) [species distribution](#) [virtual research environment](#)

Fig. 2. Web page of a service in the EOSC Marketplace. Tags are listed in the bottom of the page.

Fig. 3. List of the services showed by the EOSC Marketplace after a customer selects the tag ‘application’.

4.1.3 Service interoperability levels reflected in the GUI

The visual linkage between services that are somehow interoperable is another mechanism, already offered by the EOSC Marketplace, that can foster service composition and integration. This feature improves the service discovery and guides users in choosing the right solutions to create a service pipeline that can fully support their data processing needs, exploiting all the potential of the EOSC service catalogue.

EOSC services are classified, with respect to the service interoperability, in the following three categories:

- 1 Not interoperable services: inputs and outputs from them cannot create a logical workflow - trying to run them together makes no sense. There is no linking at the GUI level, user should have a clear message that those services, even if added to the same project, will not be exploitable together.
- 2 Compatible services: there is no low-level integration, but it makes sense to use those services together. Their inputs and outputs can create a logical workflow. The Marketplace GUI provides visible links to those services on a service entry level.

Suggested compatible services

The screenshot displays five service cards under the heading 'Suggested compatible services':

- HADDOCK**: Integrative modeling of biomolecular complexes with the user-friendly, EGI HTC-enabled HADDOCK portal. By: Bijvoet Center, Utrecht University. For: Researchers.
- AMBER**: Web portal for Nuclear Magnetic Resonance (NMR) structure refinement. By: Magnetic Resonance Center of the University of Florence - CERM, Interuniversity consortium CIRMP. For: Researchers.
- EGI Cloud compute**: Run virtual machines on-demand with complete control over computing resources. By: EGI Federation. For: Researchers, Research organisations.
- PowerFit**: The PowerFit server allows you to fit. (Description partially obscured).
- FANTEN**: (Description partially obscured).

Fig. 4. List of compatible services showed by the EOSC Marketplace.

- 3 Services integrated at a technical level: as a result of this integration, new solution based on a service bundle can be made available in the EOSC Marketplace and offered as a unique service in the GUI. Ordering a bundle (two or more services) will trigger the creation of one order for multiple services.

4.2 Access

In order to offer a better user experience and properly guide the user, it became clear that the user should undertake different paths depending on the access policies of the selected services.

The analysis identified three different types of service access:

- 1 **Open access** (services open to everyone, no login required): the user is immediately redirected to the services and no orders are created. The access 'requests' are tracked for accounting purposes.
- 2 **Orderable services with External ordering / catalogue type in the MP**: orders are created and stored in the EOSC Marketplace backoffice, but they are managed externally. The access 'requests' are tracked for accounting purposes.
- 3 **Orderable services with Internal MP ordering**: within this group we can distinguish different ordering models that need to be further defined. Currently, we are investigating these possibilities:
 - 3.1 *Lightweight integration*: e.g. communication to support the orders via dedicated email registered with Jira/Ops Portal;
 - 3.2 *Medium integration*: the order is recorded in the EOSC-hub order management system, but it is forwarded to the system of the related service provider without any further processing. Status of the order is kept up to date in the EOSC-hub system through an automatic communication with the service providers.
 - 3.3 *Tight integration*: automated as much as possible the access to the service after the 'order' button has been clicked.

In the coming months these three types of access model will be further defined and made available in the EOSC Marketplace.

4.3 Marketplace projects

Marketplace Projects is a lightweight approach to assist users in organising their services and related orders into logical blocks which share a common purpose and gaining support in the scope of the created project. Its goal is to provide a user-friendly and helpful UI where EOSC services of interest can be gathered and managed. The user is guided step by step through the service order management process, being at the same time freed from the complex operational side of this process. A first prototype implementation of this concept has been already created, discussed and it is available in the current version of the EOSC Marketplace.

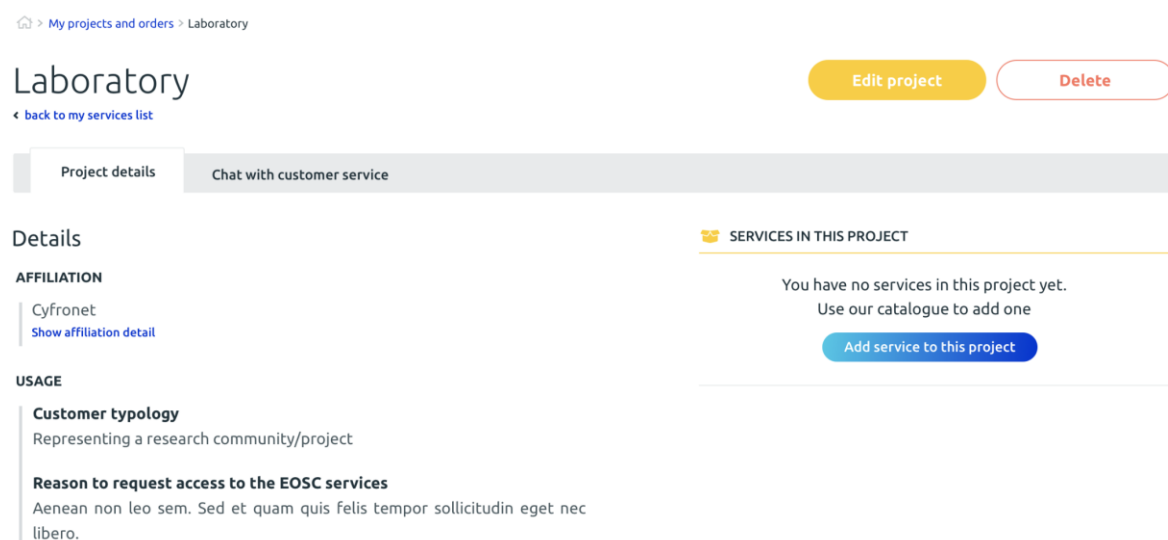


Fig. 5. Main page of a sample project in the EOSC Marketplace.

The project feature will be further enhanced in the second reporting period with developments in the scope of other user-facing features and integrational aspects:

- 1 Map a Project to a unique hierarchical order. Introduce the support for new use cases based on this feature. Improve the support for the operations behind order management.
- 2 Implementation of an early user support in the Marketplace Project: after a project is created, the user can start a chat with an EOSC Hub expert to ask for guidance.
- 3 Evolution of other features supporting the Customer Relationship Management process of the EOSC Service Management System.
- 4 Enhancements in the GUI.

5 EOSC-hub services roadmap

5.1 Individual plans for service evolution

In this section the evolution planned for each thematic area is presented.

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 take into account the technical requirements in terms of integration among EOSC-hub services. In this respect, the developers are improving and evolving their services to support interoperability leveraging standards or widely used APIs/protocols. This approach will guarantee that in the future combined usage of more services of the EOSC portfolio will be much easier.

5.1.1 TCOM area 1: Data Platforms for Processing

Onedata

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 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 OAI-PMH protocol.

The Onedata platform will be further extended to achieve better interactions with EOSC-hub services, and to provide more functionality to end users.

First, direct integration of Onedata and Jupyter Notebooks will be implemented, in order to enable Jupyter users' easy access to Onedata directly from the notebooks, as well as storing the notebooks themselves in Onedata data spaces. This functionality will be achieved by implementing a Python OnedataFS library, giving direct access to Onedata virtual filesystem from Python scripts, without the necessity to create a FUSE mountpoint on the host using the Oneclient. Furthermore, a dedicated Jupyter Contents Manager²³ plugin will be implemented, based on the OnedataFS library, allowing to run Jupyter notebooks on top of OnedataFS, thus enabling automatic synchronization of notebooks in Onedata spaces. Finally, based on this, integration with EGI JupyterHub²⁴ will be done, which will allow users to transparently create notebooks running on federated cloud

²³ <https://jupyter-notebook.readthedocs.io/en/stable/extending/contents.html>

²⁴ <https://notebooks.egi.eu>

resources, with data provisioned by Onedata virtual filesystem, using a single sign-on through EGI Check-in.

An integration with B2NOTE semantic annotation service is planned, which will extend existing Onedata metadata mechanism with support for semantic tags managed by B2NOTE. This will be achieved either by integrating the B2NOTE widget in Onedata data management interface, or by extending the Onedata metadata interface with requests to B2NOTE REST API directly.

Additionally, the WebDAV storage driver will be modified in order to improve performance through buffering of read and write requests to WebDAV storages with high network latency.

5.1.2 TCOM area 2: Data Publishing and Open Data

B2DROP

B2DROP provides a personal cloud storage in a trusted environment in which researchers can easily maintain, share and exchange research data with fellow researchers. In this respect, B2DROP plays a central role as a workspace within the research environment of a researcher where he or she can store and maintain valuable content.

From an architectural point of view, the main focus in the evolution of B2DROP is on the integration with other services and on positioning itself as the European personal cloud storage service for communities, researchers and “long tail of science” scientists who do not have direct access to a community, National and/or Regional comparable service.

For the integration of B2DROP with other services, work on the following use cases is being planned:

- CLARIN Language Resource Switchboard (LRS): a central service offered by the CLARIN ERIC to support language research via a rich set of tools. The integration of B2DROP with the Language Resource Switchboard will be twofold:
 - In LRS, B2DROP will be enabled as an input location which researchers can directly upload data from, directly into the Dropzone of the LRS.
 - In B2DROP, the LRS app will be enabled. When a researcher selects a file in B2DROP and selects the LRS app, the file will be automatically uploaded to the Dropzone of the LRS and the LRS is opened in the browser, where the researcher can start processing.
- WeNMR suite for Structural Biology: a suite of computational tools for structural biology. It consists of eight individual platforms (i.e. DISCVIS, POWERFIT, HADDOCK, GROMACS, AMPS-NMR, CS-ROSETTA, FANTEN and SPOTON). After running a computational job in one of the tools, due to limited storage resources, researchers can store the output for a limited time period (e.g. two weeks) within the WeNMR environment. The WeNMR community supports users from all over the world. To allow researchers to store the scientific output for periods longer than two weeks, the plan is to integrate B2DROP within the West-Life Virtual Folder, thus allowing WeNMR users to store their scientific output in a trusted environment.

The integration of B2DROP within the CLARIN LRS and WeNMR suite of computational tools must be seen as an example for other communities and service providers willing to adopt B2DROP within their services.

B2SHARE

B2SHARE is a user-friendly, reliable and trustworthy way for researchers, scientific communities and citizen scientists to store and publish small-scale research data from diverse contexts. It provides a solution that facilitates research data storage, guarantees long-term persistence of data and allows worldwide sharing of data, results and ideas.

In the current era, in which communities and researchers are required to make their research data and scientific artifacts available in a FAIR (i.e. Findable, Accessible, Interoperable and Reusable) way, the B2SHARE service plays a central role, as it is offered as a technology to build up FAIR compliant data repositories.

From an architectural point of view the focus of the evolution of B2SHARE is on optimising the support of the FAIR principles and the OpenAIRE guidelines for data providers, on optimising the deployment and maintenance procedures, on addressing community requirements, and on further integration with other services offered via EOSC-hub.

Concerning B2SHARE, the following use cases are planned:

- Make B2SAFE datasets discoverable: in order to make data discoverable via B2FIND, descriptive metadata of datasets stored in B2SAFE will be automatically stored in B2SHARE.
- B2SHARE community data repositories: many smaller and less mature communities and research groups struggle to comply with the FAIR principles to make their research data available. The B2SHARE technology can be deployed as a data repository services in which communities can easily define metadata templates to properly describe their research data, generate persistent identifiers and DOIs to reference datasets as a whole and the individual data objects separately and automatically discoverable via B2FIND. To lower the barrier to adopt B2SHARE as a community data repository the deployment and maintenance procedures should be optimised and targeted at inexperienced system administrators.
- Data repository for sensitive data: due to their nature, sensitive data cannot be made open and accessible. Therefore, for many researchers working with sensitive data it is challenging to comply with the FAIR principles. In collaboration with the sensitive data service task, a pilot will be conducted to enable B2SHARE as a data repository for sensitive data. Researchers will be able to properly describe sensitive datasets and make metadata discoverable via B2FIND, while the owners will maintain full control of their data in a secure environment (e.g. TSD <https://www.eosc-hub.eu/services/Services%20for%20sensitive%20data>).

The integration of B2SHARE with the TSD sensitive data service will be a solution exploitable by other communities working with sensitive data as well.

5.1.3 TCOM area 3: Data Preservation/Curation/Provenance

B2HANDLE

Research communities are struggling with data management tasks, which can ultimately interfere with the ability of researchers to work efficiently. A huge effort is spent in managing files and folders on file systems, and this can be an issue when the number, heterogeneity and complexity of files

increases. This problem can be addressed by introducing the concept of Digital Objects, which bring data and metadata together into single entities, which bear persistent identifiers.

B2HANDLE offers the necessary support for managing such identifiers and parts of the high-level metadata for Digital Objects. Currently, B2HANDLE offers identifier management in a supporting role to other services (e.g., B2SAFE, B2SHARE, B2FIND) and directly towards communities for their own usage scenarios. The number of services that fully employ the Digital Object concept is currently limited and a major evolutionary step will be to realize Digital Object support for most of the relevant EOSC services.

In addition, in order to provide community and EOSC stakeholders with a better service, future evolution may include:

- a comprehensive high-level web dashboard for community data managers and service operators to assess the status of identifiers across all the objects they manage. The dashboard should specifically be able to accommodate the huge numbers of identifiers and offer tools to manage community profiles as well.
- a detailed low-level web interface and command-line tools or services for data managers, system administrators or researchers to interact with individual identifiers or small groups of them. This should involve both end-user communities and services relying on B2HANDLE functionalities.
- an extension of the capabilities of the Central PID Catalog to support more detailed filtering, based on community metadata profiles, with additional interfaces to better support management tasks with large numbers of objects.

5.1.4 TCOM area 4: Metadata Management and Data Discovery

B2FIND

The service for data discovery is B2FIND and is intended to be the central search index and portal of EOSC-hub. With the new version B2FIND 2.4, the whole catalogue and user interface has been overhauled: the graphical user interface has been redesigned, search performance has been improved, harvesting and mapping functionalities have been enhanced, and the metadata schema has been extended.

B2FIND is aimed at research data distributed over memory areas within EOSC-hub and beyond. It is a basic service of the pan-european infrastructure EUDAT CDI and will become the central indexing tool for EOSC-hub. For this purpose, a comprehensive joint metadata catalogue has been built, which spans a widespread scope of scientific outcome - from Climate Research to Social Science, Particle Physics, and Economics. Therefore, different metadata formats, schemas and standards are homogenised on the B2FIND metadata schema, allowing users to search and find research data across scientific disciplines and research areas - thus enabling an interdisciplinary perspective.

B2FIND provides powerful search functionalities: metadata records can be found with a freetext-search, results may be narrowed down using different facets (e.g. geospatial or timeline search). Referencing data identifiers will either link to a landing page or to the data itself. B2FIND will evolve with respect to improvements of the mapping, validation and features of the user interface. To mention just a few planned developments in the roadmap:

- Community customization: provide additional, domain specific facets or adopted search templates
- Hierarchical search and categorisation: elaborate the taxonomical vocabulary for research ‘Disciplines’ and other facets.

5.1.5 TCOM area 5: HTC/HPC Compute

HTP & HPC

Regarding the support for High-Throughput and High-Performance Computing workloads, we identify several potential evolutions of the architecture that could ease the provisioning of suitable resources and the running of workloads with minimal resource management effort.

With respect to the first point, we identify areas that do not use the EGI scheduling mechanisms for HTC (such as DIRAC4EGI) due to specific processing software (e.g. ENES or DODAS) requirements. In those cases, it will be important to be able to easily deploy a customised virtual appliance taking into account four main aspects:

- Network topologies: the capability of conveniently specifying through TOSCA a virtual appliance with one or several private subnetworks plus one or several frontends, as well as the whole configuration process, will ease the deployment of complex processing tools such as OPHIDIA.
- HPC-specific resources: information systems already support the definition of specific resources such as MPI-enabled Infiniband networks or GPUs, but integration with the cloud compute part is not clear. This will enable deploying the above virtual appliances.
- Support of Function as a Service (FaaS) computing model: this model fits well the behaviour of many user communities. This should be supported as a native service, although a temporary solution could be to provide an easy deployment of FaaS custom services.
- Extend the on-demand cluster catalogue to Kubernetes: Applications on Demand provide the users with self-managed and automatically configured dynamic clusters. The support of Kubernetes clusters will be a great possibility for the communities that could express their workload as a set of Docker-container jobs.

With respect to HPC resources, the support of containerised application workloads is a major interest nowadays. However, HPC supports seldom containers and typically Singularity. The development of Docker containers in user space (uDocker) is of great interest. Automating the creation of containers for HPC jobs could provide customised execution environments with minimum overhead.

5.1.6 TCOM area 6: Cloud Compute (including containerisation and orchestration)

EGI Cloud Compute & Cloud Container Compute

The Cloud Compute services are expected to evolve their architecture to ease the integration of new providers, including those from commercial sector, which are foreseen to have an increased usage from use cases. The discovery of the capabilities and capacity of the providers of the federation should also improve to allow users and higher-level tools (such as PaaS solution and Workflow Management) can take informed decisions to make use of the best provider for every use

case. The service will adopt GlueSchema 2.1 and use novel transport mechanisms for the information based on messaging queues that will allow clients to consume information asynchronously and reduce latency on propagation using a subscription model when needed.

Management of images in the service will expand to include automatic security checks and the management dashboards will be improved to cover all the policy-related aspects to ensure the provisioning of secure services.

Container support will evolve to make Kubernetes the default container orchestration platform of the service, to ensure the portability of applications and to ease the porting of new use cases. An EOSC-hub container image registry (for both private and public images) should be provided for users to avoid relying on external parties.

Overall, all the components and tools should ensure proper integration with EOSC-hub AAI across all layers of the stack and the compatibility with the PaaS solutions developed within the project.

5.1.7 TCOM area 7: Software Release and SQA

The evolution of the services that fall under this technical area aims at improving the reliability and timely delivery of software products, i.e. the core services of the EGI repositories and the research community applications stored in the AppDB. This implies:

- Automating the SWPP process of any registered software in the EGI repositories, from artifact generation to the subsequent functional and integration testing.
- Expanding the software validation to the early stages of software development, by extending the continuous integration (CI) pipelines (Jenkins) at the level of the source code.
- Integrating source code repositories (GitHub, GitLab) with the software repositories to implement a CI/CD approach.

Software developers will then be able to store, maintain and test their code in a coherent and systematic way. Once ready for release, the new validated versions of their applications will be seamlessly available at the relevant repositories (EGI software, AppDB) in an automated way. As a result, the reliability of the EOSC services will be improved, allowing a faster delivery of bug fixes and new features, thus opening them to a new range of users.

5.1.8 TCOM area 8: Federation Tools

Details of the technical roadmap of the EOSC Hub Federation Tools are available in the deliverable D5.3 “1st Report on maintenance and integration of federation and collaboration services”²⁵.

5.1.9 TCOM area 9: PaaS Solutions

The user stories identified in the DB of Community Requirements describe the following requirements related to orchestration:

- ELIXIR: federation of cloud sites providing storage and CPUs/GPUs computing capacity for researchers

²⁵ <https://documents.egi.eu/document/3503>

- DODAS: automated deployment accessing a pool of heterogeneous computing resources, including private and public clouds
- ECAS/ENES: use of scalable cloud resources
- OPENCoastS: automated service deployment on cloud
- DARIAH: use of cloud infrastructures to support an increased number of users.

Therefore, in order to fulfil these requirements, the services of the PaaS Orchestration Thematic Area are expected to evolve with respect to the following aspects:

- both the Orchestrator and the Infrastructure Manager (IM) will provide enhanced support for deployments across federated cloud sites. The IM can currently provision virtual infrastructures across several cloud sites, but the wide variety of Cloud Management Frameworks and public cloud providers require that further testing under different conditions should be made.
- The different IP addressing schemes provided by different cloud sites (e.g. Virtual Private Clouds in Amazon Web Services) requires introducing novel approaches for automated contextualization on Cloud resources that feature no inbound connectivity.
- In order to better support GPU resources, appropriate support in the application description languages supported by the core services (TOSCA and RADL) will be introduced. This will allow users to specify their GPU requirements as part of the application architecture and allow the Orchestration system to choose the most appropriate cloud sites that satisfy these requirements.
- New TOSCA templates will be developed to support the community use-cases providing fully automated deployment of the required services.
- Towards a better compliance with modern security approaches we plan to further work on the codebase in cooperation with external tools such as SWAMP (Software Assurance Marketplace) in order to minimize the chances of security breaches and improve the quality of the code.

Moreover, to improve the integration in the EOSC-hub ecosystem, the orchestration tools will be evolved, with the following goals:

- integrating with the EOSC Hub solutions for AAI,
- leveraging the same Configuration Management information sources (e.g. integrate CMDB with AppDB-IS),
- integrating with a common EOSC Hub monitoring and accounting infrastructure.

5.1.10 TCOM area 10: Workflow management and user interfaces and Data analytics

The following EOSC-hub services will be included in the catalogue:

- **FutureGateway – science gateway service developed in the framework of the INDIGO DataCloud Project**
- **Thematic services:**

- Portals and science gateways:
 - **GEOSS portal**
 - **EISCAT_3D portal**, which provides services for data cataloguing, discovery and pre-defined analysis
- Analytics services:
 - **Earth Observation Data**
 - **MEA**, a geospatial data analysis tool empowered with OGC standard interfaces.
 - **Datacube**: Data Analytics Service proposes a multi-sensor, -scale and -purpose datacube approach. Geohazards Exploitation Platform is focused on the integration of Ground Segment capabilities and ICT technologies to maximise the exploitation of EO data.

5.1.11 TCOM area 11: Security

Details of the technical roadmap of the EOSC-hub Federation Tools are available in the deliverable D5.3 “1st Report on maintenance and integration of federation and collaboration services”²⁶.

5.1.12 TCOM area 12: AAI

In support of AAI solutions for the communities, EOSC-hub will work on implementing AARC recommendations and other relevant alignments that facilitate the use of eduGAIN and federated AAI for research and thematic services, both for technical interoperability (standardized protocols, identifiers and attribute harmonisation) as well as for policy and assurance (such as multi-factor authentication services and acceptable usage policy alignment). EOSC-hub will ensure that its AAI services will be interoperable among each other and towards all thematic services, therefore users will be able to seamlessly access the services without any obstacles. Emphasising the need for a consistent AAI experience by the communities, this area will work closely with the security area in the field of management of security incidents.

5.2 Planned integrations among EOSC-hub services

Thanks to a more detailed analysis of the use cases within the project, and given the provided studies about the status of each available service, the product teams have already defined an integration plan that will allow the end users to seamlessly access more services in a combined manner to exploit storage and computational resources.

All those activities are described in the deliverables D5.3²⁷ and D6.2²⁸. A short summary of the main integration tasks that will be worked out during the next few months is the following:

- AAI integration: end users are requiring the possibility to exploit EOSC services independently from the AAI services.

²⁶ <https://documents.egi.eu/document/3503>

²⁷ <https://documents.egi.eu/document/3503>

²⁸ <https://documents.egi.eu/document/3480>

-
- The Marketplace that will be built during the project lifetime will be tightly integrated with already available services from both EUDAT and EGI infrastructure (e.g. Service Portfolio Management Tool, Operational Tools, GOCDB) depending on the requirements from the user communities and from the service/resource providers.
 - The monitoring solution will be migrated to exploit a message bus solution in order to improve scalability, reliability and flexibility of supporting new monitoring information.
 - Both EGI and EUDAT infrastructures already have their own accounting solution, which have to be integrated in order to provide seamless access to the accounting information for all the EOSC-hub services and resources.
 - The same applies to ticketing systems as well: EGI was using xGUS, while EUDAT is exploiting EUDAT-RT. The integration of these two services is planned, in order to be able to have common APIs and interfaces to manage the operational tickets.
 - AppDB will be integrated with the ticket management systems so that the end users could easily interact with the site via ticket with the same web application used to deploy services/applications.
 - AppDB will be enhanced in order to better integrate all the information about the resources that are usually included into the information system. In this way the AppDB could better support the distribution of services and applications on top of the available resources.
 - AppDB will be further improved to support the ability to provide PIDs to identify appliances and software that could be deployed over the EOSC-hub resources.
 - EGI DataHub developers and the team of B2STAGE/B2SAFE are actively working together in order to enable the end users with a seamless data-exchange between EGI/INDIGO service and EUDAT ones.
 - B2SAFE will provide the possibility to automatically extract metadata and to ingest them into B2FIND accordingly to data policy that the end user could impose.
 - Data exchange from B2SAFE to EGI DataHub and vice-versa according to data policy driven by the end users.
 - Expose, harvest and index metadata from EGI DataHub in B2FIND so that also the data storage in EGI services could be searchable via the B2FIND service.
 - Use B2DROP to prepare input data for B2STAGE or/and retrieve and store (small sized) data.
 - Access B2DROP Storage from EGI Cloud Compute VMs as transparent as possible from the end user point of view.
 - The service INDIGO CMDB and SLAM used by the INDIGO PaaS orchestrator to exploit cloud resources will be integrated with AppDB in order to enable the PaaS orchestrator to use the same resources AppDB is able to exploit.

5.3 Long term evolution of the EOSC-hub services

During the whole project lifetime, WP10 has a key role in defining the roadmap and the evolution of the service architectures, the roadmap of the services themselves and the evolution of the Service Catalogue.

This is done basically through two main drivers:

-
- the activity of the task 10.3 that is studying and analysing the requirements coming from the user communities;
 - the activity of the TCOM that provides the steering of the technical evolution of the EOSC Hub Service Catalogue.

For what concerns the first point, a description of the process to gather and analyse specific technological requirements has been provided in Section 2. The results from the requirement evaluation have been already summarized to the developer teams in order to drive the evolution of their services.

Concerning the second item, the TCOM associated to task 10.1 is the technical board where all the technical areas are represented.

The TCOM guides and works with the AMB with responsibility for:

- Defining, maintaining and evolving the technical roadmap;
- Stimulating and supervising the contribution to open source community projects and standards bodies;
- Ensuring compliance to standards and security in defining and assessing the acceptance criteria for the evolution of the service catalogue;
- Coordinating with external initiatives and projects of technical and strategic relevance, when applicable, for example as a result of a collaboration agreement with another project.

The user communities are expected to contribute to the definition of the technical roadmap not only with requirements, but also with actual services joining the EOSC-hub: this is the case of the Thematic Services, that after the initial phase of implementation published their services in the EOSC Services Catalogue and Marketplace.

5.4 Thematic Services roadmap

In this section all the Thematic Services are described with a brief introduction to their scientific goals and a deeper definition of their roadmap. Each technical roadmap highlights the integration activities with EOSC-hub Federation and Common services.

5.4.1 CLARIN

This service proposes the component metadata framework (CMDI) to stimulate the discoverability of data sets, making the citation of these more convenient, and to take away the barriers to automated processing of data. The following infrastructure integration activities will be accomplished:

- The development of a uniform and robust workflow to (1) gather metadata descriptions from the various scientific communities, (2) convert these files into a suitable CMDI-based equivalent, (3) perform a highly parallelised indexing of all the resulting metadata files.
- Extending the reach of the Language Resource Switchboard to data types outside the scope of the humanities and to use language processing to make the research workflow more efficient

5.4.1.1 Roadmap

Currently, the Virtual Language Observatory (VLO) is the first out of three services that has been released under the EOSC-hub project. Its roadmap is focussed on improving the integration with other EOSC-hub products. For the other two services, the main focus is on the first EOSC-hub release to work on improvements and integrations. The roadmap for the CLARIN thematic services is available in the CLARIN document archive, a second iteration is planned for M12/M13.

Virtual Language Observatory

For M24 a milestone has been defined for user community specific deployments. The current VLO implementation is focused on metadata coming from the CLARIN domain. The underlying CMDI schema is sufficiently flexible to support other domains as well, however the VLO must be updated to support this kind of flexibility. Integration with the EOSC-hub accounting and reporting services has already been finished under the VA (Virtual Access) framework. More advanced integration with the Virtual Collection Registry is also planned to be implemented for the M24 milestone. The aim is to make it easier for the end user to add resources found in search results to a new or existing virtual collection, and to feature this integration more prominently.

Virtual Collection Registry

The main goal for the virtual collection registry (VCR) is the first EOSC-hub release, scheduled for M13. For this release the plan consists in improving the integration with both the VLO and LRS. Another requirement for this release is to collect feedback from users. The same approach as the VLO will be followed using Mopinion. The VCR is the only service among these three that requires authentication. Currently, authentication is implemented using SAML via the CLARIN Service Provider Federation. The best approach for integration with the EOSC-hub must be evaluated. This is not a technical but an organisational issue, since the best way to comply with the security policies adopted by the CLARIN's community must be found. After the release of this service, the focus will shift towards the integration with B2SHARE.

Language Resource Switchboard

The main focus for the Switchboard is the first EOSC-hub release, scheduled for M14. This release should already include the integration with B2DROP (roadmap LRS.3) as demonstrated at the EOSC Portal Launch event.

5.4.2 DODAS

The Compact Muon Solenoid (CMS) is one of the two general-purpose experiments at the Large Hadron Collider (LHC) at CERN in Geneva. CMS relies on the distributed computing capacities of the Worldwide LHC Computing Grid in order to process and analyse the collision data. Solutions having the potential to provide additional computing capacity to the LHC experiments (and hence to CMS) are of extreme interest to address the future data and computing challenges of the next generation High Luminosity LHC.

DODAS (Dynamic On Demand Analysis Service) provides the end user with an automated system that simplifies the process of provisioning, creating, managing and accessing a pool of heterogeneous (possibly opportunistic) computing resources. At the state of the art, DODAS

provides multi-virtual platforms support for cloud systems such as EGI Federated Cloud, OpenStack, OpenNebula, Amazon AWS and Microsoft Azure. Offering the support to a wider range of infrastructure providers requires the integration of new plugins into the Infrastructure Manager or for the PaaS Orchestrator.

Scope of this service is offering the support to a wider range of infrastructure providers improving the following features through the integration with EOSC-hub: Data Management, Data Caching, PaaS level cross-site cluster deployment, Web User interface, Authentication and Authorisation, and Accounting.

5.4.2.1 Roadmap

The integration and training activities of DODAS will continue during the next year. DODAS is already operational since M4. However, following an agile approach, new features are made available when commissioned, and gathering feedback allows to better evolve the service continuously. Integration with other EOSC-hub services will continue and will surely focus on extending the AAI integration by:

- Integration of the IAM service as solution for a dynamic user mapping management of the HTCondor batch on demand.
- Federation, through IAM, with EOSC-hub AuthN/Z services.

Integration of the Accounting toolkit: this will require training the DODAS team first, and then a design phase to be sure that all the identified cases are covered. Another priority is to further improve and integrate the MLaaS (Machine Learning as a Service) features of DODAS. New use cases based on this flavour of the Thematic Service are approaching and, as a matter of fact, this implies that integration of new features will be necessary. In this respect, EOSC-hub might be of help, providing already available solutions. The dynamic extension of cluster through clouds is a rather important aspect not mentioned in the original plan. In fact, it is actually needed, and it will be integrated in the next year. This integration relates to TOSCA, PaaS Orchestrator and Mesos configuration.

A huge amount of effort will be dedicated to support the communities using DODAS.

- Continuous support is required for user and communities as mentioned above. The amount of effort required for this task cannot be underestimated. Support is meant not to be at technical level only, but also in terms of training, design and strategic decisions needed while porting static models to a geo distributed and federated architecture.
- There are ongoing requests from user to integrate commercial providers. This will require further integration in EOSC-hub IM and PaaS orchestrator.

5.4.3 ECAS

The ENES Climate Analytics Service (ECAS) will enable scientific end users to perform data analysis experiments on large volumes of climate data, by exploiting a PID-enabled, server-side, and parallel approach. The task aims at providing a paradigm shift for the ENES community with a strong focus on data intensive analysis, provenance management, and server-side approaches as opposed to the

current ones mostly client-based, sequential and with limited/missing end-to-end analytics workflow/provenance capabilities.

5.4.3.1 Roadmap

The integration and training activities of ECAS will continue during the second year of the project. The goal is to have a pre-operational service ready and opened for wide usage by M18 and complete all integration activities by M22 latest.

Some ECAS components are already integrated with one of the EOSC-HUB AAI providers. Effort is being put on providing a common AAI solution that can possibly support token propagations between ECAS components (e.g. Ophidia and JupyterHub).

ECAS will integrate B2SHARE most likely via the existing integration between B2DROP and B2SHARE. This will require to define the necessary metadata and a process to acquire it in the ECAS-Lab environments.

Regarding IM/Orchestrator an Ansible role to support automated deployment of ECAS will be provided with the latest Ophidia release, by leveraging previous efforts. Such activity will be tested and validated in the EGI-FedCloud. In this respect, future follow ups will relate to integration of automated deployment procedures into the Marketplace framework.

Over the next weeks, technical integration activities regarding OneData will proceed further. In particular an OneProvider instance will be set up at CMCC and integrated at file-system level with the local ECASLab instance.

In PY2, ECAS will integrate B2HANDLE, putting persistent identifiers (PIDs) on output results and connecting them with input data, recording data lineage in the most basic way. This will require support for PID assignment and profiles by B2SHARE and OneData. Interaction with the services concerning these features has already been started.

In terms of future training, ECAS proposal for the EGU2019 is accepted as a short course session. Presentations and hands-on will be prepared to cover ECAS components and how to use them.

Inter-thematic-service collaboration will be supported over the next months to build 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 may be planned/organised, pending evaluation of the use cases for possible common approaches.

5.4.4 GEOSS

The GEOSS (Global Earth Observation System of Systems) services support the implementation of the Sustainable Development Goals (SDGs) defined by the United Nations. The goal is to help SDG monitoring and assessing by providing the necessary Indicators and Essential Variables (EVs) defined by the Community.

The core of the service, GEO DAB (Discovery and Access Broker), will be able to access the virtual IaaS and PaaS provided by EGI via open APIs for the following objectives:

- EGI Federated Cloud: to access computing and storage resources currently not available to end users, who currently need to procure and manage their own compute infrastructure for data exploitation;
- EGI DataHub for advanced data management services and “Cloud Workload optimisation”.
- EGI Core Infrastructure platform services (e.g. monitoring, AAI, etc.).

5.4.4.1 Roadmap

The plan for next year is to enhance present integration. This will include exploring solutions to fill the identified gaps in terms of required Cloud services; defining a clear use case for AAI integration; and implementing the AAI integration.

5.4.5 OPENCoastS

The OPENCoastS (On-demand Operational Coastal Circulation Forecast Service) service builds on-demand circulation forecast systems for user-selected sections of the North Atlantic coast and keeps them running operationally for the time frame defined by the user. This daily service generates forecasts of water levels, 2D velocities and wave parameters over the spatial region of interest for periods of 72 hours, based on numerical simulations of all relevant physical processes. Currently, the service is deployed at a single computing site (NCG-INGRID-PT). Expanding the service to international access requires additional computing capacity for scaling and resiliency (service continuity). Furthermore, it requires access to distributed data. Consequently, integrations with more sophisticated distributed data management services are necessary.

5.4.5.1 Roadmap

During 2019, the following actions will be performed:

- OPENCoastS@IFCA: the OPENCoastS services will be created in the IFCA/CSIC cloud, first as a duplicate instance of the whole service, and later interconnected with the INCD service, to achieve high availability, to have replica/synchronization of files and databases.
- Integration with European forcing services for waves to permit wave-current interaction in the European deployments.
- Integration with other EOSC-hub core services for data preservation.
- Integration with EOSC-hub computing (Grid) services, through the DIRAC4EGI.

5.4.6 WeNMR

WeNMR (Worldwide e-Infrastructure for Nuclear Magnetic Resonance and structural biology) envisions a model in which the following services are offered as thematic services: DISVIS, POWERFIT, HADDOCK, GROMACS, AMPS-NMR (AMBER), CS-ROSETTA, UNIO, FANTEN. In this context, the planned integration activities encompass:

- Consolidation of the submission machineries of the various portals building on DIRAC4EGI and INDIGO (e.g. PaaS Orchestrator, Infrastructure manager and FutureGateway solutions)
- Integration of distributed data storage solutions for users (Onedata)
- Implementation of AAI solutions developed by EGI and INDIGO-DataCloud
- Provisioning of new portals when suitable/community-requested software is identified.

5.4.6.1 Roadmap

One goal consists in connecting some of the available portals to data repositories, such as the ones offered by EUDAT, in order to allow users to directly upload and/or download data/results. However, the data generated by the WeNMR services are very specific to a user/application and not globally reusable by third parties. This is very different for example from sky images collected by telescopes. Nevertheless, it will be important to facilitate data deposition directly into public repositories where relevant.

5.4.7 EO PILLAR

The scope of this task is the Earth Observation (EO) domain and it encompasses the following services: MEA, EPOSAR, Sentinel Playground, Datacube Data Analytics Service, Geohazards Exploitation Platform (GEP), OSS-X Sentinel Service, EO Cloud, EODC SIDP. They are named EO Pillar and they will be integrated with the EOSC-hub infrastructure at different levels and categorised into three main classes: data access and computing services, data exploitation services, general user services. A synergic activity with the task T7.4 GEOSS will allow to exploit one of the services, Datacube, also to serve the GEOSS community.

In particular, task activities will be carried out to integrate:

- Data access and computing services: aiming to augment e-Infrastructures Compute and Storage services with EO data and computing resources with direct co-located access to EO data, which is a key requirement for users dealing directly with EO data and the other EO Pillar services.
- EO data exploitation services: aiming to augment e-Infrastructure thematic services portfolio with services tailored for EO scientists, to support their work and foster production of value-added EO products;
- EO general user services: aiming to augment e-Infrastructure thematic service portfolio with services coming from EO data and tailored for non-EO experts and the general public, to foster exploitation of EO satellite data.

5.4.7.1 Roadmap

The Geohazards Thematic Exploitation Platform (GEP)

In order to enable the EOSC-HUB AAI for the Geohazards TEP, there are plans to develop/integrate an identity broker to allow users to authenticate against one or more EOSC-HUB AAI providers (EGI Check-in, Indigo IAM and B2ACCESS). It is still under discussion which ones will be tackled and how they will be prioritised.

The EPOSAR Service

During the next few months the full integration of the EPOSAR service within EOSC-hub is envisaged. This means to create the final configuration of all the cloud resources needed for the production phase, to consolidate the access to EO data catalogue and test its robustness, and to set up the EPOSAR service for the automatic and updated processing of the Sentinel-1 dataset acquired over

the areas of interest. The production phase will address some areas relevant for the EPOS community and the achieved products will be accessible through the EPOS central hub and the EPOS Thematic Core Service Satellite Data access point.

EODC JupyterHub for global Copernicus data

This service is currently not fully operational. One major goal is to make the service operational and provide access to it via the EOSC marketplace. The current proof of concept should be transferred to a more elastic compute environment (PaaS). Accordingly, a second goal is to migrate the entire service to the EODC PaaS currently running OpenShift to be able to scale the service up/down as requested by the users. Additional effort will be given on improving the usability of JupyterHub with respect to earth observation use cases by providing additional needed software libraries and viewing possibilities.

EODC Data Catalogue Service

Because of the enormous number of objects stored in the metadata database, several optimisations are foreseen to reduce the DB query response time. Those optimisations will focus on a load balancing DB infrastructure and on optimised DB schemas.

Rasdaman EO Datacube

The service will become part of the emerging European Datacube Federation.

CloudFerro Data Collections Catalog

Service is going to be added to the EOSC-hub Marketplace content (see: <https://discovery.creodias.eu/dataset>).

CloudFerro Infrastructure

The Service Catalogue, Marketplace and B2FIND integration are planned. Further expansions of the available products and data collections are planned in line with data acquisition roadmap.

CloudFerro Data Related Services - EO Finder

Service is going to be added to the EOSC-hub Marketplace content (<https://finder.creodias.eu/www>).

CloudFerro Data Related Services - EO Browser

Service is going to be added to the EOSC-hub Marketplace (<https://browser.creodias.eu>).

Sentinel Hub

- Inclusion of additional data sources.
- Datacube like capabilities.

5.4.8 DARIAH

The DARIAH (Digital Research Infrastructure for the Arts and Humanities) Thematic Service (TS) aims to enhance and improve the usage of the cloud-based services and technologies in the domain of the digital arts and humanities research. It will enable end users coming from the digital arts and humanities domains to seamlessly store, describe (metadata) and share their datasets, discover, browse and reuse datasets shared by others and to perform analysis on various data volumes.

The DARIAH TS is providing the following services:

- DARIAH Science Gateway,
- Invenio-based repository in the cloud,
- DARIAH repository (based on CDSTAR).

The **DARIAH Science Gateway** is a web-oriented portal, developed during the EGI-Engage project (DARIAH Competence Centre) and is specially tailored for the researchers coming from digital arts and humanities disciplines. It currently offers several cloud-based services and applications: Semantic and Parallel Semantic Search Engines, DBO@Cloud, Workflow Development and supports several file transfers protocols.

The **Invenio-based repository** is a service that allows researchers and scholars to easily create, deploy and configure their own Invenio-based repository and host it on the cloud infrastructure (Federated Cloud). The service is aimed at smaller research groups lacking adequate technical support and budget to acquire their own infrastructure to host data repositories.

DARIAH repository is a new service based on the Common Data Storage ARchitecture (CDSTAR), a system for storing and searching objects in research projects. The repository is operated by DARIAH-DE as a digital long-term archive for human and cultural-scientific research data. The DARIAH repository is a central component of the DARIAH-DE Research Data Federation Infrastructure, which aggregates various services and applications and can be used comfortably by DARIAH users.

5.4.8.1 Roadmap

Science Gateway

During the second year of the project, the plan foresees the work on integrated login with the Marketplace; in other words, if the user navigates to the Gateway from or requests the service via the EOSC Marketplace, no authentication should be required. This integration step requires a more detailed technical consideration. The users of the DARIAH SG will benefit from this integration since they will not need to authenticate multiple times.

Invenio-based repository in the Cloud

There are two ideas concerning integration with the EOSC Marketplace:

1. *Basic integration.* The Invenio-based repository in the cloud will be published in the EOSC Marketplace such that the users landing there can directly navigate the Invenio-based repository service. This integration includes the mutual authentication, which will allow

users to log in only once. The deadline for this action is June 2019. This integration will produce an increased visibility of and simpler access to the service for end users.

2. *Advanced integration.* The Invenio-based repository will be integrated with the Marketplace allowing users to parameterize and configure a new repository instance (e.g. storage size, number of CPUs) from the Marketplace, without a need to redirect to the service configuration (remote FutureGateway site). This integration requires a more detailed technical consideration. At the end of this integration process, the Invenio-based repository in the cloud would be offered via Marketplace as an Application-as-a-Service with possible parameterization before launching. From the end user point of view, this integration will present a more transparent and simple process of deploying a new repository instance in the cloud, without an unnecessary navigation between Marketplace and the service's landing page. From the service provider point of view, this solution will decrease the maintenance effort, since there is no need to operate a separate, remote site (FutureGateway) that only provides a template for parameterizing a new repository instance.

Repository

With the growing number of services integrated it is essential for the thematic services to continuously evaluate integration potential. Regarding the DARIAH TS there is specifically potential regarding complementary services from the CLARIN TS.

6 Collaboration with external projects

This section summarises the EOSC-hub technical collaborations with other projects working on the EOSC establishment.

6.1 Collaboration with OpenAIRE Advance

EOSC-hub and OpenAIRE Advance signed a collaboration agreement that includes several technical activities with the aims to make interoperable and integrate services provided by the two projects and work together on key topics to offer effective tools to the EOSC users.

To achieve interoperability across EOSC services for the Open Science, the two projects are developing practices to expose, access and define citation metadata and link metadata for literature, datasets, and software in a FAIR way. The aim of this activity is to define and promote common guidelines for scientific product content providers in order to facilitate the publishing and exchange of scientific products over different tools. Furthermore, this activity is crucial to enable definition of new citation/quality indexes for science and come up with Open Science-flavoured research impact. Concretely, it has been agreed to endorse and foster the adoption of OpenAIRE guidelines, investigate the possibility of introducing new guidelines for specific classes of “other products” (e.g. services, protocols, workflows, virtual appliances) and define a framework for identifying and describing scientific communities. The last two activities are conducted in the context of existing RDA groups. EOSC-hub services are being enhanced to include the support of these guidelines to enable the metadata harvesting and, then, the exchange of metadata from scientific products to the OpenAIRE Research Community Dashboard. In this way researchers while performing their scientific process using EOSC-Hub services can automatically publish and report their scientific products. Work to define guidelines for enabling the exchange of usage statistics is also progressing in collaboration with the RDA WG Make Data Count. This activity will allow researchers to access scientific products together with their usage statistics, scientific products different from literature to benefit from new impact measures and scientific communities and funding agencies to have quality metrics on scientific impact on research done and funded. These metrics can be used within the decision-making process for the long-term preservation of scientific products.

Furthermore, a Data Management Plan tool (DMP) is being jointly developed by EOSC-hub and OpenAIRE-Advance to offer to Horizon 2020 projects an easy way to define DMPs that comply with the H2020 guidelines for DMPs, and to find and request services offered via the Hub. Collaboration is on-going also on data annotation, with enabling annotation functionality in OpenAIRE services (e.g. OpenAIRE Research Community Dashboard, Zenodo) leveraging the B2NOTE EOSC-hub service, and in the anonymisation of sensitive data, with the enhancing and promotion of the OpenAIRE Amnesia service.

Finally, the OpenAIRE AAI infrastructure and services will be integrated with the EOSC-hub federated AAI infrastructure. Purpose of this is to allow customers to have seamless access across OpenAIRE and EOSC-hub services.

6.2 Collaboration with GN4-2/GN4-3

Technical collaboration with the GEANT lead projects GN4-2 and GN4-3 is focusing on three main areas:

- **AAI interoperability:** the EOSC-hub Authentication and Authorisation Infrastructure (AAI) is expected to provide a consistent system with which communities can integrate. Following the approach of the AARC Blueprint Architecture (BPA), the components of which it is composed, may vary depending on the community, its structure, and its use of e-Infrastructure services. GN4-2/GN4-3 is collaborating with EOSC-hub on the design, implementation and evolution of the EOSC-hub AAI Service Components and the available AAI service offerings leveraging pertinent combinations of these components for the respective EOSC-hub communities.
- **Cloud Interoperability and Service Availability and Integration:** the two projects built an expert group for the clouds operators and users based on the existing EOSC-hub and GN4 teams that provides a support network for building, operating and using these clouds with a focus on the needs and characteristics of the research clouds. A continuous exchange of technical solutions (PaaS layer, data management in a distributed environment, network resources management in cloud, etc.) has also been setup. A pilot activity about creating an high-level federation with a PaaS layer for orchestration of computing and storage in a way that hides the complexity of provisioning resources from multiple heterogeneous services from the EOSC-hub and GEANT catalogues is ongoing.
- **Technical coordination:** this activity is ensuring technical alignment and coordination between the two projects through the participation of experts from EOSC-hub and GN4-2/GN4-3 to technical groups as requested by the execution of the Joint Activity Plan of the Collaboration Agreement. This includes the set-up of the AAI joint expert group, a cloud joint expert group and ad-hoc task forces when needed. As a result, the two projects are harmonising technical roadmaps and investigating options of interoperability across both projects.

6.3 Collaboration with other EOSC projects

6.3.1 Development projects (EINFRA-21 call)

6.3.1.1 *DEEP-HybridDataCloud*

DEEP Hybrid DataCloud is an EU H2020 funded project which aims at supporting intensive computing techniques that require specialized HPC hardware to explore very large datasets. The primary outcome of the project is the **DEEP as a Service (DEEPaaS)** solution through which a set of building blocks are provided in order to ease the deployment of applications requiring cutting-edge techniques, such as deep learning using neural networks, parallel processing of very large data, and analysis of massive online data streams.

The first release of the DEEPaaS has been announced in January 2019, codenamed DEEP Genesis²⁹. In parallel with the active development that will result in the second and final release, DEEP managers had started the preliminary steps to integrate the DEEPaaS solution in the EOSC portal, under the supervision of EOSC-hub. Thus, research communities within the EOSC context will be able to leverage DEEPaaS to develop, build and deploy complex models as a service in a scalable way.

6.3.1.2 *eXtreme-DataCloud*

The eXtreme DataCloud (XDC) is an EU H2020 funded project which aims at developing scalable technologies for federating storage resources and managing data in highly distributed computing environments. The provided services will be capable of operating at the unprecedented scale required by the most demanding, data intensive, research experiments in Europe and Worldwide.

XDC will be based on existing tools, whose technical maturity is proved and that the project will enrich with new functionalities and plugins already available as prototypes (TRL6+) that will be brought to production level (TRL8+) at end of XDC.

The targeted platforms are the current and next generation e-Infrastructures deployed in Europe, such as the European Open Science Cloud (EOSC), the European Grid Infrastructure (EGI), the Worldwide LHC Computing Grid (WLCG) and the computing infrastructures funded by other public and academic initiatives.

The main high-level topics addressed by the project include:

- federation of storage resources with standard protocols
- smart caching solutions among remote locations
- policy driven data management based on Quality of Service
- data lifecycle management
- metadata handling and manipulation
- data preprocessing and encryption during ingestion
- optimized data management based on access patterns.

All the developments will be community-driven and tested against real life use cases provided by the consortium partners representing research communities belonging to a variety of scientific domains: Life Science, Astrophysics, High Energy Physics, Photon Science and Clinical Research.

The XDC project aims at opening new possibilities to scientific research communities in Europe and worldwide by supporting the evolution of e-Infrastructure services for Exascale data resources.

The XDC software will be released as Open Source platforms available for general exploitation.

6.3.2 National EOSC projects

6.3.2.1 *EOSC-synergy (CSIC)*

EOSC-synergy extends the EOSC coordination to nine participating countries by harmonizing policies and federating relevant national research e-Infrastructures, scientific data and thematic services,

²⁹ <https://deep-hybrid-datacloud.eu/2019/01/18/deep-genesis-first-software-release-is-out/>

bridging the gap between national initiatives and EOSC. The project introduces new capabilities by opening national thematic services to European access, thus expanding the EOSC offer in the Environment, Climate Change, Earth Observation and Life Sciences. This will be supported by an expansion of the capacity through the federation of compute, storage and data resources aligned with the EOSC and FAIR policies and practices.

EOSC-synergy builds on the expertise of leading research organizations, infrastructure providers, NRENs and user communities from Spain, Portugal, Germany, Poland, Czech Republic, Slovakia, Netherlands, United Kingdom and France, all already committed to the EOSC vision and already involved in related activities at national and international level.

Furthermore, we will expand EOSC's global reach by integrating infrastructure and data providers beyond Europe, fostering international collaboration and open new resources to European researchers. The project will push the EOSC state-of-the-art in software and services life-cycle through a quality-driven approach to services integration that will promote the convergence and alignment towards EOSC standards and best practices. This will be complemented by the expansion of the EOSC training and education capabilities through the introduction of an online platform aimed at boosting the development of EOSC skills and competences. EOSC-synergy complements ongoing activities in EOSC-hub and other related projects liaising national bodies and infrastructures with other upcoming governance, data and national coordination projects.

6.3.2.2 EOSC-Pillar (INFN)

EOSC-Pillar gathers representatives of the fast-growing national initiatives for coordinating data infrastructures and services in Italy, France, Germany, Austria and Belgium.

The project aims to propose the initiatives for the national coordination of data infrastructures and service recently started in many Member States as one of the founding pillars for the development and long-term sustainability of the EOSC.

EOSC-Pillar starts with an initial group of neighbouring countries who are active in open science, to define and set a model to harmonise and interfederate the initiatives.

EOSC-Pillar will aim at:

- Harmonise related policies in Europe and facilitate alignment with international initiatives
- Aggregate and achieve wide adoption of processes and practices for optimal use of resources
- Contribute to FAIR data uptake in Europe
- Progressive removal of technical and organisational barriers to ensure findability, accessibility, interoperability and re-use of research data
- Piloting and establishing future (co)funding strategies and business/usage models to ensure long-term sustainability
- Delivery of European added value of EOSC
- Widespread dissemination of European achievements

6.3.3 ESFRI Clusters

6.3.3.1 ESCAPE

ESCAPE (European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures) brings together the astronomy, astroparticle and particle physics communities putting together a cluster with ESFRI projects with aligned challenges of data-driven research, with demonstrated capabilities in addressing various stages of data workflow and concerned with fundamental research through complementary approaches.

ESPACE aims at producing versatile solutions, with great potential for discovery, to support the improvement of EOSC thanks to open data management, cross-border and multi-disciplinary open environment, according to FAIR (Findable, Accessible, Interoperable and Reusable) principles. The ESCAPE foundations lay on the capacity building of the ASTERICS project work towards enabling interoperability between the facilities, minimising fragmentation, encouraging cross-fertilisation and developing joint multiwavelength/multi-messenger capabilities in astronomy, astrophysics and astroparticle physics communities.

ESCAPE main goals:

- Improving access to data and tools to unlock innovation for the society at large
- Providing data with FAIR principles to increase efficiency of researchers thanks to scientific data interoperability
- Establishing new methodological approaches and rules for quality certified data and science tools sharing
- Building an European cross-border and multi-disciplinary open innovation environment for research data, knowledge and services, while connecting EOSC and ESFRI
- Facilitating interdisciplinary and networked research between different sciences, through research infrastructure ecosystem and by supporting data publishing, analytics, computational capacity, virtual analysis environments and workflow systems
- Creating economies of scale through the adoption of common approaches for data management
- Educating and training the scientific and wider user communities, to ensure the uptake of the results of the project

6.3.3.2 EOSC-Life (MU/CESNET)

EOSC-Life brings together the 13 Biological and Medical ESFRI research infrastructures (BMS RIs) to create an open collaborative space for digital biology.

It is a joint response to the challenge of analysing and reusing the huge amounts of data produced by life-science. Managing and integrating this data is beyond the capabilities of most individual end users and institutes. By publishing data and tools in an Europe-wide cloud EOSC-Life aims to bring the capabilities of big science projects to the wider research community. Federated user access (AAI) will allow transnational resource access and authorisation. EOSC-Life establishes a novel access model for the BMS RI: through EOSC scientists would gain direct access to FAIR data and tools in a cloud environment available throughout the European Research Area.

EOSC-Life will make BMS RIs data resources FAIR and publish them in the EOSC following guidelines and standards (e.g. EDMI). Overall, this will drive the evolution of the RI repository infrastructure for EOSC and integration of the BMS RI repositories. EOSC-Life will implement workflows that cross disciplines and RI boundaries and address the needs of interdisciplinary science. Through open hackathons and bring-your-own-data events EOSC-Life will be created from user communities, providing a blueprint for how the EOSC supports wide-spread and excellent data-driven life science research. EOSC-Life will address the data policies needed for human research data under GDPR. Interoperable provenance information describes history of sample and data to ensure reproducibility and adherence to regulatory requirements.

6.3.3.3 *PaNOSC*

European Photon and Neutron sources and their instruments are essential tools for scientists of a plethora of fields (biology, materials, chemistry, technology, nuclear physics, high-energy physics, cultural heritage ...) and their industrial counterparts. The Photon and Neutron community provides PBs of curated Open Data from a large variety of experiments, only a small portion of which is currently openly accessible. Anyway, the initial embargo period for the first institute implementing an open data policy has already ended and more and more data is going to be open. The Open Data are curated in data catalogues following community standards for metadata. The data catalogues will be exposed to general search machines and EOSC linked catalogues to facilitate their access to a larger community.

The goal consists not only in providing such open data, but also analysis and simulation services that allow the transition from measurements to insight and new science. A coherent set of services is expected, not only for existing user community (mainly experimentalists), but also for a wider public and a new level of coherency amongst the whole scientific ecosystem. To reach this full potential and level of impact, the project needs to work with the other EOSC builders and in particular with the EOSC-hub partners.

Since the PaNOSC kick-off, the partners already had fruitful exchanges about the best technical and organisational solutions, for instance to transfer data to EOSC-hub services (e.g.: notebook and data transfer). The expectation is to be able to deploy user-friendly EOSC-hub notebook services for the Photon and Neutron user community by the end of 2019. Scientists should be able to get transparent access to their data when using these services, perform analysis or simulation and archive their results seamlessly.

6.3.3.4 *SSHOC*

Social Sciences and Humanities Open Cloud is the ESFRI cluster project for the SSH. It combines the Social Sciences ERICs CESSDA, SHARE and ESS with the Humanities ERICs CLARIN and DARIAH and the Cultural Heritage Network E-RIHS in a project aimed at exploiting their synergies and preparing to move the SSH services into the cloud. The SSH research infrastructures are very diverse with respect to IT and research infrastructure usage and also have different attitudes collaborating with eInfra projects. SSHOC is discussing the options in becoming a conduit for SSH services (beyond SSHOC) for integration in EOSC.

Amongst the SSHOC partners, CLARIN has been involved in the eInfrastructure collaboration: in EUDAT and in EOSC-hub the CLARIN metadata infrastructure (VLO, LRS) are thematic services. DARIAH is currently involved in EOSC-hub with thematic services based on the EGI DARIAH Gateway.

The SSHOC activities most relevant for the EOSC-hub are the offering of NLP services for the broad Humanities and Social Sciences planned in WP3 “Lifting SSH Services and Technology into the Cloud”, developing a SSHOC Marketplace for the SSH from WP7 “Developing a SSHOC Marketplace”, developing a SSHOC interoperability-hub for (meta-) data conversions, and linking data and services via the SSHOC Switchboard generalising the CLARIN technology that is already a EOSC thematic service.

Currently, SSHOC is conducting an analysis of the to-be-developed and/or adopted services and technology in order to decide if and how these could be integrated in EOSC. Confirmed candidates are:

- A series of NLP solutions (services) to address specific challenges in SSH research e.g. machine translation for specific domains, automatic occupational title extraction etc. Depending on their success and application prospects such services should be offered on a permanent basis, using compute and storage resources from either the stakeholder infrastructures or EOSC.
- SSHOC Interoperability Hub. A set of services and tools that offers conversions between the most used (meta-)data formats in the SSH.
- A more general version of the CLARIN Language Resource Switchboard that can connect SSH catalogues and data storage facilities with a broad set of SSH services. Currently the CLARIN Switchboard is one of the EOSC thematic services.
- SSHOC Marketplace (<https://sshopencloud.eu/marketplace>). This is a one-stop-shop where SSH researchers should find a broad set of resources and solutions for the digital aspects of their research. SSHOC is aware of the EOSC Marketplace and the need to be interoperable. Relevant SSHOC Marketplace materials should be also accessible via EOSC e.g. B2FIND.
- A number of important SSH vocabularies and ontologies should be made broadly available through a vocabulary platform.
- Curated questionnaires will be gathered in central repositories, although this can be considered a community solution, the questionnaires can be made discoverable & accessible from EOSC.
- A repository for training materials is needed to manage those also beyond the end of the project.

The need for sustainable data archiving and access provisioning is particularly evident in those cases when there is no obvious candidate from the SSHOC stakeholder infrastructures.

SSHOC is currently discussing also the role it can play in the registration and integration of other SSH services that are not part of the SSHOC project itself.

With respect to the SSHOC and EOSC-hub liaison, SSHOC’s coordinator Ron Dekker is also chair of the EOSChub strategic board and SSHOC has an EOSC alignment task-force responsible for the SSHOC service inventory and analysis which is in contact with EOSChub.

7 References

No	Description/Link
R1	EOSC-hub Grant Agreement
R2	EOSC-hub Collaboration Agreement
R3	D1.1 Quality and Risk Management Plan
R4	D1.2 Data Management Plan
R5	D3.1 Communications and Stakeholder Engagement Plan
R6	The Plan for the Exploitation and Dissemination of Results in Horizon 2020 https://www.iprhelpdesk.eu/sites/default/files/newsdocuments/FS-Plan-for-the-exploitation-and-dissemination-of-results_1.pdf
R7	Confluence Innovation Management related pages https://confluence.egi.eu/display/EOSC/Project+Results
R8	https://confluence.egi.eu/display/EOSC/Dissemination+Activities
R9	The European Innovation Management Standard CEN/TS 16555 https://standards.cen.eu/dyn/www/f?p=204:110:0::::FSP_PROJECT,FSP_ORG_ID:35932,671850&cs=13A816A57184977C465944D2F2E2C5645
R10	Catalogue of Project Results https://wiki.eosc-hub.eu/display/EOSC/Catalogue+of+Project+Results
R11	Catalogue of Aggregate Project Results https://wiki.eosc-hub.eu/display/EOSC/Catalogue+of+Aggregate+Project+Results
R12	EOSC-hub Deliverables Page https://confluence.egi.eu/display/EOSC/Deliverables
R13	EOSC-hub Milestones Page https://confluence.egi.eu/display/EOSC/Milestones