



EOOSC-hub

D6.3 Second release of common services

Lead Partner:	MPG
Version:	1
Status:	Under EC review
Dissemination Level:	Public
Document Link:	https://documents.egi.eu/document/3558

Deliverable Abstract

This deliverable provides an overview of the release of second maintenance package for common services required by the thematic and specialized services. These common services are a key component of the EOOSC-hub KER “Internal Services Provided in the Hub Portfolio” and their continued maintenance and evolution is a core activity in the EOOSC-hub project. The maintenance package addresses common services across all WP6 tasks. It comprises advanced software interoperability adaptations regarding federation and collaboration services as well as adaptations required by thematic and specialized services. Finally, it provides future plans for the evolution of the services.



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:	John Alan Kennedy Germán Moltó Miguel Caballer Enol Fernández	MPG UPV UPV EGI	
Moderated by:	Malgorzata Krakowian	EGI.eu/WP1	
Reviewed by:	Pavel Weber Matti Heikkurinen	KIT /WP5 EGI/WP2	
Approved by:	AMB		

DOCUMENT LOG

<i>Issue</i>	<i>Date</i>	<i>Comment</i>	<i>Author</i>
v0.1	10/06/2019	Draft table of contents	Germán Moltó (UPV), Miguel Caballer (UPV), Enol Fernández (EGI)
v0.2	26/06/2019	Revised table of contents	Germán Moltó (UPV)
v0.3	19/07/2019	Contributions from Task Leaders	Heinrich Widmann (DKRZ), Enol Fernández (EGI), Germán Moltó (UPV), Miguel Caballer (UPV), Michele Carpena (CINECA), Olivier Rouchon (CINES), Abdulrahman Azab (SIGMA2 - UiO)
v0.4	30/09/2019	Contributions from Service Owners	Heinrich Widmann (DKRZ), Enol Fernández (EGI), Germán Moltó (UPV), Miguel Caballer (UPV), Michele Carpena

			(CINECA), Olivier Rouchon (CINES), Abdulrahman Azab (SIGMA2 - UiO)
v0.5	10/10/2019	Harmonised text, added summary	Germán Moltó (UPV)
v0.6	18/10/2019	Final draft for review - all services described	John Alan Kennedy (MPCDF)
v0.7	08/11/2019	Feedback from external Review	Pavel Weber (KIT), Matti Heikkurinen (EGI)
Final	25/11/2019	Final document - incorporating comments from External Review.	John Alan Kennedy (MPCDF), Heinrich Widmann (DKRZ), Enol Fernández (EGI), Germán Moltó (UPV), Miguel Caballer (UPV), Michele Carpena (CINECA), Olivier Rouchon (CINES), Abdulrahman Azab (SIGMA2 - UiO)
Final	09/01/2020	Addressed issues raised by PO regarding info about licenses etc for eTDR and Sensitive Data services.	John Alan Kennedy (MPCDF)

TERMINOLOGY

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

<i>Terminology/Acronym</i>	<i>Definition</i>
API	Application Programming Interface
CKAN	Comprehensive Knowledge Archive Network
CMD	Cloud Middleware Distribution
DOI	Digital Object Identifier
IaaS	Infrastructure as a Service
IdP	Identity Provider
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting

PaaS	Platform as a Service
CaaS	Computing as a Service
PAM	Pluggable Authentication Module
PID	Persistent Identifier
RCD	Research Community Dashboard
REST	REpresentational State Transfer
SAML	Security Assertion Markup Language
VM	Virtual Machine
VO	Virtual Organization
WebDAV	Web Distributed Authoring and Versioning
WMS	Workload Management System
YAML	Yet Another Markup Language

Contents

1	Introduction	8
2	Discovery and Access	9
2.1	EGI DataHub	9
2.2	B2FIND	12
2.3	B2STAGE	14
2.4	B2DROP	17
3	Federated Compute	19
3.1	EGI Cloud Compute	20
3.2	EGI Cloud Container	23
3.3	EGI Workload Management	25
3.4	EGI Online Storage	27
3.5	Advanced IaaS	28
3.6	EGI High-Throughput Compute	31
4	Processing and orchestration	33
4.1	TOSCA for Heat	34
4.2	Infrastructure Manager	36
4.3	PaaS Orchestration system	38
4.4	Future Gateway	42
5	Data and Metadata Management	44
5.1	B2HANDLE	44
5.2	B2SAFE	47
5.3	B2SHARE	49
5.4	B2NOTE	51
6	Preservation	53
6.1	eTDR	53
7	Sensitive Data	55
7.1	TSD	55
7.2	ePouta	57
8	Summary and Outlook	61

Executive summary

This document contains a report on the second maintenance release of common services from EOSC-hub service catalogue, supplementing the information reported in deliverable D6.1. The common services in EOSC-hub provide middleware layer enabling thematic services to efficiently use the resources provided by leading European infrastructure providers. These common services are a key component of the EOSC-hub KER “Internal Services Provided in the Hub Portfolio” and their continued maintenance and evolution is a core activity in the EOSC-hub project. The main effort in the second year of the project focused on service maintenance and further integration with other EOSC-hub common services as well as functional improvements to address the requirements coming from the user communities.

This deliverable covers the services in 6 categories, reflecting the different tasks of WP6

- **Discovery and access:** The services in this category enable consistent interfaces to services in the area of the discovery and the access of digital resources. It will create a common access and discovery layer for depositing, exchanging, retrieving and staging data and metadata.
 - **Services Provided:** EGI Datahub, B2FIND, B2STAGE, B2DROP
- **Federated Compute:** Focuses on the corrective, adaptive and perfective maintenance of the EGI Cloud Compute, Cloud Container and High-Throughput Compute services. It aims to satisfy the emerging needs from the user communities and to support the most recent technology evolution
 - **Services Provided:** EGI Compute Cloud, EGI Cloud Container, EGI Workload Management, EGI Online Storage, Advanced IaaS, EGI High Throughput Compute
- **Processing and Orchestration:** Focuses on the maintenance and integration of orchestration services with the Cloud Compute and Cloud Container services. This allows to build complex virtual infrastructures based on standards.
 - **Services Provided:** TOSCA for HEAT, Infrastructure Manager, PaaS Orchestrator, Future Gateway
- **Data and Metadata Management:** Integration and maintenance of the EOSC-hub common repository services and policy-driven data management/stewardship services with particular regard to registered data.
 - **Services Provided:** B2HANDLE, B2SAFE, B2SHARE, B2NOTE,
- **Preservation:** The integration of certified Trusted Digital Repository (TDR) in the catalogue, resulting in a sustainable long-term data preservation service: the European Trusted Digital Repository (ETDR).
 - **Services Provided:** eTDR
- **Sensitive Data:** Focusing on integration activities for services with data whose access is restricted either by national or European regulations or by other confidentiality policies (e.g. business confidentiality).
 - **Services Provided:** TSD, ePouta

This deliverable presents an update on the status of each of the services detailed above including release notes and plans. The services for data preservation (eTDR) and sensitive data (TSD and ePouta) are new to the EOSC-hub catalogue and are detailed here for the first time.

1 Introduction

This document presents the portfolio of EOSC-hub common services, focusing on reporting the status of the services and their evolution since the last reporting period (reported in deliverable D6.1). It includes appropriate references to the documentation and source code repositories of common services, as well as brief reports on their current release major features and future plans, which are of interest to EOSC-hub user communities, for the second release.

The services covered by this deliverable represent the key functional components of major e-Infrastructures from Europe in the areas of data discovery, access and management, federated computing, processing and orchestration, data and metadata management, preservation and sensitive data. The coordination of integration and maintenance activities of these services is the main goal of WP6 in the EOSC-hub project.

The main effort in the second year of the project was focused on maintenance, further integration with other EOSC-hub common services and functionality improvement according to the requirements coming from the user communities. Thus, this deliverable provides second release and integration notes about the services along with description of developments and improvements planned for the remaining part of the project.

This document establishes a baseline for the full range of common services that support the integration of the different services infrastructures as well as the planned thematic services. The services are/will be represented in the EOSC-hub service catalogue¹, which is also used to maintain information about updates and enhancements of the service components.

The document is organized as follows. Section 2 contains information about services related to data discovery and access. Section 3 covers services in the federated computing category. Section 4 contains services from the area of processing and orchestration. Section 5 lists the services in the data and data management category. Section 6 focuses on services aimed at data preservation, while section 7 covers the sensitive data services. Finally, section 8 provides a brief summary and outlook of the work carried out.

For the sake of completeness, each service in this document is presented in a brief form including references to online documentation, reference deployments, licenses, major features, responsible teams and current release major updates. This subsumes the information reported in D6.1, supplemented with the developments carried out in the last year. The information on future plans for each service in the scope of EOSC-hub project is included.

¹ <https://eosc-hub.eu/catalogue>

2 Discovery and Access

The objective of this activity is to create a common discovery and access level and interface through which end-users can locate, localize and use data resources within EOSC-hub for their own purposes.

The metadata service B2FIND plays a crucial role in data discovery as a central search indexing tool of EOSC-hub. The service was extended and enhanced to cover data from all of the storage services and data archives within EOSC-hub:

- The already existing bridge between the publishing service B2SHARE and B2FIND is expanded by adoption of the unified and common metadata schema and stabilizing metadata ingestion of B2SHARE records
- An OAI-PMH provider on top of EGI-DataHub has been set up to enable B2FIND to harvest metadata from there
- Progress was made with the proof-of-concept plan of metadata extraction from B2SAFE (iRODS) data collections and indexing these datasets in B2FIND
- In parallel there was a continued uptake of internal and external research communities and data providers into B2FIND.

In addition to the above activities, a solution to enable seamless data access between storage services, in this case EGI datahub and B2SAFE, was developed and demonstrated. The implementation and a demo test bed are described in detail in the milestone M6.6².

2.1 EGI DataHub

2.1.1 Service description

Service/Tool name	EGI DataHub
Service/Tool url	https://datahub.egi.eu
Service/Tool information page	https://onedata.org
Description	Onedata is a globally distributed storage solution, integrating storage services from various providers using possibly heterogeneous underlying technologies, such as NFS or other POSIX-compliant filesystems as well as Ceph, S3, GlusterFS, WebDAV and OpenStack SWIFT and provides to clients' interfaces based on CDMI, REST API and virtually mounted POSIX filesystem. Onedata architecture consists of two major components: Oneprovider and Oneclient. The former is installed within data

²

<https://confluence.egi.eu/display/EOSC/M6.6+Demo+of+managing+data+transfers+across+EGI%2C+EUDAT%2C+INDIGO+data+and+compute+services>

	<p>center and provides a unified interface to multiple filesystems used in the center. Servers can scale to thousands of instances in order to improve performance.</p> <p>The main functional components include:</p> <ul style="list-style-type: none"> ● Onezone - the federation and authentication service, each Onezone instance (e.g. EGI DataHub) provides a single-sign on to a network of connected storage providers ● Oneprovider - is the main data management component of Onedata, which is deployed in the data centers and is responsible for provisioning the data and managing transfers ● Oneclient - provides access to the virtual filesystem on a VM or host directly via a Fuse mountpoint
Value proposition	<ul style="list-style-type: none"> ● Discovery of data via a central portal. This will include a search mechanism plus a rating system which may be based on, e.g. number of accesses ● Access to data conforming to required policies which may be: 1) unauthenticated open access; 2) access after user registration or 3) access restricted to members of a Virtual Organization (VO). This access may be via a GUI (e.g. a webpage) or an API (e.g. programmatic access to the data) ● Replication of data from data providers for resiliency and availability purposes. Replication may take place either on-demand or automatically. Replication will require the introduction of a file catalogue to enable tracking of logical and physical copies of data. ● Access to data from the AppDB to enable VOs to associate appropriate data with matching Virtual Appliances ● Authentication and Authorization Infrastructure (AAI) integration between the EGI DataHub and with other EGI components and with user communities existing infrastructure ● File catalogue to track replication of data: logical and physical file
User of the service/tool	Users or communities with data intensive applications or interested in open data publishing/sharing
User Documentation	https://onedata.org/#/home/documentation/doc/user_guide.html

Technical Documentation	https://onedata.org/#/home/documentation/doc/admin_guide.html
Product team	ACC CYFRONET-AGH
License	MIT/Apache 2.0
Source code	https://github.com/onedata
Testing	Onedata platform, which is the technological solution used for EGI-DataHub undergoes continuous integration and functional tests, constituting of over 500 test cases.

2.1.2 Release notes

In the reporting period, EGI DataHub has been upgraded to a new release of Onedata 18.02.3, which includes the following new features:

- Integration with B2HANDLE service enabling direct generation of PID identifiers for datasets which are being published through EGI DataHub
- Integration with B2FIND service which allows discovery of data published via EGI DataHub through the B2FIND interface, the integration has been performed using the OAI-PMH implementation of the EGI DataHub
- Integration of EGI DataHub authentication with B2ACCESS allowing direct login to EGI DataHub using B2ACCESS credentials and automatic token refresh necessary for long-term data transfers
- Development of WebDAV driver for Onedata enabling direct access from EGI-DataHub services to the B2SAFE WebDAV endpoints
- Integration of the B2SAFE WebDAV endpoint with EGI DataHub allowing registration of existing B2SAFE instances as third-party storages for EGI DataHub
- Development of OnedataFS library enabling direct access to the Onedata distributed filesystem from Python without creating a system mountpoint using FUSE: https://onedata.org/#/home/documentation/doc/using_onedata/onedatafs.html
- Provisioning of Oneclient and OnedataFS using Anaconda:
 - <https://anaconda.org/onedata/onedatafs>
 - <https://anaconda.org/onedata/oneclient>

Detailed changelog for each component of Onedata can be found at:

- Onezone - <https://github.com/onedata/oz-worker/blob/release/18.02.3/CHANGELOG.md>
- Oneprovider - <https://github.com/onedata/op-worker/blob/release/18.02.3/CHANGELOG.md>
- Oneclient - <https://github.com/onedata/oneclient/blob/release/18.02.3/CHANGELOG.md>

2.1.3 Future plans

For the next year, the following developments and integration activities are planned:

- Upgrade of the EGI DataHub service to the new version of Onedata (19.02.1)
- Improvements in performance in the WebDAV driver, mostly related to internal buffering of requests
- Integration with B2STAGE/B2SAFE services allowing accessing the data stored in EGI DataHub by these services (i.e. the opposite direction then has been implemented in this reporting period)

2.2 B2FIND

2.2.1 Service description

Service/Tool name	B2FIND
Service/Tool url	http://b2find.eudat.eu
Service/Tool information page	https://eudat.eu/catalogue/B2FIND
Description	<p>B2FIND provides a discovery portal which allows users to find data collections within an international and inter-disciplinary scope. It is based on a comprehensive metadata catalogue of research data collections stored in EUDAT data centres as well as metadata that are steadily harvested from community specific repositories. Harmonization of the metadata descriptions collected from heterogeneous sources enables not only the presentation in a consistent form, but as well the faceted search across scientific domain boundaries.</p> <p>The backbone of B2FIND is its ingestion process that consists of three steps. First the metadata records - provided by various research communities - are harvested. Afterwards the raw metadata records are converted and mapped to unified key-value dictionaries as specified by the B2FIND schema. The main challenge here is the divergence in metadata standards and schemes; these differences correspond to the specific communities' needs. To assure and improve metadata quality this mapping process is accompanied by a) iterative and intense exchange with community representatives, b) usage of controlled vocabularies and community specific ontologies and c) formal and semantic mapping and validation.</p>

	Finally, the mapped and checked records are uploaded as datasets to the catalogue which is based on CKAN, an open source data portal software that provides a rich RESTful JSON API and uses Apache SOLR for indexing.
Value proposition	<ul style="list-style-type: none"> • Allows users to search and browse datasets via keyword searches • Supports faceted, geospatial and temporal metadata searches • Results displayed in user-friendly format and listed in order of relevance • Access to the scientific data objects is given through references provided in the metadata • Available for communities in the EUDAT registered domain of data
User of the service/tool	B2FIND is openly accessible to everyone
User Documentation	https://eudat.eu/services/userdoc/b2find http://b2find.eudat.eu/guidelines/index.html http://b2find.eudat.eu/guidelines/providing.html http://b2find.eudat.eu/guidelines/harvesting.html http://b2find.eudat.eu/guidelines/mapping.html
Technical Documentation	https://github.com/EUDAT-B2FIND
Product team	DKRZ
License	GNU Affero General Public License version 3 (AGPLv3)
Source code	https://github.com/EUDAT-B2FIND/ckanext-b2find https://github.com/EUDAT-B2FIND/md-ingestion
Testing	B2FIND offers several test machines that are used for testing a) metadata ingestion of new integrated scientific communities, b) improvements of the ingestion workflow and c) improvements of CKAN. A training test instance is offered, as well as a training module: https://github.com/EUDAT-Training/B2FIND-Training

2.2.2 Release notes

B2FIND Metadata Schema was enhanced according to the community requirements, including now metadata elements <Contributor> and <ResourceType>. Mandatory elements were expanded to <Publisher> and <Discipline> (besides <Title> and one <Identifier>, whereas an internal ranking describes a preferred consistency of persistent identifiers: DOI, PID or any other URL or URN). Publication of full B2FIND Metadata Schema v1.0 (including XSD schema implementation) is planned before 2020.

Already integrated communities were reingested due to improved mapfiles. Several new scientific Communities integrated first on a testing machine, then on a production instance. Therefore, new VMs have been deployed at DKRZ (for testing metadata ingestion, technical development and search functionalities).

The upgrades of the software stack include the following:

- Upgrade to CKAN 2.4
- Upgrade to SolR 6.4
- B2FIND v. 2.4

Detailed change history can be tracked at:

<https://github.com/EUDAT-B2FIND/ckanext-b2find/commits/master>

<https://github.com/EUDAT-B2FIND/md-ingestion/commits/master>

2.2.3 Future plans

Emphasis is put on the development of a generic Classification of Research Areas (CLARA) which will be presented for the first time at RDA 14th plenary in Helsinki October 2019. Future plans include the broader dissemination of CLARA in order to establish it as a useful standard as well as the continuously uptake of new scientific communities within B2FIND.

It is planned to integrate (harvest) datasets of High Energy Physics Research via research infrastructure projects PaNOSC (Photon and Neutron Open Science Cloud) and ExPANDS (European Open Science Cloud Photon and Neutron Data Services) in mid-2021 as well as further the uptake of projects in Astronomy and Astrophysics via ESCAPE (European Science Cluster of Astronomy & Particle Physics ESFRI research infrastructures) respectively. Similar integration activities will follow for the IVOA (International Virtual Observatory Alliance), within the next three years.

2.3 B2STAGE

2.3.1 Service description

Service/Tool name	B2STAGE
--------------------------	---------

Service/Tool url	https://www.eudat.eu/b2stage
Service/Tool information page	https://github.com/EUDAT-B2STAGE/B2STAGE-GridFTP https://github.com/EUDAT-B2STAGE/http-api
Description	<p>The B2STAGE service allows data transfer into and out of EUDAT data nodes and supports the assignment of unique Persistent IDentifiers (PID) to staged data. EUDAT exposes two protocols for staging data: GridFTP and HTTP-API.</p> <p>The B2STAGE GridFTP service (via the EUDAT Data Storage Interface) is aimed at supporting large data transfer and many files. The key functionality of the B2STAGE GridFTP service is transferring relevant data sets between HPC centers and EUDAT in order to store them, process them and, possibly, move the results back. Data could also be already stored in one or more EUDAT data centers, as a result of the Safe Replication activity. Eventually, output data are identified through a Persistent Identifier (PID).</p> <p>The B2STAGE HTTP-API service is a set of RESTful endpoints that allow access to both B2SAFE data and metadata and it is aimed at small to medium datasets. This service offers programmatic access to data and thus allows for smooth integration of such data into other applications and data services.</p>
Value proposition	<ul style="list-style-type: none"> • Transfer large data collections from EUDAT storage facilities to external HPC facilities for processing • Ingest computation results onto the EUDAT infrastructure • Access stored data sets through associated PIDs • In conjunction with B2SAFE, replicate community data sets, ingesting them onto EUDAT storage resources for long-term preservation.
User of the service/tool	B2STAGE GridFTP is mainly aimed for power users who need to access data in B2SAFE and move them to compute sites. B2STAGE HTTP-API is mainly conceived for community developers who want to integrate data and features from B2SAFE into their community-specific applications.
User Documentation	https://github.com/EUDAT-B2STAGE/B2STAGE-GridFTP https://github.com/EUDAT-B2STAGE/http-api#user-guide

Technical Documentation	https://github.com/EUDAT-B2STAGE/B2STAGE-GridFTP https://github.com/EUDAT-B2STAGE/http-api#get-started https://github.com/EUDAT-B2STAGE/http-api#administration
Product team	CINECA
License	HTTP API: MIT License
Source code	https://github.com/EUDAT-B2STAGE/B2STAGE-GridFTP https://github.com/EUDAT-B2STAGE/http-api
Testing	<p>B2STAGE http-api can be tested by accessing the online prototype available at https://b2stage-test.cineca.it/api/status.</p> <p>Prototype dedicated instructions: https://github.com/EUDAT-B2STAGE/http-api/blob/master/docs/prototype.md</p>

2.3.2 Release notes

During the second reporting period we worked to extend the integration of B2STAGE HTTP-APIs with other EOSC-hub services, towards B2SHARE. To achieve the goal, we introduced the support of URL encoded access tokens as an alternative to canonical Authorization headers. This way, once authenticated to B2STAGE, users can access protected resources by using normal GET requests, e.g. from a browser. This new option can also be adopted if users are authenticated by means of the B2ACCESS OAuth2 workflow. Furthermore, to enhance the integration from web interfaces, we implemented the support for HEAD methods that can be used to verify resources accessibility without the overhead of a GET operation.

To ease the integration of the B2STAGE HTTP-APIs into community projects (in particular SeaDataCloud³ and ORFEUS EPOS⁴) we implemented the ability to extend the B2STAGE base project by creating customized modules. By adopting this approach, communities can inherit the development framework used by B2STAGE and only focus on custom endpoints and functionalities.

We also worked on extending the deployment options by testing a full installation on a native (non-container) OS to complement the default deployment method adopted in B2STAGE, which is based on Docker.

Detailed change history can be tracked at:

<https://github.com/EUDAT-B2STAGE/http-api/releases>

³ <https://www.seadatanet.org/About-us/SeaDataCloud>

⁴ <https://www.orfeus-eu.org/>

2.3.3 Future plans

The enhancement of B2STAGE will continue with more focus on the integration with other services, in particular with B2SHARE, and with additional deployment methods. A first deployment not based on Docker has been successfully completed as proof-of-concept, but more tests are needed to achieve the required level of reliability. We are also working on the implementation of an optional landing page to which redirect users, to avoid automatic downloads in case of large resources.

2.4 B2DROP

2.4.1 Service description

Service/Tool name	B2DROP
Service/Tool url	https://b2drop.eudat.eu
Service/Tool information page	https://eudat.eu/services/b2drop
Description	<p>B2DROP is an easy-to-use, user-friendly and trustworthy storage environment which allows users to synchronise their active data across various desktops and to easily share this data with collaborators. The service is targeted at all European researchers whose institute does not host such storage. B2DROP is based on Nextcloud. The service is intended for the long-tail and still volatile data which can change and are still subject to active research, e.g. drafts of research papers. Beside the functionalities of sync and share services, B2DROP has advanced functionalities with several extensions.</p> <p>The B2SHARE bridge extension allows users to publish the final documents on B2SHARE. The users select files, the metadata schema and the publication name within B2DROP and the publication draft, containing the supplied information is created in the background. The user finishes the draft on B2SHARE and receives a persistent identifier for the publication.</p>
Value proposition	<ul style="list-style-type: none"> ● Sharing a file with local users ● Share publicly across B2DROP instances ● Accessing shared artefacts ● Desktop synchronization ● Mount B2DROP using the WebDAV client ● Publishing files to B2SHARE

User of the service/tool	Users who need to share/access small data sets from their desktops
User Documentation	https://eudat.eu/services/userdoc/b2drop
Technical Documentation	https://eudat.eu/services/userdoc/b2drop
Product team	FZ Juelich
License	GNU Affero General Public License v3.0
Source code	https://github.com/EUDAT-B2DROP
Testing	Internal continuous integration test suite.

2.4.2 Release notes

During the report period, we made two major updates of the software stack. Through these updates a hardening of the security was made, and the service performance was enhanced by several database optimizations. The integration to B2SHARE was enhanced by highlighting communities, which have restricted access on B2SHARE. The old local login flow was completely removed at the beginning of the reporting period and now only the federated AAI mechanism is supported for authentication.

2.4.3 Future plans

We are going to adjust the graphical user interface and the integration with B2SHARE service (B2SHAREbridge) to every release of the underlying software. These are, on the one hand, security changes and, on the other hand, changes for a better usability of the service. We are continuously updating the functionality and the usability of the B2SHAREbridge. With these changes we want to offer users an easier workflow from collaborative working on documents to the publication of the results. As one part of these improvements, we are going to investigate the possibility to support multiple B2SHARE servers at the same time.

3 Federated Compute

Federate Compute focuses on the EGI services providing resources for the execution of user applications on the EGI e-infrastructure as virtual machines (Cloud Compute), as containers (Cloud Container Compute), or as jobs (High-Throughput Compute). These services are summarised in Table 1. Users who need more flexibility and control on the resources, including efficient allocation, should use Cloud Compute, users with existing containerised applications following a cloud-native approach are better served with Cloud Container Compute, and for those users with the need to run parallel computing tasks at scale that can be modelled as jobs in a batch system, High Throughput Compute will better meet their needs.

Table 1. Federated Compute services

	Cloud Compute	Cloud Container Compute	High Throughput Compute
What is it?	Multi-cloud IaaS	Kubernetes on top of EGI Cloud Compute	The grid, a scalable batch system
What you run?	VMs	(Docker) Containers	Jobs
Typical workloads	Lift and shift existing applications Specific OS (kernel) requirements	Cloud-native containerised applications.	Execution of parallel computing tasks to analyse large datasets.
Pros / Cons	[+] Complete control on resources, run (almost) anything you'd like [-] Complex operation	[+] Industry standard [+] Hides complexity of Kubernetes setup [-] Some Kubernetes features not available	[+] No management of resources, just submit jobs [-] Legacy interfaces [-] Porting of applications

The Workload Manager (formerly known as DIRAC4EGI) service⁵ allows users to manage and distribute computing tasks in an efficient way while maximising the usage of computational resources of Cloud Compute and High-Throughput Compute. It was added to the EGI Catalogue as beta during October 2018.

CREAM is the main component of High-Throughput Compute included for maintenance and evolution within EOSC-hub WP6 Task 6.2. This task also includes three components that further extend the functionality of High-Throughput Compute:

- Advanced IaaS offers a way to execute containerised applications on traditional batch systems with the udocker tool
- CVMFS is an Application software distribution service that provides users an easy way to deploy software on the distributed computing infrastructure powering the High-Throughput Compute

⁵ Service description at EGI's website: <https://www.egi.eu/services/workload-manager/>

- StoRM, a Storage Resource Manager compliant with the SRM v2.2 standard, is one of the technology options of the EGI Online Storage service

3.1 EGI Cloud Compute

3.1.1 Service description

Service/Tool name	Cloud Compute
Service/Tool url	https://www.egi.eu/services/cloud-compute
Service/Tool information page	https://www.egi.eu/services/cloud-compute
Description	<p>Cloud Compute gives you the ability to deploy and scale virtual machines on-demand. It offers guaranteed computational resources in a secure and isolated environment with standard API access, without the overhead of managing physical servers.</p> <p>Cloud Compute offers the possibility to select pre-configured virtual appliances (e.g. CPU, memory, disk, operating system or software) from a catalogue replicated across all EGI cloud providers.</p> <p>With Cloud Compute you can:</p> <ul style="list-style-type: none"> • Execute compute- and data-intensive workloads (both batch and interactive) • Host long-running services (e.g. web servers, databases or applications servers) • Create disposable testing and development environments on virtual machines and scale your infrastructure needs • Select virtual machine configurations (CPU, memory, disk) and application environments to fit your requirements • Manage your Cloud Compute resources in a flexible way with integrated monitoring and accounting capabilities
Value proposition	Manage IaaS resources in a multi-cloud environment in a flexible way with single sign on and integrated image catalogue.
User of the service/tool	Research Communities, individual researchers, VREs, SME/Industry, Resource providers

User Documentation	https://wiki.egi.eu/wiki/Federated_Cloud_user_support
Technical Documentation	https://wiki.egi.eu/wiki/Federated_Cloud_Technology
Product team	Partners involved: CESNET, CSIC, EGI.eu, GRNET, IASA, SRCE
License	Apache 2.0
Source code	https://github.com/EGI-Foundation/fedcloud-integration-documentation https://github.com/EGI-Foundation/VMI-endorsement https://github.com/EGI-Foundation/cloud-info-provider https://github.com/IFCA/caso https://github.com/IFCA/keystone-voms https://github.com/the-cloudkeeper-project/cloudkeeper https://github.com/the-cloudkeeper-project/cloudkeeper-one https://github.com/the-cloudkeeper-project/cloudkeeper-os https://github.com/the-rocci-project/keystorm https://github.com/ARGOeu/nagios-plugins-fedcloud https://github.com/IASA-GR/appdb-core https://github.com/IASA-GR/appdb-is-publisher https://github.com/the-oneacct-export-project/oneacct-export
Testing	<p>All products in CMD (Cloud Middleware Distribution) are verified against EGI's current Quality Criteria. Reports are available in EGI document DB under "Software Provisioning and Technology" category. Installation and configuration notes are available in the EGI repository.</p> <p>Images are tested automatically by the AppDB/SECANT integration.</p>

3.1.2 Release notes

CMD-OS (Components of Cloud Compute for OpenStack):

The CMD-OS components released during the last reporting period include:

- cloudkeeper 2.0 / cloudkeeper-one 2.0. A complete redesign of the cloudkeeper tool that enhances the image handling and allows the operation of the software scoped to VOs. This simplifies the migration to a centralised operations model of the EGI Cloud service. Details available at <https://github.com/the-cloudkeeper-project/>.
- cloud-info-provider 0.11.[1-5] This series of releases introduced GlueSchema 2.1 support, support for messaging as transport layer and better VO-scoped operation of the software. Several bug fixes were also included. Details available at <https://github.com/EGI-Foundation/cloud-info-provider/releases>.
- cASO 1.3.x releases fix the long-running VM accounting issues found in production. Details available at <https://github.com/IFCA/caso/releases>.

AppDB

The main changes in the AppDB during this period are listed below:

- VMOps:
 - Integration with native (OpenStack) APIs
 - Integration with OpenID Connect
 - Support for user quotas
- Cloud Marketplace:
 - Integration with OpenStack sites
 - Integration with GlueSchema 2.1 for information discovery
 - Integration with Argo Messaging System for information discovery
 - Improved SECANT integration
 - Usability improvements

EGI images

All the images provided by EGI in AppDB were regularly updated by using latest versions from upstream distribution. A new Ansible-based creation mechanism was introduced to improve the reliability of the building process. Updated images are:

- [CentOS 6](#), [CentOS 7](#)
- [EGI FedCloud clients](#)
- [EGI Small Ubuntu 16.04 for monitoring](#), [Ubuntu 14.04](#), [Ubuntu 16.04](#)

The appliance for integration of OpenStack providers was also updated to cover the latest version of integration tools available in the infrastructure and to fix several issues found by sites in production:

- [FedCloud integration appliance for OpenStack](#)

Monitoring

The monitoring probes for OpenStack have been subject to a complete refactoring that facilitates the introduction of new features, namely:

- Testing for correct exposure of OpenStack endpoints (Glance, Neutron, Nova)
- Testing for correct propagation of metadata from AppDB
- Improved deletion of VMs from previous executions

- Improved detection of parameters of the providers
- Enhanced authentication with OpenID Connect

3.1.2.1 Documentation

As with the integration documentation, the user-facing documentation is now maintained at github and hosted at readthedocs: <https://egi-federated-cloud.readthedocs.io/en/latest/>. The documentation for integration has been reviewed and updated to cover suggestions from new providers joining the federation. Detailed changelog of the documentation is available in the git log⁶.

3.1.3 Future plans

The changes introduced during the current period by the different components are focused on paving the road for a centralised operation of these components. For the next period, the EGI Cloud service development will focus on completing this transition to a central operations model. This will enable a more agile operation of the federation and will facilitate new providers joining.

3.2 EGI Cloud Container

3.2.1 Service description

Service/Tool name	Cloud Container
Service/Tool url	https://www.egi.eu/services/cloud-container
Service/Tool information page	https://www.egi.eu/services/cloud-container
Description	<p>Cloud Container Compute (in beta phase) gives you the ability to deploy and scale Docker containers on-demand. It offers guaranteed computational resources in a secure and isolated environment with standard API access, without the overhead of managing the operating system.</p> <p>Main characteristics:</p> <ul style="list-style-type: none"> ● On-demand provisioning ● Lightweight environment for maximised performance ● Standard interface to deploy on multiple service providers ● Interoperable and transparent

⁶ <https://github.com/EGI-Foundation/fedcloud-integration-documentation/commits/master> and <https://github.com/EGI-Foundation/fedcloud-documentation/commits/master>

	<ul style="list-style-type: none"> Removes friction between development and operations environments.
Value proposition	Deploy, manage, and scale containerized applications using Kubernetes on EGI Cloud providers.
User of the service/tool	Research Communities, individual researchers VREs, SME/Industry, Resource providers
User Documentation	https://wiki.egi.eu/wiki/Federated_Cloud_Containers
Technical Documentation	https://wiki.egi.eu/wiki/Federated_Cloud_Containers
Product team	EGI Foundation, UPV
License	Apache 2.0
Source code	https://github.com/grycap/ansible-role-kubernetes https://github.com/EGI-Foundation/VMI-endorsement
Testing	Images are tested automatically by the AppDB/SECANT integration. Travis integration for Kubernetes: https://travis-ci.org/grycap/ansible-role-kubernetes

3.2.2 Release notes

The base docker image in AppDB includes the following changes:

- Upgrade the base system to Ubuntu 18.04
- Upgrade to Docker v18.06
- Addition of Kubelet, Kubectl and Kubeadm v1.15

The grycap.kubernetes ansible role includes several new features and fixes:

- Support for NVIDIA docker
- Updated Docker version
- Better CNI (Container Networking Interface) support
- Improved node management
- Support for ingress
- Support for NFS persistent storage provisioner

3.2.3 Future plans

The EGI Cloud Container service will focus on the automated deployment of Kubernetes for users by relying on the `grycap.kubernetes` Ansible role and the Infrastructure Manager component developed in Task 6.3

3.3 EGI Workload Management

3.3.1 Service description

Service/Tool name	DIRAC4EGI
Service/Tool url	https://dirac.egi.eu/DIRAC
Service/Tool information page	https://wiki.egi.eu/wiki/Workload_Manager
Description	<p>Workload management service to distribute and manage centrally thousands of computational tasks on cloud and HTC. The DIRAC4EGI service is a cluster of instances to allow the users to submit computational tasks to the HTC and Cloud services, monitor the status and retrieve the results. Users interact with DIRAC both using GUI and API.</p> <p>The component of the DIRAC service is:</p> <ul style="list-style-type: none"> • DIRAC WMS server (multiple high-performance machines) • DIRAC DB (MySQL) server (high performance, large memory machine) • DIRAC REST server (medium sized machine) • DIRAC Web server (low CPU, high memory machine) • DIRAC configuration server (low CPU, high memory machine) • DIRAC data manager service (low CPU, high memory machine)
Value proposition	<ul style="list-style-type: none"> • Retain communities and attract new communities into the <code>dirac.egi.eu</code> service • Provide more flexible VRE simplifying access to EGI resources and accelerate the ability of researchers to undertake excellent science with DIRAC technical innovations • Transfer DCI skills and know-how to other medium and big communities and resource providers in the context of EGI

User of the service/tool	Users who need to deploy parallel applications in federated infrastructures or Cloud platforms.
User Documentation	https://wiki.egi.eu/wiki/HOWTO22
Technical Documentation	https://dirac.readthedocs.io/en/latest/DeveloperGuide/Overview/index.html?highlight=workload%20manage
Product team	CNRS
License	GNU General Public License v3.0
Source code	https://github.com/DIRACGrid/DIRAC
Testing	Internal continuous integration test suite: https://travis-ci.org/DIRACGrid/DIRAC

3.3.2 Release notes

The current production version is v6r22 which was out in July 2019. It was followed by a series of minor patches. The main features relevant to the EGI Workload Manager service are the following: enhanced functionality of pilot jobs including better support for multi-core jobs; enhanced packaging of DIRAC binary dependencies (DIRACOS package); enhanced procedure for user data synchronization with VOMS services to take into account Suspended user status. The next released v7r0 is already available and should be installed in production on the EGI Workload Manager hosts. This version has enhanced support for the complex workflow management.

3.3.3 Future plans

The next major version v7r1 is now in the process of being certified by the development team of the DIRAC Consortium. The certification is done against the set of comprehensive tests defined in the DIRAC project. These tests will be followed by tests done against the LHCb/CERN set of tests which cover a large part of the general DIRAC behaviour. This version will have tools to support EGI Check-In based user identification for both the Web Portal and the command line clients. It also enables automated provisioning of the RAuth proxies for the users.

3.4 EGI Online Storage

3.4.1 Service description

Service/Tool name	EGI Online Storage
Service/Tool url	https://www.egi.eu/services/online-storage/
Service/Tool information page	https://www.egi.eu/services/online-storage/
Description	Online Storage allows you to store data in a reliable and high-quality environment and share it across distributed teams. Your data can be accessed through different standard protocols and can be replicated across different providers to increase fault-tolerance. Online Storage gives you complete control over the data you share and with whom.
Value proposition	Easily share and organise your data, and control the data you share
User of the service/tool	Researchers, VREs, SME/Industry
User Documentation	http://italiangrid.github.io/storm/documentation.html
Technical Documentation	http://italiangrid.github.io/storm/documentation.html
Product team	INFN
License	Apache 2.0
Source code	https://github.com/italiangrid/storm
Testing	Quality Criteria Report: https://documents.egi.eu/document/3335 StageRollout Report: https://documents.egi.eu/document/3335

3.4.2 Release notes

In the reporting period, StoRM has been upgraded to version 1.11.15.

Concerning the StoRM backend and frontend services, this release introduced:

- proper limits for the incoming requests of frontend;
- improvements to backend's startup logic through a huge refactoring;
- minor fixes to the info provider;

Regarding the StoRM WebDAV service, this release introduced:

- support for third-party copy transfers implemented by extending the semantic of the WebDAV copy method;
- support for token-based authentication and authorization, by introducing an internal OAuth authorization server that can be used to issue tokens to client authenticated with VOMS credentials;
- support for OpenID connect authentication and authorization on storage areas;

3.4.3 Future plans

In the immediate future StoRM will be upgraded to version 1.11.16. This release:

- introduces Conscript JSSE provider support for StoRM WebDAV;
- introduces the support for CKSUM GridFTP command;
- fixes several minor bugs on StoRM services.

We are focusing on porting all the services to CentOS 7 and providing a StoRM Puppet module.

3.5 Advanced IaaS

3.5.1 Service description

Service/Tool name	Advanced IaaS
Service/Tool url	https://github.com/indigo-dc/udocker
Service/Tool information page	https://github.com/indigo-dc/udocker/blob/master/SUMMARY.md
Description	<p>udocker is a user-level tool that enables users to run containerized applications without needing root privileges on the machine, and without having to install any additional software.</p> <p>udocker offers a subset of the Docker functionalities with a very similar interface and can be used to pull containers from Docker hub and other Docker like repositories and execute them without requiring the presence of Docker. udocker can also import</p>

	<p>standard directory tarballs making possible the execution of non-Docker containers.</p> <p>udocker can be easily transferred to the target Linux system by the end user and simply executed there without any need of compilation or installation and without requiring the system administrator intervention. It relies on several user level mechanisms to provide a chroot-like environment to execute applications and services, but without requiring privileges. udocker integrates several drivers (execution modes) to offer the best possible environment for the execution of containers without requiring administrator privileges.</p>
Value proposition	<ul style="list-style-type: none"> ● Running Linux containers on machines without Docker or another container runtime locally installed ● Running Linux containers on machines without root privileges ● Running Linux containers without having to compile additional software. ● Using Linux containers in batch, interactive or cloud systems including grid, HTC and HTP clusters ● Running containers in systems where kernel features needed to run containers are unavailable or disabled ● Enables execution of applications with GPGPUs. ● Empower users to execute their applications everywhere through container encapsulation
User of the service/tool	<p>Users with complex applications and software dependencies who need to run applications or services encapsulated in containers, but which do not have administrator privileges on the execution nodes. This includes computing clusters many of which may not be willing to support conventional container execution due to security concerns and related policies. udocker is also being used successfully to support containers execution in cloud computing namely to exploit function-as-a-service.</p>
User Documentation	<p>https://github.com/indigo-dc/udocker/blob/master/doc/user_manual.md</p>
Technical Documentation	<p>https://github.com/indigo-dc/udocker/blob/master/SUMMARY.md</p>

Product team	LIP
License	Apache License 2.0
Source code	https://github.com/indigo-dc/udocker
Testing	<p>Functional, integration and unit tests available at: https://github.com/indigo-dc/udocker/tree/master/tests</p> <p>The udocker team uses a Jenkins CI system to perform code style verification, execution of tests and identification of potential security issues.</p>

3.5.2 Release notes

The latest release is v1.1.3, which includes the following new features:

- Add support for nvidia-docker like functionality
- Add support for runC to v1.0.0-rc5
- Command line interface improvements
- Improve metadata handling
- Add kill-on-exit to proot on Pn modes
- Force python2 as interpreter
- Directory absolute symlinks in Fn modes not being converted
- Package distribution in Pypi
- Support udocker execution inside a Docker container where the Docker image is "Alpine"
- Support "ARM" architectures

It also includes several improvements and the following bug fixes:

- Improve parsing of quotes in the command line
- Issues in file downloads
- Mitigation for upstream limitation in runC without tty
- Fix detection of executable with symlinks in container
- Improve pathname translation in Fn modes for mounted dirs

3.5.3 Future plans

The future plans for udocker in EOSC-Hub are:

- Improve support for MPI applications.
- Support for Python3.
- Modularization of the udocker source code: sustainability, code clarity, re-usability.
- Improve support for bind mount directories/volumes.
- Integration of additional execution engines.

- Improve OCI support.

3.6 EGI High-Throughput Compute

3.6.1 Service description

Service/Tool name	CREAM
Service/Tool url	CREAMCE is a distributed service, with multiple endpoints.
Service/Tool information page	https://www.egi.eu/services/high-throughput-compute
Description	<p>The CREAM (Computing Resource Execution And Management) Service is a simple, lightweight service that implements all the operations at the Computing Element (CE) level. The service is a basic component for a federated service-oriented architecture managing distributed processes (jobs). In order to guarantee the interoperability among different applications it implements a standard Web Service interface based on WS-I specification.</p> <p>The CREAM service accesses and operates local resource management systems. The current version of the application supports the following management systems: TORQUE, LSF, SLURM, HTCondor, Grid Engine (partially).</p> <p>For the Federated Compute package of EOSC-Hub a new installation and configuration tool has been developed for the CREAM and the resource BDII services. The tool is based on the Puppet framework and replaces the old reference application (YAIM) for the Grid environment and it is distributed as a Puppet module through the Puppet Forge portal: https://forge.puppet.com/infnpd/creamce</p>
Value proposition	<ul style="list-style-type: none"> • Workload and data management tools to manage all computational tasks • Large amounts of processing capacity over long periods of time • Faster results for your research • Shared resources among users, enabling collaborative research

User of the service/tool	Users of large-scale data intensive computational parallel applications.
User Documentation	https://cream-guide.readthedocs.io/en/latest
Technical Documentation	https://cream-guide.readthedocs.io/en/latest
Product team	INFN
License	Apache Software License, v.2.0
Source code	https://github.com/italiangrid
Testing	The platform chosen for the continuous integration is Jenkins; the build system takes advantage of its advanced features, like pipeline processing and Docker container support, for compiling and testing the code efficiently. The system at the moment makes use of the Jenkins platform hosted at INFN; it will be completely integrated into UMD workflow as soon as the compatibility is certified.

3.6.2 Release notes

In the reported period no major updates of the CREAM service or client have been delivered. No new features have been implemented and the communities didn't require any enhancement. The software can be considered completely mature.

3.6.3 Future plans

There's no plan to implement new features during the remaining project period. The minor issues reported in the GGUS ticketing system will be fixed by the end of this year (2019). The service maintenance team will deliver only security updates starting from January 2020.

4 Processing and orchestration

This task focuses on the maintenance and integration of orchestration services with the Cloud Compute and Cloud Container services. This allows to build complex virtual computing infrastructures based on the OASIS TOSCA Simple Profile YAML standard⁷ and integrate the INDIGO-DataCloud PaaS components as orchestrator for the EOSC-hub services.

Figure 1 provides an overview of the architecture and interrelation of the different components that are part of task T6.3 “Processing and Orchestration”. It also includes additional components that, even though they are not strictly included in T6.3 since they are not expected to be evolved in the context of EOSC-HUB, they are part of the PaaS Orchestration layer.

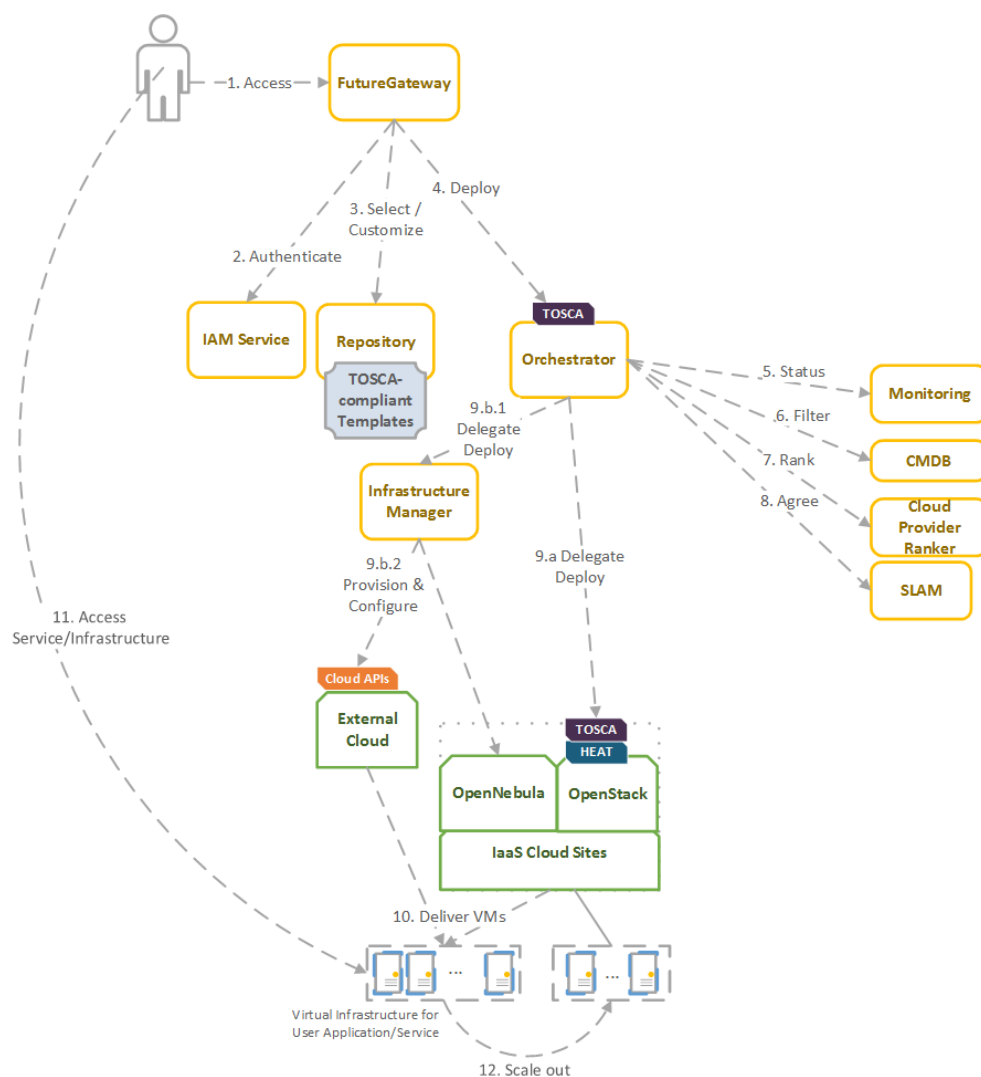


Figure 1. Architecture and workflow among the Processing and Orchestration services in EOSC-hub.

⁷ Crandall, John, and Paul Lipton. “OASIS Topology and Orchestration Specification for Cloud Applications (TOSCA) TC.” https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=tosca.

End users are expected to access the PaaS Orchestration layer via high-level graphical user interfaces, such as the portlets provided by the *FutureGateway*. These are web-based components that are customized for specific user applications and are responsible for performing the authentication with the *IAM service* and interacting with the *Orchestrator* by submitting a TOSCA template. TOSCA stands for the Topology and Orchestration Specification for Cloud Applications and it is a YAML-based domain-specific language (DSL) to describe application architectures to be deployed on a Cloud. Advanced end-users could also interact with the Orchestrator via the authenticated REST APIs provided.

Once the Orchestrator receives the TOSCA template it is in charge of interacting with different services in order to identify the most appropriate IaaS Cloud site on which to perform the execution. This decision depends on the monitoring state of the underlying Cloud sites (information managed by the *Monitoring service*), the SLAs (Service Level Agreements) agreed between the user and the sites (information managed by the *SLAM service*) and the availability of the VMIs (Virtual Machine Images) on each site (information managed by the *CMDB service*). With all of this information, the *CloudProviderRanker* service is employed in order to apply a set of rules in order to obtain a ranked list of Cloud sites.

The Orchestrator delegates on the *Infrastructure Manager* to perform the deployment on public Cloud sites (Amazon Web Services, Microsoft Azure, Google Cloud Platform and Open Telekom Cloud) or on other external Clouds managed by popular Cloud Management Platform (CMPs) such as OpenNebula and OpenStack. The IM can also be configured with single-site mode in order to provide a TOSCA-enabled endpoint and support local-site orchestration of complex application architectures. In order to achieve a similar functionality in OpenStack, the *HEAT Translator* component can be used in order to translate from a TOSCA template into a HOT template, the native language employed by the HEAT service in OpenStack.

The following section provides an overview of the services involved in task T6.3 “Processing and Orchestration in EOSC-hub”. For each service, a brief description is included and, in a separate section, the added value in EOSC-HUB for each service is identified.

4.1 TOSCA for Heat

4.1.1 Service description

Service/Tool name	TOSCA for Heat
Service/Tool url	https://wiki.openstack.org/wiki/Heat-Translator
Service/Tool information page	https://wiki.openstack.org/wiki/Heat-Translator
Description	TOSCA for Heat service translates OASIS TOSCA templates in order to produce valid HOT (Heat Orchestration Template)-compliant

	<p>documents ready to be deployed by the OpenStack Orchestration program, Heat.</p> <p>The main component is the heat-translator command-line tool, which, in the case of incoming TOSCA templates, makes use of in-memory graphs provided by TOSCA Parser to obtain the resultant HOT.</p>
Value proposition	Deployment of cloud topologies in OpenStack Heat described using OASIS TOSCA standard.
User of the service/tool	User communities interfacing directly with an OpenStack IaaS frameworks to deploy complex applications using TOSCA templates.
User Documentation	https://docs.openstack.org/heat-translator/latest/
Technical Documentation	https://docs.openstack.org/heat-translator/latest/
Product team	CSIC
License	Apache License 2.0
Source code	https://github.com/openstack/heat-translator https://github.com/indigo-dc/heat-translator
Testing	https://zuul.openstack.org/job/heat-translator-tox-py27-tp https://jenkins.indigo-datacloud.eu:8080/job/Pipeline-as-code/job/heat-translator/

4.1.2 Release notes

No major or relevant feature has been added to the service in the reported period. Only maintenance effort to support minor modifications in the current set of TOSCA templates related to OpenStack deployments currently in use by the communities in the project.

4.1.3 Future plans

Continue with the service maintenance plan and integration with EOSC-hub services.

4.2 Infrastructure Manager

4.2.1 Service description

Service/Tool name	Infrastructure Manager
Service/Tool url	URL REST API: https://appsgrycap.i3m.upv.es:31443/im URL Web portal: https://appsgrycap.i3m.upv.es:31443/im-web/
Service/Tool information page	http://www.grycap.upv.es/im/index.php
Description	<p>IM is a tool that orchestrates the deployment of complex and customized virtual infrastructures on multiple back-ends. The IM automates the deployment, configuration, software installation, monitoring and update of virtual infrastructures. It supports a wide variety of back-ends, including both public IaaS Clouds (Amazon Web Services, Microsoft Azure, etc.), on-premises Cloud Management Platforms (OpenNebula, OpenStack, etc.) and Container Orchestrators (Kubernetes), thus making user applications Cloud agnostic.</p> <p>In addition, it features DevOps capabilities, based on Ansible enabling the contextualization of the infrastructure at run-time by installing and configuring all the required software that may not be available in the Virtual Machine Images, thus providing the user with a fully functional virtual infrastructure.</p> <p>The IM provides both XML-RPC and REST APIs to enable high-level components to access its functionality. These APIs provide a set of functions for clients to create, destroy, and get information about the infrastructures. It also enables the elasticity management, both horizontal (adding/removing nodes) and vertical (growing/shrinking the capacity of nodes).</p> <p>The cloud topologies can be defined using its native language called RADL (Resource and Application Description Language) or TOSCA OASIS standard (YAML version 1.0).</p>

Value proposition	Deployment of complex and customized virtual infrastructures on multiple Cloud back-ends using standard specification languages.
Customer of the service/tool	Research Communities, VREs, individual researchers.
User of the service/tool	Research Communities, individual researchers.
User Documentation	https://imdocs.readthedocs.io/
Technical Documentation	https://imdocs.readthedocs.io/
Product team	GRyCAP - UPV
License	GPL 3.0
Source code	https://github.com/indigo-dc/im
Testing	https://jenkins.indigo-datacloud.eu/job/Pipeline-as-code/job/im/job/master/

4.2.2 Release notes

The IM has been improved to enable the creation of private networks, in those sites where the policies enable the user to do it, instead of selecting from the ones previously created by the administrator. This enhances the security on the virtual infrastructures as the networks are not shared with other users/infrastructures. Also, the addition of static routes has been integrated in order to use some instances (Virtual Machines) as routers for all the nodes of the created network. These new functionalities are oriented to improve the creation of hybrid deployments among different set of cloud providers, thus enhancing network connectivity.

A public highly available instance of the IM is maintained at UPV. It is deployed on a dedicated Kubernetes cluster of three nodes (one master and two working nodes). It has deployed ten instances of the IM container using a HAProxy to balance the load among them. Also, the IM web portal has been deployed in this cluster using a nginx service to separate the application deployed. Both services has been added at the EOSC portal and marketplace: <https://eosc-portal.eu/infrastructure-manager>.

4.2.3 Future plans

Continue with the maintenance of the service and integration with EOSC-hub services.

4.3 PaaS Orchestration system

4.3.1 Service description

Service/Tool name	PaaS Orchestrator
Service/Tool url	https://indigo-paas.cloud.ba.infn.it/orchestrator
Service/Tool information page	https://indigo-dc.gitbooks.io/indigo-paas-orchestrator/content
Description	<p>The Orchestrator is the core component of the INDIGO PaaS system and the entry point for the users' requests: it allows to coordinate the provisioning of virtualized compute and storage resources on Cloud Management Frameworks, both private and public (like OpenStack, OpenNebula, AWS, etc.), and the deployment of dockerized long-running services and batch jobs on Apache Mesos clusters.</p> <p>It receives the deployment requests, expressed through templates written in TOSCA (Simple Profile in YAML version 1.0), and orchestrates the deployments on the best available cloud sites. In order to select the best site, the Orchestrator implements a complex workflow: it gathers information about the SLAs signed by the providers with the user, the monitoring data about the availability of the compute and storage services, the location of the data requested by the user (if any). Hybrid deployments spanning multiple sites are supported.</p> <p>Using the PaaS Orchestrator and the TOSCA templates, the end users can exploit computational resources without any knowledge about the IaaS details.</p> <p>The end users can submit their request to the Orchestrator using different clients, e.g. the web dashboard or the command-line interface or using directly the REST APIs. Anyway, the Orchestrator is the only service the users need to interact with for creating, monitoring and deleting their deployments.</p> <p>In order to coordinate the complex deployment workflow, the Orchestrator relies on the following INDIGO PaaS services:</p>

CMDB (Configuration Management Database) is a database system in INDIGO PaaS environment used by other platform components as a central point of technical information such as location, types and other metadata about sites, services and virtual machine images. CMDB provides a REST API for that purpose. The interface allows to manage schema of the entities stored in the database. The schema is enforced in operations such as CREATE, READ, UPDATE and DELETE. Security of these operations is ensured by different access levels for operators and other users as well as hierarchy of entity ownership.

SLAM (SLA manager) is a component of the INDIGO PaaS system that handles negotiations between infrastructure customers and infrastructure providers. *Customer* is an entity representing a group of people (system users) sharing common interests and entitled to share computing and storage resources in the scope of the PaaS system. *Provider* is an entity delivering computing and storage resources for customers, represented by systems users with provider role given. SLAM allows users to create SLA drafts that express customer requirements for the underlying infrastructure such as computing volume restrictions, public IP restrictions for computing resources, as well as time and volume restrictions for storage resources. Each SLA is created for service available on a given site.

The SLA draft is created by customer representative and subjected to provider's approval. The Provider can either agree or deny customer's proposal, as well as to reflect their own suggestions for the given restrictions. The customer reviews these suggestions. The process of negotiations continues until the consensus is met between the customer and the provider. Once SLA is signed it is made available for other systems to consume (Orchestrator). SLAM also allows to reflect customer preferences for the priority for the use of resources - in case specific sites or services should be used in favour to others. In order to access SLAs as well as customer preferences REST API was made available.

CPR (Cloud Provider Ranker) is a component of the INDIGO PaaS system that ranks cloud provider services implementing a rule-based algorithm. It's a stateless microservice providing REST APIs to request the ranking. The ranking is performed taking into account SLAs targets and monitoring metrics for the different

	<p>services; configurable rules are applied to the input data in order to normalize them and use specific weights. The CPR computes the scores for the different services and assigns a rank to each of them, thus returning an ordered list of cloud services. The Orchestrator consumes this service in order to obtain the ordered list of providers/services where it will try to schedule the deployment. The first of the list will be considered the “best” service for fulfilling the user request; the other ones will be considered in case of failure.</p> <p><i>Monitoring</i> system is another component of the INDIGO PaaS system. Its architecture consists of three sub-components: the Zabbix server that stores the metrics, the Zabbix wrapper that provides a REST API interface to Zabbix to retrieve the metrics (e.g. this is used by the Orchestrator to get the monitoring data about the different services) and a set of Zabbix probes that collect the relevant metrics for the different cloud services.</p>
Value proposition	<ul style="list-style-type: none"> ● high level orchestration of heterogeneous Cloud resources ● provisioning of virtualized compute and storage resources ● SLA negotiation
User of the service/tool	Users who want to deploy complex applications on multiple Cloud resources, with SLA support.
User Documentation	https://indigo-dc.gitbooks.io/indigo-paas-orchestrator/content/how_to_deploy.html
Technical Documentation	https://indigo-dc.gitbooks.io/indigo-paas-orchestrator/content REST APIs docs: https://indigo-dc.github.io/orchestrator/restdocs
Product team	INFN, Reply
License	Apache License 2.0: https://github.com/indigo-dc/orchestrator/blob/master/LICENSE
Source code	https://github.com/indigo-dc/orchestrator (main component) https://github.com/indigo-dc/cmdb https://github.com/indigo-dc/slam

	https://github.com/indigo-dc/CloudProviderRanker https://github.com/indigo-dc/monitoring
Testing	https://jenkins.indigo-datacloud.eu:8080/job/Pipeline-as-code/job/orchestrator

4.3.2 Release notes

The new version (2.2.0) of the Orchestrator integrates an improved TOSCA parser that ensures better compliance with the standard.

Several improvements have been implemented in the auxiliary services of the PaaS Orchestration system, in particular:

- the Cloud Provider Ranker (CPR) has been re-engineered in order to provide a better ranking of the sites where the Orchestrator can schedule the deployments;
- the Monitoring probes have been improved and new ones implemented;
- the CMDBv1 has been extended to publish additional information about the images and the flavors (e.g. the GPU support).

Moreover, the Orchestrator now supports the following new features:

- GPU resource scheduling on Clouds. The user can specify in the TOSCA template the number of GPUs, the vendor/model, and eventually the driver version for the Virtual Machines to be deployed.
- Secrets management for Marathon applications (long-running services). The Orchestrator is able to store the secrets provided in the TOSCA template into Vault for applications deployed on Marathon clusters with enabled secrets support.
- HPC integration through QCG-Computing (first implementation). The Orchestrator now includes a plugin for submitting jobs to HPC through TOSCA exploiting the QCG-Computing interfaces (<http://www.qoscosgrid.org/trac/qcg-computing/>).

Finally, the support for hybrid deployments and the retry strategy have been improved in order to ensure a better user-experience (e.g. the log in case of retry failure is now returned by the Orchestrator).

The Orchestrator now comes with a user-friendly dashboard that can be used to submit deployment requests in a simple and straightforward way. The dashboard hides the complexity of the TOSCA templates providing simple forms for the deployment configuration and customization. The user can retrieve the list of her/his deployments with details like the creation time, the status, the provider hosting the deployment resources, the log, etc.

4.3.3 Future plans

Follow with the maintenance of the services and integration with EOSC-hub services.

4.4 Future Gateway

4.4.1 Service description

Service/Tool name	Future Gateway
Service/Tool url	https://futuregatewayframework.github.io
Service/Tool information page	https://github.com/FutureGatewayFramework/fgDocumentation
Description	<p>FutureGateway is a set of components enabling end-users to access functionality such as authentication/authorization or deployment management. It exposes a REST interface to a service called fgAPIServer. The REST calls can be used to Create, Read, Update or Delete (CRUD) entities such as tasks to be executed on remote computing resources. For high-level actions (e.g. new task submission), fgAPIServer delegates to APIServerDaemon, a backend responsible for direct interaction with different middlewares. This solution is flexible to allow largely different scenarios to be handled. On one hand, fgAPIServer has been shown to work seamlessly with mobile applications, Java-based scientific workflows and with direct web-browser REST call generation. At the same time, APIServerDaemon handles single job submission as well as Virtual Machine deployment in a private cloud.</p>
Value proposition	An easy to use frontend for complex backend scenarios such as: remote job submission or virtual machine and container deployment.
User of the service/tool	Research communities, individual researchers
User Documentation	https://github.com/FutureGatewayFramework/fgDocumentation/blob/master/usage.md
Technical Documentation	https://github.com/FutureGatewayFramework/fgDocumentation/blob/master/installation.md
Product team	INFN Catania, PSNC
License	Apache 2.0

Source code	https://github.com/FutureGatewayFramework https://github.com/indigo-dc/indigo-parent https://github.com/tzok/fg-docker-compose https://github.com/tzok/eosc-fgapiserver
Testing	<p>A set of tests is available in the main repository: https://github.com/FutureGatewayFramework/fgAPIServer/tree/master/tests</p> <p>Additionally, the Docker Compose / Kubernetes version of the service has been validated with integration tests from the following repository: https://github.com/indigo-dc/indigoclient</p>

4.4.2 Release notes

This component has been mostly updated in terms of maintenance in response to changes in libraries and tools used upstream. Namely, [tzok/eosc-futuregateway](https://github.com/tzok/eosc-futuregateway) project which contains specifications on how to build and run FutureGateway service with Docker images required changes of the main tool used for this purpose. Previously, Ansible Container was used, but due to its further development being discontinued, the new tool Buildah was selected for the process of Docker images preparation. Project [tzok/eosc-futuregateway](https://github.com/tzok/eosc-futuregateway) was released as version 2.0 and additionally it contains example specifications of Kubernetes deployment and Docker Compose YAML file.

This release covers also libraries updates with security fixes in [indigo-dc/indigo-parent](https://github.com/indigo-dc/indigo-parent) project and switch from deprecated MySQL-python to official mysql-connector in [FutureGatewayFramework/fgAPIServer](https://github.com/FutureGatewayFramework/fgAPIServer) project.

4.4.3 Future plans

Follow the changes dictated by requirements from user communities.

5 Data and Metadata Management

The EOSC-hub common repository services and the policy-driven data management/stewardship services, with particular regard to registered data (i.e. data associated to a persistent identifiers), are described in detail in the following paragraphs and shown in the picture (Figure 2).

These services allow datasets to be stored in a geographically distributed repository and enable the association of persistent identifiers with the data, making the data set's physical location independent from the logical references pointing to it. The identifier is globally resolvable, and the data set is replicated in multiple copies, which are tracked in the metadata associated with the identifier. Data can be published, and community specific metadata may be associated with it, this metadata can then be harvested and indexed by a discovery service to make the data findable. Data can also be annotated, manually or programmatically via API. And last, but not least, data are curated through a set of policies that each data manager can define.

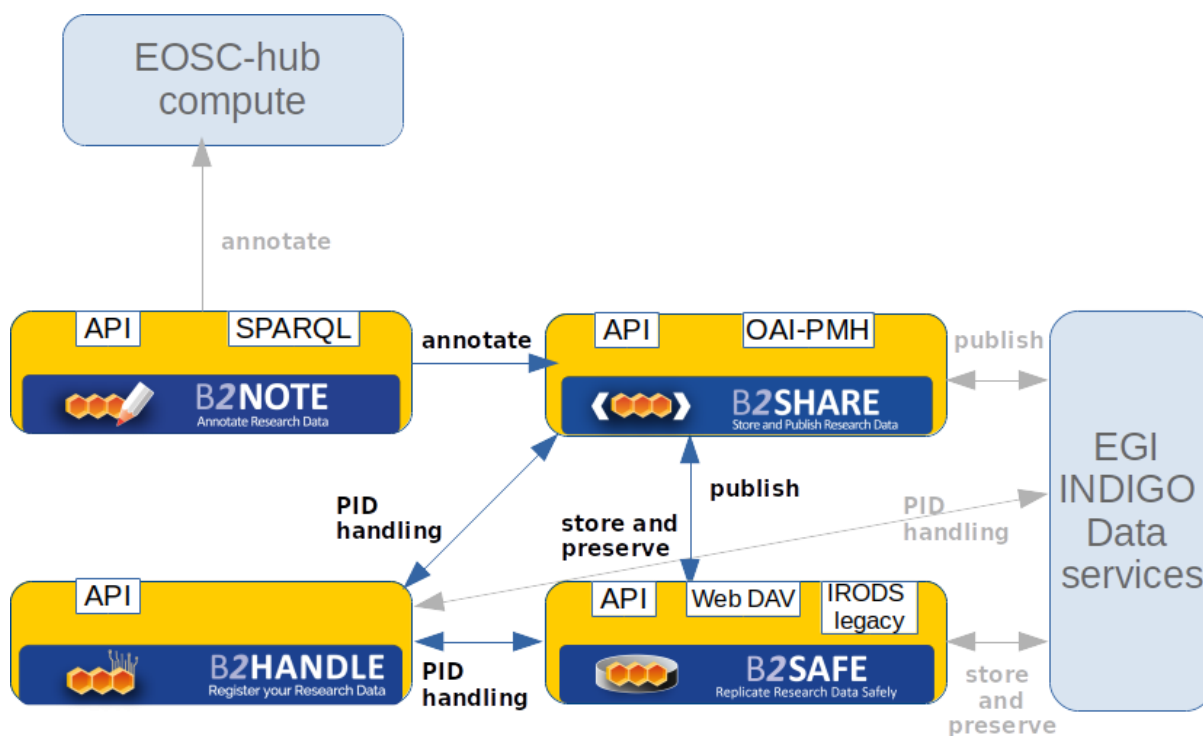


Figure 2. Data and metadata management services

5.1 B2HANDLE

5.1.1 Service description

Service/Tool name	B2HANDLE
-------------------	----------

Service/Tool url	Not applicable as it is a distributed service.
Service/Tool information page	https://eudat.eu/services/b2handle
Description	<p>B2HANDLE is a distributed service, designed to contribute to data persistency by maintaining opaque, globally unique persistent identifiers (PIDs). PIDs are used in other user-facing e-infrastructure services such as B2SAFE and B2SHARE to reliably identify data objects over long periods of time, possibly beyond object lifetime, and thereby provide easy, stable references for use by service components and end-users. B2HANDLE identifiers bear metadata (kernel information), which provides the critical information about an identified object for use by other services. The B2HANDLE service offers management of identifier namespaces (Handle prefixes), supports object policies, and maintains stable business workflows to provide a reliable, trustworthy and scalable service. B2HANDLE also offers the Central PID Catalog, which enables reverse lookups on B2HANDLE PIDs and simple searches/filtering. For this, B2HANDLE maintains the necessary software components. To simplify automated management of PIDs, B2HANDLE provides a set of Python libraries (b2handle, pyhandle). B2HANDLE also includes the HRLS component to provide reverse-lookups and searching, which is essential to run the Central PID Catalog.</p>
Value proposition	<p>B2HANDLE facilitates persistent identification of research assets, particularly data, independent from and throughout changes of current location or ownership. It provides a trustworthy, scalable, reliable service. Usage of PID Profiles, the Central PID Catalog and the B2HANDLE support libraries can facilitate efficient automated data management at early stages of the data life cycle.</p>
Customer of the service/tool	Research infrastructures, disciplinary or generic, and their hosting institutions/service providers
User of the service/tool	Infrastructure services; end-users interested in identifying research assets before publication phases.
User Documentation	https://www.eudat.eu/services/userdoc/b2handle

	https://eudat.eu/services/userdoc/b2handle-for-communities https://eudat.eu/services/userdoc/b2handle-for-end-users
Technical Documentation	http://eudat-b2safe.github.io/B2HANDLE http://eudat-b2safe.github.io/PYHANDLE https://github.com/EUDAT-B2SAFE/B2HANDLE-HRLS
Product team	SURFsara, GRNET, SNIC/KTH, DKRZ
License	Source code available under Apache 2.0 Open Source license: https://github.com/EUDAT-B2SAFE/B2HANDLE/blob/master/LICENSE
Source code	https://github.com/EUDAT-B2SAFE/B2HANDLE https://github.com/EUDAT-B2SAFE/PYHANDLE https://github.com/EUDAT-B2SAFE/B2HANDLE-HRLS
Testing	Continuous integration testing is configured for B2HANDLE via Jenkins service (GRNET)

5.1.2 Release notes

The Handle Reverse-Lookup Servlet (HRLS) has been released as version 1.0.4 providing the following fixes and new features:

- Several dependencies (Java libraries) were updated to the latest version to address security issues uncovered since the previous release.
- The log rotation mechanism was updated and extended to better work in operational environments with large numbers of servers and high request rates.
- Licensing information was added to comply with EOSC-hub policy.

In addition, monitoring of the overall B2HANDLE stack was updated to comply with new monitoring mechanics set up by EOSC-hub.

5.1.3 Future plans

A new release of the b2handle/pyhandle library is planned for end of 2019 to address already documented issues and add new features, including support for new Python versions (i.e., beyond Python 3.5).

5.2 B2SAFE

5.2.1 Service description

Service/Tool name	B2SAFE
Service/Tool url	Not Applicable: it is a distributed service.
Service/Tool information page	https://www.eudat.eu/services/b2safe
Description	<p>B2SAFE is a highly available multi-purpose service that allows community repositories to implement data management policies on their research data that is distributed across multiple administrative domains. The following components are part of B2SAFE.</p> <p>The service core: it offers functionality for the long term data preservation. The main feature is the function to replicate data sets across different data centres in a safe and efficient way while maintaining all information required to easily find and query information about the replica locations. The information about the replica locations and other important information is stored in a PID registry (B2HANDLE). The B2SAFE Service is implemented as a package on top of iRODS, providing a set of iRODS rules and scripts.</p> <p>Data transfer: B2SAFE supports different transport protocols in combination with B2STAGE. It provides endpoints for those protocols and enable third-party transfers.</p> <p>Data Policy Management: data management policies are abstract descriptions of data management, stewardship and curation tasks. These policies, which are stored in a database and handled via a Data Policy Manager (DPM), translate into concrete operational tasks that are event-triggered and executed within the B2SAFE rule engine for instance when data is ingested, stored and must be curated or transferred.</p>

Value proposition	For the communities who need to guard against data loss; a customer facing service that allows data replication and policy-driven data management using different storage services provided by geographically distributed centres in the EUDAT CDI.
User of the service/tool	Teams of researchers or single users
User Documentation	https://eudat.eu/services/userdoc/b2safe https://eudat.eu/services/userdoc/configure-b2safe http://eudat.eu/services/userdoc/using-b2safe https://eudat.eu/services/userdoc/joining-b2safe
Technical Documentation	https://github.com/EUDAT-B2SAFE/B2SAFE-core/wiki https://github.com/EUDAT-B2SAFE/pam-oauth2 https://github.com/EUDAT-B2SAFE/B2SAFE-DPM
Product team	eudat-safereplication@postit.csc.fi
License	https://github.com/EUDAT-B2SAFE/B2SAFE-core/blob/master/LICENSE
Source code	https://github.com/EUDAT-B2SAFE/B2SAFE-core https://github.com/EUDAT-B2SAFE/pam-oauth2 https://github.com/EUDAT-B2SAFE/B2SAFE-DPM
Testing	A test suite, which is executed manually is available here: https://github.com/EUDAT-B2SAFE/B2SAFE-core/tree/master/scripts/tests

5.2.2 Release notes

Major Achievements during the last year are: B2SAFE improvement including bug fixing, new rules and PID microservices, integration of the B2SAFE with the B2SHARE service (currently in progress).

A new release version of B2SAFE (4.2.1) will be scheduled for early 2020.

5.2.3 Future plans:

- Finalise the integration with B2SHARE through the iBridge component, including management of access tokens
- Adding Apache Airflow as a framework to improve the B2SAFE/B2SHARE connection component
- Introducing iRODS Python rule engine
- Freezing snapshot of collections
- Integration with B2SHARE (MD5 checksum, file size, upload/download from B2SHARE UI)
- New testbed based on Jenkins, including a configuration for the integration tests + Github/Gitlab as source code repository
- Integration with DMP (Data Management Policy)

5.3 B2SHARE

5.3.1 Service description

Service/Tool name	B2SHARE
Service/Tool url	https://b2share.eudat.eu
Service/Tool information page	https://www.eudat.eu/services/b2share
Description	Repository for shareable digital objects to improve your data sharing and publishing and guarantee long-term persistence of your locally stored data. Increased findability and discoverability.
Value proposition	<p>For the individual researchers who do not have adequate facilities for storing, preserving and sharing data, B2SHARE service is a customer-facing service which provides a safe repository for scientific data and an easy way to share it in the research community.</p> <p>For research communities that want to centrally manage publications and enforce policies like review workflows.</p>
User of the service/tool	Individual researcher, scientific community, scientific institution (e.g. university)
User Documentation	https://eudat.eu/services/userdoc/b2share-usage https://www.eudat.eu/b2share-training-suite

	https://github.com/EUDAT-Training/B2SHARE-Training
Technical Documentation	https://github.com/EUDAT-B2SHARE/b2share https://www.eudat.eu/b2share-training-suite
Product team	CSC - IT-center for Science
License	GPLv2
Source code	https://github.com/EUDAT-B2SHARE/b2share
Testing	Training instance of B2SHARE where users can try and test out functionality provided by B2SHARE is available at: https://trng-b2share.eudat.eu Automatic unit test suite can be found at: https://github.com/EUDAT-B2SHARE/b2share/tree/master/tests

5.3.2 Release notes

A new version of B2SHARE has been [released](#) early October. This release includes:

- Bug fixes
- Small UI fixes and improvements
- Display community's recent records on community details page
- List of communities on search page sorted alphabetically
- Remove custom domain certificate for B2ACCESS staging instance

A next released is targeted for early 2020.

5.3.3 Future plans

Future plans will also be discussed at the next EUDAT Hackathon in October 2019 at JSC. Examples of activities will include:

- PID metadata updates from B2SHARE
- B2NOTE integration
- Software stack updates

5.4 B2NOTE

5.4.1 Service description

Service/Tool name	B2NOTE
Service/Tool url	https://b2note.eudat.eu
Service/Tool information page	https://eudat.eu/catalogue/B2NOTE
Description	Data annotation service to enrich the description of shared/published datasets without modifying the underlying metadata schemata. Service allows to add semantic tag coming from existing semantic resources (controlled vocabularies, thesauri, ontologies), user-defined keywords and comments that could be used to curate, enrich and retrieve data elements of relevance. Improve reusability and findability.
Value proposition	<p>For the individual researchers who are reusing shared/published data for their scientific project, B2NOTE service is a customer-facing service which integrates with data repositories and data services user interface to provide a simple tool for annotating data with semantic tags, keywords or comments and reusing these annotations to retrieve and aggregate relevant data elements across heterogeneous and distributed data repositories and services.</p> <p>For research communities that want to foster the reuse of their datasets, extend the metadata description with semantics, curate datasets.</p>
User of the service/tool	Individual researcher, scientific community, scientific institution (e.g. university)
User Documentation	<p>Textual documentation: coming soon</p> <p>Webinar: https://youtu.be/EkvUrBPBjck</p> <p>Slides: https://eudat.eu/sites/default/files/WebinarB2NOTE_presentation.pdf</p>

Technical Documentation	In revision- coming soon
Product team	e-Science Data Factory
License	MIT licence
Source code	https://github.com/EUDAT-B2NOTE/b2note
Testing	<p>Integrated with the training instance of B2SHARE where users can try and test out functionality provided by B2NOTE is available at: https://trng-b2share.eudat.eu</p> <p>Automatic unit test suite can be found at: https://travis-ci.org/EUDAT-B2NOTE/b2note</p>

5.4.2 Release notes

A new version of B2NOTE has been [released](#) end of July. This release includes:

- Bug fixes
- Upgrade to Python 3 for the backend (based on Flask)
- Full javascript client which can be integrated as webcomponent
- Simplified architecture
- Updated API to allow POST and PUT requests
- Authentication integrated with API

This release is a first version of the new B2NOTE architecture. A next release is targeted for early 2020.

5.4.3 Future plans

- Integration with B2SHARE
- Integration with OpenAIRE explore and OpenAIRE Research Community Dashboard
- Integration with Onedata
- Integration with Zenodo
- Development of dedicated user interface to extend the capabilities of B2NOTE
- Improvement of the javascript client

6 Preservation

This task aims at integrating certified Trusted Digital Repository (TDR) in the EOSC-hub catalogue, resulting in a sustainable long-term data preservation service: the European Trusted Digital Repository (eTDR).

6.1 eTDR

6.1.1 Service description

Service/Tool name	eTDR
Service/Tool url	Service to be added to EOSC-hub service catalogue before M36
Service/Tool information page	https://confluence.egi.eu/display/EOSC/EUDAT+-+Long-term+preservation+services (internal)
Description	eTDR is a service that ensures digital information remains findable, accessible, interoperable and reusable over time. This includes capacity and resource planning, as well as application of long-term preservation techniques and technologies. It also combines policies, processes and actions to ensure access to "born-digital" and re-formatted data, regardless of the challenges of technological changes or failures (metadata, file format, media).
Value proposition	The service relies on mature, "ready-to-use" EUDAT services, which provide data transferring functionalities. These include B2SAFE, B2HANDLE, B2FIND.
Customer of the service/tool	Communities willing to preserve large collections of datasets over time, with no limit of duration.
User of the service/tool	Institutions representing a research community
User Documentation	https://drive.google.com/drive/folders/1MZehVTZ9aGc5s0xdkqovVEY7TLEDy5Mi
Technical Documentation	This is an internal service within CINES/DANS which is exposed via EUDAT services. The technical documentation of the eTDR itself is internal to CINES/DANS.

Product team	CINES
License	This is an internal tool which is not intended for deployment outside the pool of existing identified EOSC-hub service providers (CINES, DANS) - thus there is no general license available, excepted for the tools on which eTDR relies on : B2SHARE, B2SAFE
Source code	This is an internal (not publicly accessible) repository for scripts, code managing: quality assurance (file format validation), workflow management, etc.
Testing	Strategy includes multiple environments (development, user-testing, production) as well as datasets for regression testing - which are not publicly available

6.1.2 Release notes

Two instances of the eTDR are part of the deployment plan: CINES and DANS.

The access interface of the CINES instance of eTDR has been strengthened by integrating B2FIND as an access a portal that provides easy access to the archived data collections (scanned images and data from plant species) from the Herbadrop consortium. In the meantime, the monitoring feature has been improved for the instance of B2SAFE that is used for uploading the data to be ingested to the eTDR. The digital preservation platform (PAC) now integrates three generic services (B2SAFE, B2HANDLE and B2FIND).

In parallel, DANS proposed a definition for the architecture of its instance of eTDR, with the objective to demonstrate with a POC (Proof of Concept) that it can interface with B2SHARE. The source code of B2SHARE has been amended to integrate access to the DANS digital repository (EASY) and the prototype is now up and running. More information is available in the EOSC-hub milestone M6.7 "Ingest points of two TDRs integrated with EOSC-hub data transfer services" document.

6.1.3 Future plans

The activities for the final period of the project will lead to the addition of eTDR to the EOSC-hub service catalogue. These will include the integration of B2ACCESS as the AAI protocol for authentication, the definition of a business model for the long-term preservation service, and the deployment of the DANS eTDR prototype into production.

7 Sensitive Data

The EOSC-hub service catalogue includes two services that offer storage and processing of data in a secure environment. These are:

TSD: provides researchers with a desktop with secure storage and software to run your collection and analysis of sensitive data. The system is built on the idea that having a robust firewall around a system that provides a full separation of projects, is the best policy. A two-step authentication is needed to gain access to the system. Inside the system, every project has its own VLAN and its own virtual file system. This means that projects cannot find any information about any other project on the system.

CSC ePouta: provides Infrastructure as a-Service for running analysis on sensitive data. The ePouta Virtual Private Cloud service allows customers to provision virtual machines and storage resources directly to their own internal networks. It provides an easy to use admin web interface and a programmable API for managing virtual machines, networks and storage. CSC ePouta meets elevated information security level regulations and is targeted for sensitive data processing. ([Terms of use](#))

7.1 TSD

7.1.1 Service description

Service/Tool name	Services for Sensitive Data (TSD)
Service/Tool url	https://www.uio.no/english/services/it/research/sensitive-data/
Service/Tool information page	https://www.uio.no/english/services/it/research/sensitive-data/about/
Description	TSD offers storage capability, computing infrastructure, analysis / visualization platforms and web-based data collection tools suitable for running complex research projects in a secure IT-infrastructure.
Value proposition	Desktop with secure storage and software to enable the collection and analysis of sensitive data.

Customer of the service/tool	Researchers, Research group, Business, Research organisations
User of the service/tool	Researchers dealing with sensitive data
User Documentation	https://www.uio.no/english/services/it/research/sensitive-data/use-tsd/usermanual/
Technical Documentation	https://www.uio.no/english/services/it/research/sensitive-data/about/description-of-the-system.html
Product team	Department of Research Support Services https://www.usit.uio.no/english/about/organisation/rc/rss/index.html
License	https://github.com/unioslo/tsd-api-client/blob/master/LICENSE
Source code	This is an internal (not publicly accessible repository) for scripts managing: TSD API, file-system, Slurm cluster, access control, accounting, etc. https://bitbucket.usit.uio.no/projects/TSD/repos/documentation TSD API client is public: https://github.com/unioslo/tsd-api-client
Testing	For the HPC cluster “Colossus”, we have an identical test cluster “Snabel” for testing the new features, and upgrades. Similar testbeds exist for TSD API, user accounting, Main HNAS file-system, VMware hypervisor, routers, and firewalls

7.1.2 Release notes

New service release:

- HPC cluster “Colossus 3.0”.
 - Major changes:
 - OS upgrade CentOS 6 to CentOS 7
 - Architecture: From Intel to AMD
 - New CI: Singularity 3 module
- Windows OS for project VMs “Windows server 2019”

New services:

- **Data upload:** file upload/download service using web and API interfaces
 - Service components:
 - Easy web-based file upload
 - API (Application Programming Interface) based file upload
 - Main service target in the SLA:
 - Real-time delivery of Nettskjema to the project folder
- **Consent service:** Provides TSD projects an easy framework for search consent forms and enables consenters (Data subjects) the ability to view and revoke given consents in compliance with the Norwegian and European data protection laws.
- **Self-service project administration:** Provides TSD project PIs with a secure web interface through which they can: add new users (with Norwegian PID), assign/revoke privileges to users, and change the HPC quota of the project
- **Data publication service:** The ability of TSD project users to security provide access to sensitive datasets for collaborators who are not members of TSD
- Internal mirrors tracking PyPi, CRAN and Bioconductor
- * All Windows VMs upgraded to Server 2019?

GDPR compliance: The data protection officer and the security officer at the University of Oslo conducted an audit of TSD. The input:

- Detailed system description (including the consent service)
- Risk analysis
- Description of all services, service components, CIs that involve input/output/analysis of personal/sensitive data

7.1.3 Future plans

- Support ELIXIR AAI full compatibility
- IoT smart homes
- cross-border workload distribution
- Browser-based video recording directly to TSD storage backend for computers / mobile
- Self-service user registration for international users

7.2 ePouta

7.2.1 Service description

Service/Tool name	Secure and cost-effective cloud computing for processing sensitive data (ePouta)
Service/Tool url	https://research.csc.fi/epouta

Service/Tool information page	https://research.csc.fi/cloud-computing
Description	This service provides infrastructure as a-service for running analysis on sensitive data. The ePouta Virtual Private Cloud service allows customers to provision virtual machines and storage resources directly to their own internal networks. It provides an easy to use admin web interface and a programmable API for managing virtual machines, networks and storage. CSC ePouta meets elevated information security level regulations and is targeted for sensitive data processing
Value proposition	IaaS for sensitive data processing combined with local secure data storage.
Customer of the service/tool	Researchers, Research organisations, Research group, Providers
User of the service/tool	Researchers dealing with sensitive data
User Documentation	https://research.csc.fi/pouta-user-guide
Technical Documentation	This is an internal service which is exposed to EOSC via EUDAT services, as such the technical documentation is internal to CSC.
Product team	CSC
License	Scripts for configuration and managing ePouta: N/A (CSC internal code) Secure B2SHARE extension for ePouta: GPLv2
Source code	Scripts for configuration and managing ePouta: CSC internal repository Secure B2SHARE extension for ePouta: https://github.com/secureb2share/secureb2share-b2share

Testing

ePouta platform: The similar community cloud CSC cPouta can be used for testing (with non-sensitive data). cPouta is available through CSC's normal usage policy.

Secure B2SHARE extension for ePouta: Specific test server currently not provided

7.2.2 Release notes

Main development involving ePouta has been done in context of Secure B2SHARE proof of concept.

Secure B2SHARE is a framework based on B2SHARE service, and includes components for storing, describing and authorized sharing of sensitive datasets as well as controlling access to the datasets without jeopardizing privacy or security.

With Secure B2SHARE we aim to demonstrate storage of sensitive data at CSC ePouta connected to B2SHARE service.

In this concept only metadata is stored within normal B2SHARE and the actual sensitive data within ePouta platform; access to sensitive data is restricted, but metadata describing the data is made publicly available to be searched and harvested by metadata catalog services such as B2FIND. All data access can only happen within ePouta secure computing environment.

Note that Secure B2SHARE is also going to be demonstrated with TSD infrastructure.

Following work has been completed:

- Use case to demonstrate the feasibility of storing datasets containing sensitive data to ePouta with the metadata stored in B2SHARE has been selected. Data in question is produced in a study about children's speech disorders in context of Finnish Language Bank (Kielipankki).
- Software architecture definition for ePouta based Secure B2SHARE implementation have been drafted. (For now implementation effort has been focused on TSD based implementation)
- Possibility to connect to ePouta via web-browser based remote desktop connection has been implemented. The Apache Guacamole based remote desktop connection enables authorized users to connect to ePouta with their web-browser. (Previously the only connection option has been a direct layer 2 virtual private network connection to ePouta, which usually is not feasible option for individual researchers.)

7.2.3 Future plans

Future efforts on development of ePouta based Secure B2SHARE implementation will focus on the following activities:

- CSC's REMS (Resource Entitlement Management System) will be integrated with Secure B2SHARE in order to enable requesting access to a sensitive data through B2SHARE user interface. REMS will be used for authorization management (review, allow, reject access requests) by the dataset owner.
- Integration of CSC's AAI system to Secure B2SHARE as an authentication mechanism. This is a central part of the whole workflow of requesting access to sensitive data and allowing authorized access to ePouta through remote desktop connection.

8 Summary and Outlook

This deliverable detail a collection of services constituting the second maintenance release of EOSC-hub common services catalogue. An increased degree of integration among the services has been achieved, as requested by the Thematic Services. As an example, the integration of the EGI DataHub with B2HANDLE, B2FIND and B2ACCESS provides bridges among different distributed e-infrastructures. Further integration activities are planned for the final reporting period in order to achieve both a high degree of technical integration as well as seamless user experience provided through a unified EOSC-hub service catalogue.

Additional detailed integration and interoperability efforts will be reported in the subsequent deliverable from WP6, i.e. *D6.5 Final release of common services software*.