



EOOSC-hub

D6.1 First release of common services software

Lead Partner:	CYFRONET
Version:	2
Status:	Under EC review
Dissemination Level:	Public
Document Link:	https://documents.egi.eu/document/3414

Deliverable Abstract

This deliverable provides overview of the release of first maintenance package for common services required by the thematic and specialized services. The package addresses common services in the area of T6.1, 6.2, 6.3 and T6.4 (without B2NOTE). It comprises first software adaptations of the common services that feature horizontal and vertical interoperability.



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

	Name	Partner/Activity	Date
From:	Bartosz Kryza	CYFRONT/WP6	
Moderated by:	Malgorzata Krakowian	EGI Foundation/ WP1	
Reviewed by	Pablo Orviz	IFCA	30 November 2018
Approved by:	AMB		

DOCUMENT LOG

<i>Issue</i>	<i>Date</i>	<i>Comment</i>	<i>Author</i>
v0.1	24/09/2018	Draft table of contents	Bartosz Kryza (CYFRONET), Łukasz Dutka (CYFRONET)
v0.2	25/10/2018	Revised table of contents	Bartosz Kryza (CYFRONET), Johannes Reetz (MPG), Andrea Ceccanti (INFN), Claudia Martens (DKRZ), Mattia D'Antonio (CINECA), Heinrich Widmann (DKRZ), Enol Fernandez (EGI), Catalin Condurache (STFC), Jorge Gomez (LIP), Andrei Tsaregorodtsev (CNRS), German Molto (UPV), Marica Antonacci (INFN), Bartosz Wilk (CYFRONET), Pablo Orviz (IFCA), Tomasz Zok (PCSS), Claudio Cacciari (CINECA), Lukasz Dutka (CYFRONET)
v0.8	10/11/2018	Combined contributions from partners	Bartosz Kryza (CYFRONET)
V0.9	15/11/2018	Added terminology, introduction and summary	Bartosz Kryza (CYFRONET)
v1	9/01/2019	Updated after external review review	Bartosz Kryza (CYFRONET)

v2	10/01/2020	Final version including project review comment	Bartosz Kryza (CYFRONET)
----	------------	--	--------------------------

TERMINOLOGY

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

<i>Terminology/Acronym</i>	<i>Definition</i>
API	Application Programming Interface
CDMI	Cloud Data Management 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
SCIM	System for Cross-domain Identity Management
VM	Virtual Machine
VO	Virtual Organization
WebDAV	Web Distributed Authoring and Versioning
WMS	Workload Management System
YAML	Yet Another Markup Language

Table of Contents

1	Introduction	8
2	Discovery and Access	9
2.1	INDIGO IAM	9
2.1.1	Service description	9
2.1.2	Release notes	10
2.1.3	Future plans	10
2.2	EGI-DataHub	11
2.2.1	Service description	11
2.2.2	Release notes	12
2.2.3	Future plans	13
2.3	B2FIND	13
2.3.1	Service description	13
2.3.2	Release notes	14
2.3.3	Future plans	14
2.4	B2STAGE	15
2.4.1	Service description	15
2.4.2	Release notes	16
2.4.3	Future plans	16
2.5	B2DROP	17
2.5.1	Service description	17
2.5.2	Release notes	18
2.5.3	Future plans	19
3	Federated Compute	20
3.1	EGI Cloud Compute	21
3.1.1	Service description	21
3.1.2	Release notes	23
3.1.2.1	CMD-OS (Components of Cloud Compute for OpenStack):	23

3.1.2.2	AppDB	23
3.1.2.3	EGI images	23
3.1.2.4	Documentation	24
3.1.3	Future plans	24
3.2	EGI Cloud Container	24
3.2.1	Service description	24
3.2.2	Release notes	25
3.2.2.1	EGI images	25
3.2.2.2	grycap.kubernetes ansible role	26
3.2.3	Future plans	26
3.3	EGI Workload Management	26
3.3.1	Service description	26
3.3.2	Release notes	27
3.3.3	Future plans	28
3.4	EGI Online Storage	28
3.4.1	Service description	28
3.4.2	Release notes	29
3.4.2.1	StoRM v. 1.11.13	29
3.4.3	Future plans	29
3.5	Advanced IaaS	30
3.5.1	Service description	30
3.5.2	Release notes	31
3.5.3	Future plans	32
3.6	EGI High-Throughput Compute	32
3.6.1	Service description	32
3.6.2	Release notes	33
3.6.3	Future plans	33
3.7	CVMFS	34
3.7.1	Service description	34
3.7.2	Release notes	35
3.7.3	Future plans	35
4	Processing and orchestration	36

4.1	TOSCA for Heat	38
4.1.1	Service description	38
4.1.2	Release notes	39
4.1.3	Future plans	39
4.2	Infrastructure Manager	39
4.2.1	Service description	39
4.2.2	Release notes	41
4.2.3	Future plans	41
4.3	PaaS Orchestrator	41
4.3.1	Service description	41
4.3.2	Release notes	44
4.3.3	Future plans	44
4.4	Future Gateway	44
4.4.1	Service description	44
4.4.2	Release notes	46
4.4.3	Future plans	46
5	Data and Metadata Management	47
5.1	B2HANDLE	47
5.1.1	Service description	47
5.1.2	Release notes	49
5.1.3	Future plans	50
5.2	B2SAFE	50
5.2.1	Service description	50
5.2.2	Release notes	52
5.2.3	Future plans	52
5.3	B2SHARE	52
5.3.1	Service description	52
5.3.2	Release notes	53
5.3.3	Future plans	54
6	Summary and Outlook	55
Appendix I.	Software Repositories	56

Executive summary

This document contains a report on the first maintenance release of common services from EOSC-hub service catalogue, which were released in the first year of the project. The common services in EOSC-hub provide middleware layer enabling thematic services to efficiently use the resources provided by main European infrastructures. The document lists the service in 4 categories, reflecting the different tasks of WP6 i.e. Discovery and Access (T6.1), Federated Compute (T6.2), Processing and Orchestration (T6.3) and Data and Metadata Management (T6.4). Other WP6 tasks are not included as their first software releases are planned for the next year of the project. 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. Where relevant, information on future plans in the scope of EOSC-hub project is included.

1 Introduction

This document presents a portfolio of EOSC-hub services, focusing on reporting the status of the services, valid references to documentation and code repositories as well as brief reports on their current release major features and future plans, which can be interesting to EOSC-hub user communities.

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 and orchestration. The coordination of integration and maintenance activities of these services is the main goal of WP6 of the EOSC-hub project.

The main effort in the first year of the project was focused on solving the most important issues related to integration of services such as enabling authentication and authorization mechanisms between EGI and EUDAT platforms, unifying application interfaces, enabling data transfers and access among different services and improving services documentation. Thus, this deliverable provides initial release and integration notes about the services with a new release in the first year of the EOSC-hub project along with description of developments and improvements planned for the remaining part of the project.

Future updates on the services will be publicly available in the EOSC-hub service catalogue¹.

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 and section 5 lists the services in the data and data management category.

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 use data resources within EOSC-hub for their own purposes.

Regarding data discovery the metadata service B2FIND is intended to play the role of the central search indexing tool of EOSC-hub. To that end, the service is extended and enhanced to cover as well data from 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 (<http://b2find.eudat.eu/guidelines/mapping.html#b2fmdschema>) and stabilizing metadata ingestion of B2SHARE records
- An OAI-PMH provider on top of EGI-DataHub, a service for publishing reference open data set, has been set up to enable B2FIND to harvest metadata from there
- Progress is made with the concept of metadata extraction from B2SAFE (iRODS) data collections and indexing these datasets in B2FIND
- In parallel B2FIND continued the uptake of internal and external research communities and data providers

Regarding data access we elaborated and assessed the possibilities and issues of seamlessly accessing data collections in the storage services EGI-DataHub and B2SAFE, as well as overcoming AAI issues using INDIGO IAM (Identity and Access Management), which is already integrated with EGI-DataHub.

2.1 INDIGO IAM

2.1.1 Service description

Service/Tool name	INDIGO IAM
Service/Tool url	https://github.com/indigo-iam/iam
Service/Tool information page	https://www.indigo-datacloud.eu/identity-and-access-management
Description	The INDIGO IAM (Identity and Access Management) service provides user identity and policy information to services so that consistent authorization decisions can be enforced across distributed services.

Value proposition	<ul style="list-style-type: none"> ● OpenID connect provider based on the MitreID OpenID connect library ● SCIM user provisioning and management APIs ● SAML authentication support ● Google authentication support ● OAuth token exchange support
User of the service/tool	Users who need to translate credentials from different infrastructures and different authentication protocols.
User Documentation	https://indigo-iam.github.io/docs/v/current/user-guide
Technical Documentation	https://indigo-iam.github.io/docs/v/current/admin-guide
Product team	INFN
License	Apache License, Version 2.0
Source code	https://github.com/indigo-iam/iam
Testing	Internal continuous integration test suite: https://travis-ci.org/indigo-iam/iam

2.1.2 Release notes

Current version of IAM is 1.4.0, which features:

- New paginated user and group search API
- Support for login hint on authorization requests: this feature allows a relying party to specify a preference on which external SAML IdP should be used for authentication
- Documentation for the IAM group request API
- New group membership requests API: this API allows user to submit requests for membership in groups, and provide administrators the ability to approve/reject such requests

Complete changelog for all versions can be found at: <https://github.com/indigo-iam/iam/blob/master/CHANGELOG.md>

2.1.3 Future plans

Integration with selected EOSC-hub services and other EOSC-hub AAI solutions is planned.

2.2 EGI-DataHub

2.2.1 Service description

Service/Tool name	EGI DataHub
Service/Tool url	https://datahub.egi.eu
Service/Tool information page	https://onedata.org
Description	<p>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 and OpenStack SWIFT and provides to clients interfaces based on CDMI, REST API and virtually mounted POSIX filesystem. Onedata architecture comprises of 2 major components: Oneprovider and Oneclient. The former is installed within data center and provides a unified interface to multiple filesystems. 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 the 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.

	<ul style="list-style-type: none"> • 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 EOSC-hub B2ACCESS • 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 undergoes continuous integration and functional testing, comprised of over 500 test cases.

2.2.2 Release notes

Current EGI-DataHub release is based on Onedata release 18.02.0-rc13, which includes the following features:

- Universal configuration file for adding new OpenID Connect and SAML Identity Providers
- Automatic Let's Encrypt support when registering new Oneprovider deployments
- Initial integration with B2FIND for automatic PID assignment for publishing open data sets
- Improved querying based on spatial metadata

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

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

2.2.3 Future plans

The future plans in the context of EOSC-hub project are mainly related to integration of various methods of data transfer between EGI-DataHub and EUDAT services (e.g. B2STAGE and B2SHARE). Initial version of WebDAV driver for Onedata is being evaluated for this purpose.

2.3 B2FIND

2.3.1 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> <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.</p>

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 any user interested in searching for open data sets.
User Documentation	https://eudat.eu/services/userdoc/b2find http://b2find.eudat.eu/guidelines/introduction.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://eudat.eu/services/userdoc/b2find-integration
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	<p>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. It also offers a trainings instance:</p> <p>https://github.com/EUDAT-Training/B2FIND-Training</p>

2.3.2 Release notes

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>

2.3.3 Future plans

Integration with selected EOSC-hub services is planned.

2.4 B2STAGE

2.4.1 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>GridFTP (via the EUDAT Data Storage Interface) is a service aimed at 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 result of the Safe Replication activity. Eventually, output data are identified through a Persistent Identifier (PID).</p> <p>The HTTP API service is a set of RESTful endpoints that allow accessing to both B2SAFE data and metadata and it is aimed for 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 GridFTP: Apache License 2.0
Source code	https://github.com/EUDAT-B2STAGE/B2STAGE-GridFTP https://github.com/EUDAT-B2STAGE/http-api
Testing	B2STAGE http-api can be tested by accessing the online prototype available at https://b2stage-test.cineca.it/api/status . Prototype dedicated instructions: https://github.com/EUDAT-B2STAGE/http-api/blob/master/docs/prototype.md

2.4.2 Release notes

The current B2STAGE HTTP-API release (1.0.3) is mainly a maintenance release with several BUG fixes and the integration of the latest RPyDo version (the python framework on which HTTP-APIs are based, currently at version 0.6.2). This release also improves the support for communities already adopting B2STAGE HTTP-API.

Detailed changelog for this component can be found at:

<https://github.com/EUDAT-B2STAGE/http-api/blob/master/CHANGELOG.md>

2.4.3 Future plans

The next release of B2STAGE-http (1.0.4) will introduce the support for user authentication through that PAM protocol. This release will integrate contributions implemented on the official python iRODS client (<https://github.com/irods/python-irodsclient>) previously lacking the support of PAM authentication. The introduction of PAM will allow the complete integration of B2ACCESS OAuth2-

based authentication workflow, which in turn will allow an easier integration with other EUDAT services.

2.5 B2DROP

2.5.1 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 background. The user finishes the draft on B2SHARE and receives a PID for the publication.</p>
Value proposition	<ul style="list-style-type: none"> ● Sharing a file to 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 (AGPLv3)
Source code	https://github.com/EUDAT-B2DROP
Testing	Internal continuous integration test suite.

2.5.2 Release notes

For this release, i.e. 12.0.4, B2DROP has harmonized the user authentication and switched from LDAP back end, which provided local B2DROP authentication, to B2ACCESS authentication. The B2ACCESS authentication allows users to use the same account for B2DROP that is already used for other services. On the other hand, the integration with other services is simplified with a common authentication mechanism. As a first part of this change B2DROP added a second authentication mechanism at the beginning of the year. Since this integration of B2DROP and B2ACCESS, new users could only authenticate via B2ACCESS. Existing users could still use their old accounts. In July 2018, the migration of old LDAP accounts to new B2ACCESS accounts was started and the end of life for LDAP authentication was announced for 1st of October. With this announcement the users received a small guide to create new accounts and additional support was offered. A migration without action of the users was not possible. To avoid a broken system for users, who did not finish the migration in time, the LDAP authentication was turned off for the web frontend on 1st of October. Sync clients were still able to authenticate via LDAP. We investigated which users were still using the outdated authentication mechanism for their sync clients and informed them about the end of life for this client authentication in the middle of October. We sent a guide to use the sync clients with B2ACCESS authenticated accounts. In mid-October, the old LDAP authentication mechanism was turned off. The migration of user data is still possible and there is no loss of data.

In July, Nextcloud, the underlying software, was updated to major version 13. This update harmonised the various authentication methods on the web UI. Previously, the Shibboleth integration required adaptations to allow for additional authentication mechanisms to be used. This functionality is now covered by Nextcloud itself, adaptations are no longer required.

2.5.3 Future plans

In the reporting period the integration of B2STAGE with B2DROP was discussed. During the discussion and investigation of the possibilities, we come to the result, that the direct integration of B2SAFE with B2DROP is a better integration. We switched from an integration of B2STAGE to an integration of B2SAFE with B2DROP. The exchange with potential users shows that the benefit is limited because of the different data sizes, which are supported by B2DROP and B2SAFE. But even with the limited effort for some users we investigated how the integration of the two services could be handled. With the extension of B2SAFE to support the WebDAV protocol the integration could be doable. The implementation of this approach must be investigated in the future. Even with the issue of supported storage technology being solved, some challenges in authentication and PID management need to be solved before the integration can be evaluated by a wider audience.

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 the following table. Users needing tighter control on the resources and how these are allocated 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 traditional jobs in a batch system, High Throughput Compute will better meet their needs.

	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

Table 1. Federated Compute services

The Workload Manager (formerly known as DIRAC4EGI) service 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 3 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

- 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 https://github.com/EGI-Foundation/fedcloud-integration-documentation
Product team	Partners involved: CESNET, CSIC, EGI.eu, GRNET, IASA, SRCE
License	Apache 2.0
Source code	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 are verified against EGI's current Quality Criteria (https://wiki.egi.eu/wiki/EGI_Quality_Assurance). Reports are available in EGI document DB under "Software Provisioning and Technology" category (https://documents.egi.eu/public/ListBy?topicid=35). Installation and configuration notes are available in the EGI repository.</p>

3.1.2 Release notes

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

CMD-OS 1.2.1 (CMD for OpenStack) includes a fix for cloudkeeper-os for CentOS7, and a new release of cASO for CentOS7 and Ubuntu.

- cloudkeeper-os 0.9.10 - fix dependency issue on CentOS7 (only)
- cASO 1.2.0 - a bug was introduced in how records are retrieved with the introduction of the "deleted" filter; this release fixes the bug, sites will need to regenerate accounting records; see cASO release notes for more details.

Additional details available in CMD-OS-1.2.1 wiki²

3.1.2.2 *AppDB*

Several subcomponents of AppDB were updated during this period. This list shows a brief overview of the main changes introduced:

- VMOps:
 - Allow to file a GGUS ticket directly to the site
 - Request and release public IP addresses
 - Attach or remove block storage at any point in the lifecycle of VMs
 - Generation of daily reports for the VMops service
 - Added support for new VOs (see complete list at AppDB wiki³)
- Cloud Marketplace:
 - Integration of AppDB with the Security Cloud Assessment Tool (SECANT)
 - Improved UX of VO wide image list panel
 - Continuous delivery mechanism for VM authors

3.1.2.3 *EGI images*

The following EGI-provided images in AppDB were updated to latest versions provided by upstream distributions fixing security issues:

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

² <https://wiki.egi.eu/wiki/CMD-OS-1:CMD-OS-1.2.1>

³ https://wiki.appdb.egi.eu/main:faq:which_vos_are_supported_by_the_vmops_dashboard

The appliance for integration of OpenStack providers was also updated to cover the latest version of integration tools available in the infrastructure:

- [FedCloud integration appliance for OpenStack](#)

Monitoring

- Probes were updated to support OpenID Connect across all providers.
- Update of the set of probes used for calculation of A/R metrics to avoid issues appearing in latest versions of OpenNebula
- Various fixes for OpenStack probes to improve native-API-only sites

3.1.2.4 Documentation

Moved documentation on integration to GitHub and readthedocs:

- <https://github.com/EGI-Foundation/fedcloud-integration-documentation>
- <https://egi-federated-cloud-integration.readthedocs.io>

Detailed changelog of documentation is available in the git log⁴

3.1.3 Future plans

- Release of CMD-ONE update with components for OpenNebula
- Complete transition to OpenID Connect at all providers
- Implementation of GlueSchema 2.1 for information discovery and transition to message-based transport
- Improved support for providers with less privileged access to the cloud management platforms

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	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

⁴ <https://github.com/EGI-Foundation/fedcloud-integration-documentation/commits/master>

	<p>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 ● 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	Travis integration for Kubernetes: https://travis-ci.org/grycap/ansible-role-kubernetes

3.2.2 Release notes

3.2.2.1 EGI images

The base Docker image in AppDB includes the following changes:

- Upgrade to docker v17.03.2

- Upgrade to docker-compose 1.21.2
- Addition of Kubelet, Kubectl and Kubeadm v1.11.2

Published in AppDB as EGI docker Ubuntu 16.04

(<https://appdb.egi.eu/store/vappliance/egi.docker.ubuntu.16.04>)

3.2.2.2 *grycap.kubernetes ansible role*

- Fix a bug related to Docker installation in RedHat systems.
- Enable the user to specify the kubernetes version to install (or latest).
- Fix an error in case of firewalld not installed in RedHat systems
- Fixed some error not enabling to detect nodes as idle. It causes that the nodes are never powered off.

3.2.3 Future plans

- Provision of Container Repositories
- Automatic provisioning of the Kubernetes on the infrastructure
- Automatic configuration of OpenID Connect integration

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)

	<ul style="list-style-type: none"> • 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 dirac.egi.eu service • Provide more flexible VRE simplifying access to EGI resources and accelerate the ability of researches 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

Detailed release notes for DIRAC4EGI can be found on the DIRACGrid page:

<https://github.com/DIRACGrid/DIRAC/wiki#release-information>

Current EOSC-hub release of this component is *v6r19p16*.

3.3.3 Future plans

Integration with selected EOSC-hub services is planned.

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

TestingQuality Criteria Report: <https://documents.egi.eu/document/3335>StageRollout Report: <https://documents.egi.eu/document/3335>

3.4.2 Release notes

3.4.2.1 StoRM v. 1.11.13

Available in EGI UMD on 17.07.2018. This release provides fixes to some outstanding bugs:

- fixes a problem on status update for tape recalls created through the REST endpoint
- fixes pool account mapping failures observed when StoRM is deployed with lcmapi-plugins-basic >= 1.6.3
- enhances the request garbage collector so that PrepareToPut requests that are stuck in the state SRM_REQUEST_INPROGRESS are automatically expired after a configurable amount of time
- fixes a bug in the garbage collector so that now only recall requests older than a configurable amount of time are garbage collected
- adds the ability to generate a storage usage JSON report, following the rules and format defined by WLCG

Bug fixes:

- Failure on updating recall task status
- Pool account mapping fails since lcmapi-plugins-basic v1.6.4
- Garbage Collector ignore timestamps on cleaning recall tasks

Enhancements:

- StoRM backend should garbage collect requests that are stuck in SRM_IN_PROGRESS for a configurable amount of time
- JSON storage usage reporting

Detailed release notes can be found on the StoRM page:

<http://italiangrid.github.io/storm/2018/02/19/storm-v1.11.13-released.html>

3.4.3 Future plans

Integration with selected EOSC-hub services is planned.

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://www.indigo-datacloud.eu/userspace-container-support
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 standard directory tar archives 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 the 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 other 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

	<ul style="list-style-type: none"> • Enables execution of applications with GPGPUs. • Empower users to execute their applications everywhere through container encapsulation
User of the service/tool	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.
User Documentation	https://github.com/indigo-dc/udocker/blob/master/doc/user_manual.md
Technical Documentation	https://github.com/indigo-dc/udocker/blob/master/SUMMARY.md
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

Detailed udocker changelog can be found at:

<https://github.com/indigo-dc/udocker/blob/master/changelog>

3.5.3 Future plans

Integration with selected EOSC-hub services. udocker is currently being integrated in the OPENCoastS thematic service. Address limitations in the mounting of host directories, improve the Docker compatibility, improve the configuration mechanism, enlarge the test coverage, support for OCI images, maintain udocker to cope with the evolution of new and existing the Linux distributions.

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>
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 take 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 will be certified.

3.6.2 Release notes

The CREAM service has been declared the main Computing Element for the grid environment by the European Grid Infrastructure (EGI). The software must be considered completely stable, no more features are planned. It has been deployed and operated in the grid environment for more than ten years. The maintenance in the recent past was mainly dedicated to the improvements of the security infrastructure, for example with the adoption of new cryptographic schemes, and changes required for keeping the compatibility with the latest releases of the batch systems.

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 puppetforge portal: <https://forge.puppet.com/infnpd/creamce>

3.6.3 Future plans

Integration with selected EOSC-hub services is planned.

3.7 CVMFS

3.7.1 Service description

Service/Tool name	CVMFS
Service/Tool url	https://cernvm.cern.ch/portal/filesystem
Service/Tool information page	https://cernvm.cern.ch/portal/filesystem
Description	<p>The CernVM File System (CernVM-FS) is a read-only file system designed to deliver scientific software onto virtual machines and physical worker nodes in a fast, scalable, and reliable way. Files and file metadata are downloaded on demand and aggressively cached.</p> <p>For the LHC experiments, CVMFS hosts several hundred million files and directories that are distributed to the order of hundred thousand client computers.</p>
Value proposition	Replaces package managers and shared software areas on cluster file systems as means to distribute the software used to process experiment data
User of the service/tool	CVMFS is actively used by small and large HEP and non-HEP collaborations working in space, natural and life Sciences).
User Documentation	https://cvmfs.readthedocs.io/en/stable/cpt-quickstart.html
Technical Documentation	https://cvmfs.readthedocs.io/en/stable/part-advanced.html
Product team	CERN
License	<p>BSD 3-Clause "New" or "Revised" License</p> <p>https://github.com/cvmfs/cvmfs/blob/devel/COPYING</p>
Source code	https://github.com/cvmfs/cvmfs

Testing

Internal continuous integration test suite: <https://travis-ci.org/cvmfs/cvmfs>

3.7.2 Release notes

The detailed changelog for CVMFS is available at:

<https://cvmfs.readthedocs.io/en/stable/cpt-releasenotes.html>

3.7.3 Future plans

CVMFS will be integrated with other selected services in EOSC-hub.

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 building complex virtual computing infrastructures, based on the OASIS standard TOSCA Simple Profile YAML Version 1.0 standard⁵ 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.

⁵ 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.

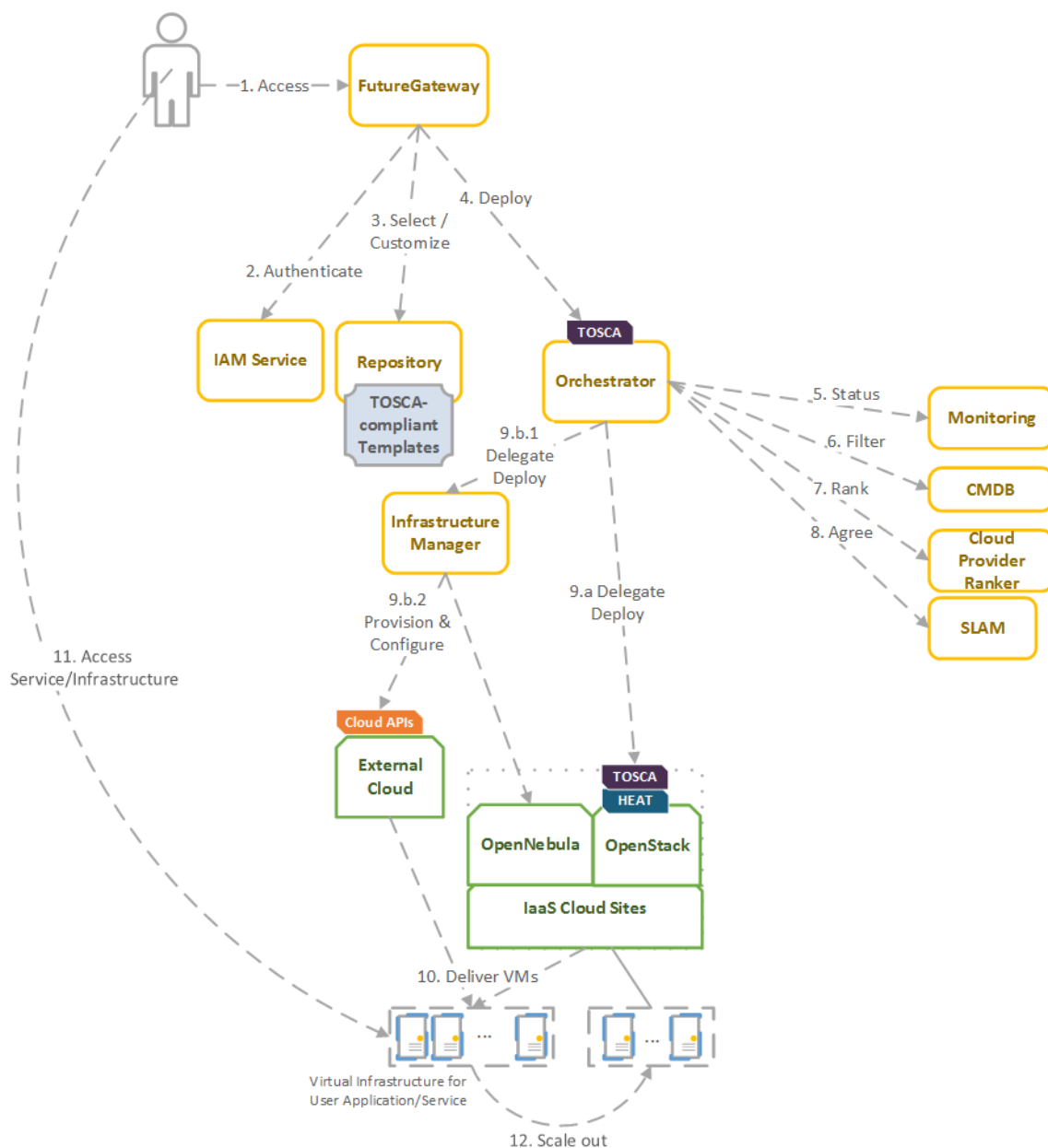


Figure 1. Architecture and workflow among the Processing and Orchestration services in EOSC-hub.

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 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 (IM)* 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.

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	<p>TOSCA for Heat service translates TOSCA templates in order to produce valid HOT (Heat Orchestration Template)-compliant 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 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

Detailed changelog can be found at:

<https://github.com/indigo-dc/heat-translator/commits/devel>

4.1.3 Future plans

Integration with selected EOSC-hub services is planned.

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 public IaaS Clouds (Amazon Web Services, Microsoft Azure, etc.), on-premises CMPs (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 back-ends.
Customer of the service/tool	Research Communities, VREs, and 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	GRCAP - UPV
License	GPL 3.0
Source code	https://github.com/indigo-dc/im
Testing	https://jenkins.indigo-datacloud.eu:8080/job/Pipeline-as-code/job/im/job/master/

4.2.2 Release notes

The IM has been evolved in the framework of the EOSC-Hub project. The first step was to test the authentication systems provided in the project. IM was already integrated with the INDIGO IAM (based on OpenID) and it has been successfully tested with EGI Check-In system (also based on OpenID) without any code modification.

As requested by the DODAS Thematic service from WP7 it has been added the support for EC2 spot instances. It required adding the `preemptible_instance` property in the TOSCA compute node definition and the proper translation in the IM orchestrator core.

Also the Exoscale provider (CloudStack API) has been added as requested by the EGI Applications on Demand service to access this platform as part of collaboration with the project HNSciCloud.

Another extension made to the IM is the ability to use SSH reverses tunnels to connect with VMs that only feature private IP addressing in different Cloud providers. This enables to contextualize hybrid deployments that are virtual infrastructures deployed across multiple IaaS Cloud sites, using only one public IP per infrastructure.

4.2.3 Future plans

Integration with selected EOSC-hub services is planned.

4.3 PaaS Orchestrator

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: it allows to coordinate the provisioning of virtualized compute and storage resources on CMPs, 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 single or hybrid deployments on the best available cloud sites. Using the PaaS Orchestrator and the TOSCA templates, the end users can exploit computational resources without any knowledge about the IaaS details.</p> <p>In order to select the best site, the Orchestrator implements a complex workflow: it gathers information about the SLAs signed by the providers with the user, the monitoring data about the availability of the compute and storage services and the location of the data requested by the user (if any). Hybrid deployments spanning multiple sites are supported.</p> <p>Several components are involved in the operation of the PaaS Orchestrator:</p> <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 managing 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.</p> <p>SLAM (SLA manager) is a component of the INDIGO PaaS system that handles negotiations between infrastructure customers and infrastructure providers. <i>Customer</i> 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. <i>Provider</i> is an entity delivering computing and</p>

	<p>storage resources for customers, represented by systems users with provider role given.</p> <p>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.</p> <p>SLA draft is created by customer representative and subjected to provider's approval. 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 reflecting 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.</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 https://github.com/indigo-dc/cmdb https://github.com/indigo-dc/slam

Testing

<https://jenkins.indigo-datacloud.eu:8080/job/Pipeline-as-code/job/orchestrator>

4.3.2 Release notes

Detailed changelog for the PaaS Orchestrator components can be found at:

Orchestrator - <https://github.com/indigo-dc/orchestrator/blob/master/CHANGELOG.md>

CMDB - <https://github.com/indigo-dc/cmdb/blob/master/CHANGELOG.md>

4.3.3 Future plans

In the scope of EOSC-hub project CMDB component will be reengineered in order to adapt it to the EOSC-hub ecosystem. As most of information needed for the PaaS to operate is already provided by AppDB system, which was made available for the project by EGI, AppDB will be treated as the main source of information for the INDIGO PaaS CMDB system.

4.4 Future Gateway**4.4.1 Service description**

Service/Tool name	Future Gateway
Service/Tool url	https://www.indigo-datacloud.eu/future-gateways-programmable-scientific-portal
Service/Tool information page	https://github.com/FutureGatewayFramework/fgDocumentation
Description	FutureGateway is comprised of a set of components enabling end-users to access functionalities such as authentication/authorization or deployment management. It exposes a REST interface to a service called fgAPIServer. The API 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 enough 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 VM deployment in private clouds.
Value proposition	An easy to use frontend for complex backend VM 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>Internal continuous integration test suite: https://github.com/FutureGatewayFramework/fgAPIServer/tree/master/tests</p> <p>The tzok/eosc-fgapiserver is an Ansible Container project which is used to build a minimal Docker image with fgAPIServer component of FutureGateway. The tzok/fg-docker-compose has been added to link a database engine with it in a common way using Docker Compose. These two new elements in EOSC-hub allow quickly building and testing any released version of fgAPIServer component.</p>

4.4.2 Release notes

The **indigo-dc/indigoclient** and **indigo-dc/indigokepler** repositories are now managed jointly as indigo-dc/indigo-parent project. This allows to share dependency and plugin management between two projects and also to streamline the release procedure. This method was used to release v1.4 of both components.

4.4.3 Future plans

- Create another Ansible Container project to be able to build APIServerDaemon component of FutureGateway and run it in an isolated Tomcat instance
- Configure it to work with Docker Compose configuration
- Support Kubernetes configuration for dockerized components
- Implement a thorough testing procedure for each component as well as for the full service

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 (data associated with PIDs), are described in detail in the following paragraphs and shown in the Figure 2.

Those services allow storing a data set in a repository, which is geographically distributed, and associated a PID to it, making the data set location independent from the references pointing to it. The PID is globally resolvable and the data set is replicated in multiple copies, which are tracked in the metadata associated to the identifier. Data can be published, and community specific metadata associated to it, then those metadata can be harvested and indexed by a discovery service to make the data findable. Data can also be annotated, manually or programmatically via an API. And last, but not least, data are curated through a set of policies that each data manager can define.

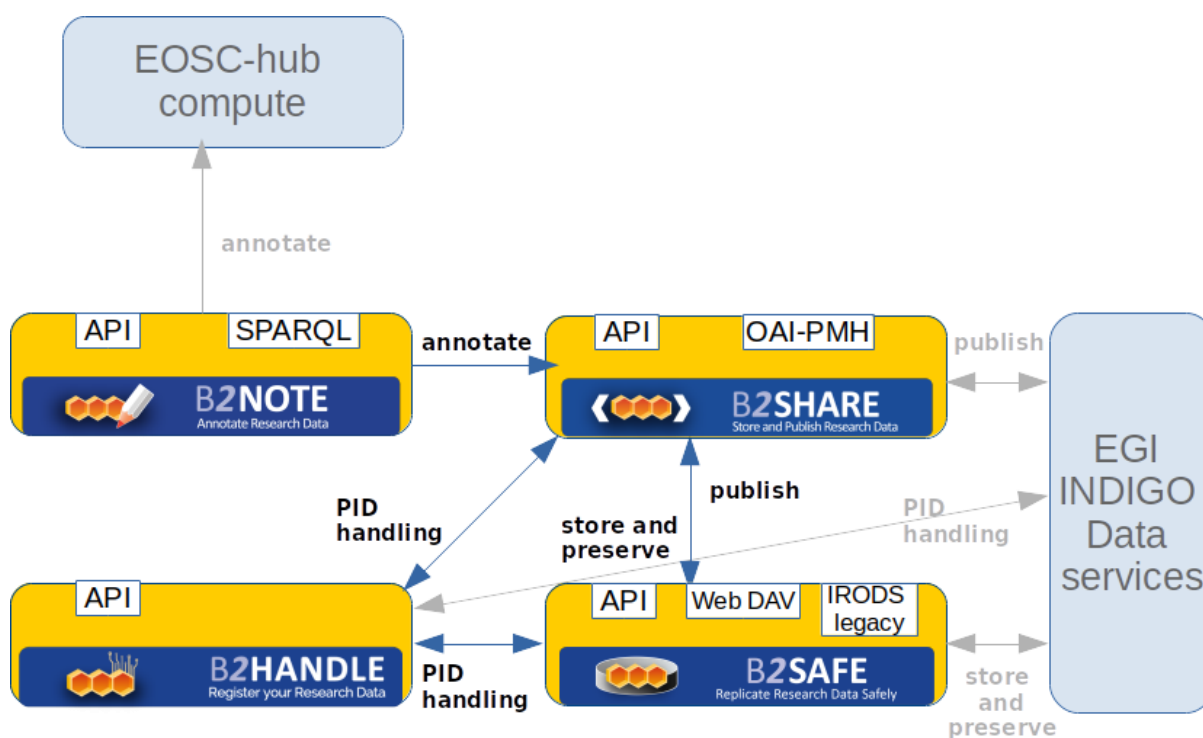


Figure 2. Data and metadata management services

5.1 B2HANDLE

5.1.1 Service description

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

Service/Tool url	Current nodes visible at https://dp.eudat.eu/operations/B2HANDLE
Service/Tool information page	https://eudat.eu/services/b2handle
Description	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,). Thereby provide easy and 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 Catalogue, 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.
Value proposition	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 Catalogue and the B2HANDLE support libraries can facilitate efficient automated data management at early stages of the data life cycle.
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	<p>Continuous integration testing is configured for B2HANDLE via Jenkins (GRNET).</p> <p>B2HANDLE has also updated its monitoring facilities to be compliant with EOSC-hub operational requirements. Probes were deployed at the B2HANDLE provider sites that do not only test regular availability, but also emulate the typical actions (create, read, update, delete) on Handles. In addition, the B2HANDLE Central PID Catalogue is also monitored.</p>

5.1.2 Release notes

B2HANDLE has started adaptation of its operational processes to be in line with FitSM procedures and best practices.

- Changes to the operational setup of B2HANDLE are managed in coordination with EOSC-hub and EUDAT **Change Management**.
- Releases of new B2HANDLE software components will be coordinated with EOSC-hub **Release and Deployment Management**.
- The specification of configuration items of B2HANDLE is maintained as part of EUDAT **Configuration Management** using the EUDAT DPMT.
- A first description of B2HANDLE has been included in the EOSC-hub service portfolio/marketplace (**Service Portfolio Management**).
- The B2HANDLE operations team has also started to remodel internal workflows and operational status pages to support **Customer Relationship Management**.
- B2HANDLE service monitoring is currently being updated by the monitoring team to accommodate EOSC-hub requirements.

B2HANDLE integration activities are two-fold: Integration and maintenance to keep the service in line with EOSC-hub requirements; and integration with additional EOSC components.

To accommodate requirements of EOSC-hub, B2HANDLE has improved some of its components already. Most importantly, an update to the key component for the Central PID Catalogue, the Handle Reverse-Lookup Servlet (HRLS), has been released (v1.0.4) to accommodate requirements for metric measurement as part of virtual access and extended monitoring. Also, the component was updated to be compatible with a recent release of the Handle System component (v9.0).

5.1.3 Future plans

Concerning integration with additional EOSC components, B2HANDLE has taken the following steps:

1. Integration with B2SHARE involves use of PID profiles and reverse-lookups, going beyond existing integration (use of identifiers only). Use of PID profiles by B2SHARE was discussed and it was agreed that an initial profile would be developed by B2HANDLE to be reviewed together with B2SHARE against a few typical use cases. The reverse-lookup service and extended filtering capabilities will be discussed as a follow-on action once profiles are established.
2. Integration with the EGI-DataHub was discussed with the EGI-DataHub team, particularly concerning the integration at the technical level. The available interfaces of B2HANDLE were explained and it was agreed that the best way forward is for the EGI-DataHub team to start integration based on a test identifier namespace (test prefix), which can be provided by B2HANDLE, followed by integration action by EGI-DataHub against the native CRUD interface of B2HANDLE. Activities on this are on-going.
3. Integration with EGI Online Storage was discussed with EGI. Further investigation on concrete steps to take is on-going, pending further clarification of the scope of the EGI Online Storage service.
4. Integration with the EGI Federated Data Manager was discussed with EGI. The discussion indicated that the Federated Data Manager is based on Onedata, which is the same solution powering the EGI-DataHub. Integration with Federated Data Manager is thus likely achieved as part of EGI-DataHub integration activities.
5. B2HANDLE has participated in a EUDAT-led workshop to observe and plan for future improvements, which can be fed into **Continual Service Improvement**.

5.2 B2SAFE

5.2.1 Service description

Service/Tool name	B2SAFE
Service/Tool url	See the number of instances listed in the status report here: http://avail.eudat.eu/lavoisier/status_report-site?accept=html

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>Service core: it offers functionality for the longterm 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 easily finding and querying replica locations. The information about the replica locations and other important information is stored in a PID registry (B2HANDLE). The B2SAFE Service core 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 enables 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 (e.g. when data is ingested, stored and must be curated or transferred).</p>
Value proposition	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 who need to guard against data loss.
User Documentation	<p>https://eudat.eu/services/userdoc/b2safe</p> <p>https://eudat.eu/services/userdoc/configure-b2safe</p> <p>http://eudat.eu/services/userdoc/using-b2safe</p> <p>https://eudat.eu/services/userdoc/joining-b2safe</p>

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	CSC
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

Detailed release information can be found at:

<https://github.com/EUDAT-B2SAFE/B2SAFE-core/releases>

5.2.3 Future plans

Next release expected before the end of the year (2019/01/01). It will be v4.2.1 and it will include some relevant features like the integration with B2ACCESS oauth2 protocol and the support to http endpoints.

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.
Value proposition	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 a easy way to share it in the research community.
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
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

Release notes available in GitHub:

<https://github.com/EUDAT-B2SHARE/b2share/releases>

<https://github.com/EUDAT-B2SHARE/b2share/blob/master/CHANGELOG.rst>

5.3.3 Future plans

- Implement more record metadata exporters (e.g. support for exporting metadata as Datacite XML)
- Make necessary changes and additions to B2SHARE to enable harvesting B2SHARE metadata to OpenAIRE RCD.
- Further development of B2SHARE to enable and support new communities to start using B2SHARE.
- Improve integration with B2HANDLE service for displaying PID metadata in B2SHARE.
- Improve integration with B2NOTE.
- Integrate with suitable, non-EUDAT services that are part of EOSC-hub service catalogue. Possibly, for example, with EGI-DataHub and EGI Online Storage.

6 Summary and Outlook

In this deliverable we have presented a collection of services constituting the first maintenance release of EOSC-hub common services catalogue. Since the services come from 2 different e-Infrastructures namely EGI and EUDAT, their interoperability is not yet complete, and will be the focus of the next year of the project. The goal of the integration will be to both functional i.e. enabling thematic services users to leverage resources from both infrastructures as well as interoperability on the user experience level, provided through unified EOSC-hub service catalogue.

More detailed integration and interoperability efforts will be reported in the consecutive deliverable from WP6, i.e. *D6.2 First report on the maintenance and integration of common services*.

Appendix I. Software Repositories

Modern open source software projects make extensive use of code repositories to openly expose their projects to their communities. Likewise, EOSC-Hub and the infrastructure projects that make up EOSC-Hub are strongly committed to openness regarding their solutions and development processes. When maintaining software repositories the EOSC-Hub project requests that providers follow best practices including, providing well defined and documented software releases, change logs etc. However, since the EOSC-Hub is reliant on multiple providers it is not feasible to enforce complete uniformity across all repos. Each provider is also requested to acknowledge the support of the EOSC-Hub project within their software repositories.