



EOSC-hub

D7.4 Second Thematic Service software release

Lead Partner:	CINECA
Version:	1
Status:	Under EC review
Dissemination Level:	Public
Document Link:	https://documents.egi.eu/document/3642

Deliverable Abstract

This second release includes all the services, which have been integrated so far with EOSC-hub. The integration happens at various levels: from the usage of cloud resources (IaaS, PaaS), to the usage of EOSC-hub common services (for example data and computing services), to the usage of EOSC-hub core services, like the authentication or the monitoring tools, to the publication of the services in the Service Catalogue or in the Marketplace.



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:	Debora Testi	CINECA/WP7	
Moderated by:	Małgorzata Krakowian	EGI Foundation/WP1	
Reviewed by:	Gergely Sipos Hien Bui	EGI Foundation/WP8 EGI Foundation/WP1	
Approved by:	AMB		

DOCUMENT LOG

<i>Issue</i>	<i>Date</i>	<i>Comment</i>	<i>Author</i>
v.0.1	25/05/2020	ToC sent to partners	Debora Testi
v.0.20	25/06/2020	All inputs from received	Dieter Van Uytvanck, Willem Elbers, Daniele Spiga, Sandro Fiore, Fabrizio Antonio, Mattia Santoro, Anabela Oliveira, Alberto Azevedo, Mário David, Alexandre Bonvin, Christian Brieze, Davor Davidovic, Sheila Izquieta Rojano
v.1.0	26/06/2020	Final version completed	Debora Testi
v.2.0	24/08/2020	Revised version circulated with comments from reviewers	Alexandre Bonvin, Willem Elbers, Sandro Fiore, Daniele Spiga, Anabela Oliveira
	28/08/2020	Final version	Debora Testi

TERMINOLOGY

The EOSC-hub glossary of terms is available at:

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

Contents

1	Introduction	6
2	CLARIN	7
2.1	Service description	7
2.2	Service architecture and integrated services	9
2.3	Software release.....	9
2.4	References.....	10
2.5	EOSC-hub Integration	11
3	DODAS	12
3.1	Service description	12
3.2	Service architecture and integrated services	13
3.3	Software release.....	18
3.4	References.....	18
3.5	EOSC-hub integration	18
4	ECAS.....	19
4.1	Service description	19
4.2	Service architecture and integrated services	20
4.3	Software release.....	29
4.4	References.....	30
4.5	EOSC-hub integration	30
5	GEOSS	31
5.1	Service description	31
5.2	Service architecture and integrated services	32
5.3	Software release.....	32
5.4	References.....	32
5.5	EOSC-hub integration	32
6	OPENCoastS	33
6.1	Service description	33
6.2	Service architecture and integrated services	34
6.3	Software release.....	37
6.4	EOSC-hub integration	38
7	WeNMR	39

7.1	Service description	39
7.2	Service architecture and integrated services	40
7.3	Software release.....	43
7.4	References.....	44
7.5	EOSC-hub integration	44
8	EO Pillar	46
8.1	Service description	46
8.2	Service architecture and integrated services	51
8.3	Software release.....	54
8.4	References.....	55
8.5	EOSC-hub integration	55
9	DARIAH	56
9.1	Service description	56
9.2	Service architecture.....	59
9.3	Software release.....	61
9.4	References.....	61
9.5	EOSC-hub integration	62
10	LifeWatch.....	63
10.1	Service description	63
10.2	Service Architecture and integrated services	71
10.3	Software release.....	72
10.4	References.....	72
10.5	EOSC-hub integration	73
	References.....	75

Executive summary

The research communities, which are partner in the project, are both service consumers and providers. They offer services to their users, which are called Thematic Services (TSs). Thematic Services are scientific services (incl. data) that provide discipline-specific capabilities for researchers (e.g. browsing and download data and apps, workflow development, execution, online analytics, result visualisation, sharing of result data, publications, applications). In some cases, in order to integrate those services with EOSC-hub (which is the main scope of WP7), the software code implementing them have been modified, by extending existing components or developing new plugins. These set of software applications forms the current software release.

This final release includes all the services, which have been integrated within EOSC-hub. The integration happens at various levels: from the usage of cloud resources (IaaS, PaaS), to the usage of EOSC-hub common services (for example data and computing services), to the usage of EOSC-hub core services (such as authentication or monitoring tools), to the publication of the services in the EOSC Marketplace. The following Thematic Services are included in this software release:

- CLARIN: only the Virtual Language Observatory (VLO) service is part of this release.
- DODAS: Dynamic On Demand Analysis Service.
- ECAS: ENES Climate Analytics Service.
- GEOSS: Global Earth Observation System of Systems
- OPENCoastS: On-demand oPERatioNal Coastal circulation forecast Services.
- WeNMR: Worldwide e-Infrastructure for NMR and structural biology.
- EO Pillar: a set of services related to Earth Observation (EO) domain.
- DARIAH: Digital Research Infrastructure for the Arts and Humanities
- Lifewatch: a set of services related to Biodiversity and Ecosystem.

All Thematic services are also now available in the EOSC Marketplace as described in the respective sections.

1 Introduction

This document describes the Thematic Services software release. There is one chapter for each Thematic Service. A Thematic Service can be a group of services or just a single one. Each service has its own components and dependencies. The description of each service includes the references to the code, the documentation and the endpoints of the production instances.

Each Thematic Service description is structured along 5 sub-sections:

- Service description
- Service architecture and integrated services
- Software release
- References
- EOSC-hub integration

This is the update from a previous version (D7.1); and it follows other two deliverables (D7.2 and D7.3) each of them containing updated information on the Thematic Services. In some sections of this document, it was considered useful (for completeness and readability) to repeat some of the information reported in previous documents.

2 CLARIN

2.1 Service description

Service/Tool name	Virtual Language Observatory
Service/Tool url	https://vlo.clarin.eu
Service/Tool information page	https://www.clarin.eu/vlo
Description	A facet browser for fast navigation and searching in huge amounts of metadata.
Value proposition	A facet browser for fast navigation and searching in large amounts of metadata. This portal enables the discovery of language data and tools, provided by over 40 CLARIN centres, other language resource providers and Europeans.
Customer of the service/tool	Researchers; Repository managers
User of the service/tool	Researchers
User Documentation	https://vlo.clarin.eu/help
Technical Documentation	https://trac.clarin.eu/wiki/CmdiVirtualLanguageObservatory (requires authentication, information available on request)
Product team	CLARIN ERIC
License	GPLv3
Source code	https://github.com/clarin-eric/VLO
Testing	Unit and integration tests using junit Vulnerability analysis provided by snyk.io

Service/Tool name	Virtual Collection Registry
Service/Tool url	https://collections.clarin.eu
Service/Tool information page	https://www.clarin.eu/content/virtual-collections
Description	A service that allows researchers to create their own citable digital bookmarks.
Value proposition	A virtual collection is a coherent set of links to digital objects (e.g. annotated text, video) that can be easily created, accessed and cited. The links can originate from different archives, hence the term virtual.
Customer of the service/tool	Researchers
User of the service/tool	Researchers

User Documentation	https://collections.clarin.eu/app/help
Technical Documentation	https://trac.clarin.eu/wiki/VirtualCollectionRegistry (requires authentication, information available on request)
Product team	CLARIN ERIC
License	GPLv3
Source code	https://github.com/clarin-eric/VirtualCollectionRegistry
Testing	Unit and integration tests using junit Vulnerability analysis provided by snyk.io

Service/Tool name	Language Resource Switchboard
Service/Tool url	https://switchboard.clarin.eu
Service/Tool information page	https://www.clarin.eu/content/language-resource-switchboard
Description	A web application that suggests language analysis tools for specific data sets.
Value proposition	The Language Resource Switchboard will automatically provide a list of available tools, based on the language and format of the input. The Switchboard can also be invoked from the Virtual Language Observatory and B2DROP (see Suggested compatible services below).
Customer of the service/tool	Researchers Tool administrators
User of the service/tool	Researchers
User Documentation	https://switchboard.clarin.eu/help
Technical Documentation	https://trac.clarin.eu/wiki/LanguageResourceSwitchboard (requires authentication, information available on request)
Product team	CLARIN ERIC
License	GPLv3
Source code	https://github.com/clarin-eric/switchboard
Testing	Unit and integration tests using junit Vulnerability analysis provided by snyk.io

CLARIN is offering three Thematic Services within the EOSC-hub: the Virtual Language Observatory (VLO)¹, the Virtual Collection Registry (VCR)² and the Language Resource Switchboard (LRS)³.

Detailed descriptions for the VLO, VCR and LRS services have already been described in D7.1, D7.2 and D7.3 [R1, R2, R3]. All services are published in the EOSC-hub marketplace (Section 2.5).

¹ <https://www.clarin.eu/vlo>

² <https://www.clarin.eu/content/virtual-collections>

³ <https://www.clarin.eu/content/language-resource-switchboard>

2.2 Service architecture and integrated services

Virtual Language Observatory

The architecture of the VLO has been described in D7.1 [R1], section 2.1.1. The VLO is integrated with the VCR, the LRS, the B2FIND service, Virtual Access (VA) usage satisfaction via Mopinion⁴.

Virtual Collection Registry

The architecture of the VCR has been described in D7.3 [R3], section 2.1.1.2. The VCR is integrated with the VLO, the LRS and VA usage satisfaction via Mopinion. Integration with B2SHARE is still waiting on changes on the B2SHARE side. This work has been discussed with the B2SHARE team and has been included in their development roadmap, however no concrete deadline has been set.

Language Resource Switchboard

The architecture of the LRS has been described in D7.3 [R3], section 2.1.1.3. The LRS is integrated with B2DROP via a NextCloud plugin⁵, the VLO, the VCR and D4Science. Integration with B2SHARE is still waiting on changes on the B2SHARE side. This work has been discussed with the B2SHARE team and has been included in their development roadmap, however no concrete deadline has been set by the B2SHARE team. The pending integration does not block the use of the service, but it will streamline the process and make it more user friendly when in place.

2.3 Software release

Within CLARIN the general direction and prioritization of issues is decided internally, considering bug reports and user feedback. The concrete items resulting from this process are managed as issues and milestones on GitHub⁶, resulting in a medium-term roadmap.

Development versions of the services are made available on a public location. After completing a milestone, a release candidate is provided and made available a public location as well. This version is tested and evaluated by a pool of testers from within the CLARIN community. Several iterations of bug fixes, grouped into one or more new release candidates, are made available until a stable instance is available. This version is then released into production.

For each of the services the following locations are used:

VLO:

- Development versions: <https://alpha-vlo.clarin.eu>
- Release candidates: <https://beta-vlo.clarin.eu>
- Production releases: <https://vlo.clarin.eu>

VCR:

- Development versions and release candidates: <https://collections.clarin-dev.eu>
- Production releases: <https://collections.clarin.eu>

LRS:

- Development versions and release candidates: <https://switchboard.clarin-dev.eu>
- Release candidates: <https://beta-switchboard.clarin.eu>

⁴ <https://www.mopinion.com>

⁵ <https://github.com/clarin-eric/B2Drop2LRSwitchboard>

⁶ <https://github.com/clarin-eric/VLO/milestones>

- Production releases: <https://switchboard.clarin.eu>

All releases are managed with major, minor and patch releases, where major releases can include breaking changes, minor version include new features and patch releases provide bug fixes. This approach is inspired by the semantic versioning⁷ principles.

The entire release is packaged as Docker images and accompanied with a Docker-compose setup for easy deployment, either locally or in the CLARIN infrastructure.

2.4 References

VLO:

- Service endpoints:
 - <https://vlo.clarin.eu>
- Software code and releases:
 - Software code and release are managed in GitHub:
 - <https://github.com/clarin-eric/VLO> (/releases)
 - <https://github.com/clarin-eric/oai-harvest-manager> (/releases)
 - Docker images and compose packages are managed in gitlab:
 - <https://gitlab.com/CLARIN-ERIC/docker-vlo-beta>
 - https://gitlab.com/CLARIN-ERIC/compose_vlo
 - <https://gitlab.com/CLARIN-ERIC/docker-oai-harvester>
 - https://gitlab.com/CLARIN-ERIC/compose_oai_harvester
- Documentation: user oriented, administrator oriented and developer oriented:
 - <https://trac.clarin.eu/wiki/CmdiVirtualLanguageObservatory>
 - This is a closed wiki. On request we can provide an export of the relevant wiki pages

VCR:

- Service endpoints:
 - <https://collections.clarin.eu>
- Software code and releases:
 - Software code and release are managed in GitHub:
 - <https://github.com/clarin-eric/VirtualCollectionRegistry> (/releases)
 - Docker images and compose packages are managed in gitlab:
 - <https://gitlab.com/CLARIN-ERIC/docker-alpine-supervisor-java-tomcat-vcr>
 - https://gitlab.com/CLARIN-ERIC/compose_vcr
- Documentation:
 - <https://trac.clarin.eu/wiki/VirtualCollectionRegistry>
 - This is a closed wiki. On request we can provide an export of the relevant wiki pages

LRS:

- Service endpoints:
 - <https://switchboard.clarin.eu>
- Software code and releases:
 - Software code and release are managed in GitHub:

⁷ <https://semver.org>

- <https://github.com/clarin-eric/switchboard/releases>
- Docker images and compose packages are managed in gitlab:
 - <https://gitlab.com/CLARIN-ERIC/docker-alpine-supervisor-java-switchboard>
 - https://gitlab.com/CLARIN-ERIC/compose_clarin_switchboard
- Documentation:
 - <https://trac.clarin.eu/wiki/LanguageResourceSwitchboard>
 - This is a closed wiki. On request we can provide an export of the relevant wiki pages

Instructions to request access to trac.clarin.eu are available on <https://www.clarin.eu/dev>

2.5 EOSC-hub Integration

The latest stable VLO, version v.4.8.2, is published via EOSC-hub:

- on the Marketplace: <https://marketplace.eosc-portal.eu/services/virtual-language-observatory>
- on the Service Catalogue: <https://www.eosc-hub.eu/catalogue/Component%20MetaData%20Infrastructure>

The latest stable VCR, version v.1.3.0, is published via EOSC-hub:

- on the Marketplace: <https://marketplace.eosc-portal.eu/services/virtual-collection-registry>
- on the Service Catalogue: <https://www.eosc-hub.eu/catalogue/Component%20MetaData%20Infrastructure>

The latest stable LRS, version v.2.2.1, is published via EOSC-hub:

- on the Marketplace: <https://marketplace.eosc-portal.eu/services/language-resource-switchboard>
- on the Service Catalogue: <https://www.eosc-hub.eu/catalogue/Component%20MetaData%20Infrastructure>

3 DODAS

3.1 Service description

Service/Tool name	DODAS
Service/Tool url	http://dodas-iam.cloud.cnaf.infn.it/
Service/Tool information page	https://dodas-ts.github.io/dodas-doc/
Description	DODAS is cloud enabler for scientists seeking to easily exploit distributed and heterogeneous clouds to process, manipulate or generate data.
Value proposition	-
Customer of the service/tool	Researcher, Resource Provider; Research Communities
User of the service/tool	Same as above
User Documentation	https://dodas-ts.github.io/dodas-apps/
Technical Documentation	https://dodas-ts.github.io/dodas-templates/
Product team	INFN
License	Apache License Version 2.0
Source code	https://github.com/dodas-ts
Testing	-

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

DODAS completely automates the process of provisioning by creating, managing, and accessing a pool of heterogeneous computing and storage resources. As a consequence, it drastically reduces the learning curve as well as the operational cost of managing community-specific services running on distributed clouds.

Currently DODAS provides support to deploy:

- Big Data Pre-Post processing facilities
 - HDFS, Spark
- Training facility for the exploitation of Machine Learning
 - Inference engine
- Batch System as a Service
 - HTCondor batch system
 - HTCondor federation solutions

DODAS provides a declarative approach which allows to create complex computing stack set-up on any cloud-based environment just by composing “lego style” any required block of the stack. The

goal is that users should experience the process of generating complex setups as easily as creating a virtual machine on any IaaS: a simple one-click solution. The DODAS added values can be summarized as follows:

- It provides a simple but complete abstraction of hybrid cloud infrastructures;
- It automates both virtual hardware provisioning and its configuration;
- It provides a cluster platform with a high level of self-healing and scalability;
- It guarantees both set-up and service customization to cope with specific scientific requirements.

DODAS is currently adopted as a solution for several use cases:

- exploitation of opportunistic computing, intended as resources not necessarily or permanently dedicated to a specific experiment and/or activity;
- elastic extension of existing facilities, to absorb peaks of resource usage;
- generation of on-demand batch systems for data processing.

The technical details are further described in D7.1, D7.2, and D7.3 [R1, R2, R3].

3.2 Service architecture and integrated services

DODAS has a highly modular architecture and its workflows are highly customisable. For this reason, it is very extensible, spanning from software dependencies up to the integration of external services, including also user-tailored code management. There are four main pillars in the DODAS architecture which can be summarized as:

- **abstraction**: both in terms of software application and dependency description, and of underlying IaaS level cloud infrastructures;
- **automation**: it refers to software and application setup in order to manage resources and orchestrate software applications;
- **multi-cloud support**: to deal with multiple heterogeneous Cloud infrastructures;
- **flexible AAI**: authentication, authorization, delegation, and credential translation primitives providing a secure composition of the various services participating in the DODAS workflow.

Regarding the technical implementation, as shown in Figure 1, these pillars have been realised mostly using services also available in the EOSC-hub portfolio: Identity and Access Management (INDIGO-IAM) and Token Translation Service (TTS), PaaS Orchestrator and Infrastructure Manager (IM). The composition of these services represents the so-called PaaS Core Service of DODAS.

IAM is the OpenID Connect (OIDC) Authorization Server, supporting the eduGAIN⁸ federation, which has the crucial role to authenticate users. It also provides an access token which is then used to authorize services to act on the users' behalf. The PaaS Orchestrator and IM are responsible to take care of the users' requests (in the form of a TOSCA template) and to prepare a cluster for containers orchestration over the IaaS. Besides the original integration with Mesos⁹ for resource management and Marathon¹⁰ for the container orchestration DODAS has been evolved integrating also Kubernetes which is now the default flavour of DODAS. The container orchestrator is the layer responsible for the execution of end-user's services. In principle, any framework and/or software application can run on such cluster.

The technical details are further described in D7.1, D7.2 and D7.3 [R1, R2, R3].

⁸ <https://edugain.org/>

⁹ <https://mesos.apache.org/>

¹⁰ <https://mesosphere.github.io/marathon/>

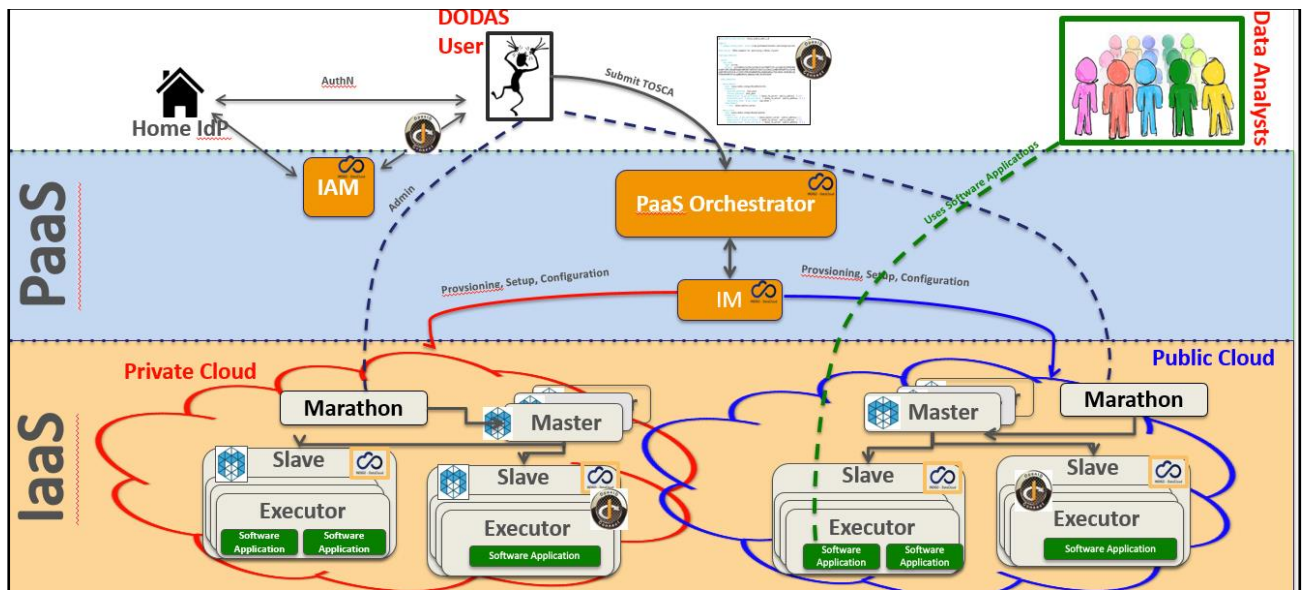


Fig. 1 - DODAS architecture

AAI Integration to federate EGI Check-in

DODAS relies on a dedicated INDIGO-IAM service instance VO management, enrolment, authentication and authorization at DODAS core services. It is also used for the integration of service like HTCondor¹¹ by implementing a DODAS specific service which enable the following workflow:

- Get access token use it as incoming auth credential;
- Implements security via IAM token exchange;
- Cache and return X509 certificates to grant access to CMS.

EGI Check-in is integrated as an external authentication mechanism of IAM as shown in Figure 2, and IAM is used as an AAI harmonization layer, so that EGI Check-in grants access to the EGI cloud resources, while the IAM harmonised identity is used to manage DODAS domain-specific security as explained above.

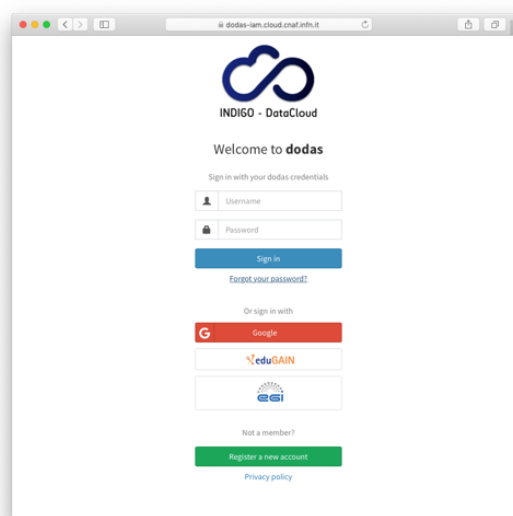


Fig. 2 - DODAS integration with EGI check-in

¹¹ <https://research.cs.wisc.edu/htcondor/>

Integration with Onedata

The data management part of DODAS has been extended to support the Onedata middleware. The main motivation for this was to provide support for data ingestion to all use cases that are not based on CERN ROOT framework. Not only the integration but also several scale tests have been performed to test performance and reliability. In the end, a very good success rate has been reached in a scale test performed using CMS analysis jobs and a comparison between local and remote data access have been evaluated. A good stability in terms of parallel client have been verified: up to 1500+ concurrent jobs. A quite comprehensive study has been carried out on pre-fetching functionalities and the conclusion is that pre-fetching strategies heavily depend on the workflow type. They can have opposite effects on different workflows.

- E.g. a given strategy allows to improve low I/O jobs, while high I/O jobs may degrade,
- An ad-hoc strategy at that level does not seem reasonable in a production environment,
- Huge fluctuations seem to suggest where to look for further improvements.



Fig. 3 - DODAS integration with Onedata

CVMFS integration

In order to accommodate the needs of communities such as AMS, Virgo and Fermi, DODAS has integrated also the support of CVMFS¹² which is needed in order to share data (libraries and user code) over a geographically distributed setup. This part also includes the support of solution for the deployment of CVMFS server as a service.

Support to data caching

Data management support has been extended to support a cache-based mechanism which includes the creation of automatic procedures to enable a fast and effortless deployment of cache clusters in a cloud environment. In fact, dynamic resource provision is a crescent use case that by construction requires a remote read mode for the jobs. So, providing an easy deployment of a scalable cache system near the computing nodes is an important tool to increase, with a minimal effort, the efficiency of jobs running on this kind of resources.

The technology used here is based on XRootD¹³ and below there is a schema of the current architecture:

¹² <https://cern.service-now.com/service-portal/service-element.do?name=cvmfs>

¹³ <https://xrootd.slac.stanford.edu/>

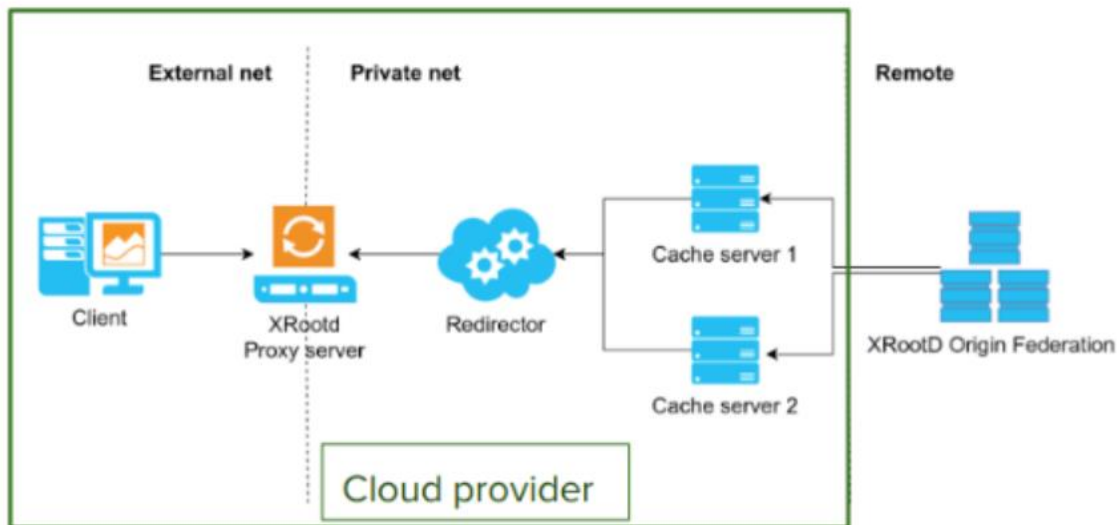


Fig. 4 - DODAS support to data caching

Integration with EGI Cloud

DODAS has been used to integrate and exploit EGI cloud compute resources. In this respect DODAS has been used as PaaS federation layer on top of the EGI Federated Cloud. A complete integration has been carried out using 5 EGI sites and PaaS federation based on HTCondor overlay. Kubernetes¹⁴ has been adopted as orchestrator solution.

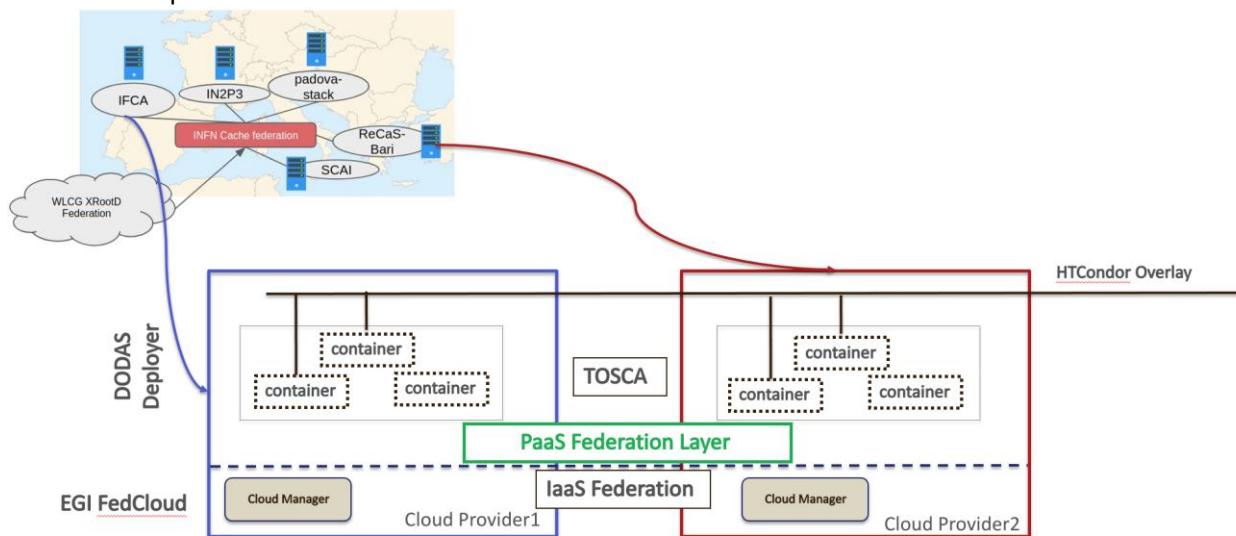


Fig. 5 - DODAS integration with EGI FedCloud

A successful validation test has been done implementing a virtual CMS site over 5 distinct providers (namely IFCA, IN2P3, Padova-Stack, ReCaS-Bari and SCAI) and the virtual infrastructure, implemented through DODAS, is seen as a single site by CMS.

¹⁴ <https://kubernetes.io/>

For the scope of this test, T3_IT_Opportunistic_dodas and CMS Analysis Workflow were used as demonstrator. The EGI Federated Cloud detailed accounting shows that DODAS pilot has consumed 760K CPU hours over 5 sites in October 2019. From CMS monitoring, it is possible to see that 400 jobs run constantly in October.

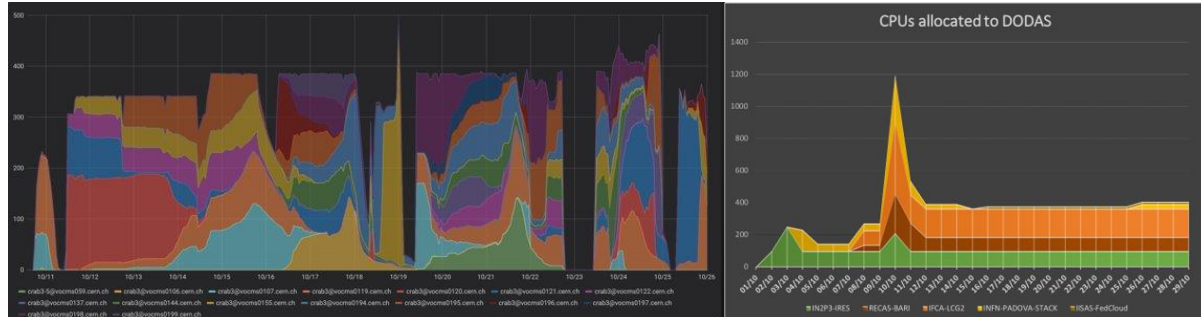


Fig. 6 - DODAS integration with EGI FedCloud - Validation test results

PaaS Orchestrator integration extension

DODAS has been integrated with the new version (2.2.0) of the Orchestrator which provides an improved TOSCA parser that ensures better compliance with the standard.

Several additional optimisations 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 flavours (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.

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.

The integration with PaaS orchestrator extension have been done and fully validated. As consequence this gives now the possibility to exploit the supported templates via the friendly dashboard.

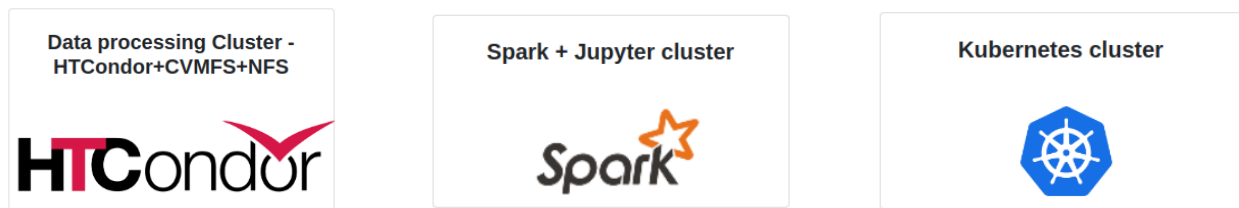


Fig. 7 - Snapshot of DODAS templates integrates within the PaaS Orchestrator's Dashboard

3.3 Software release

Being a composition of services, there is not a single release, nor a single procedure. Also about the roadmap, not a single procedure is foreseen. Core services are kept aligned to the latest stable version (provided by sub services). Experiment/community software is not under DODAS responsibility. TOSCA Templates are those who actually represent the software, from the user perspectives. From a technical point of view TOSCA templates need to guarantee coherence between underlying services actually not directly exposed to the users. Having said that, DODAS provides releases of the TOSCA templates, but a well-defined procedure is still under definition.

3.4 References

Link and references to Service endpoints:

- DODAS login page: <https://dodas-iam.cloud.cnaf.infn.it/login>.

Software code and releases:

- Community images: <https://github.com/DODAS-TS>.
- Tosca templates: <https://github.com/indigo-dc/tosca-templates/blob/master/dodas>.

Documentation:

- <https://dodas-ts.github.io/dodas-doc>

3.5 EOSC-hub integration

DODAS is currently integrated with several EOSC-hub services namely: Infrastructure Manager, PaaS Orchestrator, Identity and Access Management (IAM) and Token Translation services are part of the core. IAM has been further integrated also in the CMS workflow supported by DODAS. In the latest months, DODAS integrated also OneData as a possible solution for the data ingestion and for transparent data access. CVMFS stratum 0 and 1 have been also integrated in order to support newer uses cases coming from AMS-02 experiment requirements. Finally, DODAS stable version is published via EOSC-hub:

- On the Marketplace:
<https://marketplace.eosc-hub.eu/51-dynamic-on-demand-analysis-service>.
- On the Service Catalogue:
<https://eosc-hub.eu/catalogue/Dynamic%20On%20Demand%20Analysis%20Service>.

4 ECAS

4.1 Service description

Service/Tool name	ENES Climate Analytics Service (ECAS)
Service/Tool url	CMCC ENDPOINT: https://ecaslab.cmcc.it/web/home.html DKRZ ENDPOINT: https://ecaslab.dkrz.de/home.html
Service/Tool information page	https://www.eosc-hub.eu/services/ENES%20Climate%20Analytics%20Service https://portal.enes.org/data/data-metadata-service/processing/ecas
Description	The ENES Climate Analytics Service is a server-side processing service which offers a virtual work environment based on Jupyter notebooks, allowing users to process and analyse data from multiple disciplines using Python. Support for fast computations is provided via the Ophidia HPDA framework
Value proposition	ECAS enables scientific end-users to perform data analysis experiments on large volumes of multidimensional data by exploiting a server-side, PID-enabled, and parallel approach and aiming to improve reusability of data and workflows (FAIR approach).
Customer of the service/tool	Climate modelling community; direct downstream usage communities
User of the service/tool	Researchers; Scientific users
User Documentation	https://ee-docs.readthedocs.io/en/latest/ http://ophidia.cmcc.it/documentation/index.html
Technical Documentation	https://github.com/ECAS-Lab/ecas-authentication https://github.com/ECAS-Lab/ecas-accounting http://ophidia.cmcc.it/documentation/admin/index.html
Product team	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (Fondazione CMCC); Deutsches Klimarechenzentrum GmbH (DKRZ)
License	The Ophidia code is available on GitHub under GPLv3 license; additional components for ECAS (Docker workflow components) available under BSD license; ECAS-B2SHARE Python client available under MIT license.
Source code	https://github.com/ECAS-Lab https://github.com/OphidiaBigData
Testing	1. ECAS single-instance VMI uploaded to the EGI AppDB: https://appdb.egi.eu/store/vappliance/ecas Deployment test using the EGI AppDB VMops dashboard: https://dashboard.appdb.egi.eu/vmops 2. Multi-node ECAS environment, dynamically provisioned on the EGI FedCloud through the EC3 LToS service: https://servproject.i3m.upv.es/ec3-ltos/index.php

The ENES Climate Analytics Service (ECAS) enables scientific end-users to perform data analysis experiments on large volumes of multidimensional data (e.g. NetCDF data format), by exploiting a PID-enabled, server-side, and parallel approach. The service is aimed at providing a paradigm shift for the ENES community with a strong focus on data intensive analysis, provenance management, and server-side approaches as opposed to the current ones that are mostly client-based, sequential and with limited/missing end-to-end analytics workflow/provenance capabilities. ECAS consists of multiple integrated components, centred around the Ophidia Data Analytics framework¹⁵, which has been integrated with B2DROP, ESGF, IAM, Onedata (DataHub), EGI FedCloud, EGI Check-in, JupyterHub, and the ECAS-Lab web portal.

4.2 Service architecture and integrated services

The Ophidia framework provides the key scalable, parallel analytics capabilities. To enable easy data provisioning, it is integrated with B2DROP, Onedata and ESGF. The integration with JupyterHub, IAM and EGI Check-in enables users to easily access the service and re-use existing Python scripts, modify them or create entirely new ones in a fully server-side approach. More details are reported in EOSC-Hub deliverable D7.1.

Figure 8 shows an overview of the final ECAS architecture with all the integrated services. Besides bug fixing, software and service maintenance, several integration and adaptation activities, as well as those related to documentation and training, were performed by the ECAS team during the project.

In order to store all the developments, the integration efforts and the training material concerning ECAS, a Github repository¹⁶ has been set up. Extensions and adaptations to other tools have been published, instead, on the related Github repositories (i.e. for the Ophidia framework¹⁷), while virtual appliances have been published on the EGI services (e.g. for ECAS VM¹⁸).

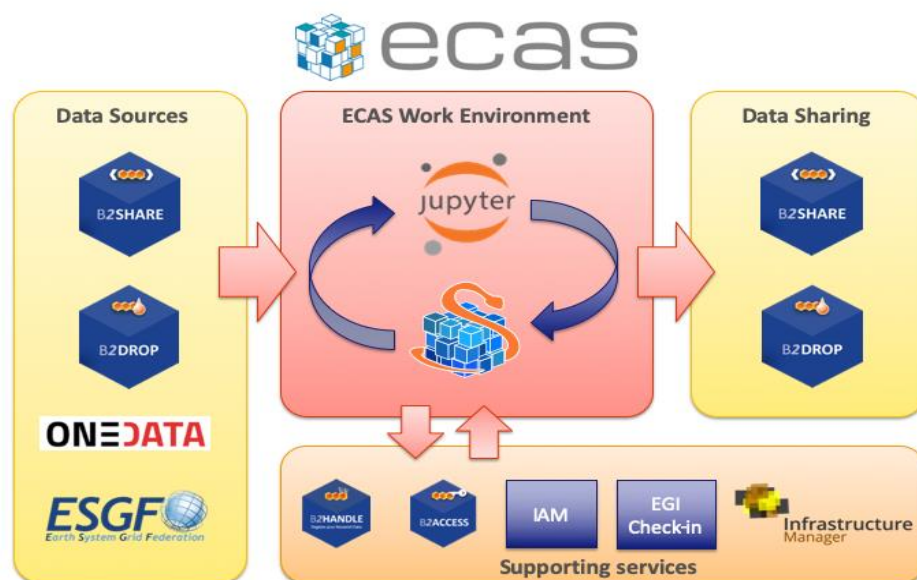


Fig. 8 - ECAS architecture overview

¹⁵ <http://ophidia.cmcc.it/>

¹⁶ <https://github.com/ECAS-Lab/>

¹⁷ <https://github.com/OphidiaBigData>

¹⁸ <https://appdb.egi.eu/store/vappliance/ecas>

Since the beginning of the project, integration has focused on enabling a first exemplary user workflow, spanning from data input over login and processing to data output sharing. The workflow is provided at the two instances setup at CMCC and DKRZ, which differ in their setup, available data and computing resources. With respect to the latest version of the deployment (described in deliverable D7.3), the CMCC instance has been updated with a more powerful node in order to better support end-user services, especially during training events or high-load periods. This node has replaced the one previously used as server and all the services have been migrated and updated from the older system to the new one. Given the higher memory and computing capacity, other similar services will be migrated from the client node to this new node in the next months.

The current state of the available resources at CMCC and DKRZ is the following:

- ECAS@CMCC:
 - 1 client node with 4 cores, 8GB memory, 12GB disk, hosting the client-side components (e.g. the Ophidia CLI and PyOphidia module) and the front-end services (e.g. JupyterHub);
 - 1 server node with 20 cores, 256GB memory, 4,7TB disk, executing the Ophidia Server and complementary administration, monitoring and accounting services;
 - 5 compute nodes, 20 cores/node, RAM 256GB/node, 60TB disk space shared via GFS plus local disks for the Ophidia framework and I/O servers.
- ECAS@DKRZ:
 - 1 client node (VM) with 4 cores, 15GB memory, 50GB local disk + 10TB NFS for the JupyterHub and Docker services;
 - 1 server node (VM) with 4 cores, 4GB memory, 30GB disk, executing the Ophidia server;
 - 1 compute node (Blade) with 40 cores, RAM 256GB, 2TB local disk, executing Ophidia I/O server + 10TB NFS disk space.
 - CMIP5 and CMIP6 data pool (read-only).

Most of the integration activities, concerning for example the integration into the ECAS environment of B2DROP and Onedata for data sharing, IAM as the AAI solution and ESGF data sources, started during the first year of EOSC-hub. Additional actions undertaken during the first-year regard software containerization of the ECAS environment, improving usability of the data analytics features from JupyterHub, providing good workflows for development and operations, integration with the EOSC-hub accounting and monitoring, documentation, implementation of multiple demonstration Jupyter notebooks, and extensive training. Integration activities for the first year of the project have already been presented in detail in deliverable D7.2.

During the second year of the project, several integration activities have been finalized. In particular, they concern a stronger integration of B2DROP and Onedata data services, the full integration of IAM for user authN/authZ, from JupyterHub to the data analytics services, and the extensions to the core services for accounting and monitoring purposes. Moreover, the Ophidia framework and its Python interface have been extended to further improve the user experience and allow the development of other climate-oriented use cases; additional Jupyter notebooks have also been produced as both demonstrators of the environment feature and self-study/training material. The views for user accounting have been also fully implemented and the ECAS environment has been successfully integrated into the EGI VA marketplace and the EGI Federated Cloud deployment services.

Furthermore, during the third year of the project additional extensions targeting a stronger integration of ECAS (through PyOphidia) with the Python ecosystem and Jupyter Notebooks have been implemented, and the integration of ECAS with EGI Check-in has also started.

This document builds on the integration activities already described in deliverable D7.3 complementing them with the latest updates in order to provide a unified view for this the software release.

A detailed description of the various activities carried out is provided as follows.

Extensions to Ophidia for supporting additional use cases

Besides the extensions developed during the first year, the Ophidia Framework and the related Python bindings (PyOphidia) have been extended with features for supporting new use cases from the climate domain. These include, for example, the possibility to concatenate data from a NetCDF file to a datacube (concatnc interface), the support for the execution in Ophidia of scripts directly defined in Python notebooks, an improved (faster and more interoperable) version of the data export function from the Ophidia server side to the Python code (and also Jupyter Notebooks). Additionally, the PyOphidia module has been extended to provide a stronger integration with the Python libraries exploited in the ECASLab environment and to improve usability from notebooks. Among the extensions, for example, new methods to export a PyOphidia Cube as a Xarray Dataset or a Pandas Dataframe have been implemented, allowing users to easily access, manipulate and visualize data. All developments are openly accessible from the Ophidia Github repository.

Additional extensions for the integration of B2DROP into ECAS

The OPH_B2DROP Ophidia operator has been further extended to also integrate, besides data upload, the download of files from a B2DROP account. This allows scientists to exploit the full data life cycle, from source data shared from B2DROP, through data processing with Ophidia and the final sharing of data again on B2DROP, within the ECAS environment. The operator interface has been hence extended with a new argument allowing the user to specify the type of operation (download/upload) on a file or dataset; the following example shows the usage of the operator to download a file (*test.nc*) from B2DROP to a folder within ECASLab (*datafolder*):

```
oph_b2drop action=get;src_path=test.nc;dst_path=datafolder/
```

whereas to upload a file from ECASLab (*datafolder/test.nc*) to the B2DROP home space, the following can be used:

```
oph_b2drop action=put;src_path=datafolder/test.nc;dst_path=test.nc
```

Furthermore, the PyOphidia module has been also updated to integrate this extended interface and a demonstrator Jupyter notebook, called *ECAS_B2DROP.ipynb*, has also been implemented (available at <https://github.com/ECAS-Lab/ecas-notebooks/>). Additional technical details about OPH_B2DROP concerning the authentication aspects of the B2DROP service and its usage can be found in D7.2 [R2].

Integration of ECAS with B2HANDLE and B2SHARE workflow

Tracing of research data as part of the workflow from data input via data processing/analysis to data sharing as facilitated by ECAS is a key challenge to make data workflows transparent, and foster re-use of both workflows and results. To support this, ECAS is integrated with B2HANDLE to detect identifiers of input data and connect results out of ECAS workflows with inputs. This was implemented in a way to minimize necessary user interaction, i.e., the user does not need to gather identifiers or other context information for inputs and make connections with output explicitly. Instead, automation scripts have been developed and included in the back-end of ECAS to make

such connection automatically. Further integration steps are necessary to enable automated recognition and exposition of the gathered information via B2SHARE. ECAS has started interaction with the B2SHARE development team to expand the B2SHARE web user interface and to define metadata profiles. Feedback from training activities and user community interaction indicates further steps will be necessary to adapt and expose metadata according to community-specific workflows and expectations.

AAI Integration of Ophidia and JupyterHub with IAM

The integration of IAM with JupyterHub and Ophidia, started during the first year, has been finalized during the second year of the project. In particular, the JupyterHub configuration has been extended to support a custom authentication procedure which allows checking the user credentials with respect to the IAM service, verifying the presence of the same user in ECASLab and matching its identity with the ECASLab account. Figure 9 shows the JupyterHub interface allowing a user to log in to the system using its IAM account. The extensions have been published on the Github repository of ECASLab¹⁹. Technical details about this integration are available in deliverable D7.3 [R3].

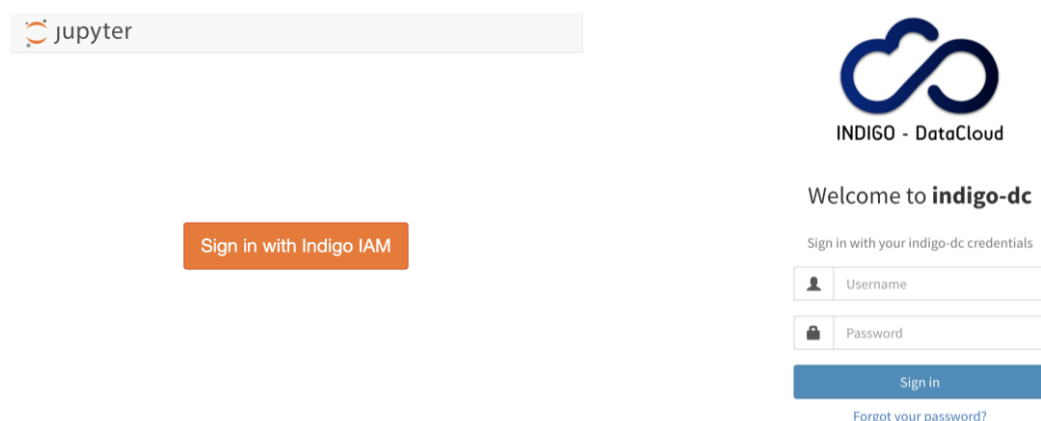


Fig. 9 - JupyterHub interface

AAI Integration of Ophidia with EGI Check-in

Integration of EGI Check-in in ECAS has also started in order to provide a stronger support AAI solution. To this end, the OpenID Connect-based authZ/authN mechanism²⁰ previously implemented in Ophidia has been extended to support also the tokens from EGI Check-in.

In the implemented extension, the Ophidia clients (i.e., PyOphidia or the terminal) can use a token generated from the EGI Check-in to authenticate at the ECAS instance and submitting the commands. Upon receiving a request from PyOphidia or the terminal, the Ophidia server, which is registered at the EGI Check-in OpenID Connect service as a client, checks the token integrity and verifies the user by interacting with the OpenID Connect service provider. Once the identity provider authorizes the user, additional authorization checks could be performed locally at the ECAS instance based on data of user profile (e.g. email provider).

Integration of ECAS into the EGI Federated Cloud

The integration of ECAS into the EGI Federated Cloud has been implemented by considering two different scenarios.

¹⁹ <https://github.com/ECAS-Lab/ecas-authentication>

²⁰ http://ophidia.cmcc.it/documentation/admin/install/components/install_openid.html

In the first scenario, an ECAS single- instance instance Virtual Machine Image (VMI), providing a ready-to-use ECAS environment, has been created and uploaded to the EGI AppDB²¹. The VMI has been assigned to a set of trusted VOs in order to be deployed on the Federated Cloud.

A deployment test has been carried out using the EGI AppDB VMops dashboard²², by which a user can deploy a pre-built VMI, get the public IP address of the running VM and download the SSH public key that will be used to access.

Name	CPUs		Memory	Disk	Connectivity		OS
	Logical	Physical			In	Out	
i b8f64e9d-d5bc-40c3-bdc6-d55fe3625c72	1	1	64 MB	1 GB	✓	✓	linux
i f1356df5-fc28-43ba-bba0-2ad0dc01b1b	1	1	512 MB	1 GB	✓	✓	linux
i 2	1	1	2048 MB	20 GB	✓	✓	linux
i 8c3bde5-f640-4870-8ed0-f73d3f8eca8e	1	1	2048 MB	20 GB	✓	✓	linux
i fea52bb1-6a69-4d6c-99fb-531c9cf73e67	2	2	4096 MB	40 GB	✓	✓	linux
i ce1d7a0e-32cf-44c8-a8ba-6a387f516b69	4	4	8192 MB	80 GB	✓	✓	linux
i 831e1465-3551-4ce3-a663-cebee7b04710	8	8	16384 MB	160 GB	✓	✓	linux

Fig. 10 - ECAS integration with EGI Federated Cloud

The second scenario refers to a multi-node elastic ECAS environment²³, which is dynamically provisioned on the Federated Cloud through the EC3 LToS service²⁴ according to the user requirements. In this case, a RADL file for the Infrastructure Manager (IM) has been implemented and tested with the support of the UPV Grycap team to define the cluster setup in terms of resources, infrastructure and software configuration, and contextualization²⁵.

²¹ <https://appdb.egi.eu/store/vappliance/ecas>

²² <https://dashboard.appdb.egi.eu/vmops>

²³ <https://www.egi.eu/about/newsletters/elastic-deployment-of-ecas-on-egi/>

²⁴ <https://servproject.i3m.upv.es/ec3-ltos/index.php>

²⁵ <https://github.com/grycap/ec3/blob/devel/templates/ophidia.radl>

Elastic deployment of ECAS on EGI

The ECAS service is available via the EGI Applications on Demand

Authors:

*Sandro Fiore, Donatello Elia, Fabrizio Antonio, CMCC and Tobias Weigel, DKRZ
Ignacio Blanquer, Miguel Caballer, UPV
Giuseppe La Rocca, EGI*

In the latest release of the Elastic Cloud Compute Cluster (EC3) platform, tailored to support the [EGI Applications on Demand](#) (AoD) service, a new cluster configuration type is now available for researchers interested to deploy ECAS cluster on the EGI Infrastructure.

The ENES Climate Analytics Service (ECAS), setup in the context of the EOSC-Hub project by CMCC and DKRZ, enables scientific end-users to perform data analysis experiments on large volumes of multidimensional data (e.g. NetCDF data format), by exploiting server-side, in-memory and parallel approach.

The service aims at providing a paradigm shift for the ENES community with a strong focus on data intensive analysis, provenance management, and server-side approaches as opposed to the current ones that are mostly client-based, sequential and with limited end-to-end analytics capabilities. ECAS consists of multiple integrated components, centered around the Ophidia High Performance Data Analytics framework, which has been integrated with B2DROP, ESGF, IAM, Onedata (DataHub), EGI Federated Cloud, JupyterHub, and the ECAS-Lab web portal.

Fig. 11 - EGI newsletter - Elastic deployment of ECAS on EGI

For aspects related to software setup and contextualization, both scenarios rely on the Ophidia Ansible Role, which has been extended to include the whole ECAS environment services, i.e. JupyterHub, PyOphidia, several Python libraries such as numpy, matplotlib and Basemap. The updated version of the role is available²⁶.

Integration of ECASLab with Onedata

Following the discussions with Onedata team, an architectural solution, addressing security and performance needs, has been identified to integrate Onedata with ECAS.

Specifically, a OneProvider service has been deployed at the CMCC SuperComputing Center (as shown in Figure 15) and attached to the ECAS cluster to reduce networking latencies with respect to the access to remote external providers.

In the proposed solution, the Onedata provider is used to support a shared read-only repository (ECAS_space), which exposes existing data collection (e.g. data published by modelling groups) to ECAS users.

²⁶ <https://github.com/OphidiaBigData/ansible-role-ophidia-cluster>

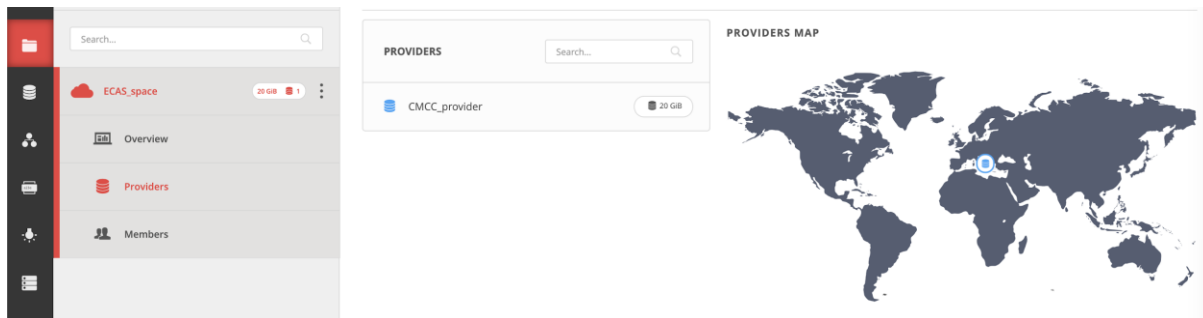


Fig. 12 -ECAS integration with Onedata

To comply with security policies in a data centre environment, a single OneClient instance has been set up to interact with the provider and the data folders have been mounted on the ECAS user's home through NFS in read-only mode (see figure below). In such a context, the data provider is well isolated from the ECAS resources and communication with other Onedata services occurs through the Onedata protocol.

Accounting management

In order to properly track resource usage on a user basis, the Ophidia framework has been extended to conveniently store the most important information in CSV format for accounting purposes. In particular, the Ophidia server stores information in two different files:

- *workflow log*, which tracks the workflow requests, including submission timestamp, username, workflow identifier, number of tasks in the workflow, execution duration, etc:

timestamp	idworkflow	name	username	ip_address	client_address	#tasks	#success_tasks	duration
2018-03-29 01:08:48	1171802	oph_list	*4E6FA15F82F3...	192.168.88.150	193.204.199.213	1	1	0.194882
2018-03-29 01:09:05	1171804	oph_randcube	*4E6FA15F82F3...	192.168.88.150	193.204.199.213	1	1	0.613289
2018-03-29 08:29:54	1171806	oph_list	*0C0EBF0FBAAF...	192.168.88.150	193.204.199.213	1	1	0.185737
2018-03-29 09:16:38	1171808	oph_list	*704101A7468E...	192.168.88.150	193.204.199.213	1	1	0.458210
2018-03-29 09:36:34	1171810	oph_list	*704101A7468E...	192.168.88.150	193.204.199.213	1	0	0.273170

- *task log*, which tracks the single tasks forming a workflow in terms of submission timestamp, task identifier, workflow identifier (link to workflow data in the previous log file), operator name, number of cores, execution duration, etc:

timestamp	idtask	idworkflow	operator	#cores	success_flag	duration
2018-03-29 01:08:48	1171803	1171802	oph_list	1	1	0.153245
2018-03-29 01:09:05	1171805	1171804	oph_randcube	1	1	0.578767
2018-03-29 08:29:54	1171807	1171806	oph_list	1	1	0.152150
2018-03-29 09:16:38	1171809	1171808	oph_list	1	1	0.223756
2018-03-29 09:36:34	1171811	1171810	oph_list	1	0	0.153697

A detailed documentation about the accounting features is published²⁷.

Starting from the information tracked by the Ophidia server about the workflow requests and the submitted tasks, a set of accounting metrics has been identified to extract useful statistics about users and computing resources usage.

²⁷ <http://ophidia.cmcc.it/documentation/admin/management/accounting.html>

A report containing the statistics about a given period has been created in the form of Jupyter Notebook (Figure 13). In particular, data has been processed and displayed using pie charts, bar charts, tables, and so on, by exploiting the matplotlib and Pandas Python libraries. The notebook, as well as the related developments for virtual access are available on the repository²⁸.

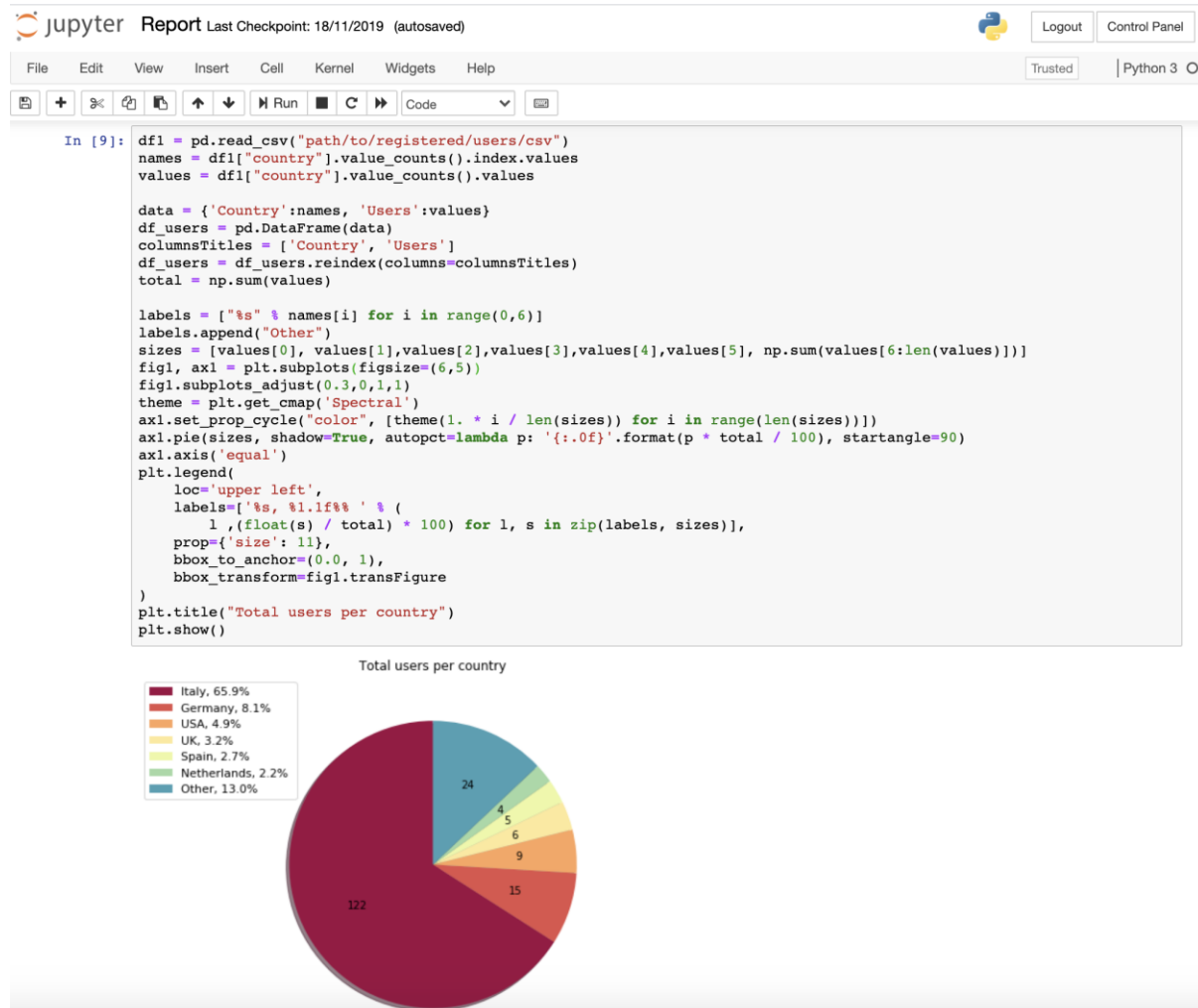


Fig. 13 - ECAS metrics reporting

The pie chart in the figure shows the user distribution concerning the CMCC deployment; in the M1-M26 period nearly 200 users from over 20 different countries and 60 affiliations (universities and research institutes) have created an account on ECASLab. During this period over one million and a half of workflows have been executed, for a total of about 2 millions of jobs and about 10000 cores/hour. Additional details about the service exploitation and the accounting management are provided in D7.3.

Extended ECASLab monitoring integration

In order to provide a more fine-grained control of the system and the usage of its resources, the Grafana-based monitoring system deployed during the first year has been extended with additional

²⁸ <https://github.com/ECAS-Lab/ecas-accounting>

views. These new views help with (i) tracking which user is currently running an Ophidia operator/workflow, (ii) identifying which operator/workflow is being executed and its current execution status, (iii) which node is being used by a user for the execution. Moreover, to allow an easier management of the migration of the service, a containerized version (Docker-based) of Grafana has been deployed in place of the previous bare-metal setup at the ECASLab instance at CMCC.

The following figure shows an example of the new dashboards implemented in Grafana²⁹.

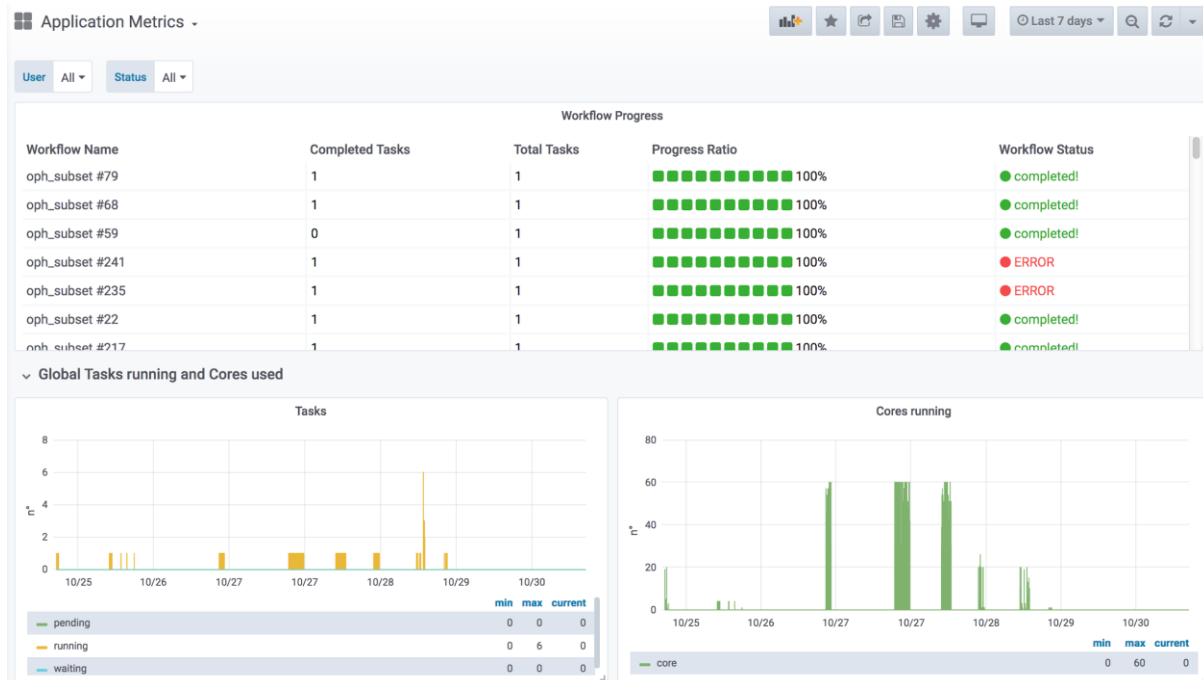


Fig. 14 - ECAS monitoring

JupyterHub notebooks for training purposes

To support user training and demonstration activities, several Jupyter Notebooks integrating some of the key features of ECAS have been implemented. The notebooks are available on the ECASLab github repository³⁰ and have been designed in different formats: (i) hands-on and exercises specifically designed for training and (ii) real use case for feature demonstration.

Some of the main examples from the first group are:

- *ECASLab-training* provides an overview of ECASLab, its main features, some step-by-step instructions and a guided exercise;
- *Linear_regression_ECAS* provides the user with an exercise regarding linear regression to be performed with the ECAS features and resources;
- *Time series extraction* shows how to plot on a map a time series from an Ophidia datacube;
- *Time series difference* shows how to perform the difference among two time series;
- *Aggregated map* shows how to create a map from an Ophidia datacube;

²⁹ <https://grafana.com/>

³⁰ <https://github.com/ECAS-Lab/ecas-notebooks>

- *Subsetted map* shows how to create a map from a subset of a datacube.

Notebooks related to the second group include:

- *ECAS_B2DROP* shows how to exploit the OPH_B2DROP extensions to download and upload data from/to B2DROP;
- *ECAS_Basics* provides an overview of how to run some of the basic and most important functionalities;
- *Tropical Nights* shows how to compute the Tropical Nights indicator;
- *Frost Days* shows how to compute the Frost Days indicator;
- *Summer Days* shows how to compute the Summer Days indicator;
- *Icing Days* shows how to compute the Icing Days indicator;
- *Daily Temperature Range* shows how to compute the Daily Temperature Range climate indicator.
- *Comparison of observational and simulation data (an obs4mips- variable related to the water cycle e.g. hfls = latent heat flux).*

4.3 Software release

ECAS is based on the different components described in the previous architectural section: in this section we will keep providing a component-wise view, with respect to the release process. ECAS release notes provides a pointer to the Ophidia, PyWPS and JupyterHub releases.

Ophidia is released as open source software (GPLv3). All the components are available on GitHub on the Ophidia project page: <https://github.com/OphidiaBigData/>. Releases of the components can be found in the single code repositories. For each source code release, RPMs and DEBs packages, respectively for CentOS7 and Ubuntu 18 (added in replacement of Ubuntu 14 packages), as well as a ready-to-use VM image, are created. Additionally, the Ophidia Python bindings (PyOphidia), is released also on Conda Forge³¹ and on the Python Package Index³². A complete administrator and end-user documentation is available for each new release at <http://ophidia.cmcc.it/documentation/>.

Additional components required for the integration that are not services or software maintained by ECAS (e.g. JupyterHub, B2DROP) include Docker images and associated documentation. These are released under an Apache Open Source license.

When the code is ready for a new release a software rollout procedure is applied. This includes the automated building of the code and creation of the binaries (RPMs and DEBs), as well as the configuration of the full ECAS software stack and the execution of some functional tests on both CentOS7 and Ubuntu 18 systems. The whole process exploits a flexible containerized (Docker-based) approach to perform Jenkins automated tasks. More details about this procedure are presented in deliverable D7.1.

The source code and building of the customized Jupyter Docker image for ECAS at DKRZ are managed through Cloud-based services: GitHub and Docker hub. For each new Docker image, a new tag is locally assigned and then pushed to Github. Next, a built is triggered automatically on Docker hub and made available with the assigned tag.

³¹ <https://anaconda.org/conda-forge/pyophidia>

³² <https://pypi.org/project/PyOphidia/>

The ECAS releases include multiple components with their own release cycles. Full ECAS software stack releases are coordinated and agreed by the ECAS team at DKRZ and CMCC, taking project deadlines and planning into account.

4.4 References

- Service endpoints:
 - CMCC: <https://ecaslaboratory.cmcc.it/web/home.html>
 - DKRZ: <https://ecaslaboratory.dkrz.de/home.html>
- Software/code repositories:
 - ECAS Website: <https://github.com/ECAS-Lab/ecas-web>
 - Ophidia framework: <https://github.com/OphidiaBigData>
 - JupyterHub: <https://github.com/jupyterhub/jupyterhub>
 - ECAS (Jupyter) Notebooks: <https://github.com/ECAS-Lab/ecas-notebooks>
- Documentation:
 - Ophidia (quickstart, user guides, administrator manuals): <http://ophidia.cmcc.it/documentation/index.html>
 - JupyterHub (installation and usage guides): <https://jupyterhub.readthedocs.io/en/stable/>
 - The ECASLab web interface with workflow experiments description, quick start guide and registration form: <https://ecaslaboratory.cmcc.it/web/home.html>

4.5 EOSC-hub integration

The current version of ECAS is integrated in EOSC-hub both at component level and at level of access channels.

Integrated components

There is an additional ECAS component called ECASLab Auth (test instance), which includes the integration of ECAS with the three major EOSC-hub AAI providers: B2ACCESS, Indigo IAM and EGI Check-in. This component is able to authenticate new users to the JupyterHub against the above-mentioned AAI providers and it is also part of the testing/release process. Its role consists of:

- Redirecting users from JupyterHub to ECAS Auth Portal.
- Brokering the authentication/authorization.
- Managing users on premise.

Access channels

The latest stable ECAS service is published via EOSC-hub:

- on the Marketplace: <https://marketplace.eosc-portal.eu/services/enes-climate-analytics-service>
- on the Service Catalogue: <https://eosc-hub.eu/catalogue/ENES%20Climate%20Analytics%20Service>

5 GEOSS

5.1 Service description

Service/Tool name	GEO Discovery and Access Broker (GEO DAB) Virtual Earth Laboratory (VLab)
Service/Tool url	GEO DAB API: http://api.geodab.eu/ VLab API: http://vlabapi.geodab.org/
Service/Tool information page	GEO DAB: https://www.geodab.net/ VLab: https://essilab.wixsite.com/vlab
Description	GEO DAB is a key component of the GEOSS Platform, transparently connecting GEOSS User's requests to the resources shared by the GEOSS Providers. VLab allows to use GEOSS datasets for the generation of new products, implementing all orchestration functionalities to ingest GEOSS data, execute workflows/models and save outputs.
Value proposition	GEO DAB goal is to simplify cross and multi-disciplinary discovery, access, and use (or reuse) of disparate data and information. GEO DAB is a brokering framework that interconnects hundreds of heterogeneous and autonomous supply systems by providing mediation, harmonization and transformation functionalities. VLab addresses the needs of scientists and modelers facilitating the generation of knowledge for an evidence-based decision-making process. It enables the execution of scientific models/workflows not only on EOSC but also on other cloud platforms, including AWS and the Copernicus DIAS.
Customer of the service/tool	Research Communities
User of the service/tool	N/A
User Documentation	GEO DAB: https://www.geodab.net/ VLab: https://confluence.geodab.eu/display/VTD/VLab+Documentation
Technical Documentation	GEO DAB: https://www.geodab.net/ VLab: https://confluence.geodab.eu/display/VTD/VLab+Documentation
Product team	CNR
License	N/A
Source code	N/A
Testing	Both GEO DAB and VLab utilize unit and integration testing. Testing framework is based on JUnit.

The GEO Discovery and Access Broker (GEO DAB) is a key component of the GEOSS (Global earth Observation System of Systems) Platform, transparently connecting GEOSS User's requests to the resources shared by the GEOSS Providers. GEO DAB was described in more detail in D7.2 and D7.3

[R2, R3]. In connection with Virtual Laboratory (VLab), it is possible to use GEOSS datasets for the generation of new products: users can discover and execute workflows to generate new products useful for her/his analysis, utilizing the GEO DAB to discover and ingest input data and VLab to orchestrate and run the workflow of interest

5.2 Service architecture and integrated services

Details of the service architecture for both GEO DAB and VLab are available in previous deliverables (D7.2 and D7.3) and in scientific literature (Nativi et al., 2015, Big Data challenges in building the Global Earth Observation System of Systems, Environmental Modelling & Software 68, 1-26) (Santoro et al., 2020, The VLab Framework: An Orchestrator Component to Support Data to Knowledge Transition, Remote Sensing 12).

5.3 Software release

Both GEO DAB and VLab frameworks apply an agile development methodology and are regularly updated with bug fixes/new features.

5.4 References

GEO DAB:

- Documentation: <https://www.geodab.net/>
- API: <http://api.eurogeoss-broker.eu/>

VLab:

- Documentation: <https://confluence.geodab.eu/display/VTD/VLab+Documentation>
- API: <http://vlabapi.geodab.org/>

5.5 EOSC-hub integration

The GEOSS service integration in EOSC-hub utilizes EOSC-hub computing cloud resources provided by CESNET. A Kubernetes cluster was installed and utilized for the deployment of the software modules.

6 OPENCoastS

6.1 Service description

Service/Tool name	OPENCoastS
Service/Tool url	https://opencoasts.ncg.ingrid.pt/
Service/Tool information page	http://opencoasts.lnec.pt/index_en.php
Description	The OPENCoastS (On-demand Operational Coastal Circulation Forecast Service) service builds on-demand circulation forecast systems for user-selected coastal systems and keeps them running operationally for the time frame defined by the user. This daily service generates forecasts of relevant variables (water levels, 2D or 3D velocities, wave parameters, 3D salinity and temperature) over the spatial region of interest for periods of 48 hours, based on numerical simulations of all relevant physical processes. Currently, the service is deployed at a two computing sites (NCG-INGRID-PT and IFCA).
Value proposition	A user-friendly web portal allows users to setup a forecast system of their coastal region of interest.
Customer of the service/tool	Coastal Research Communities, Coastal end-users, Coastal authorities
User of the service/tool	researchers, people in charge of emergency and daily planning at coastal infrastructures
User Documentation	Manual: http://opencoasts.lnec.pt/OPENCoastS_manual.htm Reference paper: Oliveira, A., A.B. Fortunato, J. Rogeiro, J. Teixeira, A. Azevedo, L. Lavaud, X. Bertin, J. Gomes, M. David, J. Pina, M. Rodrigues, P. Lopes, 2020. OPENCoastS: An open-access service for the automatic generation of coastal forecast systems, Environmental Modelling and Software, 124: 104585. DOI: 10.1016/j.envsoft.2019.104585
Technical Documentation	not public at this stage
Product team	LNEC, LIP, CNRS/LIENS, UNICAN
License	Apache License Version 2.0
Source code	not public at this stage
Testing	a suite of testing deployments, covering all options, was defined and is conducted every time a new release is performed

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

This freely available web service builds forecasts systems and maintains them in operation using the European Open Science Cloud infrastructure. Forecasts are based on the SCHISM community model's simulations using a computational grid uploaded by the user, to predict the 2D barotropic, 2D Waves & Currents or 3D baroclinic circulation in the coastal area of interest. The 2D Waves & Currents circulation option was added in the last few months and the second official release occurred at the beginning of July.

The web platform includes the whole forecast workflow: system configuration, system management and forecasts viewer. Forecast accuracy can be controlled by the user at configuration stage by: i) the selection of the forcings from the several sources available (including Meteo France and NOAA's atmospheric predictions and ocean boundary conditions from CMEMS and FES2014, among others); and ii) the choice of several model parameters. The quality of the predictions can be assessed by automatic comparison with user-selected EMODNET data stations. Default settings are also provided to facilitate the uptake of the OPENCoastS platform by people less familiar with the use of the SCHISM model. In the past months, an open grid repository in Zenodo/Github is also available for testing the platform and to promote shared work.

This reporting period was also used to publish OPENCoastS in peer-reviewed journals and books, including several presentations at international conferences. Two training events were also promoted targeting coastal managers, Africa and East Europe: 14th Silusba and the 14th MEDCOAST conferences.

More information can be retrieved from EOSC-hub deliverable 7.1, 7.2 and 7.3 [R1, R2, R3], on the several papers published in 2019: OPENCoastS core paper (published in open access), OPENCoastS grid repository paper and OPENCoastS 3D version papers [R4-R6]. A paper on the 2D wave and currents version is being prepared.

6.2 Service architecture and integrated services

The OPENCoastS service architecture includes a frontend with a user interaction component for forecast systems configuration and management, via a web application, and a backend where models and mapping services run and a storage tier for preservation. The following figure illustrates this architecture with more detail.

OPENCoastS service

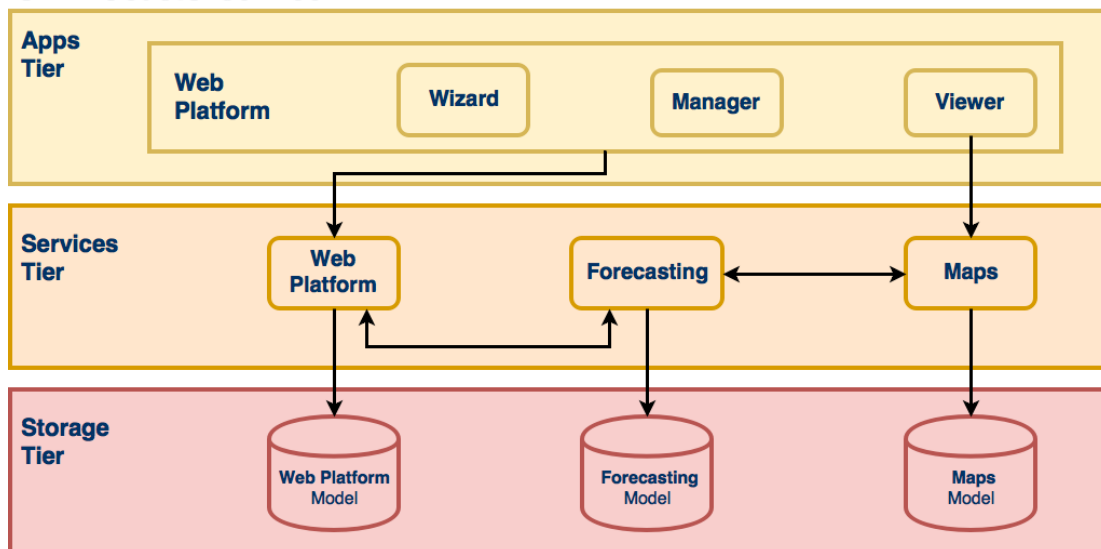


Fig. 14 - OPENCoastS service architecture

Web application

This component provides access to the service, through web pages hosting wizard, manager and viewer applications (apps). Each of these apps allows the users to interact with the different aspects of the service while keep them independent of each other. The Django Web Framework is used as the development basis of this component, which follows Django's design philosophy of having a project composed by applications, each with a set of concerns and functionalities.

Forecasting

While the web application, as the user facing component, allows the users to interact with the service and manage their operational real-time forecast systems, it is the forecasting component that is responsible for producing the forecast results. As a central piece, it interacts with all other components, directly or indirectly, to be able to gather all the necessary information to run simulations and make their results available.

Mapping

The mapping services complement the web application ones, by providing WMS services, which are then consumed by the viewer application. The ncWMS2 server is used to publish the spatial forecasting results on the web, sourced from netCDF files.

Storage

The storage component keeps the state of all services and is a requirement for all of them. The storage technologies range from typical relational databases servers, to lower level shared file systems. The relational database software used is the PostgreSQL, with PostGIS support, storing most structured information about the service. The object storage service, reviewed to be connected to EUDAT B2SHARE and B2DROP, makes available files used in and resulting from the simulations. Shared file systems, in this case NFS, are used to share folders and files among the computational jobs submission and mapping host resources.

Computation

The computational demands for the Opencoasts schisms requires an infrastructure able to answer the HPC requirements for the forecasts.

EGI High Throughput Compute Thematic Service provides the distributed network of computing centres that allows to analyse the datasets and execute the parallel computing tasks leveraging the forecasts before the next day. The service delivers the workload, data management, integrated monitoring and accounting tools, allowing to manage all computational tasks and reporting the information about the availability and resource consumption.

Subsequently, the EGI Workload Manager Thematic Service provides the solution to manage and distribute the computing tasks maximising the usage of the computational resources. The service is based on DIRAC technology and is the main interface for submitting the jobs with a RESTful API.

Using this implementation Opencoasts can use more computational cores, not only on the Cloud infrastructure, but also from the computing GRID network.

The StoRM WebDAV service is used for data transfer between the Cloud and the computation infrastructure. It allows the access from the Opencoasts deployed services in the Cloud to the external computing environment, as defined in figure 2. With the RESTful interface provided by WebDAV implementation was possible to access the storage with common http requests, hiding the complexity of the storage services and the heterogeneity between different providers.

The udocker tool is used to run the processing scripts in the computing nodes with the required software dependencies and environment. The software is packaged inside a docker image to be launched with udocker in user space without requiring root privileges. The tool can be downloaded and executed entirely by the end user without requiring any type of privileges nor deployment of services by system administrators.

After setting into production all required services, EGI Accounting Thematic Service stores the user accounting records from the various services offered by EGI, in particular the Cloud resources, computing and storage usage. This information can be consulted online through the EGI Accounting Portal.

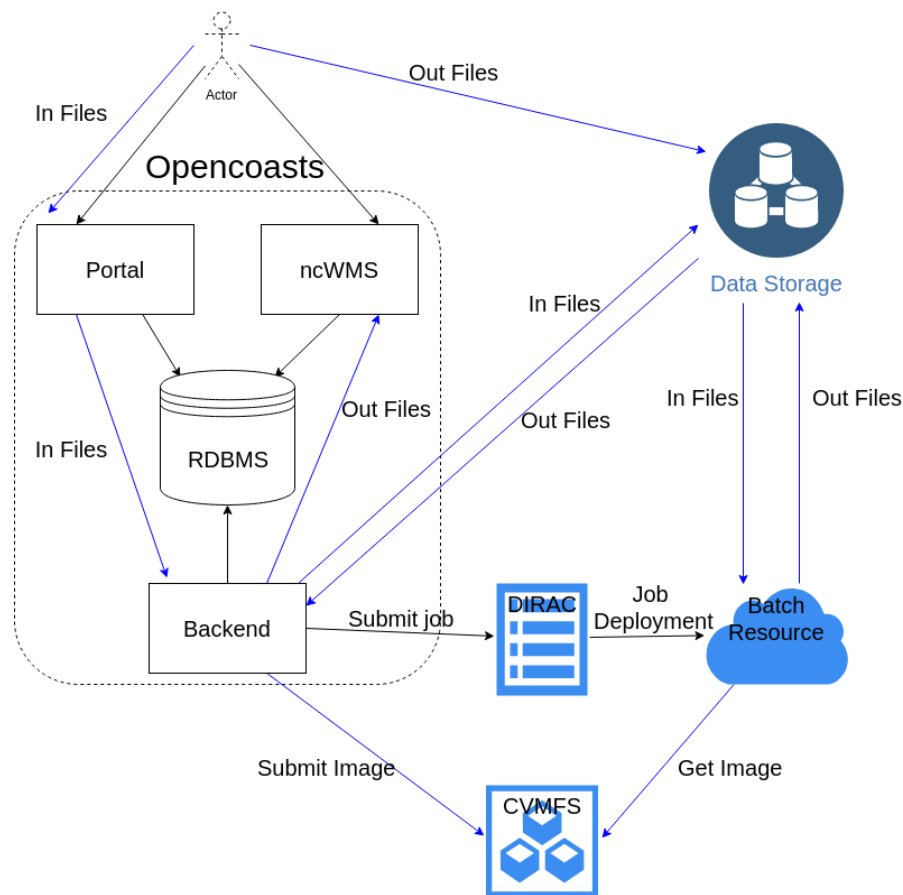


Fig. 15 - OPENCoastS integration for computational resources

6.3 Software release

OPENCoastS integration was planned considering a set of milestones, starting from the simplest version, which integrates the basic functionalities necessary to configure and deploy a forecast system on demand, to more complex versions that add more choices for users during the configuration process, such as different model settings and forcing sources to apply. These milestones are set by our scientific team members that are simultaneously part of the testers' team and also end users.

OPENCoastS development considers three environments:

- local development environment;
- local production environment;
- official environment.

Newly developed features, along with bug fixing and feature improvements, are always published first to the local development environment where they are fully tested and validated by developers and testers. As soon as the testers feedback is positive these changes are published and tested in the local production environment, that simulates more accurately the official environment; only then, and with the Project's manager approval, updates are published to all end users.

References (from D7.1)

Service endpoints:

-
- <https://opencoasts.ncg.ingrid.pt> - Opencoasts main portal
 - <https://opencoasts.ncg.ingrid.pt:9443> - Minio object store
 - <https://opencoasts.ncg.ingrid.pt:8443/ncWMS> - ncWMS (NetCDF files)

Software code and releases: the project's versioning is managed by a Git service, more specifically using Gogs. At the moment, the service is hosted on local servers. However, it is planned to move it to a public server in the future.

The user manual is available at:

- https://opencoasts.ncg.ingrid.pt/static/OPENCoastS_manual.pdf
- http://opencoasts.lnec.pt/OPENCoastS_manual.htm

Additional information for users is available at:

- http://opencoasts.lnec.pt/pdfs/IMUM_AFortunato.pdf
- http://opencoasts.lnec.pt/pdfs/aoliveira_final_imum_hands_on.pdf

Documentation for administrators and developers will be made available when the software code is placed in a public repository.

6.4 EOSC-hub integration

The latest stable version of the service is published via EOSC-hub:

- on the Marketplace: <https://marketplace.eosc-hub.eu/44-opencoasts>;
- on the Service Catalogue: <https://eosc-hub.eu/catalogue/OPENCoastS>.

7 WeNMR

7.1 Service description

Service/Tool name	WeNMR
Service/Tool url	https://www.eosc-hub.eu/services/WeNMR%20suite%20for%20Structural%20Biology
Service/Tool information page	http://www.wenmr.eu
Description	The WeNMR Thematic Services are a suite of web portals providing user-friendly access to complex computational workflows and tasks in the structural biology field.
Value proposition	The WeNMR portals allow inexperienced and experienced structural biologists to make use of state-of-the-art software for their research while benefiting from the computational infrastructure provided through the EOSC-hub project, in a transparent way.
Customer of the service/tool	Structural biology and bioinformatics communities, INSTRUCT-ERIC
User of the service/tool	Main users of the WeNMR are structural biologists of any degree of experience, with the aim of modeling and refining NMR structures, complexes of proteins and other biomolecules or fitting cryo-electron microscopy maps.
User Documentation	http://www.wenmr.eu
Technical Documentation	http://www.wenmr.eu
Product team	Utrecht University (UU), Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP), INFN Padova
License	All portals are freely accessible to non-profit users upon registration
Source code	N.A.
Testing	The portals are heavily used and have support mechanisms in place (e.g. via ask.bioexcel.eu). Updates to the portal machinery itself is subjected (in most cases) to continuous integration via GitHub/Jenkins.

The WeNMR Thematic Services are **a suite of web portals** providing user-friendly access to complex computational workflows and tasks in the **structural biology field**. The goal of these portals is to allow inexperienced and experienced structural biologists to make use of state-of-the-art software for their research while benefiting from the computational infrastructure provided through the EOSC-hub project, in a transparent way. As already described in Deliverables 7.1, 7.2 and 7.3 [R1, R2, R3], the WeNMR services make use of high-throughput computing (HTC) resources and some even of accelerated computing (GP-GPUs) grid resources and cloud computing. Seven services vertebrate the WeNMR suite: AMPS-NMR (Nuclear Magnetic Resonance structure refinement), CS-ROSETTA (3D structure modelling of proteins), DISVIS (visualizing and quantifying accessible

interaction space in macromolecular complexes), FANTEN (computing the magnetic susceptibility anisotropy tensor for paramagnetic metalloproteins), HADDOCK (modelling complexes of proteins and other biomolecules), POWERFIT (rigid body fitting of atomic structures into cryo-EM density maps) and SPOTON (identification and classification of interfacial residues as hot-spots in protein-protein complexes).

7.2 Service architecture and integrated services

Most of the WeNMR portals have been in operation since several years. They are geographically located in Utrecht in the Netherlands and at the University of Florence in Italy. All portals are web-based, built on a variety of technological solutions (e.g. Python, Flask, Apache, ...), but all present a unified and well-recognizable front end to users. They make use of the EGI HTC resources to distribute jobs to the sites supporting the enmr.eu VO using in most cases DIRAC4EGI for job submission, but also gLite in some specific cases, e.g. like the GPGPU-grid enabled DISVIS and POWERFIT portals that are sending jobs to specific CEs in Florence. These two applications make also use of udocker, a basic user tool to execute simple Docker containers in user space without requiring root privileges, developed by the INDIGO-DataCloud project and currently supported by EOSC-Hub.



Fig. 16 - Front end of the Utrecht WeNMR portals

Description of the WeNMR portals

AMPS-NMR

AMPS-NMR³³, (AMBER-based Portal Server for NMR structures) is a web interface to set up and run calculations with the AMBER package. The interface allows the refinement of NMR structures of biological macromolecules through restrained Molecular Dynamics (rMD). Some predefined protocols are provided for this purpose, which can be personalized; it is also possible to create an entirely new protocol. AMPS-NMR can handle various restraint types. As an ancillary service, it provides access to

³³ <http://py-enmr.cerm.unifi.it/access/index>

a web interface to AnteChamber, enabling the calculation of force field parameters for organic molecules such as ligands in protein–ligand adducts.

The AMPS-NMR grid-enabled web server, originally developed under the WeNMR e-Infrastructure project³⁴ uses resources provided by the EGI and the associated National Grid Initiatives (NGIs) [R7].

CS-Rosetta3

CS-ROSETTA is a protocol which generates 3D models of proteins, using only the ¹³CA, ¹³CB, ¹³C', ¹⁵N, ¹HA and ¹HN NMR chemical shifts as input. Based on these parameters, CS ROSETTA uses a SPARTA-based selection procedure to select a set of fragments from a fragment-library (where the chemical shifts and the 3D structure of the fragments are known). The fragments are assembled using the Rosetta protocol. The generated models are rescored based on the difference between the back-calculated chemical shifts of the generated models and the input chemical shifts, and when available, with a post-scoring procedure based on unassigned NOE lists.

The CS-Rosetta3 grid-enabled web server³⁵, originally developed under the WeNMR e-Infrastructure project uses resources provided by the EGI and the associated National Grid Initiatives (NGIs) [R8, R9].

Note that the CS-Rosetta3 server will be decommissioned in September 2020, due to very low usage in combination with old code that causes security issues.

DisVis

DisVis³⁶ allows visualizing and quantifying the information content of distance restraints between macromolecular complexes. It performs a full and systematic 6-dimensional search of the three translational and rotational degrees of freedom to determine the number of complexes consistent with the restraints. In addition, it outputs the percentage of restraints being violated and a density that represents the center-of-mass position of the scanning chain corresponding to the highest number of consistent restraints at every position in space [R10].

FANTEN

Pseudocontact shifts (PCSs) and residual dipolar couplings (RDCs) arising from the presence of paramagnetic metal ions in proteins as well as RDCs due to partial orientation induced by external orienting media are nowadays routinely measured as a part of the NMR characterization of biologically relevant systems. PCSs and RDCs can be used: 1) to determine and/or refine protein structures in solution, 2) to monitor the extent of conformational heterogeneity in systems composed of rigid domains which can reorient with respect to one another, and 3) to obtain structural information in protein-protein complexes. The use of both PCSs and RDCs proceeds through the determination of the anisotropy tensors which are at the origin of these NMR observables. A new user-friendly web tool, called FANTEN (Finding ANisotropy TENSors), has been developed for the determination of the anisotropy tensors related to PCSs and RDCs and has been made freely available through the WeNMR³⁷ gateway. The program has many features not available in other existing programs, among which the possibility of a joint analysis of several sets of PCS and RDC data and the possibility to perform rigid body minimizations [R11].

³⁴ www.wenmr.eu

³⁵ <http://haddock.science.uu.nl/enmr/services/CS-ROSETTA3/>

³⁶ <http://milou.science.uu.nl/services/DISVIS/>

³⁷ <http://fanten-enmr.cerm.unifi.it:8080>

HADDOCK

HADDOCK2.2 (High Ambiguity Driven protein-protein DOCKing) is an integrative, information-driven flexible docking approach for the modeling of biomolecular complexes. HADDOCK distinguishes itself from ab-initio docking methods in the fact that it encodes information from identified or predicted protein interfaces in ambiguous interaction restraints (AIRs) to drive the docking process. HADDOCK can deal with a large class of modelling problems including protein-protein, protein-nucleic acids and protein-ligand complexes [R12].

The HADDOCK2.2 grid-enabled web server, originally developed under the WeNMR e-Infrastructure project uses resources provided by the EGI and the associated National Grid Initiatives (NGIs).

HADDOCK2.4

A new version of HADDOCK, version2.4, has been officially released in May 2020. It's associated web portal has been in testing phase since several months and is in full operation as of May 2020 as well. The new portal is accessible from: <https://wenmr.science.uu.nl/haddock2.4>

The HADDOCK2.4 portal allows the modelling of up to 20 molecules and supports new types of experimental data, in particular from cryo-electron microscopy experiments. In the current COVID-19 pandemic, all HADDOCK portals are now allowing users to tag their submissions as COVID-related, which allows to monitor them and direct HTC jobs to resources specifically supporting COVID research.

PowerFit

PowerFit³⁸ automatically fits high-resolution atomic structures into cryo-EM densities. To this end it performs a full-exhaustive 6-dimensional cross-correlation search between the atomic structure and the density. It takes as input an atomic structure in PDB- or mmCIF-format and a cryo-EM density with its resolution and outputs positions and rotations of the atomic structure corresponding to high correlation values. PowerFit uses the local cross-correlation function as its base score. The score is by default enhanced with an optional Laplace pre-filter and a core-weighted version to minimize overlapping densities from neighbouring subunits [R13].

SpotON

SpotOn³⁹ is a robust algorithm developed to identify and classify the interfacial residues as Hot-Spots (HS) and Null-Spots (NS) with a final accuracy of 0.95 and a sensitivity of 0.95 on an independent test set. The predictor was developed using an ensemble learning algorithm with up-sampling of the minor class and was trained on a large number of complexes and on a high number of different structural- and evolutionary sequence-based features [R14].

General architecture of the WeNMR portals

All WeNMR portals are built on the same philosophy of shielding the end user from the complexity of accessing/using HTC (grid or cloud) resources. From a user perspective, a user only interacts with web-based portals, filling in forms and uploading data. Upon successful submission those data are processed through complex workflows calling typically a variety of software and using a combination of both local and EOSC HTC resources. Finally, the results are post-processed and presented to the user in a user-friendly manner, facilitating their interpretation. The general architecture is illustrated in the following figure.

³⁸ <http://milou.science.uu.nl/services/POWERFIT/>

³⁹ <https://milou.science.uu.nl/services/SPOTON/>

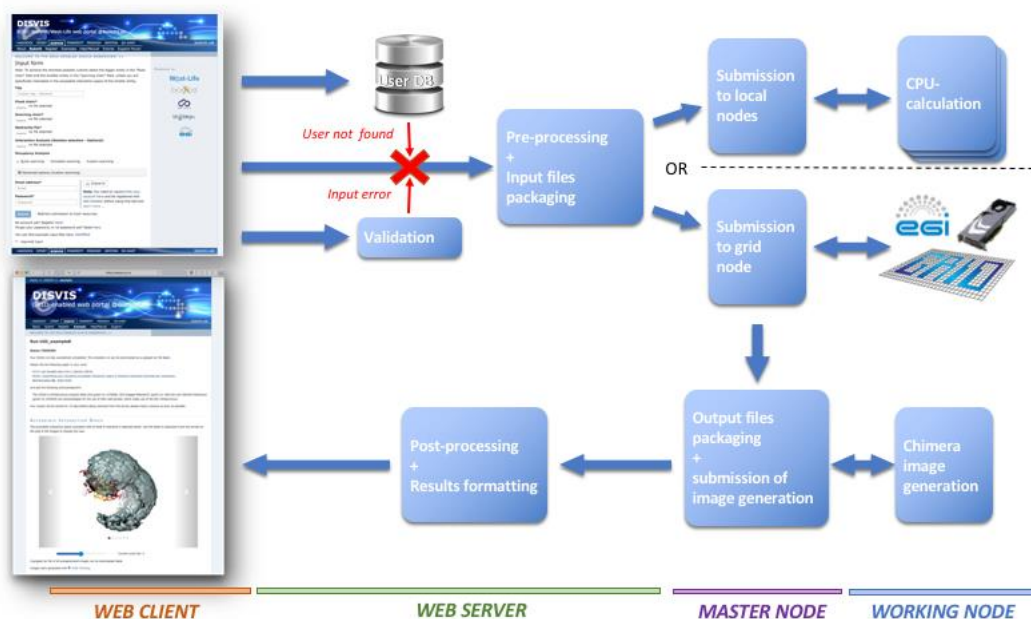


Fig. 17 - Illustration of the general workflow behind the WeNMR web portals

7.3 Software release

There is no general software release associated with the WeNMR Thematic Services at this time. In some cases, the software run behind the portal is from third parties and not developed within the WeNMR partners. In other cases, WeNMR partner are the developers of the software behind the portals and those are usually freely available and downloadable directly from GitHub. Other do require a specific license. Note that none of these software has been developed in the context of EOSC-hub (or previous related projects). They are the result of research work financed by a variety of national and international project.

Freely downloadable software:

- DISVIS <https://github.com/haddock/disvis>
also available as a Docker container from the INDIGO Docker hub:
 - <https://hub.docker.com/r/indigodatacloudapps/disvis/>
- POWERFIT <https://github.com/haddock/powerfit>
also available as a Docker container from the INDIGO Docker hub:
 - <https://hub.docker.com/r/indigodatacloudapps/powerfit/>

Software requiring a license (free for non-profit):

- HADDOCK <http://www.bonvinlab.org/software/haddock2.2>
- HADDOCK2.4 <http://www.bonvinlab.org/software/haddock2.4>

Third party software:

- AMBER <http://ambermd.org>
- CS-Rosetta <https://www.rosettacommons.org/docs/latest/CS-Rosetta>
- CNS (required for HADDOCK) <http://cns-online.org/v1.3/>

7.4 References

Portal-related publications:

- Bertini, I., Case, D. A., Ferella, L., Giachetti, A. & Rosato, A. A Grid-enabled web portal for NMR structure refinement with AMBER. *Bioinformatics* **27**, 2384–2390 (2011).
- Wassenaar, T. A. et al. WeNMR: Structural Biology on the Grid. *J Grid Computing* **10**, 743–767 (2012).
- Schot, G. & Bonvin, A. M. J. J. Performance of the WeNMR CS-Rosetta3 web server in CASD-NMR. *Journal of Biomolecular NMR* **62**, 497–502 (2015).
- G.C.P. van Zundert, M. Trellet, J. Schaarschmidt, Z. Kurkcuoglu, M. David, M. Verlato, A. Rosato and A.M.J.J. Bonvin. The DisVis and PowerFit web servers: Explorative and Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.*, **429**, 399-407 (2017).
- van Zundert, G. C. P. et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* **428**, 720–725 (2015). I.S. Moreira, P.I. Koukos, R. Melo, J.G. Almeida, A.J. Preto, J. Schaarschmidt, M. Trellet, Z.H. Gümüş, J. Costa and A.M.J.J. Bonvin. SpotOn: High Accuracy Identification of Protein-Protein Interface Hot-Spots. *Sci. Reports*. **7**:8007 (2017).

WeNMR web site with services, description, support and tutorials:

- <http://www.wenmr.eu>

EOSC-Hub related pages:

- <https://eosc-hub.eu/services/WeNMR%20suite%20for%20Structural%20Biology>
- <https://eosc-hub.eu/demos/demo-wenmr-suite-structural-biology>
- <https://eosc-hub.eu/news/haddock-support-covid-19-research>
- <https://eosc-hub.eu/news/eosc-practice-eosc-hub-wenmr-demonstrator>
- <https://eosc-hub.eu/eosc-in-practice-wenmr>
- <https://eosc-hub.eu/training-material/wenmr-suite-structural-biology>

7.5 EOSC-hub integration

From day 1 of the project, the WeNMR Thematic Services have been in operation, sending over the first eight months of the project over 5 million jobs to the EOSC HTC resources, most of which through the DIRAC4EGI service. According to the EGI Accounting Portal, these account for over 15 million HS06 CPU Time hours consumed. Although the WeNMR services operate making use of opportunistic computing resource, the access to the federated resources of EGI has been formalized by a Service Level Agreement (SLA) between EGI.eu and the enmr.eu VO (represented by the Faculty of Science – Department of Chemistry of Utrecht University). This SLA was established in 2016 and has been renewed in 2018, granting to enmr.eu VO until 31/12/2020 an amount of opportunistic computing time up to 53 Million of normalized CPU hours and opportunistic storage capacity up to 54 TB⁴⁰. Five resource centres signed this last version of the SLA: INFN-PADOVA (Italy), TW-NCHC (Taiwan), SURFSara and NIKHEF (The Netherlands), NCG-INGRID-PT (Portugal).

⁴⁰ <https://documents.egi.eu/public/ShowDocument?docid=2751>

During the first nine month of the project, a number of WeNMR portals have been migrated from the old gLite-based job submission to the EOSC-hub **DIRAC4EGI** service. Further, all portals are now offering Single Sign On (SSO), either through the West-Life SSO which connects to both ARIA (the access management solution of the Structural Biology infrastructure INSTRUCT-ERIC) and the old legacy WeNMR SSO⁴¹, or directly through the **EGI Check-in**. Users can now register and use the WeNMR services using the Check-in, allowing them to use a variety of identity providers. AMPS-NMR is currently using only the West-Life SSO, but is in process of including EGI Check-in too.

West-Life SSO integration

INFN hosts an instance of the EOSC-hub OneData service, offering up to 10 TB of storage space to WeNMR community. WeNMR users can request a storage space at the Oneprovider service hosted at INFN-Padova data center by connecting to the Onezone server (onezone.cloud.cnaf.infn.it) hosted at INFN-CNAF data center (located in Bologna). INFN developed a new plugin to enable West-Life SSO as authentication method for Onedata.

West-Life Virtual Folder integration with Onedata

Virtual Folder (VF) is a tool developed in the context of West-Life project and now maintained by INSTRUCT-ERIC⁴². Currently integrated in several WeNMR portals, it acts as a gateway for many storage systems, such as Dropbox, EOSC B2Drop and any other system accessible through the WebDAV protocol. Each protocol requires a specific plugin loaded by the engine of the VF framework for retrieving metadata for files and directories of the storage back-end. INFN developed a plugin for integrating VF with Onedata, i.e. to enable Onedata storage system as additional back-end. The current integration of VF with Onedata can only be considered as a prototype. It will become "production ready" only when a final release of Onedata is available.

⁴¹ https://indico.egi.eu/event/3903/sessions/2838/attachments/8596/9939/EOSC_HUB_Malaga18.pdf

⁴² <https://instruct-eric.eu/>

8 EO Pillar

The EO-Pillar within the EOSC provides access to different services established in the field of Earth Observation (EO). The services are categorised into three main classes: data access and computing services, data exploitation services, general user services.

The task T7.7 EO Pillar already ended in M15 (i.e.03/2019). The services are already completely described in D7.2 and D7.3; here the sections have been copied again for completeness.

8.1 Service description

These services are:

- The Geohazard Exploitation Platform (GEP)
- EODC JupyterHub for global Copernicus data
- EODC Data Catalogue Service
- Sentinel Hub
- OSX-Sentinel
- MEA Platform (Data access and exploitation service)
- Rasdaman EO Datacube
- CloudFerro Data Collections Catalog
- CloudFerro Infrastructure
- CloudFerro Data Related Services - EO Finder
- CloudFerro Data Related Services - EO Browser
- EPOSAR service

Further details are summarised in the following tables:

Service/Tool name	Geohazards Exploitation Platform (GEP)
Service/Tool url	https://geohazards-tep.eu/#! https://geohazards-tep.eu/geobrowser/?id=eoschub-landslide-app https://geohazards-tep.eu/geobrowser/?id=eoschub-alpsmonitoring-app
Service/Tool information page	https://geohazards-tep.eu/#!
Description	<p>The Geohazards TEP (GEP) is an enhancement of the precursor platforms (G-POD, SSEP), and is designed to support the Geohazard Supersites (GSNL) and the Geohazards community via the CEOS WG Disasters. GEP is an ESA originated R&D activity on the EO ground segment to demonstrate the benefit of new technologies for large scale processing of EO data. Its goal is to apply a complementary operations concept, i.e. moving User activities to the Data: users access a work environment containing the data and resources required, as opposed to downloading and replicating the data 'at home'. The Geohazards TEP platform is a complex system composed by the following main components:</p> <ul style="list-style-type: none"> - Web portal

	<ul style="list-style-type: none"> - Data Agency including <ul style="list-style-type: none"> - Catalogue service - Data Gateway - Production Center - Development environment <p>In the context of EOSC_Hub it exposes two main services:</p> <ul style="list-style-type: none"> - A service providing an interferometric & coherence product at 50m resolution and 25m pixel spacing systematically generated every 6-days, for each new Sentinel-1 SLC pair, over the Alpine Region. It supports ground deformation monitoring, as well as rapid response to earthquakes occurring within the processing mask, by automatic generation of co-seismic interferograms that are published in a dedicated GeoBrowser and made available for visualization and download. - a service providing access to a set of on-demand terrain motion applications supporting: interferogram generation, co-seismic displacement mapping, landslide rapid mapping and landslide displacement field monitoring with Sentinel-1 and Sentinel-2 data.
Value proposition	Exploitation of ICT resources from multiple providers federated in EOSC-Hub for enhanced processing capabilities and improved scalability. Integration with EOSC-Hub AAI to allow seamless access to the platform services to a growing number of research communities.
Customer of the service/tool	Research Communities
User of the service/tool	large research groups
User Documentation	https://terradue.github.io/doc-tep-geohazards/ https://geohazards-tep.eo.esa.int/#!blog
Technical Documentation	https://docs.terradue.com/t2-api/
Product team	Terradue Srl
License	-
Source code	https://github.com/geohazards-tep
Testing	-

Service/Tool name	EODC Data Catalogue Service and EODC JupyterHub for global Copernicus data
Service/Tool url	https://csw.eodc.eu/ https://jupyterhub.eodc.eu

Service/Tool information page	https://marketplace.eosc-portal.eu/services/eodc-data-catalogue-service https://marketplace.eosc-portal.eu/services/eodc-jupyterhub-for-global-copernicus-data
Description	<p>The EODC Data Catalogue service allows querying the Copernicus Sentinel satellite data hosted at EODC. The service is available through a simple Web GUI, eomEX+, as well as an API. The back-end of eomEX+ is the EODC pycsw server, an implementation of an OGC CSW server. As a consequence, the eomEX+ API is accompanied by an expert level API provided by the EODC CSW server, located at link.</p> <p>EODC JupyterHub provides access to the global EODC Copernicus Data Archive. Eliminate the barrier to get access to and to start to develop algorithms in view of remote sensing data. Full access to EODC Copernicus satellite data archive. No need to setup a VM to access data archive. Simple python dev. environment.</p>
Value proposition	The service provides a data catalogue service that allows to search Copernicus data available at EODC based on an up-to-date metadata catalogue and the EODC JupyterHub provides direct access to the global EODC Copernicus Data Archive.
Customer of the service/tool	Research Communities
User of the service/tool	Research groups
User Documentation	https://eomex.eodc.eu/manual
Technical Documentation	https://eomex.eodc.eu/manual
Product team	EODC
License	Different licenses for the individual components
Source code	git.eodc.eu
Testing	-

Service/Tool name	Sentinel Hub
Service/Tool url	https://services.sentinel-hub.com/
Service/Tool information page	https://sentinel-hub.com
Description	Sentinel Hub is a cloud platform that provides versatile access to Earth observation data and enables its processing with user-defined algorithms. It is optimised for on-the-fly computation, producing the result whenever

	and in such form as the user requires, directly from original satellite data files without replication or pre-processing.
Value proposition	E.g. Improved portal design that allows new and easier way to access and visualise data for the final users. Third parties can now gather monitoring data from the system through a complete API. A central deployment of the ARGO monitoring engine can serve a large infrastructure reducing the maintenance costs.
Customer of the service/tool	Various companies providing access to EO data and their analysis, e.g. precision farming, security, oil&gas, etc. Governments for monitoring of agriculture, land administration.
User of the service/tool	Application developer Data scientists Machine learning experts (EO)
User Documentation	https://docs.sentinel-hub.com https://www.sentinel-hub.com/develop/documentation/api/ogc_api
Technical Documentation	https://docs.sentinel-hub.com https://www.sentinel-hub.com/develop/documentation/api/ogc_api
Product team	Sinergise
License	Proprietary
Source code	Proprietary
Testing	Automated testing based on thousands of random requests created by users. In the last 30 days, comparing the outputs with production and development version of the service.

Service/Tool name	ADAM (former MEA Platform)
Service/Tool url	https://explorer.adamplatform.eu
Service/Tool information page	https://adamplatform.eu
Description	The Advanced geospatial Data Management platform (ADAM) is a tool to access a large variety and volume of global environmental data.

Value proposition	ADAM allows you extracting global as well as local data, from the past, current time, as well as short term forecast and long-term projections. Most of the data are updated daily to allow users having always the most recent data to play with.
Customer of the service/tool	E.g. RI; Resource Provider; Research Communities
User of the service/tool	Earth scientists; Professionals;
User Documentation	https://adamplatform.eu Short videos focused on specific applications (e.g. study of agricultural parcels using Landsat 8, MODIS and NOAA precipitation data, burnt areas detection using Sentinel 2 data, computation of statistical parameters using processing functions) are available on the application on the ADAM's YouTube channel .
Technical Documentation	https://adamplatform.eu
Product team	MEEO

Service/Tool name	Rasdaman datacubes
Service/Tool url	http://eoschub.rasdaman.com/rasdaman/ows
Service/Tool information page	http://eoschub.rasdaman.com
Description	The rasdaman datacube engine provides fast and flexible access, analytics, and fusion on massive spatio-temporal CreoDIAS data assets.
Value proposition	Thanks to the zillions of files homogenized into few datacubes data wrangling becomes substantially easier for both humans and machines, such as Machine Learning tools. Additionally, the CreoDIAS rasdaman instance is member of the EarthServer federation thereby making available the data from all EarthServer partners for location-transparent mix & match.
Customer of the service/tool	Value-adding remote sensing businesses; researchers; authorities
User of the service/tool	Earth data analysts
User Documentation	http://eoschub.rasdaman.com/eosc/index.html?page=demonstrations
Technical Documentation	http://www.earthserver.eu/webinars
Product team	Rasdaman

Detailed descriptions of the services are already provided in D7.2 (Section 8, T7.7 EO-Pillar) and here: <https://wiki.eosc-hub.eu/display/EO+Pillar+VA>

8.2 Service architecture and integrated services

The different services utilise different architectures, protocols and standards.

8.2.1 Multi-sensor Evolution Analysis (MEA) platform

MEA platform implements the concept of Digital Earth making global environmental geospatial data Findable, Accessible, Interoperable and Reusable (FAIR). MEA platform provides an effective sub-setting functionality that accesses the data only when requested and serves to the client only the data amount that is really needed. It is a cross-domain application that supports various application fields: climate health, land cover, pollution, agriculture, weather, oceanography.

The MEA platform is data exploitation (via GUI) and data discovery and access (via DAS component) services:

- Graphic User interface: <https://eodataservice.org>. This UI provides access to data available through DAS components distributed across several infrastructures. Three types of user interfaces are provided:
 - web based graphic user interface
 - Jupyter notebook (python console with possible extension to 40 different programming languages)
 - APIs: OGC-compliant (OpenSearch, WCS2.x) RESTful interfaces
- Data Access System (DAS), this component enables service discovery and access via OGC standard interfaces (OpenSearch, WCS):
 - <http://snf-823526.vm.oceanos.grnet.gr/> (access to Regional data(cubes): ground measurements, numerical simulations, ...)
 - <https://sentinel.eodataservice.org/> (access to Regional and Global data(cubes): Sentinel-1, Sentinel-2, Sentinel-3, Sentinel-5P)
 - <https://landsat.eodataservice.org/> (access to European data(cube): Landsat-8)
 - <http://dar05.eodataservice.org> (access to Global thematic data(cube): MODEL Atmospheric Level 2 and Level 3 products, ...)

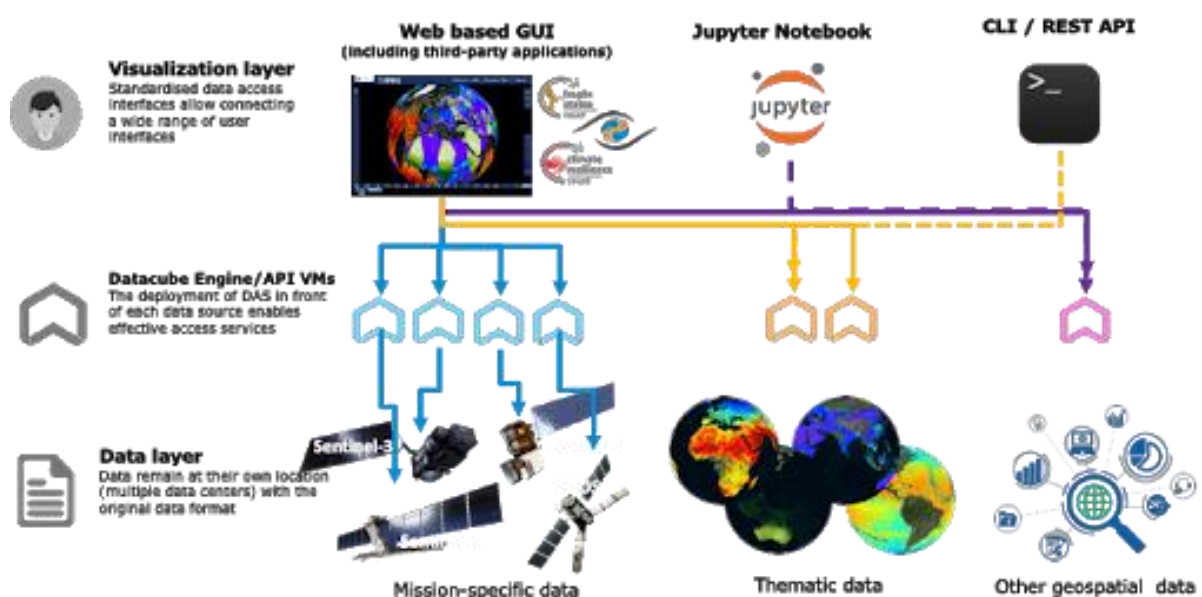


Fig. 18 - MEA platform

8.2.2 Geohazards Exploitation Platform (GEP)

The Geohazards TEP (GEP) is an enhancement of the precursor platforms (G-POD, SSEP), and is designed to support the Geohazard Supersites (GSNL) and the Geohazards community via the CEOS WG Disasters. GEP is an ESA originated R&D activity on the EO ground segment to demonstrate the benefit of new technologies for large scale processing of EO data. Its goal is to apply a complementary operations concept, i.e. moving User activities to the Data: users access a work environment containing the data and resources required, as opposed to downloading and replicating the data ‘at home’. The Geohazards TEP platform is a complex system composed by the following main components:

- Web portal
- Data Agency including
 - Catalogue service
 - Data Gateway
- Production Center
- Development environment

8.2.3 EODC Data Catalogue Service

The EODC Data Catalogue service allows querying the Copernicus Sentinel satellite data hosted at EODC. The service is available through a simple Web GUI, eomEX+⁴³, as well as an API⁴⁴. The back-end of eomEX+ is the EODC pycsw server⁴⁵, an implementation of an OGC CSW server⁴⁶. As a consequence, the eomEX+ API is accompanied by an expert level API provided by the EODC CSW server, located at <https://csw.eodc.eu>. Further details can be found here: <https://eomex.eodc.eu/manual>.

8.2.4 EODC JupyterHub for global Copernicus data

This service is based on an implementation of JupyterHub at EODC. The service functions as a starting point to get free access to Copernicus Sentinel satellite data hosted at EODC. In addition, the service facilitates the development and realizations of algorithms. Functions and algorithms can be executed directly on the data by utilizing well-known environments such as Jupyter notebooks and a simple terminal interface. It can be accessed via <https://jupyterhub.eodc.eu/hub/login>. Support is provided via support@eodc.eu.

8.2.5 Sentinel Hub: Real-time processing of EO data

Sentinel Hub OGC API is optimized for on-demand on the fly processing of raw (unchanged) EO data. The following steps are typically performed within a standard request:

- query Catalogue for chosen AOI, time range, cloud coverage, mission, etc.
- download necessary data from on-line storage
- decompress data
- application of pre-mosaic filters (e.g. DOS-1, statistical atmospheric correction, etc.)
- creation of mosaic based on priority (e.g. most recent data on top), cloud replacement, etc.

⁴³ <https://eomex.eodc.eu/>

⁴⁴ <https://eomex.eodc.eu/manual>

⁴⁵ <http://pycsw.org/>

⁴⁶ <http://www.opengeospatial.org/standards/cat>

- compositing relevant bands to chosen EO product (true color, false color, NDVI, etc.) using chosen style (greyscale, RGB, red temperature, etc.)
- application of post-mosaic filters (color balance, HDR, midtone, gamma correction)
- re-projection to chosen CRS (e.g. Popular Web Mercator, WGS 84, national CRS systems).
- output creation in chosen file type (JPG, PNG, JP2, JPX, GeoTiff, etc.)
- compression of the output for faster download, based on user settings

Note that user can choose based on the parameters, in case some steps should be performed or not. A typical scenario described above takes about 1-2 second for an area of interest with a size of 1000*1000 px (depending on the chosen scale this may represent either 100 sq.km up to 25.000 sq.km).

8.2.6 Rasdaman datacube

The rasdaman datacube engine is a multi-parallel, federated array database system optimized for flexibility, performance, and horizontal/vertical scalability. Its interfaces consist of easy-to-use OGC WMS and WCS APIs plus high-level, declarative, standardized query languages (OGC WCPS and ISO Array SQL) allowing any query, any time, on any size.

8.2.7 Sentinel Data Hub

GRNET offers a dedicated instance for the Sentinel Data hub (OSX – Sentinel⁴⁷), that retrieves Sentinel Products from the National Sentinel Data Mirror Site⁴⁸ which is a web based system designed to provide EO data users with Search - Cataloguing - Order and Dissemination capabilities for the Sentinel products. Furthermore, GRNET offers upon request cloud resources (VMs & Storage) from its primary IAAS facility ~Okeanos. GRNET is currently working to integrate OSX-Sentinel with the EOSC-hub AAI ecosystem via EGI Check-in. This will allow users to have seamless access to the service using their credential from their home institution.

8.2.8 CloudFerro Data Collections Catalog (CloudFerro)

The Catalogue is based on CKAN open source software, which is widely used for open data publications like European Data Portal, data.org.uk or danepubliczne.gov.pl. CKAN provides a user-friendly web interface for all activities associated with data publication and subscription. It is capable of advanced data management. All datasets are organized and described with metadata, which allows them to be easily discoverable, with the use of search phrases and customizable filters (e.g.: tags, categories, data formats). It is possible to publish one dataset in different data formats, not only as downloadable files but also as links to web service, web API or links to external WWW resources. Datasets can be stored along with version history and dataset statistics, which allows to monitor the interest in datasets. CKAN also provides functionalities for collaboration, community participation and providing feedback, such as comments, ratings and sharing. It is highly customizable in both terms of Look&Feel and functionalities. Besides, it provides very rich RESTful JSON API, which allows other applications to discover and access the datasets, and it can be integrated easily with Semantic Web technologies such as RDF data model and SPARQL.

8.2.9 CloudFerro Infrastructure (CloudFerro)

CloudFerro infrastructure covers full set of virtual resources available in the solution:

⁴⁷ <https://sentinel.eosc.grnet.gr>

⁴⁸ <https://sentinels.space.noa.gr/>

- VM – Virtual Machines (or virtual computing servers) with several operating systems available (both free like CentOS, Ubuntu, Debian, Scientific Linux, and commercial like RedHat, SUSE, Microsoft Windows Server),
- virtual storage volumes that can be easily mounted to the VMs together with object storage solution,
- virtual networks,
- virtual appliances like firewalls (FWaaS) and VPN concentrators (VPNaaS),
- physical servers (baremetal) that can be integrated to the virtual world,
- Single Server VMs – full physical server with a single VM and very fast pass-through NVMe storage – a combination of advantages of a dedicated server and a cloud VM (high capacity, storage speed, no noisy neighbour problem).

8.2.10 CloudFerro Data Related Services - EO Finder (CloudFerro)

Catalog and a search engine dedicated to Earth Observation products. Its main purpose is to handle EO satellite imagery but it can be used to store any kind of geospatialized data.

EO Finder search API is compliant with the CEOS OpenSearch Best Practice Document and is mentioned in ESA's "Exploitation Platform Common Core Components" as the closest implementation of a catalogue component according to the requirements specified in ESA's "Exploitation Platform Open Architecture". It provides API-s in OpenSearch⁴⁹, GEOJson⁵⁰ and ATOM standards allowing for easy data search. The interface exposes the operational metadata catalogue.

8.2.11 CloudFerro Data Related Services - EO Browser

<https://github.com/sentinel-hub/EOBrowser>

8.2.12 EPOSAR service

EPOSAR is one of the services of the EPOS infrastructure (www.epos-ip.org). EPOSAR, based on the Small BASeline Subset (SBAS) DInSAR technique, is targeted to generate Earth surface displacement maps and time series with sub-centimetric accuracy by exploiting Sentinel-1 (S1) data of the Copernicus Programme. The service implements the whole processing chain, from SAR data retrieval until the generation of geocoded products. It is able to efficiently manage very large SAR datasets (hundreds or thousands of acquisitions) in very large computing environments (it has been used with 280 AWS instances). The products are generated in automatic and systematic way over several selected areas of the Earth and are continuously updated when new S1 acquisitions are available. The achieved products can be discovered, visualized and downloaded by users through the web interface of the EPOS infrastructure.

8.3 Software release

The different services utilise different release mechanisms. No common procedure is defined for the different services. We can expect individual roadmaps. This decision is performed by the individual provider.

⁴⁹ <http://www.opensearch.org>

⁵⁰ <http://geojson.org/>

8.4 References

- <https://geohazards-tep.eo.esa.int/#!>
- <https://geohazards-tep.eo.esa.int/geobrowser/?id=eoschub-alpsmonitoring-app>
- <https://geohazards-tep.eo.esa.int/geobrowser/?id=eoschub-landslide-app>
- <https://jupyterhub.eodc.eu>
- <https://csw.eodc.eu/>
- <http://sentinel-hub.com/>
- <https://eodataservice.org>
- <https://discovery.creodias.eu/dataset>
- <https://creodias.eu>
- <https://finder.creodias.eu/www>
- <https://browser.creodias.eu>
- <https://github.com/cloudferro>
- <https://github.com/sentinel-hub/EOBrowser>
- rasdaman: [geo-services guide](#); [WCS, WCPS and WMS tutorial](#); [webinars](#)
- <https://sentinel.eosc.grnet.gr>

8.5 EOSC-hub integration

The EO Pillar services are available via the EOSC Marketplace:

- <https://marketplace.eosc-portal.eu/services/cloudferro-data-related-services-eo-browser>
- <https://marketplace.eosc-portal.eu/services/gep-eo-services-for-earthquake-response-and-landslides-analysis>
- <https://marketplace.eosc-portal.eu/services/eodc-data-catalogue-service>
- <https://marketplace.eosc-portal.eu/services/cloudferro-data-related-services-eo-finder>
- <https://marketplace.eosc-portal.eu/services/eodc-jupyterhub-for-global-copernicus-data>
- <https://marketplace.eosc-portal.eu/services/cloudferro-infrastructure>
- <https://marketplace.eosc-portal.eu/services/cloudferro-data-collections-catalog>
- <https://marketplace.eosc-portal.eu/services/sentinel-hub>
- <https://marketplace.eosc-portal.eu/services/gep-eo-services-for-earthquake-response-and-landslides-analysis>
- <https://marketplace.eosc-portal.eu/services/rasdaman-eo-datacube>

9 DARIAH

9.1 Service description

Service/Tool name	DARIAH Science Gateway
Service/Tool url	https://dariah-sg.irb.hr
Service/Tool information page	https://dariah-sg.irb.hr
Description	The DARIAH Science Gateway is a web-oriented portal, developed during the EGI-Engage project (DARIAH Competence Centre) and is specially tailored for the researchers coming from digital arts and humanities disciplines. It currently offers several cloud-based services and applications: Semantic and Parallel Semantic Search Engines (SSE, PSSE), <u>DBO@Cloud</u> , Workflow Development and supports several file transfers protocols.
Value proposition	Centralized portal for various digital humanities services connected with AAI EGI. Easy integration of new services and high-level abstraction for accessing and exploiting Cloud computing infrastructure (IaaS) for inexperienced users.
Customer of the service/tool	Digital Humanities research groups and projects
User of the service/tool	Individual researchers, small to large research groups
User Documentation	DARIAH Science Gateway help pages: https://dariah-sg.irb.hr/help2 Simple Cloud access service: https://ltos-gateway.lpsd.sztaki.hu/web/wizard/help Gutenberg use case for Simple Cloud access service: https://bitbucket.org/davordavidovic/textanalysis/src/master/
Technical Documentation	-
Product team	RBI, SZTAKI
License	-
Source code	gUSE Grid and Cloud Science Gateway: https://sourceforge.net/projects/guse/
Testing	-

Service/Tool name	Invenio-based repository
Service/Tool url	https://dariah-portal.cloud.ba.infn.it/
Service/Tool information page	https://dariah-portal.cloud.ba.infn.it/
Description	The Invenio-based repository is the cloud is a service that enables researchers and scholars to easily create, deploy and configure their own

	Invenio-based repository and host it on the cloud infrastructure (Federated Cloud). The service aims to a smaller research groups lacking in adequate technical support and budget to acquire their own infrastructure for hosting data repositories.
Value proposition	Easy creation of a new (empty) instance of the Invenio-based repository in the Cloud. The initialization and setup of a new repository instance does not require technical knowledge on configuring Invenio nor Cloud infrastructure.
Customer of the service/tool	Digital and art and humanities research groups and communities. General research communities
User of the service/tool	Small research group, small project in the domain of digital arts and humanities
User Documentation	Deploying open-data repository in the Cloud (video): https://www.indigo-datacloud.eu/deploy-open-data-repository-cloud-using-marathon
Technical Documentation	-
Product team	RBI, Indigo-DataCloud
License	GNU General Public Licence 2
Source code	DockerHub container: https://hub.docker.com/r/indigodatacloudapps/dariah-repository Source code: https://github.com/indigo-dc/dariah-repository Ansible roles: https://github.com/indigo-dc/ansible-role-dariahrepo
Testing	-

Service/Tool name	DARIAH Repository
Service/Tool url	https://projects.gwdg.de/projects/dariah-de-repository
Service/Tool information page	https://dariah-sg.irb.hr
Description	DARIAH repository is a service based on the Common Data Storage ARchitecture (CDSTAR), a system for easy-to-use storage system that can store, modify, annotate, search, and access a large amounts of diverse data, both structured and unstructured.
Value proposition	The system integrates the ability of storing metadata along the research data in a flexible metadata schema that can be tailored for the specific use in different scientific disciplines. Additionally, the data objects that are stored in CDSTAR can be registered automatically at the EPIC Persistent Identifier (PID) service. The EPIC service gives data sets a unique, globally resolvable identifier as an additional abstraction layer that allows citing data sets in scientific publications. A role-based security concept is also integrated into CDSTAR, which allows the protection of data sets with an individual set of permissions and rights for each user. Additionally, CDSTAR is capable to use SAML infrastructures such as the one provided by DARIAH in order to verify data access. This allows to use the data permissions for the infrastructure at a central data point provided by DARIAH and to use the DARIAH access credentials.

Customer of the service/tool	DARIAH, Digital arts and humanities communities
User of the service/tool	Individual researchers, research groups, institutions, research projects
User Documentation	https://info.gwdg.de/docs/doku.php?id=en:services:storage_services:gwdg_cdstar:start
Technical Documentation	Rest API: https://www.gwdg.de/documents/20182/31188/gwdg-bericht-78.pdf
Product team	GWDG
License	-
Source code	https://github.com/clld/pycdstar
Testing	-

The DARIAH Thematic Services is a set of web-based services that provide user-friendly solutions (tools and applications) addressing the needs of different research groups within the DARIAH community and digital arts and humanities research domain in general. Although the DARIAH TS cannot address all the requirements, it provides a set of services enabling end-users to seamlessly store, describe and share their datasets as well as to discover, browse and reuse datasets shared by others and to perform elemental analysis on those data. The DARIAH TS services are aiming at the DARIAH-ERIC community providing free access to its members. For more details on the DARIAH TS services refer to the deliverables D7.1, D7.2 and D7.3 [R1, R2, R3].

Currently, a new web-oriented service for interactive analysis of digital humanities data using Python is under development and integration. The service is based on the EGI Notebook service.

The DARIAH Thematic Service provides three independent services:

- DARIAH Science Gateway,
- Invenio-based repository in the cloud,
- DARIAH repository.

The DARIAH Science Gateway is a web-oriented portal offering several cloud-based services and applications: Semantic and Parallel Semantic Search Engines, DBO@Cloud, Workflow Development and supports several file transfer protocols.

The Invenio-based repository is a service that enables easy creation, deployment and configuration of Invenio-based repository instances and hosting on the cloud infrastructure.

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

9.2 Service architecture

DARIAH Science Gateway

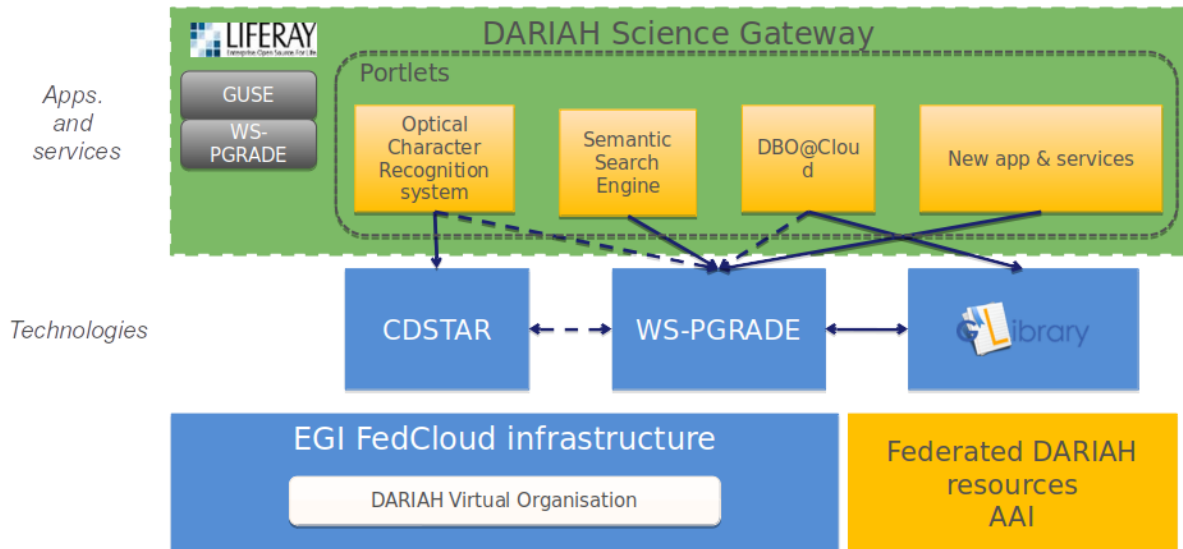


Fig. 19 - DARIAH Architecture

DARIAH Science Gateway is based on the WS-PGRADE/gUSE, an open-source science gateway platform. The gateway consists of a set of portlets that provide specific functionality or services integrated into the portal. The gateway currently provides 5 services: DBO@Cloud, Semantic Search Engine, Parallel Semantic Search Engine, Simple Cloud Access and workflow management. Each of the services is integrated as a portlet and is based on different underlying frameworks: WS-PGRADE, gLibrary and CDSTAR.

Invenio-based repository

The Invenio-based repository services main functionality is to provide easy deployment of a document repository in the Cloud. The user interface is based on FutureGateway, which provides functionality such as the basic configuration of the repository, e.g. required resources in terms of the number of CPUs and storage size and fine-tuned (optionally) configuration of the Mesos cluster size and per-sub-service resource allocation. The user then starts the process of creating a new instance of the repository in the cloud. The user request is submitted as a TOSCA template to the PaaS Orchestrator that coordinates resource provisioning and virtual infrastructure deployment. Once deployed virtual machines are automatically configured running dedicated Ansible roles shared on Ansible Galaxy. The Invenio components are installed via Ansible and executed as long-running services on top of a Mesos cluster through its framework, Marathon, to ensure fault tolerance and scalability. The IP address of the newly created repository instance is returned to the user via FutureGateway. Afterwards, the user can access the repository using the provided IP.

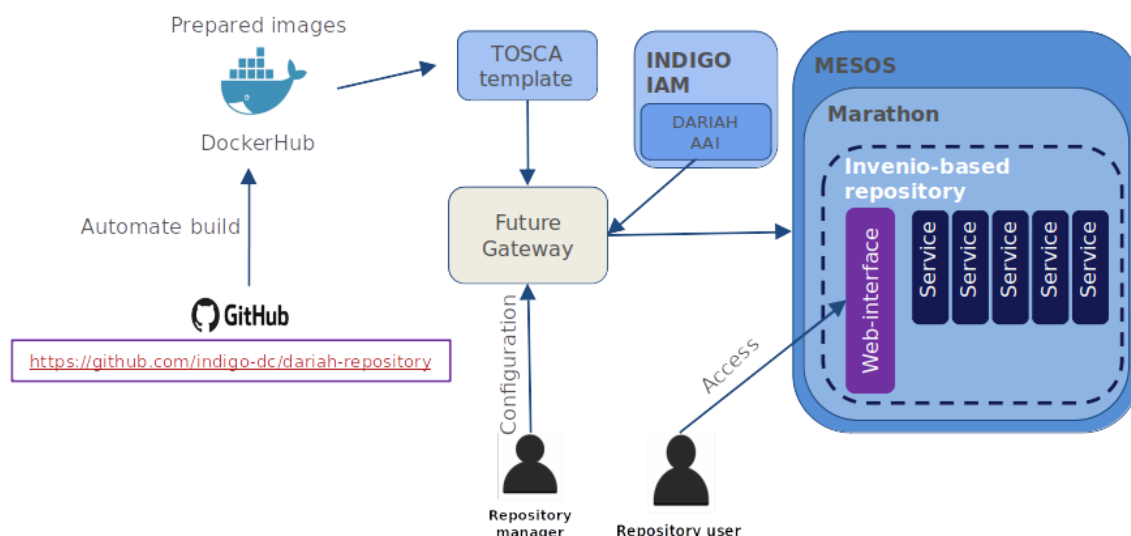


Fig. 20 - Invenio-based repository in the cloud

DARIAH repository

DARIAH-DE repository is a digital long-term archive for human and cultural-scientific research data. The repository is a central component of the DARIAH-DE Research Data Federation Infrastructure, which aggregates various services and applications and can be used comfortably by DARIAH users. The repository allows users to save their research data in a sustainable and secure way, to provide it with metadata and to publish it. The collections as well as individual files are available in the DARIAH-DE Repository in the long term and gets a unique and permanently valid Digital Object Identifier (DOI) with which the data can be permanently referenced and cited. In addition, one can register his collections within the Collection Registry, which are then also found in the Generic Search. The overall architecture of the DARIAH-DE repository is depicted in figure 10. An in-depth technical description can be found here: <https://repository.de.dariah.eu/doc/services>.

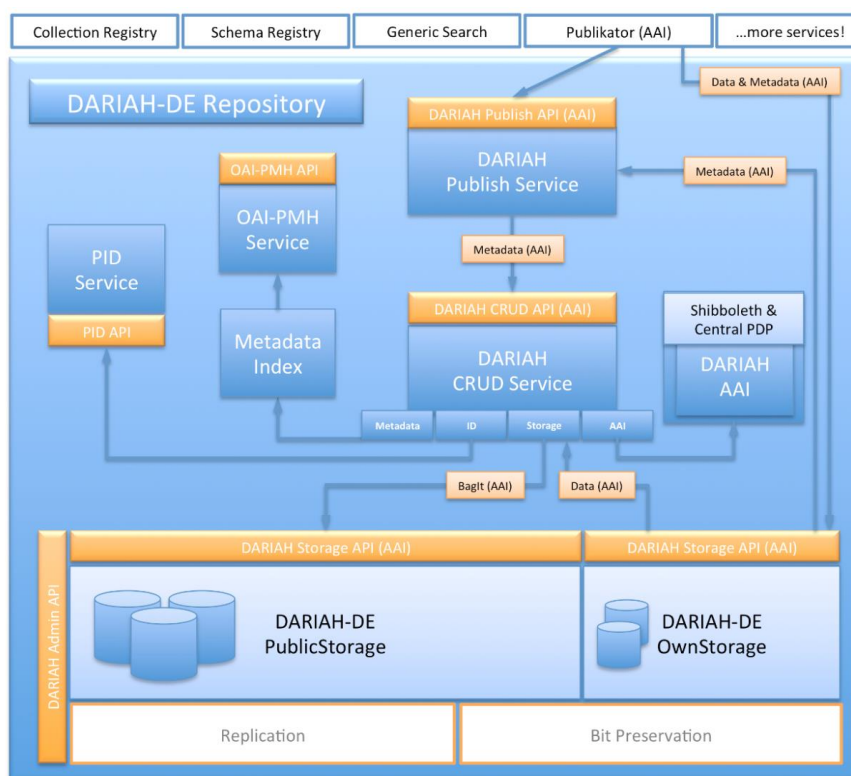


Fig. 21 - Overall scheme of the DARIAH repository

9.3 Software release

Currently, there is no strict concept of releases in DARIAH TS and each service has its own release policy, update and versioning policy. The DARIAH Science Gateway updates are connected with new releases of the gUSE/WS-PGRADE system, though, the Gateway does not have any code versioning. The code changes are verified manually on local machines before applying them into the production gateway instance. The Invenio-based service (not Invenio itself) does not have versioning.

The Invenio-based repository also does not have any code versioning, although, with a new Invenio framework release, the repository service can be updated (based on the users' request). In this particular case, the new software release does not affect the already running repository instances, only the new ones that will be launched based on the new Invenio release. In that case, the deployment process has to be validated and tested manually, to test for the correctness of the deploying particular Invenio sub-components and its inter-dependencies.

The DARIAH repository is developed, tested, maintained and released based on the release-management processes defined by DARIAH-DE. Technically, the CI-system is based on Jenkins and Puppet. The complete process is documented in the DARIAH-DE Wiki.

9.4 References

- Service endpoints:
 - DARIAH Science Gateway - <https://dariah-gateway.lpds.sztaki.hu/> (a new web address will be provided soon)
 - Invenio-based repository - under deployment

-
- DARIAH repository - <https://repository.de.dariah.eu/publikator>
 - Software code and releases:
 - **DARIAH Science Gateway:** guse/WSPGrade portal: <https://sourceforge.net/projects/guse/>
 - **Invenio-based repository:**
 - Invenio-based repository source code: <https://github.com/indigo-dc/dariah-repository>
 - Ansible role to start Dariah Repository containers on Marathon: <https://github.com/indigo-dc/ansible-role-dariahrepo>
 - **GWDG DARIAH repository:**
 - CDSTAR: <https://github.com/gwdg/cdstar>
 - Documentation, user oriented, administrator oriented and developer oriented:
 - **DARIAH Science Gateway:**
 - WSPGRADE usage and examples for end-users: <https://sourceforge.net/projects/guse/files/WS-PGRADE-Cookbook.pdf/download>
 - Documentation for developers: [https://sourceforge.net/projects/guse/files/Cookbook for Science Gateway Developers.pdf/download](https://sourceforge.net/projects/guse/files/Cookbook%20for%20Science%20Gateway%20Developers.pdf/download)
 - **Invenio-based repository in the cloud:** local repository installation for administrators/developers: <https://github.com/indigo-dc/dariah-repository/blob/master/INSTALL.rst>
 - **GWDG DARIAH repository:**
 - CDSTAR (end-user documentation): https://info.gwdg.de/docs/doku.php?id=en:services:storage_services:gwdg_cdstar:start
 - DARIAH repository (architecture): <https://repository.de.dariah.eu/doc/services>
 - DARIAH repository (end-user documentation): <https://wiki.de.dariah.eu/display/public/Das+DARIAH-DE+Repository>

9.5 EOSC-hub integration

Integrated resources

The DARIAH Science Gateway and the Invenio-based repository are integrated with the EGI FedCloud compute and storage resources. The authentication of DARIAH Science Gateway is integrated with the EGI Check-in.

Access channels

The latest stable Science Gateway service is published via EOSC-hub:

- on the Marketplace: <https://marketplace.eosc-hub.eu/49-dariah>
- on the Service Catalogue: <https://eosc-hub.eu/catalogue/DARIAH%20Science%20Gateway>

10 LifeWatch

10.1 Service description

Service/Tool name	Plant Classification
Service/Tool url	http://deep.ifca.es/plants/
Service/Tool information page	https://github.com/IgnacioHeredia/plant_classification
Description	This service has been developed by IFCA-CSIC. It consists of a large-scale plant classification algorithm based on the ResNet convolutional neural network architecture.
Value proposition	This tool can definitely open the field to active contributions of non-expert users including citizen scientists.
Customer of the service/tool	Researchers, citizens
User of the service/tool	It is a citizen collaboration tool. It is an open API that allows access to any user: those with a special interest in nature and botany, but also those curious people that want to know about a specific plant at some moment. Of course, its use is also aimed at developers and researchers
User Documentation	https://github.com/IgnacioHeredia/plant_classification
Technical Documentation	https://github.com/IgnacioHeredia/plant_classification
Product team	IFCA-CSIC, LifeWatch
License	-
Source code	https://github.com/IgnacioHeredia/plant_classification
Testing	-

Service/Tool name	Remote Monitoring and Smart Sensing Analysis Service
Service/Tool url	https://remote-sensing.ifca.es/hub/login
Service/Tool information page	https://remote-sensing.ifca.es/hub/login
Description	The Remote Monitoring and Smart Sensing Analysis Service is a web server designed to cover the entire process (from the selection, downloading to the view and analysis) required to work with satellite data products. First, the web server provides an Interface to search, find and download Sentinel & Landsat satellite products easily, and after that provides different tools to manage and work with the products.

Value proposition	<p>This service covers the whole process required to work with satellite data products.</p> <p>During the downloading process, the user can perform a valid search for different zones, and also restrict the queries by different keywords: cloud coverage, date, platform name (S1, S2, L8). In case of interruptions or other exceptions, downloading will restart from where it left off.</p> <p>At the same time, a geospatial integration with Smart Sensing data (where applicable, mainly from isolated areas) will be performed</p> <p>In terms of data treatment, the following products are available to be processed: • Sentinel-1 • Sentinel-2 • Landsat 8</p>
Customer of the service/tool	Environmental researchers and authorities
User of the service/tool	Scientists and policy and decision makers related to the following research areas: Computer sciences, Biological sciences, Environmental engineering, or other natural sciences
User Documentation	https://remote-sensing.ifca.es/hub/login
Technical Documentation	https://remote-sensing.ifca.es/hub/login
Product team	IFCA-CSIC, Lifewatch
License	https://remote-sensing.ifca.es/hub/login
Source code	https://remote-sensing.ifca.es/hub/login
Testing	

Service/Tool name	Glacier Lagoons of Sierra Nevada
Service/Tool url	https://lagunasdesierranevada.es/
Service/Tool information page	https://lagunasdesierranevada.es/
Description	<p>This service focuses on citizen collaboration and bio-conservation, and constitutes the end result of a Citizen Science Campaign ("74 high mountain glacier oasis") created and coordinated by the Department of Ecology of the University of Granada with the collaboration of the Sierra Nevada National Park, and the Global Change Observatory of Sierra Nevada. Its main aim is to involve society in the investigation and protection of the high mountain sites of Sierra Nevada.</p>
Value proposition	<p>The potential users of this service are mountaineers and citizens with a general interest in nature preservation. Counting on the voluntary participation, the researchers from the UGR seeks to enlarge the historical record of lagoon photographs and share all the scientific and practical information about the conservation, science and practices around these vulnerable ecosystems</p>
Customer of the service/tool	Researchers, management authorities

User of the service/tool	It is a platform for citizen collaboration in which anyone can participate by sharing their photos. The photos are from an area located in Granada, Sierra Nevada, so the area of collaboration is limited. The collaborators (photographers) are mainly lovers of mountain and ecology.
User Documentation	-
Technical Documentation	-
Product team	University of Granada, Lifewatch
License	https://lagunasdesierranevada.es/condiciones-de-uso/
Source code	-
Testing	-

Service/Tool name	GBIF Spain Spatial Portal
Service/Tool url	https://espacial.gbif.es/
Service/Tool information page	https://espacial.gbif.es/
Description	GBIF.es data access under biogeographic context provides access with advanced facets to GBIF biodiversity data under a biogeographic context. Map service, to visualise and analyse relationships between species, location and environment.
Value proposition	Map service, to visualise and analyse relationships between species, location and environment.
Customer of the service/tool	Biodiversity and Ecology research communities, Environmental authorities
User of the service/tool	This tool allows, through a map, to access the species of an area, load lists of species, generate graphs, make classifications and even create data sets.
User Documentation	https://www.gbif.es/en/portal-nacional-de-datos/
Technical Documentation	https://www.gbif.es/en/portal-nacional-de-datos/
Product team	GBIF Spain, LifeWatch
License	MPL (Mozilla Public License)
Source code	https://github.com/AtlasOfLivingAustralia/spatial-hub
Testing	https://demo.gbif.es/?lang=en

Service/Tool name	GBIF Spain Species
Service/Tool url	https://especies.gbif.es
Service/Tool information page	https://especies.gbif.es
Description	GBIF.es data access under biogeographic context provides access with advanced facets to GBIF biodiversity data and it aggregates data on species description, common names, taxonomy, image gallery, sequence data and bibliography.
Value proposition	Aggregates data on species description, common names, taxonomy, image gallery, sequence data and bibliography.
Customer of the service/tool	Biodiversity and Ecology research communities, Environmental authorities
User of the service/tool	It is a search tool where researchers can search for descriptions, common names, images, ... The content can be downloaded. It is also useful for students. The database contains species from all over the world, not only from Spain; therefore, it is not limited to focusing on research on Spanish species.
User Documentation	https://www.gbif.es/en/portal-nacional-de-datos/
Technical Documentation	https://www.gbif.es/en/portal-nacional-de-datos/
Product team	GBIF Spain, LifeWatch
License	MPL (Mozilla Public License)
Source code	https://github.com/AtlasOfLivingAustralia/bie-plugin/
Testing	https://demo.gbif.es/?lang=en

Service/Tool name	GBIF Spain Collections
Service/Tool url	https://colecciones.gbif.es
Service/Tool information page	https://colecciones.gbif.es
Description	GBIF.es data access under biogeographic context provides access with advanced facets to GBIF biodiversity data. Institutions, collections and datasets list, description and linked to occurrences.
Value proposition	This service allows users to search for data using DOIs. The database contains information on institutions, collections and species records. Currently there are only data from entities that collaborate with GBIF
Customer of the service/tool	Biodiversity and Ecology research communities, Environmental authorities
User of the service/tool	The access to these collections is open, this service is aimed both at people who want to learn and research. Serves as a record.

User Documentation	https://www.gbif.es/en/portal-nacional-de-datos/
Technical Documentation	https://www.gbif.es/en/portal-nacional-de-datos/
Product team	GBIF Spain, LifeWatch
License	MPL (Mozilla Public License)
Source code	https://github.com/AtlasOfLivingAustralia/collectory-plugin
Testing	https://demo.gbif.es/?lang=en

Service/Tool name	GBIF Spain Regions
Service/Tool url	https://regiones.gbif.es
Service/Tool information page	https://regiones.gbif.es
Description	GBIF.es data access under biogeographic context provides access with advanced facets to GBIF biodiversity data. Map tool to browse states, territories, local government areas and biogeographic regions.
Value proposition	Map tool to browse states, territories, local government areas and biogeographic regions.
Customer of the service/tool	Biodiversity and Ecology research communities, Environmental authorities
User of the service/tool	This open tool shows, through a map, the species of the different regions of Spain. It is mainly oriented to research and academic use. It is also useful for any citizen who lives in an area of which this service has information.
User Documentation	https://www.gbif.es/en/portal-nacional-de-datos/
Technical Documentation	https://www.gbif.es/en/portal-nacional-de-datos/
Product team	GBIF Spain, LifeWatch
License	MPL (Mozilla Public License)
Source code	https://github.com/AtlasOfLivingAustralia/regions
Testing	https://demo.gbif.es/?lang=en

Service/Tool name	GBIF Spain Occurrences
Service/Tool url	https://registros.gbif.es
Service/Tool information page	https://registros.gbif.es

Description	GBIF.es data access under biogeographic context provides access with advanced facets to GBIF biodiversity data under a biogeographic context. List of occurrence data, georeferenced records and data set search statistics. Allows access to the information of each record and the quality tests performed.
Value proposition	List of occurrence data, georeferenced records and data set search statistics. Allows access to the information of each record and the quality tests performed
Customer of the service/tool	Biodiversity and Ecology research communities, Environmental authorities
User of the service/tool	Research use. Access to occurrence biodiversity data published by Spanish providers
User Documentation	https://www.gbif.es/en/portal-nacional-de-datos/
Technical Documentation	https://www.gbif.es/en/portal-nacional-de-datos/
Product team	GBIF Spain, LifeWatch
License	MPL (Mozilla Public License)
Source code	https://github.com/AtlasOfLivingAustralia/biocache-hubs
Testing	https://demo.gbif.es/?lang=en

Service/Tool name	GBIF Spain Images
Service/Tool url	https://imagenes.gbif.es
Service/Tool information page	https://imagenes.gbif.es
Description	GBIF.es data access under biogeographic context provides access with advanced facets to GBIF biodiversity data under a biogeographic context. Image storage and web services application. Images can be measured, calibrated, zoomed or downloaded
Value proposition	Image storage and web services application. Images can be measured, calibrated, zoomed or downloaded
Customer of the service/tool	Biodiversity and Ecology research communities, Environmental authorities
User of the service/tool	Research and academic use. Visualize images found in the data portal for a taxon.
User Documentation	https://www.gbif.es/en/portal-nacional-de-datos/
Technical Documentation	https://www.gbif.es/en/portal-nacional-de-datos/
Product team	GBIF Spain, LifeWatch
License	MPL (Mozilla Public License)

Source code	https://github.com/AtlasOfLivingAustralia/image-service/
Testing	https://demo.gbif.es/?lang=en

Service/Tool name	GBIF Spain eLearning
Service/Tool url	https://elearning.gbif.es
Service/Tool information page	https://elearning.gbif.es
Description	GBIF.es data access under biogeographic context provides access with advanced facets to GBIF biodiversity data under a biogeographic context. The GBIF.ES e-learning platform offers online workshops on data quality, data publication, data use, etc. open to whole GBIF community
Value proposition	The GBIF.ES e-learning platform offers online workshops on data quality, data publication, data use, etc. open to whole GBIF community
Customer of the service/tool	Biodiversity and Ecology research communities, Environmental authorities
User of the service/tool	Platform offers online workshops open to whole GBIF community.
User Documentation	https://www.gbif.es/en/portal-nacional-de-datos/
Technical Documentation	https://www.gbif.es/en/portal-nacional-de-datos/
Product team	GBIF Spain
License	GNU/GPLv3+
Source code	https://github.com/chamilo/chamilo-lms
Testing	-

LifeWatch ERIC is a European Infrastructure Consortium providing e-Science research facilities to scientists seeking to increase our knowledge and deepen our understanding of Biodiversity organisation and Ecosystem functions and services in order to support civil society in addressing key planetary challenges.

In EOSC-hub, the University of Sevilla (USE) is the organization in charge of managing the integration of the e-services linked to LifeWatch into the Marketplace. Due to some issues related to the service level agreements (SLAs) between LifeWatch ERIC – The Service Providers – EOSC-hub, the activity of the USE in this project has been delayed with regard to the initial plan. After some meetings with the leaders of WP7, it was agreed to move the dates for the planned activities. As a consequence, the roadmap has been modified in order to get the same level of development as the other service providers by December 2019 (M24).

In this regard, and considering the original proposal, there have been some changes in the list of e-services to be integrated into the marketplace. These modifications have arisen from causes outside of our control such as the decision of the service providers of not going on with the development or

deployment of some e-services (e.g. PAIRQRS, CINDA). In addition, although all the services we are integrating were planned to be part of the LifeWatch ERIC catalog in the near term, this action is still pending. In any case, all the services considered are directly related to LifeWatch, since their ownership belongs to entities that have signed the Joint Research Unit LifeWatch Spain (JRU LW.ES) agreement. This is the case for the Institute of Physics of Cantabria – Spanish National Research Council (IFCA-CSIC), and University of Granada (UGR). Founded in 2014, the JRU LW.ES represents the Spanish National Node supporting LW ERIC. It brings together several organizations and institutions carrying out research, development and innovation activities in biodiversity, supporting knowledge-based strategic solutions for environmental preservation.

On the other hand, it has been a while since the Global Biodiversity Information Facility (GBIF) and the e-infrastructure LifeWatch ERIC started conversations for establishing a close and fruitful relationship between them. At present, both organizations are working on the definition of a collaboration agreement that satisfies both parts. In the framework of the EOSC-hub project, this willingness is evidenced by the integration of the GBIF Spain (GBIF.ES) services into the EOSC-hub marketplace under the umbrella of LifeWatch Thematic Services (although none of GBIF.ES's or GBIF's services are part of the LifeWatch ERIC's catalogue currently).

Ultimately, and as part of our commitments in the WP7, the following services are being integrated into the EOSC marketplace:

Glacier Lagoons of Sierra Nevada ("Lagunas de Sierra Nevada")

This service focuses on citizen collaboration and bio-conservation, and constitutes the end result of a Citizen Science Campaign ("74 high mountain glacier oasis") created and coordinated by the Department of Ecology of the University of Granada with the collaboration of the Sierra Nevada National Park, and the Global Change Observatory of Sierra Nevada. Its main aim is to involve society in the investigation and protection of the high mountain sites of Sierra Nevada.

The potential users of this service are mountaineers and citizens with a general interest in nature preservation. Counting on voluntary participation, the researchers from the UGR seek to enlarge the historical record of lagoon photographs and share all the scientific and practical information about the conservation, science and practices around these vulnerable ecosystems. Through the interactive platform, internet users are able to:

- Interact with the scientific teams that carry out their research in these ecosystems
- Share and visualize the historical record of photographs
- Access to all scientific and information related to Sierra Nevada lagoons
- Participate in the activities, courses and visits organized for the protection of these sites.

GBIF.es

GBIF.es provides access with advanced facets to biodiversity data under a biogeographic context. Considering the framework of this project, 7 different services have been considered:

- Collections: Institutions, collections and datasets list, description and linked to occurrences. This service allows users to search for data using DOIs. The database contains information on institutions, collections and species records. Currently there are only data from entities that collaborate with GBIF.
- Species: Aggregates data on species description, common names, taxonomy, image gallery, sequence data and bibliography

- Regions: Map tool to browse states, territories, local government areas and biogeographic regions.
- SpatialPortal: Map service, to visualise and analyse relationships between species, location and environment.
- eLearning: The GBIF.ES e-learning platform offers online workshops on data quality, data publication, data use, etc. open to whole GBIF community
- Occurrences: List of occurrence data, georeferenced records and data set search statistics. Allows access to the information of each record and the quality tests performed.
- Images: Image storage and web services application. Images can be measured, calibrated, zoomed or downloaded.

Plant Classification

This service has been developed by IFCA-CSIC. It consists of a large-scale plant classification algorithm based on the ResNet convolutional neural network architecture. This tool can definitely open the field to active contributions of non-expert users including citizen scientists.

Remote Monitoring and Smart Sensing

This service has also been developed by IFCA-CSIC. It is a web server designed to cover the entire process of working with Sentinel data products. First, the web server provides an Interface to search, find and download Copernicus Sentinel satellite products easily, and after that it provides different tools to manage and work with the products.

During downloading process, the user can perform a valid search for different zones, and also restrict the queries by different keywords: cloud coverage, date, platform name (S1, S2, S3). In case of interruptions or other exceptions, downloading will restart from where it left off.

At the same time, a geospatial integration with Smart Sensing data (where applicable, mainly from isolated areas) will be performed in terms of data treatment. The following products are available to be processed:

- Sentinel-1
- Sentinel-2
- Landsat 8

10.2 Service Architecture and integrated services

At present, all the Thematic Services intended to be integrated are available on the EOSC-hub marketplace (additionally to their corresponding provider's webpage).

Only the Plant Classification service needs to be updated since the currently available version in the marketplace is obsolete and has several errors that prevent its use from users (the version on the platform is deprecated, much of the information is missing and has a low integration level since it only has a link that redirects to the LifeWatch home page. Thus, this version should be replaced as it was conveniently indicated to the integration team).

Regarding the integration with EOSC common services, the Remote Sensing service is going to be integrated with: XDC and Indigo IAM. The other services, so far, have their own storage system and are open or easy to access:

- GBIF.ES, note we work with the Spanish node. This node is part of a huge worldwide organization, which has its own storage and authentication systems.
- The Plant Classification service is an open API and no login is required. This service does not require storing any new information.
- Glacier Lagoons of Sierra Nevada is a citizen collaboration service. To make a contribution, users need to have a registered account. However, for visualization and other functionalities, no requirement has been placed.

10.3 Software release

Section not applicable to LifeWatch related services.

10.4 References

Links of interest:

- <http://deep.ifca.es/plants/>
- https://github.com/IgnacioHeredia/plant_classification
- <https://remote-sensing.ifca.es/hub/login>
- <https://lagunasdesierranevada.es/>
- <https://elearning.gbif.es/>
- <https://registros.gbif.es/>
- <https://colecciones.gbif.es/>
- <https://imagenes.gbif.es/>
- <https://regiones.gbif.es/>
- <https://auth.gbif.es/>
- <https://especies.gbif.es/>

Scientific publications linked to the above-mentioned services:

GBIF Spain

- Ramírez-Valiente, J. A., López, R., Hipp, A. L., & Aranda, I. (2019). Correlated evolution of morphology, gas exchange, growth rates and hydraulics as a response to precipitation and temperature regimes in oaks (*Quercus*). *New Phytologist*.
- Pappalardo, P., Morales-Castilla, I., Park, A. W., Huang, S., Schmidt, J. P., & Stephens, P. R. (2020). Comparing methods for mapping global parasite diversity. *Global Ecology and Biogeography*, 29(1), 182-193.
- Rodríguez-Merino, A., Fernández-Zamudio, R., García-Murillo, P., & Muñoz, J. (2019). Climatic Niche Shift during *Azolla filiculoides* Invasion and Its Potential Distribution under Future Scenarios. *Plants*, 8(10), 424.

- Mezghani, N., Khoury, C. K., Carver, D., Achicanoy, H. A., Simon, P., Flores, F. M., & Spooner, D. (2019). Distributions and conservation status of carrot wild relatives in Tunisia: A case study in the western Mediterranean basin. *Crop Science*, 59(6), 2317-2328.
- Exposito-Alonso, M., Burbano, H. A., Bossdorf, O., Nielsen, R., & Weigel, D. (2019). Natural selection on the *Arabidopsis thaliana* genome in present and future climates. *Nature*, 573(7772), 126-129.
- Remote Monitoring and Smart Sensing
- Cesini, D., Donvito, G., Costantini, A., Gomez, F. A., Duma, D. C., Fuhrmann, P., ... & Dell'Agnello, L. (2019). Advancements in data management services for distributed e-infrastructures: the eXtreme-DataCloud project. In *EPJ Web of Conferences* (Vol. 214, p. 04044). EDP Sciences.
- Aguilar, F. (2020). DataCloud infrastructure to manage FAIR environmental data. *Journal of Instrumentation*, 15(04), C04009–C04009. <https://doi.org/10.1088/1748-0221/15/04/C04009>

Dissemination materials:

Glacier Lagoons of Sierra Nevada (in Spanish)

- Glacier Lagoons of Sierra Nevada. European Researchers' Night promotional video <https://www.youtube.com/watch?v=-lvQBg9N1ZU&feature=youtu.be%2F>
- Boat sampling on the Genil River. European Researchers' Night <https://lanochedelosinvestigadores.fundaciondescubre.es/actividades/muestreo-en-barca-en-el-rio-genil/>
- The best pictures of the Sierra Nevada lagoons. El Independiente de Granada <http://www.elindependientedegranada.es/economia/mejores-fotos-lagunas-sierra-nevada>
- This is the bottom of the Sierra Nevada lagoons. El Independiente de Granada <https://www.elindependientedegranada.es/economia/asi-es-fondo-lagunas-sierra-nevada>
- Photographic and informative exhibition on the lagoons of Sierra Nevada. Faculty of Science of the UGR. <https://www.ugr.es/visitantes/agenda-cultural/exposicion-fotografica-lagunas-sierra-nevada>

10.5 EOSC-hub integration

Currently, all the agreed services have been integrated into the EOSC-hub marketplace. As stated in the corresponding section, the Plant Classification service must be updated as soon as possible in order to be used by the community (the available version is deprecated).

In summary, the Lifewatch-related services integrated into and available from the EOSC-hub marketplace are the following:

- Plant Classification (IFCA-CSIC): <https://marketplace.eosc-portal.eu/services/lifewatch-eric-plants-identification-app>
- Remote monitoring and smart sensing (IFCA-CSIC): <https://marketplace.eosc-portal.eu/services/remote-monitoring-and-smart-sensing>

- Glacier Lagoons of Sierra Nevada (University of Granada):
<https://marketplace.eosc-portal.eu/services/lagunas-de-sierra-nevada-glacier-lagoons-of-sierra-nevada>
- GBIF.ES services:
<https://marketplace.eosc-portal.eu/services/e-learning-platform-of-gbif-spain>
<https://marketplace.eosc-portal.eu/services/gbif-spain-occurrence-records>
<https://marketplace.eosc-portal.eu/services/gbif-spain-collections-registry>
<https://marketplace.eosc-portal.eu/services/gbif-spain-images-portal>
<https://marketplace.eosc-portal.eu/services/gbif-spain-regions-module>
<https://marketplace.eosc-portal.eu/services/gbif-spain-spatial-portal>
<https://marketplace.eosc-portal.eu/services/gbif-spain-species-portal>

References

No	Description/Link
R1	D7.1 “First Thematic Service software release”, https://documents.egi.eu/document/3411
R2	D7.2 “First report on Thematic Service architecture and software integration”, https://documents.egi.eu/document/3412
R3	D7.3 “First report on Thematic Service exploitation”, https://documents.egi.eu/document/3577
R4	Oliveira, A., A.B. Fortunato, J. Rogeiro, J. Teixeira, A. Azevedo, L. Lavaud, X. Bertin, J. Gomes, M. David, J. Pina, M. Rodrigues, P. Lopes, 2020. OPENCoastS: An open-access service for the automatic generation of coastal forecast systems, Environmental Modelling and Software, 124: 104585. DOI: 10.1016/j.envsoft.2019.104585
R5	Oliveira, A.; Rodrigues, M.; Rogeiro, J.; Fortunato, A. B.; Teixeira, J.; Azevedo, A.; Lopes, P. 2019, Opencoasts: an open-access app for sharing coastal prediction information for management and recreation, 794-807pp, Computational Science-Lecture Notes-ICCS2019, Vol. 11540, Computational Science-Lecture Notes-ICCS2019
R6	Marta Rodrigues, João Rogeiro, S. Bernardo, A. Oliveira, A. B. Fortunato, J. Teixeira, Pedro Lopes, A. Azevedo, J. Gomes, M. David, J. Pina, 2019.OPENCoastS: an operational, on-demand forecast tool to support coastal management, Proceedings of the 14th International MECOAST Congress on Coastal and Marine Sciences, Engineering, Management and Conservation MEDCOAST 2019, E. Ozhan (Editor), vol. 1, 57-67.
R7	Bertini, I., Case, D. A., Ferella, L., Giachetti, A. & Rosato, A. A Grid-enabled web portal for NMR structure refinement with AMBER. Bioinformatics 27, 2384–2390 (2011).
R8	Wassenaar, T. A. et al. WeNMR: Structural Biology on the Grid. J Grid Computing 10, 743–767 (2012).
R9	Schot, G. & Bonvin, A. M. J. J. Performance of the WeNMR CS-Rosetta3 web server in CASD-NMR. Journal of Biomolecular NMR 62, 497–502 (2015).
R10	G.C.P. van Zundert, M. Trellet, J. Schaarschmidt, Z. Kurkcuoglu, M. David, M. Verlato, A. Rosato and A.M.J.J. Bonvin. The DisVis and PowerFit web servers: Explorative and Integrative Modeling of Biomolecular Complexes. J. Mol. Biol., 429, 399-407 (2017).
R11	Rinaldelli, M., Carlon, A., Ravera, E., Parigi, G. & Luchinat, C. FANTEN: a new web-based interface for the analysis of magnetic anisotropy-induced NMR data. J Biomol NMR 61, 21–34 (2014).
R12	van Zundert, G. C. P. et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol 428, 720–725 (2015).

R13	G.C.P. van Zundert, M. Trellet, J. Schaarschmidt, Z. Kurkcuoglu, M. David, M. Verlato, A. Rosato and A.M.J.J. Bonvin. The DisVis and PowerFit web servers: Explorative and Integrative Modeling of Biomolecular Complexes. <i>J. Mol. Biol.</i> , 429, 399-407 (2017).
R14	I.S. Moreira, P.I. Koukos, R. Melo, J.G. Almeida, A.J. Preto, J. Schaarschmidt, M. Trellet, Z.H. Gümüş, J. Costa and A.M.J.J. Bonvin. SpotOn: High Accuracy Identification of Protein-Protein Interface Hot-Spots. <i>Sci. Reports.</i> 7:8007 (2017).