



# **EOSC-hub**

## **D7.3 First report on Thematic Service exploitation**

<b>Lead Partner:</b>	CINECA
<b>Version:</b>	1
<b>Status:</b>	FINAL
<b>Dissemination Level:</b>	Public
<b>Document Link:</b>	<a href="https://documents.egi.eu/document/3577">https://documents.egi.eu/document/3577</a>

### **Deliverable Abstract**

This document provides an update on the development and integration status of the thematic services part of WP7 activities. It also describes the target users and a first report on the usage of the services. Each thematic services provider has then analysed the gaps present in the current version and the steps to close these gaps in the next period.



## 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>
<b>From:</b>	Debora Testi	WP7/CINECA	
<b>Moderated by:</b>	Malgorzata Krakowian	WP1/EGI	
<b>Reviewed by:</b>	Malgorzata Krakowian Gergely Sipos	WP13/EGI WP8/EGI	
<b>Approved by:</b>	AMB		

## DOCUMENT LOG

<i>Issue</i>	<i>Date</i>	<i>Comment</i>	<i>Author</i>
<b>v.0.1</b>	24/10/2019	Outline sent to partners for completion	D. Testi
<b>v. 0.2</b>	10/12/2019	First draft ready for internal review	All
<b>v. 0.3</b>	03/02/2020	Draft with request for updates from reviewers circulated	D. Testi
<b>v. 1</b>	12/02/2020	Final version addressing reviewers' comments	D. Testi

## TERMINOLOGY

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

---

## Contents

1	Introduction.....	5
2	CLARIN .....	6
2.1	Service description .....	6
2.2	Services exploitation .....	13
3	DODAS .....	16
3.1	Service description .....	16
3.2	Services exploitation .....	23
4	ECAS .....	26
4.1	Service description .....	26
4.2	Services exploitation .....	37
5	GEOSS .....	41
5.1	Service description .....	41
5.2	Services exploitation .....	42
6	OPENCoastS .....	44
6.1	Service description .....	44
6.2	Services exploitation .....	46
7	WeNMR .....	49
7.1	Service description .....	49
7.2	Services exploitation .....	51
8	EO Pillar .....	55
8.1	Service description .....	55
8.2	Services exploitation .....	60
9	DARIAH .....	61
9.1	Service description .....	61
9.2	Services exploitation .....	66
10	LifeWatch.....	69
10.1	Service description .....	69
10.2	Services exploitation .....	79
11	Conclusions.....	82
12	References .....	83

## **Executive summary**

The EOSC-hub thematic services, part of WP7 activities, are now being integrated with a number of EOSC services. The services are also being provided via the EOSC-hub marketplace so to be accessible by other researchers in the same or other communities. This document provides a short description of each thematic service together with an update on the integration with other EOSC services, and a first analysis on the usage and the users' feedbacks collected so far. This is the first document of this type; updates on usage and exploitation will be provided in future deliverables.

# 1 Introduction

This document is structured with a section for each of the thematic services contributing to WP7 activities. For each thematic service, the following information are provided:

- A short description of the service/s.
- An update, with respect to the previous D7.1 and D7.2 deliverables [R1, R2], on the achieved integration and future plans.
- A short description on the main categories of users that are or will use the service.
- A summary of the current usage, availability statistics, and users' feedback.
- Analysis of the missing features or issues present in the current release and which are the plans for upgrades or development.

As not all thematic services are at the same level of development and integration, the descriptions contain different level of details based on the advancement in the WP7 activities.

## 2 CLARIN

### 2.1 Service description

<b>Service/Tool name</b>	Virtual Language Observatory
<b>Service/Tool url</b>	<a href="https://vlo.clarin.eu">https://vlo.clarin.eu</a>
<b>Service/Tool information page</b>	<a href="https://www.clarin.eu/vlo">https://www.clarin.eu/vlo</a>
<b>Description</b>	A facet browser for fast navigation and searching in huge amounts of metadata.
<b>Value proposition</b>	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 Europeana.
<b>Customer of the service/tool</b>	Researchers Repository managers
<b>User of the service/tool</b>	Researchers
<b>User Documentation</b>	<a href="https://vlo.clarin.eu/help">https://vlo.clarin.eu/help</a>
<b>Technical Documentation</b>	<a href="https://trac.clarin.eu/wiki/CmdiVirtualLanguageObservatory">https://trac.clarin.eu/wiki/CmdiVirtualLanguageObservatory</a> (requires authentication, information available on request)
<b>Product team</b>	CLARIN ERIC
<b>License</b>	GPLv3
<b>Source code</b>	<a href="https://github.com/clarin-eric/VLO">https://github.com/clarin-eric/VLO</a>
<b>Testing</b>	Unit and integration tests using junit Vulnerability analysis provided by snyk.io

<b>Service/Tool name</b>	Virtual Collection Registry
<b>Service/Tool url</b>	<a href="https://collections.clarin.eu">https://collections.clarin.eu</a>
<b>Service/Tool information page</b>	<a href="https://www.clarin.eu/content/virtual-collections">https://www.clarin.eu/content/virtual-collections</a>
<b>Description</b>	A service that allows researchers to create their own citable digital bookmarks.
<b>Value proposition</b>	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.

<b>Customer of the service/tool</b>	Researchers
<b>User of the service/tool</b>	Researchers
<b>User Documentation</b>	<a href="https://collections.clarin.eu/app/help">https://collections.clarin.eu/app/help</a>
<b>Technical Documentation</b>	<a href="https://trac.clarin.eu/wiki/VirtualCollectionRegistry">https://trac.clarin.eu/wiki/VirtualCollectionRegistry</a> (requires authentication, information available on request)
<b>Product team</b>	CLARIN ERIC
<b>License</b>	GPLv3
<b>Source code</b>	<a href="https://github.com/clarin-eric/VirtualCollectionRegistry">https://github.com/clarin-eric/VirtualCollectionRegistry</a>
<b>Testing</b>	Unit and integration tests using junit Vulnerability analysis provided by snyk.io

<b>Service/Tool name</b>	Language Resource Switchboard
<b>Service/Tool url</b>	<a href="https://switchboard.clarin.eu">https://switchboard.clarin.eu</a>
<b>Service/Tool information page</b>	<a href="https://www.clarin.eu/content/language-resource-switchboard">https://www.clarin.eu/content/language-resource-switchboard</a>
<b>Description</b>	A web application that suggests language analysis tools for specific data sets.
<b>Value proposition</b>	The <a href="https://switchboard.clarin.eu">Language Resource Switchboard</a> 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 <a href="#">Virtual Language Observatory</a> and B2DROP (see Suggested compatible services below).
<b>Customer of the service/tool</b>	Researchers Tool administrators
<b>User of the service/tool</b>	Researchers
<b>User Documentation</b>	<a href="https://switchboard.clarin.eu/help">https://switchboard.clarin.eu/help</a>
<b>Technical Documentation</b>	<a href="https://trac.clarin.eu/wiki/LanguageResourceSwitchboard">https://trac.clarin.eu/wiki/LanguageResourceSwitchboard</a> (requires authentication, information available on request)
<b>Product team</b>	CLARIN ERIC
<b>License</b>	GPLv3
<b>Source code</b>	<a href="https://github.com/clarin-eric/switchboard">https://github.com/clarin-eric/switchboard</a>

<b>Testing</b>	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)<sup>1</sup>, the Virtual Collection Registry (VCR)<sup>2</sup> and the Language Resource Switchboard (LRS)<sup>3</sup>.

A detailed description for the VLO service has already been described in D7.1 and D7.2 [R1, R2].

The VCR is a registry provided by CLARIN where scholars can create and publish virtual collections. 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. A virtual collection is suitable for manual access (using a web-browser) as well as automated processing (e.g. by a web service). The VCR is closely integrated with the CLARIN federated authentication and authorization infrastructure and provides persistent identifiers for easy citation. The collection metadata is openly available and accessible via the VLO.

The LRS is a tool that helps to find a matching language processing web application for data. After uploading a file, providing a persistent identifier or entering a URL, the resource referenced is analysed and the user is provided a list of tasks he/she can perform on the specified resource. After selecting a task, the user is provided with a list of available CLARIN tools to analyse the input. By selecting a tool, the LRS will make sure the user is sent to the service and the service is properly instrumented with the provided resource.

All services are published in the EOSC hub marketplace: VLO<sup>4</sup>, VCR<sup>5</sup> and LRS<sup>6</sup>.

## 2.1.1 Architecture and integrated services

### 2.1.1.1 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, VA usage satisfaction via Mopinion<sup>7</sup>.

<sup>1</sup> <https://www.clarin.eu/vlo>

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

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

<sup>4</sup> <https://marketplace.eosc-portal.eu/services/virtual-language-observatory>

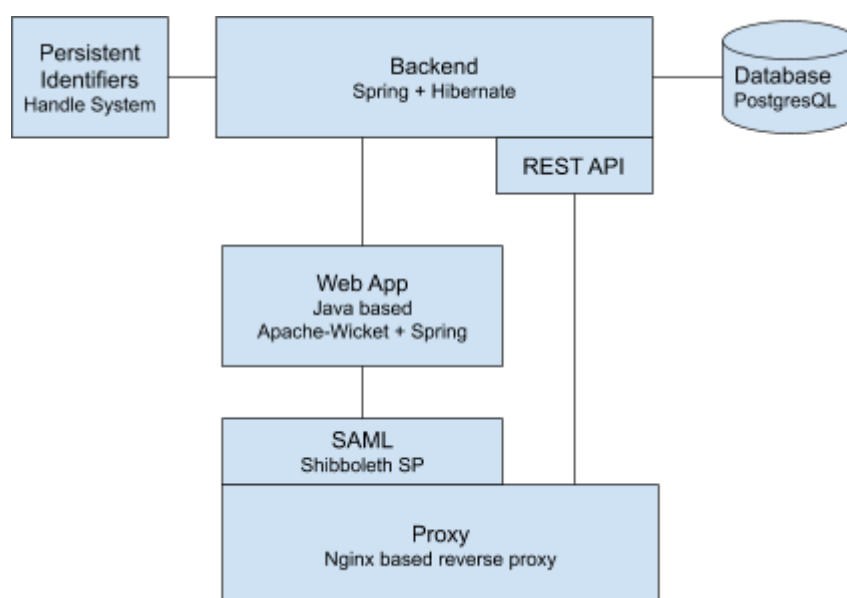
<sup>5</sup> <https://marketplace.eosc-portal.eu/services/virtual-collection-registry>

<sup>6</sup> <https://marketplace.eosc-portal.eu/services/language-resource-switchboard>

<sup>7</sup> <https://www.mopinion.com>

### 2.1.1.2 Virtual Collection Registry

The VCR is a Java<sup>8</sup> web application, using Apache Wicket<sup>9</sup> and Spring<sup>10</sup> for the frontend. The backend is also implemented in Java and is making use of Spring, Hibernate<sup>11</sup> and a PostgreSQL<sup>12</sup> database backend. The backend offers a REST API<sup>13</sup>. Persistent identifiers for the collections are minted via the Handle system. Source code is available on GitHub<sup>14</sup>.



**Fig. 1 - Virtual Collection Registry architecture**

Users can login through their home organization (enabled by using SAML federated login) in order to create, edit and delete virtual collections. Initially, a collection is not published and therefore not publicly available. A user can then publish a virtual collection. Upon publishing, the collection is assigned a persistent identifier and made publicly available. A regular user cannot remove published collections.

Integrations: the VCR is integrated with the LRS as shown in Figure 2, VA usage satisfaction via Mopinion and with the VLO. B2SHARE integration has been implemented on VCR side but it has not been released in production because of pending changes on B2SHARE side.

<sup>8</sup> <https://www.java.com/>

<sup>9</sup> <https://wicket.apache.org/>

<sup>10</sup> <https://spring.io/>

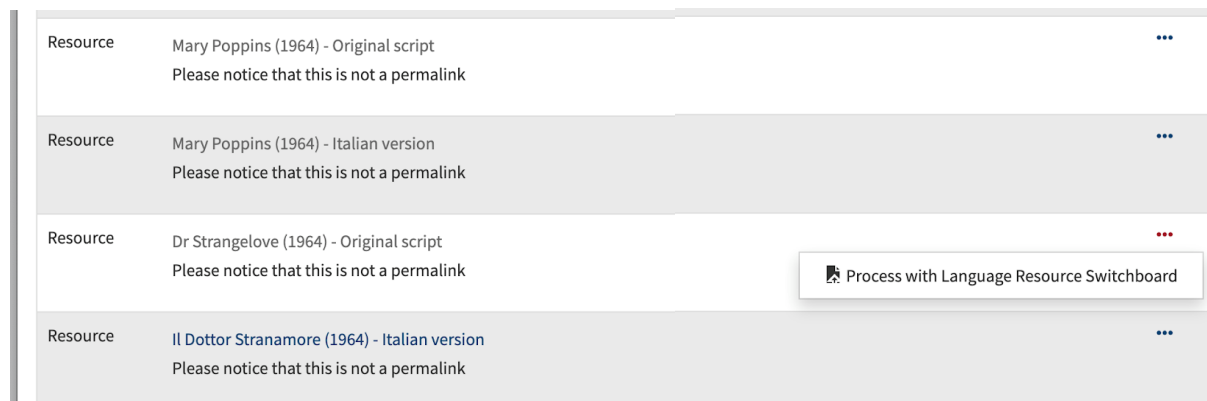
<sup>11</sup> <https://hibernate.org/>

<sup>12</sup> <https://www.postgresql.org/>

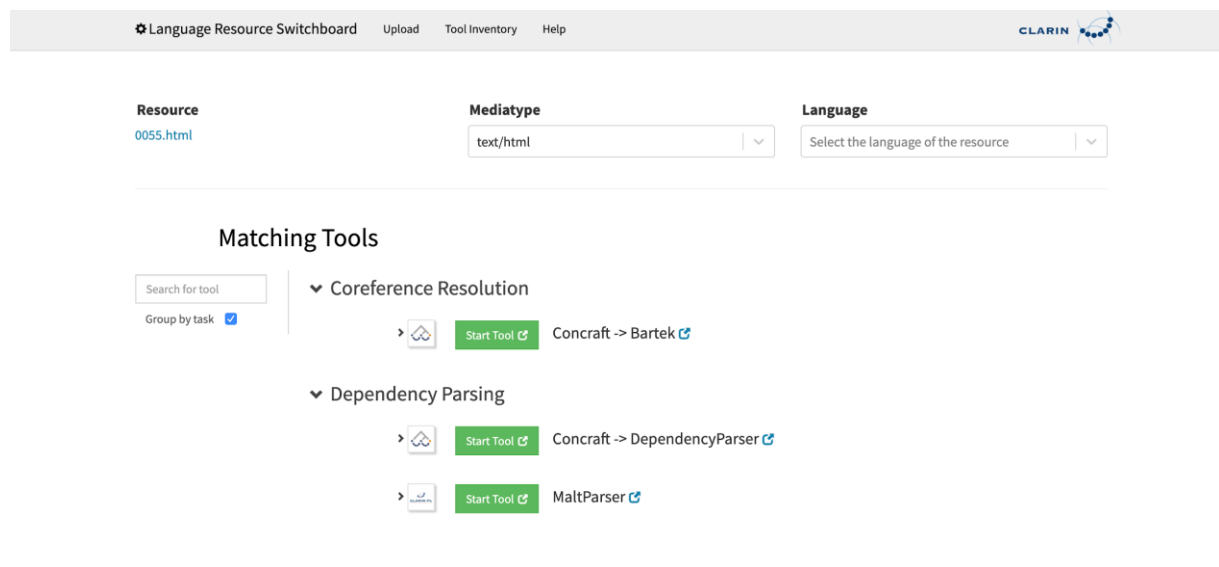
<sup>13</sup> <https://github.com/clarin-eric/VirtualCollectionRegistry/tree/master/doc>

<sup>14</sup> <https://github.com/clarin-eric/VirtualCollectionRegistry>

Step 1. Open a virtual collections resource in the LRS.



Step 2. After transferring the resource to the LRS, a list of tools capable of processing the resource is shown.



Step 3. After clicking on a 'start tool' button, the selected tool is opened with the resource.

**Fig. 2 - VCR - LRS integration**

### 2.1.1.3 Language Resource Switchboard

The LRS is composed of a Java backend and a JS<sup>15</sup>+React<sup>16</sup>+Redux<sup>17</sup> frontend.

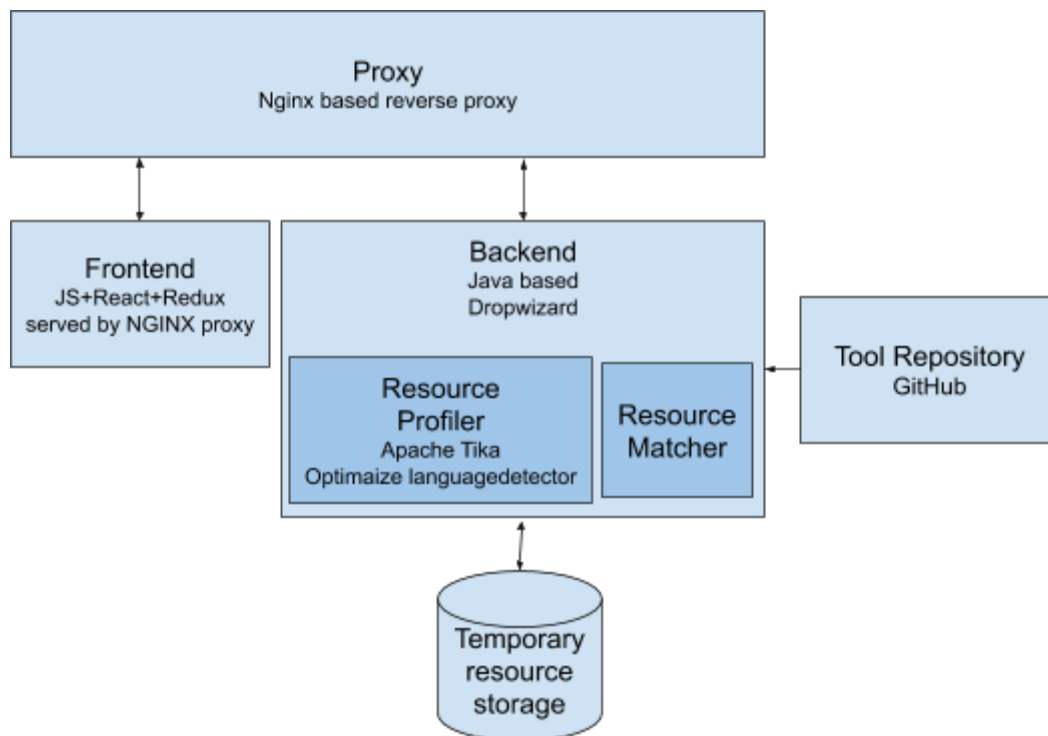
The backend serves the HTML, JS, CSS and font resources to a browser, handles the list of all tools, handles the temporary storage of user resources, and profiles the user data to detect its media type and language. The frontend (running in the user's browser) renders the User Interface (UI), handles user interaction and makes API calls to the backend for getting or updating the user's data and the

<sup>15</sup> ECMAScript 2015, <http://www.ecma-international.org/>

<sup>16</sup> <https://reactjs.org/>

<sup>17</sup> <https://redux.js.org/>

tools. In addition, the frontend opens the tools selected by the user, instrumenting them with the user's specified resources. Source code is available on GitHub<sup>18</sup>.



**Fig. 3 - LRS architecture**

Tool availability and configuration is managed in a separate repository on GitHub<sup>19</sup>. Any updates made here are automatically pulled by the LRS within 5 minutes.

The LRS tools repository is setup to enable tool managers to propose new tools and changes to the configuration of the existing ones via git pull requests. Pull requests are validated automatically via a continuous integration system and, after testing and maturity assessment, merged by the LRS operators.

The LRS has been integrated with:

- B2DROP (via a NextCloud plugin, source code on GitHub<sup>20</sup>) as shown in Figure 4.
- The VLO
- The VCR
- D4Science

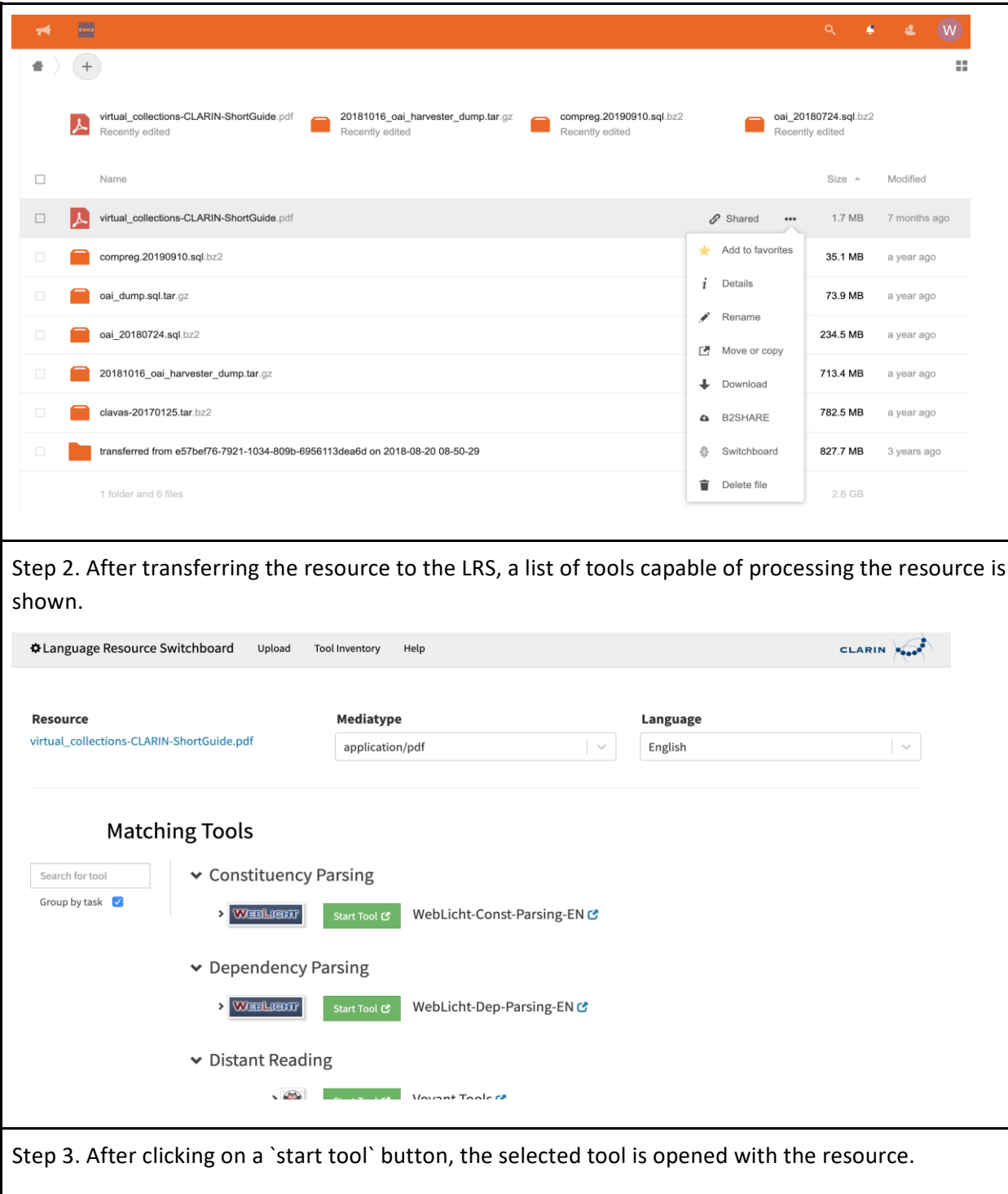
Step 1. Open a B2DROP resource in the LRS.

<sup>18</sup> <https://github.com/clarin-eric/switchboard>

<sup>19</sup> <https://github.com/clarin-eric/switchboard-tool-registry>

<sup>20</sup> <https://github.com/clarin-eric/B2Drop2LRSwitchboard>

Step 2. After transferring the resource to the LRS, a list of tools capable of processing the resource is shown.



The screenshot displays the B2DROP LRS integration interface. At the top, there is a navigation bar with a home icon, a plus icon, and a search icon. Below the navigation bar, a list of files is shown, including 'virtual\_collections-CLARIN-ShortGuide.pdf', '20181016\_oai\_harvester\_dump.tar.gz', 'compreg.20190910.sql.bz2', and 'oai\_20180724.sql.bz2'. A context menu is open over the first file, showing options like 'Add to favorites', 'Details', 'Rename', 'Move or copy', 'Download', 'B2SHARE', 'Switchboard', and 'Delete file'. Below the file list, the 'Matching Tools' section is visible, showing a search bar and a list of tools categorized by task. The tools listed are 'WebLicht-Const-Parsing-EN' and 'WebLicht-Dep-Parsing-EN' under the 'Dependency Parsing' category, and 'WebLicht-Const-Parsing-EN' under the 'Constituency Parsing' category. The 'Language' dropdown is set to 'English' and the 'Mediatype' dropdown is set to 'application/pdf'.

Step 3. After clicking on a 'start tool' button, the selected tool is opened with the resource.

**Fig. 4 - B2DROP - LRS integration**

### 2.1.2 Main users

### 2.1.2.1 Virtual Language Observatory

Details have been provided in D7.2 “First report on Thematic Service architecture and software integration” [R2].

### 2.1.2.2 Virtual Collection Registry

Researchers can login to create, update and delete virtual collections which are otherwise publicly available after being published.

### 2.1.2.3 Language Resource Switchboard

Researchers can provide an input resource to the LRS. After analysis, the LRS will propose a set of tools that can operate on the input. The user can select a tool after which the LRS will redirect to the tool instrumented with the specified resource.

Tool Providers can manage tool definitions and propose new tools so that the LRS offers them if they are able to process a certain resource.

## 2.2 Services exploitation

### 2.2.1 Usage statistics

The usage is measured using the following Virtual Access metrics, as reported also to WP13:

Metric Name	Baseline	Definition of measurement	Period 1 M7 - M8	Period 2 M9 - M17
Number of visits to metadata search portal	2017: 5'103 visits/year  = 425 visits/month	No registration, reported is number of visits over a certain timespan – measured using Matomo <sup>21</sup> .	16 July 2018 until 31 August 2018: 389 visits  = 259 visits/month	1 September 2018 until 31 May 2019 4'749 visits  = 528 visits/month
Usage: Number of harvested metadata records	1 August 2018: 896'473	The number of metadata records harvested via OAI-PMH and inserted into the Virtual Language Observatory.	896'473	13 June 2019: 909'388

<sup>21</sup> <https://matomo.org/>

Usage: Number of virtual collections registered	2017: 0	The number of virtual collections made publicly available via the VCR over a certain timespan. Note: the baseline is low since not much publicity was made before the EOSC-hub release.	Data not available (service was not launched yet)	7
Usage: Number of connected processing tools via the LR Switchboard	1 January 2018: 60 ( <a href="#">source</a> ) <a href="#">List of tools</a>	The number of web applications registered at the Language Resource Switchboard that can process incoming requests.	Data not available (service was not launched yet)	13 June 2019: 70 <a href="#">Source and list of tools</a>
Number and names of the countries reached (measurement for metadata search portal)	2017: 89 countries (list available upon request)	The evaluation is based on IPs, measured using Matomo.	45	101

In addition, three virtual servers have been provided to CLARIN. Each virtual server has 12 cores, 24GB of ram and 1TB of storage.

### 2.2.2 Users' feedbacks

For the VLO and VCR, Mopinion has been used to collect users' feedback. In the period from 1 January 2019 to 31 October 2019 the:

- VLO was rated 3.8 out of 5 on average with a total of 133 ratings.
- VCR was rated 4 out of 5 on average with a total of 2 ratings.

For the LRS it is decided not to collect users' feedback via a pop-up. Given the nature of the service, it will be difficult to measure satisfaction of the LRS itself instead of the actual tool performance.

### 2.2.3 Identified gaps

Currently for the Virtual Collection Registry, B2SHARE integration activities are on hold because of B2SHARE development capacity. For this integration, some modifications have to be made on

B2SHARE side. We have had a couple of video conferences with the B2SHARE team to discuss the technical details. As a result, we have created a Virtual Collection Registry integration document<sup>22</sup>.

#### **2.2.4 Future improvements**

For all three services, we are working towards final releases within EOSC-hub. The VCR release is scheduled for M27, the VLO and LRS releases are scheduled for M30. All final releases will focus on further improving the service usability and especially strengthen and further develop the next iteration of the existing integrations. We are also working on an evaluation of the added value of the EOSC-hub AAI solutions for the VCR. Based on this evaluation we will decide if the current CLARIN-provided AAI would require any further modification.

---

<sup>22</sup> <https://docs.google.com/document/d/1HYNDhtNlmcNP3kQg4IgLhcOk7uH60Ij5PJVksJLADM>

## 3 DODAS

### 3.1 Service description

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

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 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 and D7.2 [R1, R2].

### 3.1.1 Architecture and integrated services

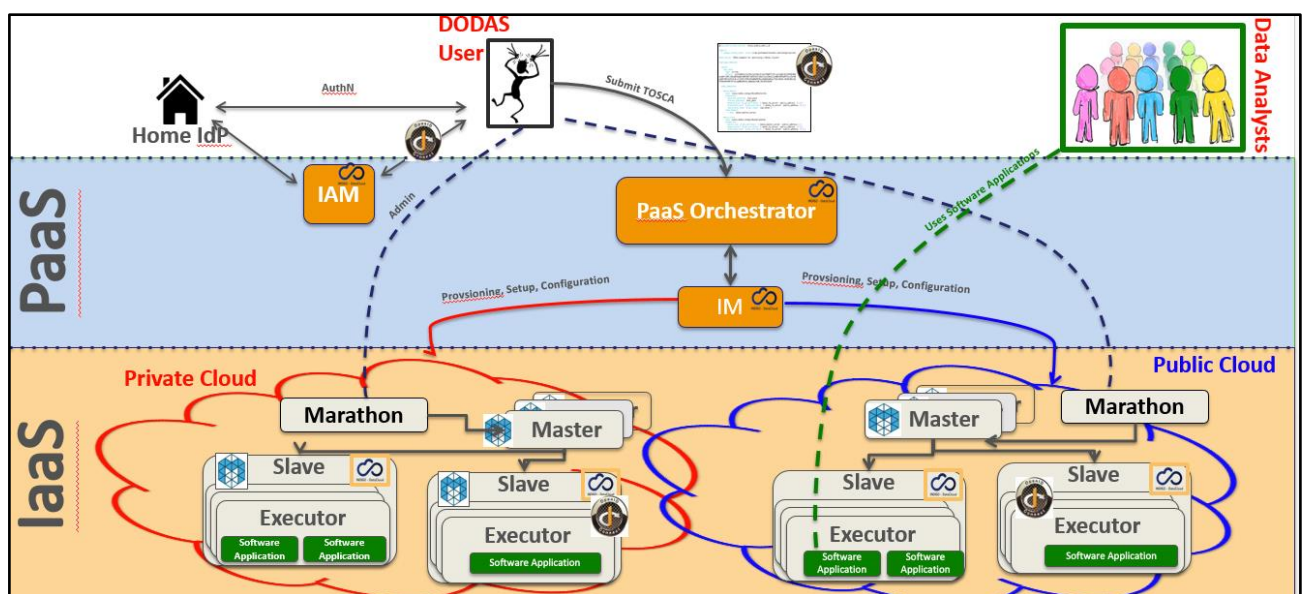
DODAS has a highly modular architecture and its workflows are highly customisable. For this reason, it is very extensible, spanning from software dependencies up to the integration of external services, including also user-tailored code management. There are four main pillars in the DODAS architecture 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 5, 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 eduGAIN<sup>23</sup> 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. DODAS relies on Mesos<sup>24</sup> for resource management and Marathon<sup>25</sup> for the container orchestration. 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 and D7.2 [R1, R2].



**Fig. 5 - 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 HTCondor<sup>26</sup> 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.

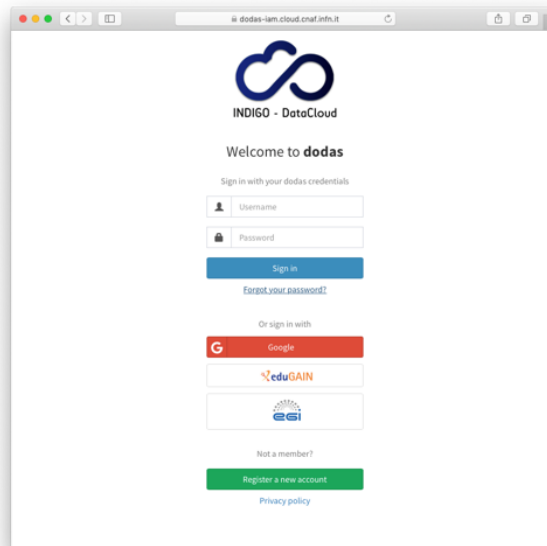
<sup>23</sup> <https://edugain.org/>

<sup>24</sup> <http://mesos.apache.org>

<sup>25</sup> <https://mesosphere.github.io/marathon/>

<sup>26</sup> <https://research.cs.wisc.edu/htcondor/>

EGI Check-in is integrated as an external authentication mechanism of IAM as shown in Figure 6, and IAM is used as an AAI harmonization layer, so that Check-in grants access to the EGI cloud resources, while the IAM harmonised identity is used to manage DODAS domain-specific security as explained before.



**Fig. 6 - DODAS integration with EGI check-in**

### 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 non ROOT based use cases. 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. 7 - 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 CVMFS<sup>27</sup> 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.

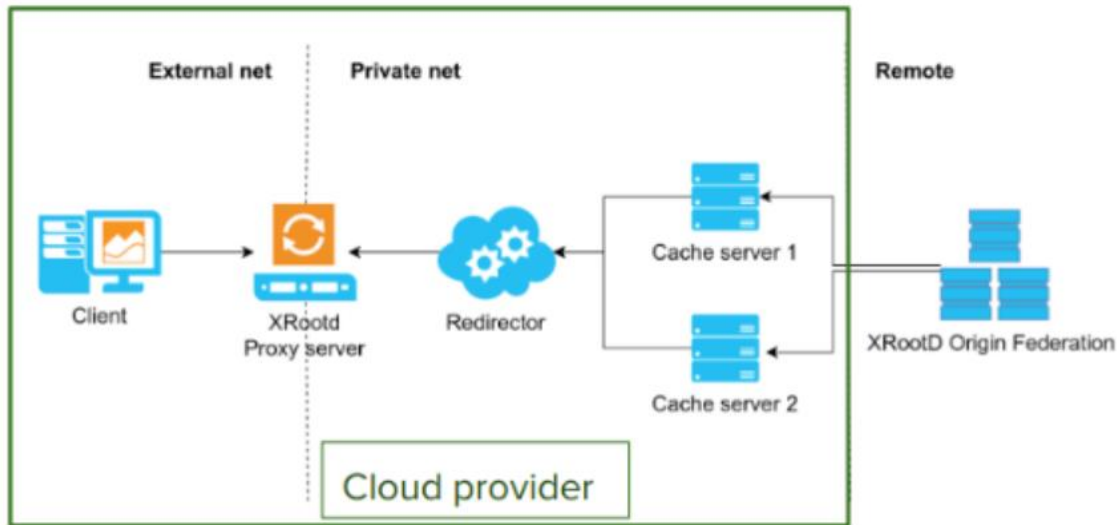
### Support to data caching

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 XRootD<sup>28</sup> and below there is a schema of the current architecture:

<sup>27</sup> <https://cern.service-now.com/service-portal/service-element.do?name=cvmfs>

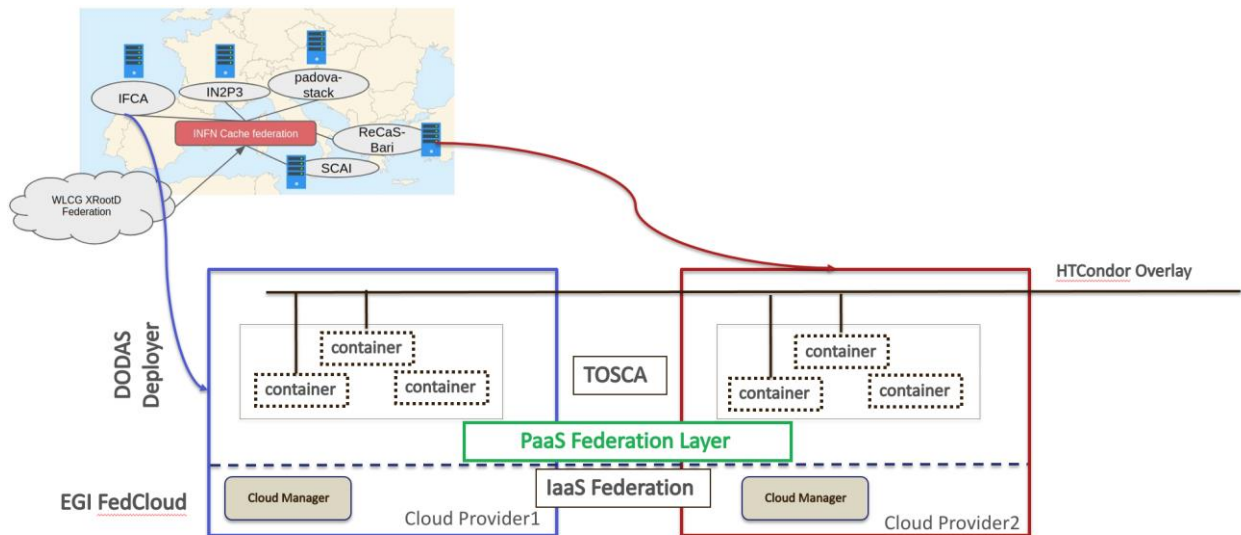
<sup>28</sup> <https://xrootd.slac.stanford.edu/>



**Fig. 8 - DODAS support to data caching**

### Integration with EGI FedCloud

DODAS has been used to integrate and exploit EGI compute resources. In this respect DODAS has been used as PaaS federation layer on top of the EGI FedCloud. A complete integration has been carried out using 5 EGI sites and PaaS federation based on HTCondor overlay. Kubernetes<sup>29</sup> has been adopted as orchestrator solution.



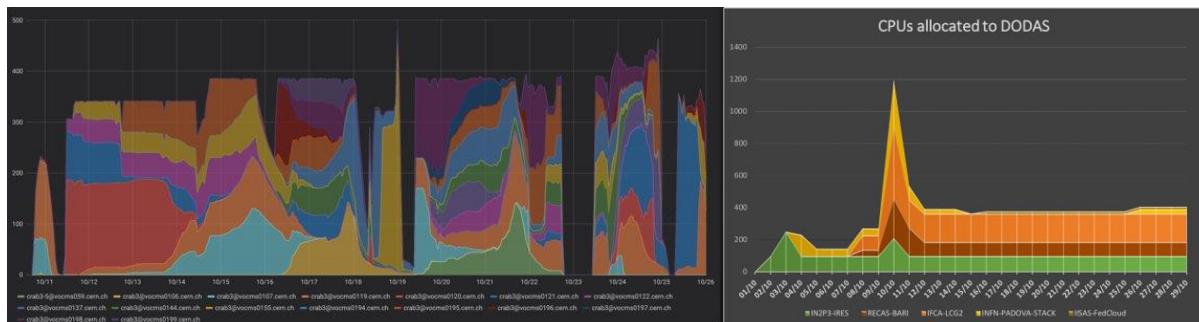
**Fig. 9 - DODAS integration with EGI FedCloud**

A successful validation test has been done implementing a virtual CMS site over 5 distinct providers and the virtual infrastructure, implemented through DODAS, is seen as a single site by CMS.

For the scope of this test, T3\_IT\_Opportunistic\_dodas and CMS Analysis Workflow were used as demonstrator. The EGI FedCloud detailed accounting shows that DODAS pilot has consumed 760K

<sup>29</sup> <https://kubernetes.io/>

CPU hours over 5 sites in October 2019. From CMS monitoring, it is possible to see that 400 jobs run constantly in October.



**Fig. 10 - 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.

### 3.1.2 Main users

Details about the various user groups are available in deliverable D7.2 [R2]. In general, any scientific community with needs to access and harvest resources geographically distributed to implement any type of analysis model spanning from batch like processing to real time processing passing through Machine Learning facilities on demand eventually exploiting specialized hardware.

## 3.2 Services exploitation

### 3.2.1 Usage statistics

DODAS invested a huge amount of effort in supporting and integrating diverse use cases coming from several communities. In particular, apart from CMS additional communities evaluated DODAS and, at different level of maturity, are adopting DODAS as baseline solution for their computing.

Namely the scientific communities are:

- Gravitational Waves: Virgo,
- Astroparticle physics: AMS and Fermi.

In the computing model of CMS DODAS has been developed and integrated in order to provide the capability of generating on demand WLCG compliant sites. DODAS is today a solution for CMS to exploit opportunistic resources and, more in general, to transparently integrate any cloud provider. Regarding AMS, which is the most advanced one after CMS, AMS does not operate a central service to manage workflows and thus it relies on batch systems and technology available on the sites where resources are made available to the collaboration. That said what AMS requires, at first stage, is to access a batch system in any Cloud to process remote data, and thus by adopting DODAS AMS will have an enabler for a Batch-system as a Service.

To accomplish this requirement a set of Ansible roles and TOSCA templates have been developed. TOSCA and Ansible allow to automate the deployment of a HTCondor batch system which in turn is executed as Marathon application over Mesos managed resources. In the case of AMS The HTCondor structure is made of a Central Manager, a submitter node and several worker nodes. The Central Manager is stateless and has the task to coordinate the jobs that the users want to do, so it connects the submitter node to the worker nodes. The submitter node is special because it has also the environment where each user can prepare his/her own jobs and of course it is similar to the environment present in the worker nodes. These ones have the task to execute the user jobs and return the results to the submitter. All the nodes are managed as Docker instances: there is a core image that is extended specifically for the HTCondor cluster. Both users compiled software and centrally managed libraries are distributed through CVMFS. Data ingestion is managed through XrootD, similarly to the CMS use case.

Concerning Virgo, the integration activity is still work in progress at the time of writing this document while Fermi just entered the stress testing phase. The peculiarity of FERMI workflow with respect the AMS one, is the need of a distributed file system for data handling. At the current stage this is provided by an NFS based system while the planned evolution foresees the migration toward a S3 based system.

In addition to the above-mentioned initiative it is worth to mention the following:

- United Kingdom initiative to adopt DODAS: In this use case, DODAS has been exploited to create a CMS Grid Tier-3 site using resources hosted at Imperial College London (ICL), UK. The primary objective of this activity has been to perform a functional test of DODAS to run requirement-specific workflows. For the functional test, a small amount of quota has been reserved on ICL public OpenStack. The new Tier3 site has been called T3\_UK\_Opportunistic\_dodas, and it relies on the Tier2 running at ILC (T2\_UK\_London\_IC) as target storage to copy the produced output data. All the use cases were successfully run at the T3, completing 100% of the jobs. Even though the statistic of the test is not sufficient to perform an accurate and exhaustive study for the general performance of the new infrastructure, the results show a total average above 90% in CPU efficiency, demonstrating the validity of the DODAS-based solution.
- The Instituto de Física de Cantabria (IFCA): IFCA participates actively in the CMS experiment sharing computing resources both within Grid computing infrastructure and with local resources, like HPC or local clusters, where national or international researchers, as well as CMS users, run their analysis. DODAS has been considered as a service where the IFCA-CMS analyses could be run on. This service has been installed on an OpenStack based cloud infrastructure, implemented by the IFCA Advanced Computing and e-Science research group. To integrate DODAS at the IFCA cloud infrastructure. Even though it is considered as a proof of concept, the service could enter the production stage as next step and test site may be turned into a local site in the HTCondor Global Pool.

### 3.2.2 Users' feedbacks

Feedback has mainly been gathered from users during the training and demonstration and hackathon events held during the project which allowed to gather information on users' experience, needs as well as suggestions to improve the services provided, both in terms of features and usability.

Concerning written feedback, it has been collected by means of the official EOSC-hub training evaluation procedure and forms<sup>30</sup>.

Considering a raw estimation of an average value in the range of 1: poor – 6: excellent based on all the feedbacks received, DODAS have got a rate of 4.83 (results from three surveys, with about 50 replies in total).

### 3.2.3 Identified gaps

There are two general areas where gaps have been identified in the DODAS integration, namely

- Unified solution for AAI,
- Advanced and dynamic systems for both Monitoring and Accounting.

<sup>30</sup> <https://documents.egi.eu/public/ShowDocument?docid=3296>

Both gaps were mitigated so far by putting in place workarounds and this caused some delay in the integration plan compared to the original one. However, the issues are by no means so severe that they would ultimately stop delivery of a fully integrated DODAS service.

### **3.2.4 Future improvements**

Almost all the integration activities foreseen by the WP7 workplan have been now carried on. The two main aspects to consider in the next months are the Accounting and Monitoring. More in general the training events and exploitations will continue throughout the project lifetime in order to engage additional users and communities.

In term of features the key is the improvements on supporting a variety of Machine Learning based workflows, primarily the Training Facility support.

The exploitation of new communities will continue by means of dissemination and training as well as hackaton session.

Concluding the assets are:

- Improve Machine Learning supports,
- Exploitations of new communities,
- Integration of new workflows and use cases.

## 4 ECAS

### 4.1 Service description

<b>Service/Tool name</b>	ENES Climate Analytics Service (ECAS)
<b>Service/Tool url</b>	CMCC ENDPOINT: <a href="https://ecaslaboratory.cmcc.it/web/home.html">https://ecaslaboratory.cmcc.it/web/home.html</a> DKRZ ENDPOINT: <a href="https://ecaslaboratory.dkrz.de/home.html">https://ecaslaboratory.dkrz.de/home.html</a>
<b>Service/Tool information page</b>	<a href="https://www.eosc-hub.eu/services/ENES%20Climate%20Analytics%20Service">https://www.eosc-hub.eu/services/ENES%20Climate%20Analytics%20Service</a> <a href="https://portal.enes.org/data/data-metadata-service/processing/ecas">https://portal.enes.org/data/data-metadata-service/processing/ecas</a>
<b>Description</b>	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
<b>Value proposition</b>	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).
<b>Customer of the service/tool</b>	Climate modelling community; direct downstream usage communities
<b>User of the service/tool</b>	Researchers; Scientific users
<b>User Documentation</b>	<a href="https://ee-docs.readthedocs.io/en/latest/">https://ee-docs.readthedocs.io/en/latest/</a> <a href="http://ophidia.cmcc.it/documentation/index.html">http://ophidia.cmcc.it/documentation/index.html</a>
<b>Technical Documentation</b>	<a href="https://github.com/ECAS-Lab/ecas-authentication">https://github.com/ECAS-Lab/ecas-authentication</a> <a href="https://github.com/ECAS-Lab/ecas-accounting">https://github.com/ECAS-Lab/ecas-accounting</a> <a href="http://ophidia.cmcc.it/documentation/admin/index.html">http://ophidia.cmcc.it/documentation/admin/index.html</a>
<b>Product team</b>	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (Fondazione CMCC); Deutsches Klimarechenzentrum GmbH (DKRZ)
<b>License</b>	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.
<b>Source code</b>	<a href="https://github.com/ECAS-Lab">https://github.com/ECAS-Lab</a> ; <a href="https://github.com/OphidiaBigData">https://github.com/OphidiaBigData</a>
<b>Testing</b>	1. ECAS single-instance VMI uploaded to the EGI AppDB: <a href="https://appdb.egi.eu/store/vappliance/ecas">https://appdb.egi.eu/store/vappliance/ecas</a> Deployment test using the EGI AppDB VMops dashboard: <a href="https://dashboard.appdb.egi.eu/vmops">https://dashboard.appdb.egi.eu/vmops</a> 2. Multi-node ECAS environment, dynamically provisioned on the EGI FedCloud through the EC3 LTOs service: <a href="https://servproject.i3m.upv.es/ec3-ltos/index.php">https://servproject.i3m.upv.es/ec3-ltos/index.php</a>

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 framework<sup>31</sup>, which has been integrated with B2DROP, ESGF, IAM, Onedata (DataHub), EGI FedCloud, JupyterHub, and the ECAS-Lab web portal.

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

#### 4.1.1 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 and IAM 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 D7.1 [R1].

Figure 11 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 first year and part of the second year of the project.

In order to store all the developments, the integration efforts and the training material concerning ECAS, a Github repository<sup>32</sup> has been set up. Extensions and adaptations to other tools have been published, instead, on the related Github repositories (i.e. for the Ophidia framework<sup>33</sup>), while virtual appliances have been published on the EGI services (e.g. for ECAS VM<sup>34</sup>).

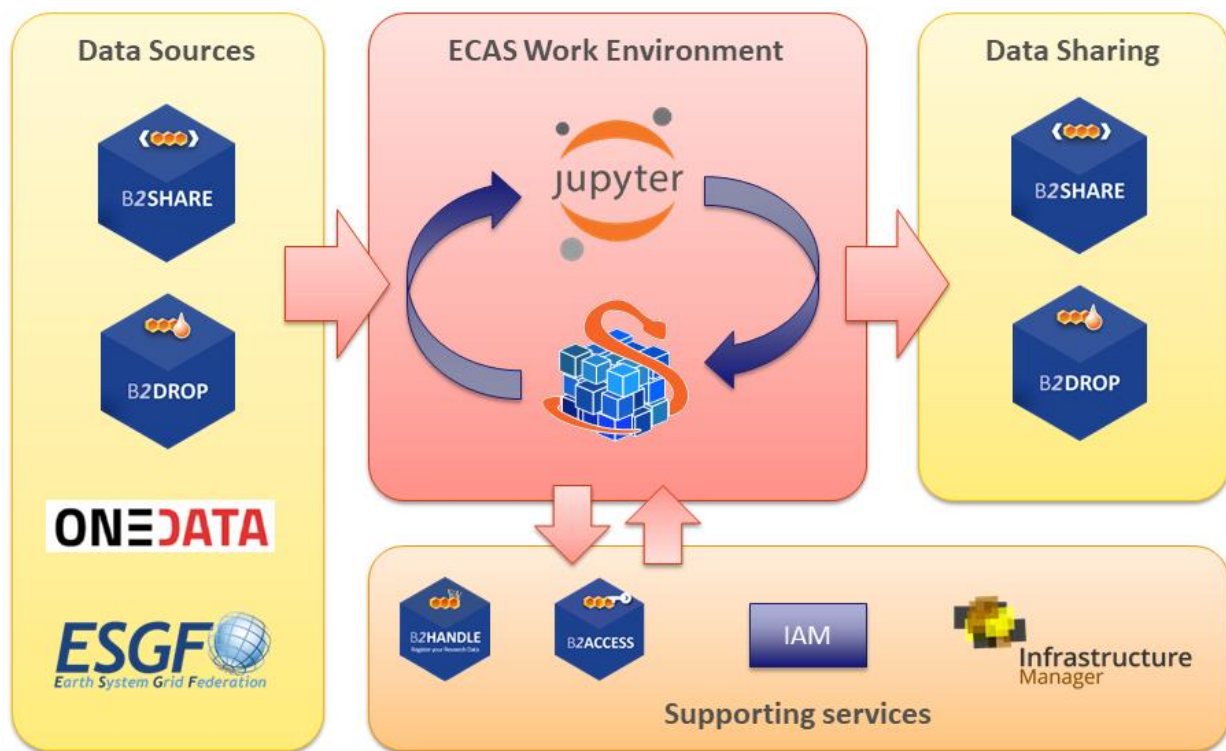
---

<sup>31</sup> <http://ophidia.cmcc.it/>

<sup>32</sup> <https://github.com/ECAS-Lab/>

<sup>33</sup> <https://github.com/OphidiaBigData>

<sup>34</sup> <https://appdb.egi.eu/store/vappliance/ecas>



**Fig. 11 - ECAS architecture overview**

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. Integration activities for the first year have already been presented in detail in D7.2 [R2]. The following subsections, after providing a brief summary of the activities started in the first year, are mainly focused on the activities finalized during the second year of the project.

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.

During the second year of the project, the 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 final 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 features and self-study/training material. Finally, the views for user accounting have been fully implemented and the ECAS environment has

been successfully integrated into the EGI VA marketplace and the FedCloud deployment services. A detailed description of the various activities is provided as follows.

### **Extensions to Ophidia for supporting additional use cases**

The Ophidia Framework and the related Python bindings (PyOphidia) have been extended with new 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, a faster version of the data export function from the Ophidia server side to the Python code (and also Jupyter Notebooks). All developments are openly accessible from the Ophidia Github repository.

### **Additional extensions for the integration of B2DROP into ECAS**

Based on the extensions developed during the first year, 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 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 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.

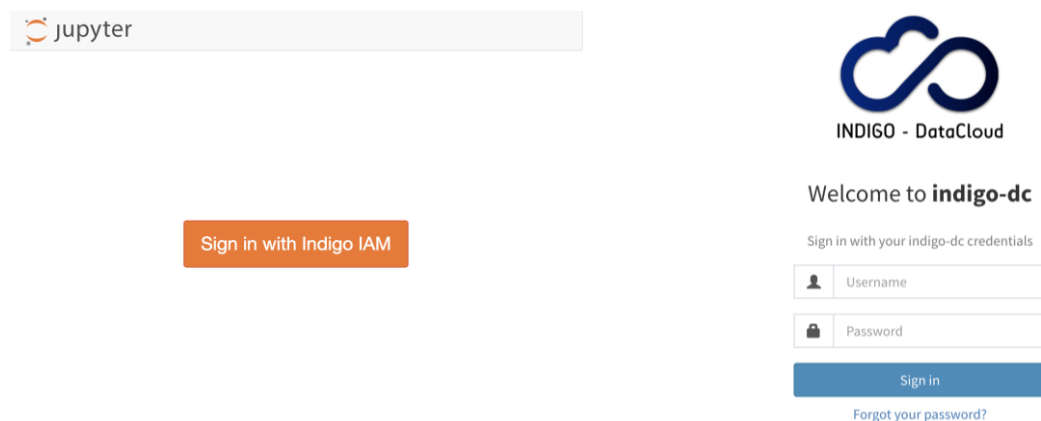
### AAI Integration of Ophidia and JupyterHub with IAM

The integration of IAM with JupyterHub and Ophidia, started during the first year, has been finalized. 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.

From a technical standpoint, the *LocalGenericOAuthenticator* class (provided by the JupyterHub OAuth authenticator<sup>35</sup>) has been extended to override the *authenticate()* method in order to check if the (*username, email*) pair is already available in the Ophidia user management database (i.e. OphidiaDB). If the user exists, the JupyterHub spawner starts a new JupyterHub instance using the proper Linux user and propagates the user token to the Ophidia environment; otherwise, a message containing the link to the registration page will be shown. Alternatively, the proposed solution also allows the automatic creation of an ECASLab account (using the user's INDIGO IAM credentials) by setting the following configuration option in the *jupyterhub\_config.py* file:

```
LocalEnvAuthenticator.ecas_user = True
```

Figure 12 shows the JupyterHub interface allowing a user to log in to the system using its IAM account.



**Fig. 12 - JupyterHub interface**

The extensions have been published on the Github repository of ECASLab<sup>36</sup>.

### Integration of ECAS into the EGI FedCloud

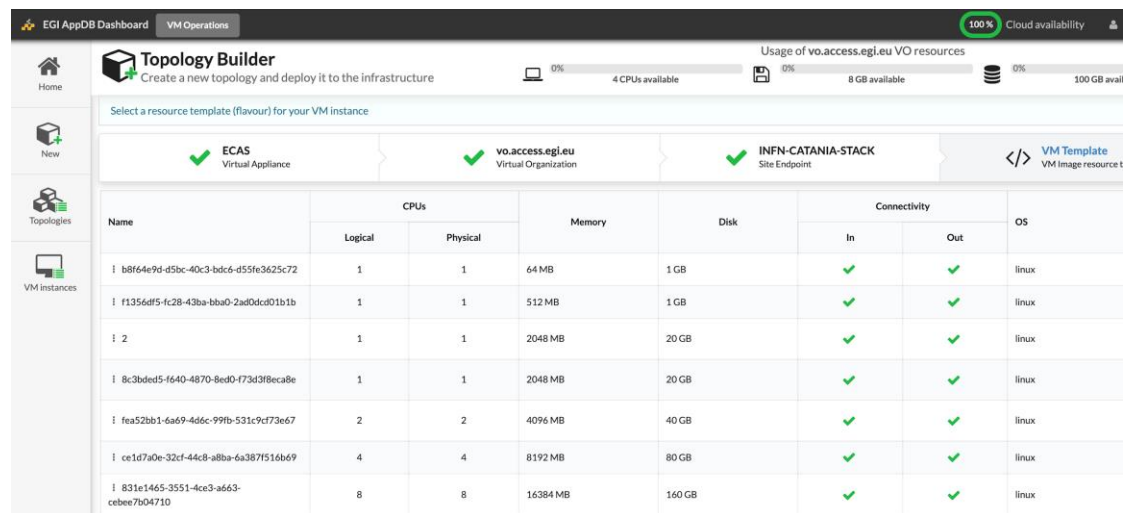
The integration of ECAS into the EGI FedCloud has been accomplished by considering two different scenarios.

<sup>35</sup> <https://github.com/jupyterhub/oauthenticator>

<sup>36</sup> <https://github.com/ECAS-Lab/ecas-authentication>

In the first scenario, an ECAS single-instance VMI, providing a ready-to-use ECAS environment, has been created and uploaded to the EGI AppDB<sup>37</sup>. The VMI has been assigned to a set of trusted VOs in order to be deployed on the FedCloud.

A deployment test has been carried out using the EGI AppDB VMops dashboard<sup>38</sup>, 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.



The screenshot shows the 'VM Operations' tab in the EGI AppDB dashboard. It features a 'Topology Builder' section with a table of VM instances. The table has columns for Name, CPUs (Logical and Physical), Memory, Disk, Connectivity (In and Out), and OS. All instances are running on Linux and have green checkmarks in the Connectivity columns.

Name	CPUs		Memory	Disk	Connectivity		OS
	Logical	Physical			In	Out	
i b8f64e9d-d5bc-40c3-bdc6-d55fe3625c72	1	1	64 MB	1 GB	✓	✓	linux
i f135d6f5-fc28-43ba-bba0-2ad0dcd01b1b	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. 13 - ECAS integration with EGI FedCloud**

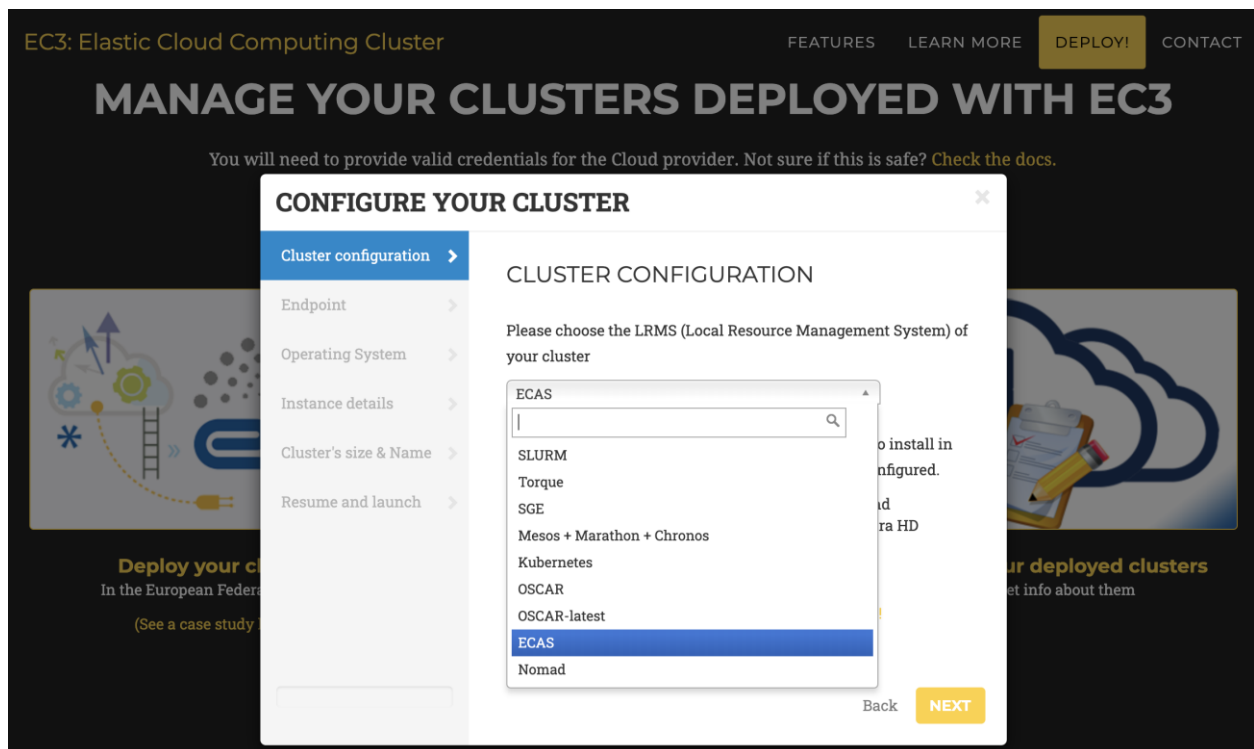
The second scenario refers to a multi-node ECAS environment, which is dynamically provisioned on the FedCloud through the EC3 LToS service<sup>39</sup> according to the user requirements. In this case, a RADL file for the Infrastructure Manager (IM) has been implemented and tested with the support of the UPV Grycap team to define the cluster setup in terms of resources, infrastructure and software configuration, and contextualization<sup>40</sup>.

<sup>37</sup> <https://appdb.egi.eu/store/vappliance/ecas>

<sup>38</sup> <https://dashboard.appdb.egi.eu/vmops>

<sup>39</sup> <https://servproject.i3m.upv.es/ec3-ltos/index.php>

<sup>40</sup> <https://github.com/grycap/ec3/blob/devel/templates/ophidia.raml>



**Fig. 14 - ECAS integration with EGI FedCloud - multinodes scenario**

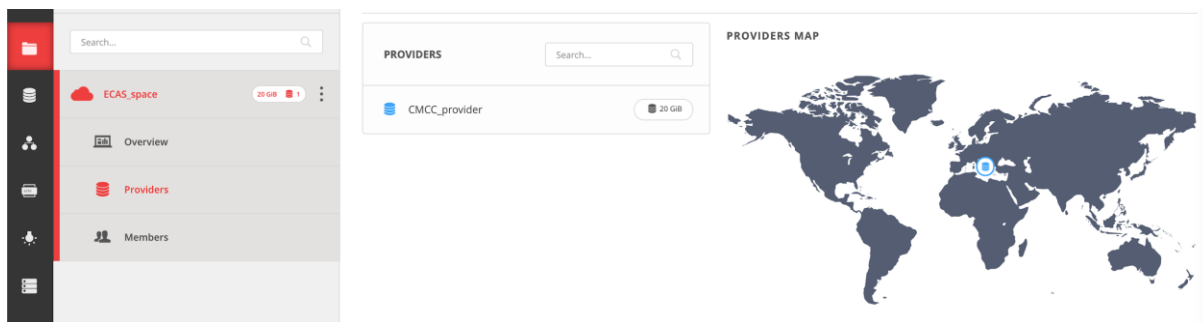
For aspects related to software setup and contextualization, both scenarios rely on the Ophidia Ansible Role, which has been extended to include the whole ECAS environment services, i.e. JupyterHub, PyOphidia, several Python libraries such as numpy, matplotlib and Basemap. The updated version of the role is available <sup>41</sup>.

### Integration of ECASLab with Onedata

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

To this end, 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.

<sup>41</sup> <https://github.com/OphidiaBigData/ansible-role-ophidia-cluster>



**Fig. 15 -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 published<sup>42</sup>.

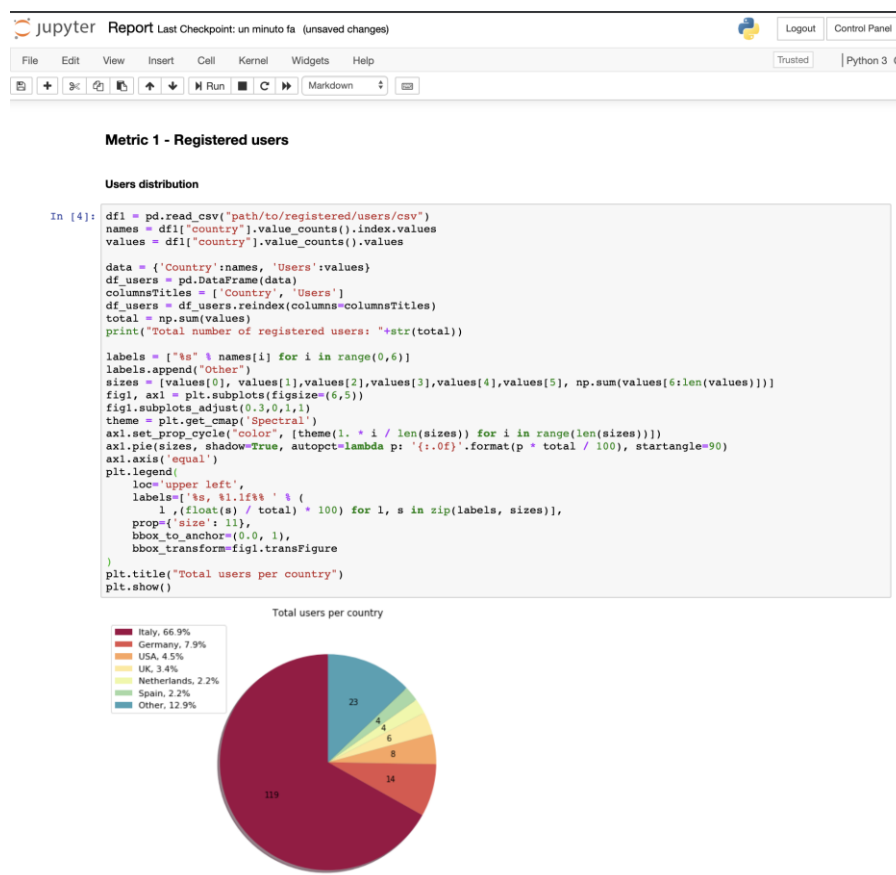
Starting from the information tracked by the Ophidia server, a set of accounting metrics has been identified to extract useful statistics about users and computing resources usage:

<sup>42</sup> <http://ophidia.cmcc.it/documentation/admin/management/accounting.html>

- total number of users and their distribution by country,
- total number of executed tasks,
- most used operators,
- total number of cores and cores/hour used,
- cumulative execution time per operator,
- total number of workflows and operators per user,
- total number of cores and cores/hour per user.

To extract the above metrics, data stored in the log files has been imported into a database using the PostgreSQL DBMS. At the same time, a set of queries and views have been defined in order to calculate each metric.

A report containing the statistics about a given period has been created in the form of Jupyter Notebook (Figure 16). 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.



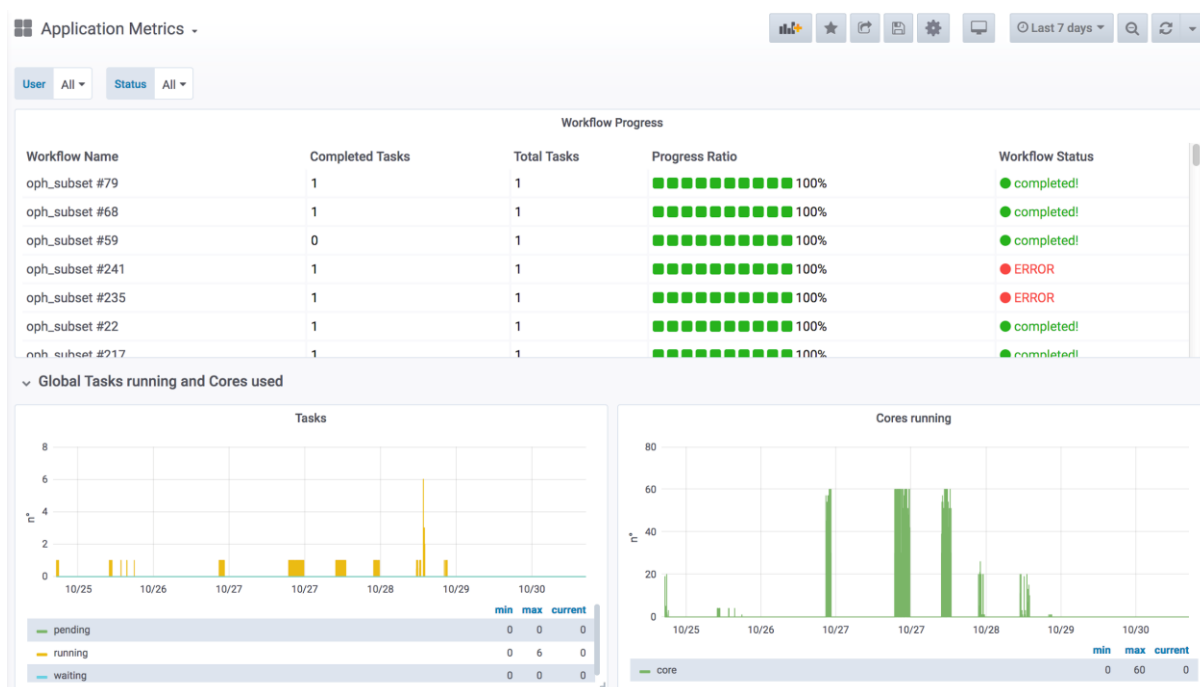
**Fig. 16 - ECAS metrics reporting**

Some examples of the output provided by this interface are given in the “Service Exploitation” section. The notebook, as well as the related developments for virtual access are available<sup>43</sup>.

### 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 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.

The following figure shows an example of the new dashboards implemented in Grafana<sup>44</sup>.



**Fig. 17 - ECAS monitoring**

### JupyterHub notebooks for training purposes

To support user training and demonstration activities, Jupyter Notebooks integrating some of the key features of ECAS have been implemented. The notebooks are available on the ECASLab github repository<sup>45</sup> 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:

<sup>43</sup> <https://github.com/ECAS-Lab/ecas-accounting>

<sup>44</sup> <https://grafana.com/>

<sup>45</sup> <https://github.com/ECAS-Lab/ecas-notebooks>

- *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;
- *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.
- In development: comparison of observational and simulation data (an obs4mips- variable related to the water cycle e.g. hfls = latent heat flux).

#### 4.1.2 Main users

ECAS provides server-side data analytics capabilities, which may be relevant for a wide range of users from many disciplines. The following is a summary of the most relevant user groups for the upcoming dissemination and, in particular, training activities that ECAS will provide. In particular, we target those who are not yet strongly accustomed in their analysis method, e.g. students from Uni-HH (Hamburg, Germany) or Unisalento (Lecce, Italy). Basically, everybody who is willing to learn new methods with Jupyter. More details about the various user groups are available in deliverable D7.2.

- Users with no directly available computing or data analysis resources,
- Users from the climate data downstream communities,
- Climate data users interested in performing cross-model ensemble analysis, and that need a “big data” approach to easily manage large-scale analysis,
- Users interested in prototyping applications with datacube concept, expecting speed-up compared to conventional approaches

## 4.2 Services exploitation

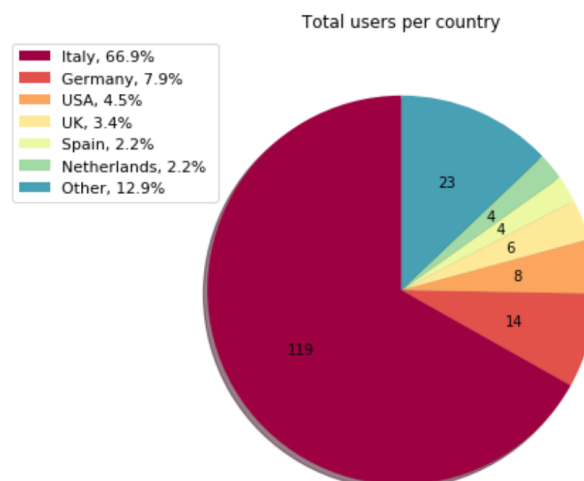
### 4.2.1 Usage statistics

ECAS is being provided by CMCC and DKRZ. The available resources, as already mentioned in details D7.1 [R1], are:

- 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 8 cores, 16GB memory, 12GB disk, executing the Ophidia Server and complementary administration 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).

As described in the “Architecture and integrated services” section, a set of metrics has been calculated using the information collected by the Ophidia server.

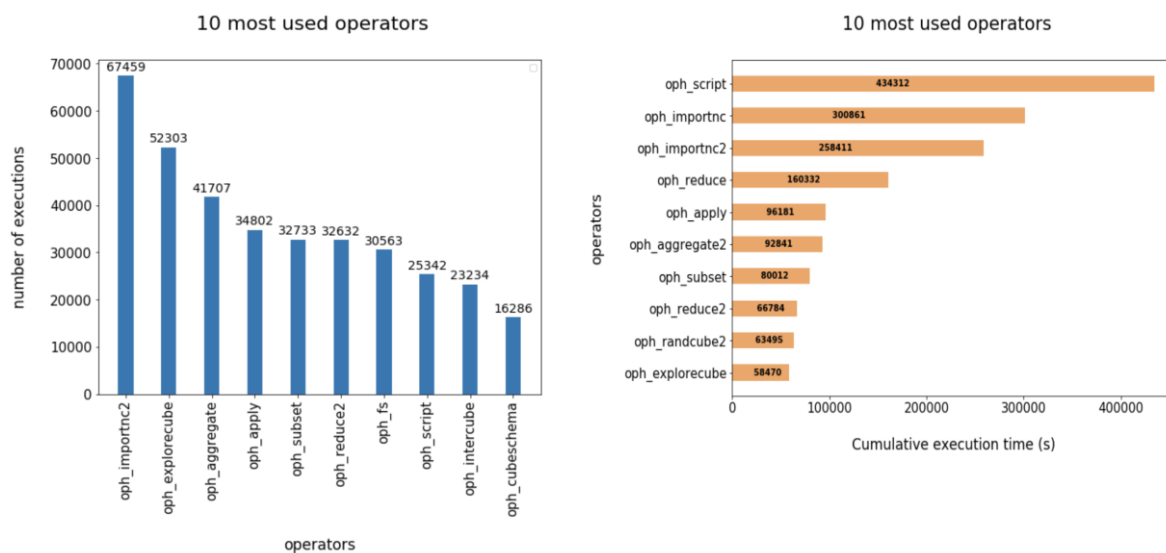
Concerning the CMCC deployment, in the M1-M18 period about 200 users from over 20 different countries and 60 affiliations (universities and research institutes) have created an account on ECASLab. Users distribution is shown in the following figure.



**Fig. 18 - ECAS usage statistics - Users per country**

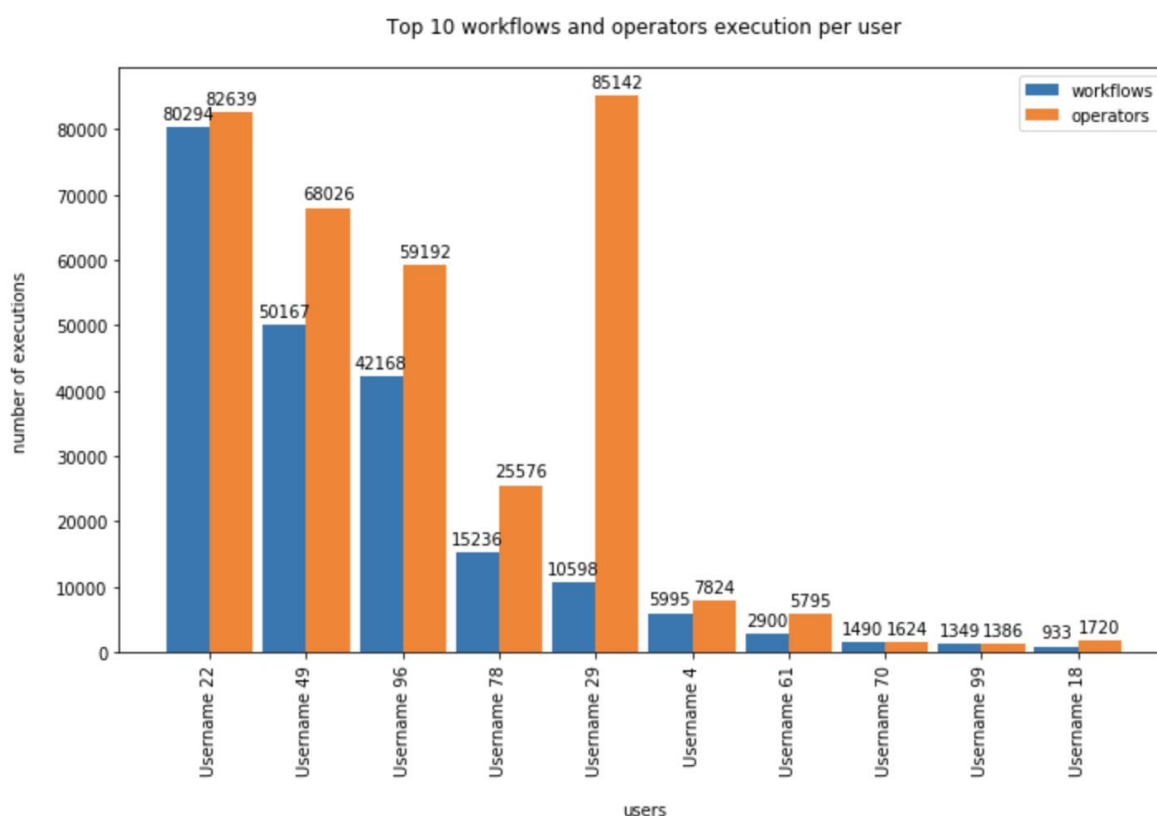
Over 500 thousand workflows have been executed, for a total of about 800 thousand jobs and about 6500 cores/hour.

The following bar charts show the 10 most used Ophidia operators, both in terms of number of executions and cumulative execution time.



**Fig. 19 - ECAS usage statistics - Most used operators**

The workflows and operators execution has also been tracked on a user basis, as depicted in the grouped bar chart below.



**Fig. 20 - ECAS usage statistics - Number of executions**

#### 4.2.2 Users' feedbacks

Feedbacks have mainly been gathered from users during the training and demonstration events held in the first part of the project. These events allowed the ECAS team to collect feedback on the users' experience, their needs as well as suggestions to improve the services provided, both in terms of features and usability.

Concerning written feedback, it has been collected by means of the official EOSC-hub training evaluation procedure and forms<sup>46</sup>.

In terms of users' evaluation, overall, the features provided by ECAS have been positively assessed. In particular, although the users reported an initially steep learning curve, mainly due the wide set of features available, the integrated environment has been praised for providing very easy access to various Jupyter datasets and multiple analytics features, jointly with computational resources, all from a single JupyterHub-based interface.

These comments have been taken into consideration during the integration activities in order to improve the users experience with ECASLab.

#### 4.2.3 Identified gaps

<sup>46</sup> <https://documents.eui.eu/public/ShowDocument?docid=3296>

The integration with B2DROP as done currently is less than ideal, ultimately resulting from not yet completed integration between IAM and B2DROP and lack of support for secured programmatic upload to B2DROP via secured webdav (the latter is a feature B2DROP would ultimately require from nextcloud). If users share results or scripts from the JupyterHub environment, they need to provide their B2ACCESS credentials separately from the original login to the ECAS environment, which seems cumbersome. A workaround was put in place that enables users to use B2DROP nonetheless if they provide their credentials via a separate file; this must only be done once by every user.

#### **4.2.4 Future improvements**

The expected integration activities, described in the previous sections, have been concluded in the 2nd year of the project, while training events will continue throughout the project lifetime in order to engage additional users.

We plan to make a webinar about the new workflow to increase the presence of ECAS in the world-wide-web. This would be an entirely new medium for ECAS and we would reach out to more potentially new communities. It'd be interesting to gather feedback from these individuals. To begin with a simple counter of webinar views will do.

ECAS will continue the B2SHARE integration that was started in PY2 to enable data tracing and foster workflow transparency by additional user groups and a wider audience. Feedback from training activities and user community interaction indicates further steps will be required to adapt and expose metadata according to community-specific workflows and expectations. Further interaction with the B2SHARE development team is required to enable full integration of the ECAS-B2SHARE workflow to automate data tracking via PIDs and metadata and expose information through the web UI adequately to different audiences (user communities and classes of users within them, i.e., different stakeholders such as data providers, consumers and PIs).

Concerning the content, we aim to develop a new notebook about the water cycle, picking examples and variables where simulation data and observations can be compared to each other. The observable(s) should be of quantitative nature (e.g. numbers or distributions).

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

## 5 GEOSS

### 5.1 Service description

<b>Service/Tool name</b>	GEO Discovery and Access Broker (GEO DAB) Virtual Laboratory (VLab)
<b>Service/Tool url</b>	GEO DAB API: <a href="http://api.geodab.eu/">http://api.geodab.eu/</a> VLab API: <a href="http://vlabapi.geodab.org/">http://vlabapi.geodab.org/</a>
<b>Service/Tool information page</b>	GEO DAB: <a href="https://www.geodab.net/">https://www.geodab.net/</a> VLab: <a href="https://essilab.wixsite.com/vlab">https://essilab.wixsite.com/vlab</a>
<b>Description</b>	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.
<b>Value proposition</b>	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.
<b>Customer of the service/tool</b>	Research Communities
<b>User of the service/tool</b>	N/A
<b>User Documentation</b>	GEO DAB: <a href="https://www.geodab.net/">https://www.geodab.net/</a> VLab: <a href="https://confluence.geodab.eu/display/VTD/VLab+Documentation">https://confluence.geodab.eu/display/VTD/VLab+Documentation</a>
<b>Technical Documentation</b>	GEO DAB: <a href="https://www.geodab.net/">https://www.geodab.net/</a> VLab: <a href="https://confluence.geodab.eu/display/VTD/VLab+Documentation">https://confluence.geodab.eu/display/VTD/VLab+Documentation</a>
<b>Product team</b>	CNR
<b>License</b>	N/A
<b>Source code</b>	N/A
<b>Testing</b>	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 [R2]. 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

### 5.1.1 Architecture and integrated services

VLab utilizes EOSC-hub computing cloud resources provided by CESNET. A Kubernetes cluster was installed and utilized for the deployment of the VLab software modules. When the generation of a new product is requested, VLab implements the following steps: (i) retrieves the necessary input data utilizing GEO DAB APIs, (ii) stores the data in the execution environment, (iii) creates a Docker container with the required libraries and model source code for the product generation, (iv) launches the Docker container, and (v) saves the generated product making it available on the Web. Besides this default workflow, when possible, VLab can launch the Docker container for execution in a different Cloud environment which already hosts required input data; this allows a more efficient execution since it allows to avoid data download time. This was experimented with three DIAS platforms: ONDA, Sobloo and CREODIAS.

### 5.1.2 Main users

The target users of the service were identified in D7.2:

- Single researcher,
- Virtual Organization (VO),
- Research project,
- Business.

In particular, the use of VLab for the generation of new products targets both scientific experts and decision makers (e.g. Protected Area managers). Scientific experts can utilize VLab to either share their model for re-use or to compare different existing models addressing the same issue for their specific use-case. Decision makers are expected to interact with VLab utilizing dedicated client applications which, via VLab APIs, allow them to seamlessly generate the products they need for an evidence-based decision-making process.

## 5.2 Services exploitation

### 5.2.1 Usage statistics

The GEO DAB offers discovery and access to data provides to GEOSS by more than 180 systems, with more than 400M discoverable metadata records. The usage statistics of the GEOSS Platform are maintained by GEOSS Platform team and will be published on the GEOSS Web Portal.

VLab provides the possibility to execute more than 20 scientific models, developed in different programming languages including: Java, Python, NetLogo, Matlab, R. The number of model executions during the last 9 months (February to end of October 2019) is 874.

### 5.2.2 Users' feedbacks

As far as GEO DAB is concerned, the GEOSS Platform team collects users' feedback through an online form where users can provide enhancement requests and/or report bugs.

VLab feedbacks were collected during webinar and workshops. As far as modelers are concerned, general feedback is positive in terms of model sharing with some exception mainly related to the use Docker technology with which not all modelers are familiar.

A VLab-based demo was developed in the context of the EuroGEOSS Sprint-to-Ministerial activity. The demo is an enhancement of the pilot presented in 2018 for EOSC Launch event. A video was produced for the demo<sup>47</sup>.

During the last GEO Plenary meeting in Camberra (November 2019), the developed demo was widely shown at the EC booth during the entire meeting, at the EuroGEO side event and during the plenary session. The general feedbacks were positive, particularly highlighting the value of being able to exploit different European cloud capacities (EOSC and DIAS platforms). It was however noted that a simplified GUI for decision-makers could improve the usability of the service.

### 5.2.3 Identified gaps

The gaps identified in D7.2 are still present. As far as integration with AAI of EOSC-hub, a use-case was defined to experiment this integration and experiments will be developed during the last year of the project. The defined use-case aims at allowing users to utilize their own AAI credentials to launch the required computing resources for VLab models execution. As far as additional cloud services, the ones identified in D7.2 would allow a better integration with EOSC-hub infrastructure.

Finally, an important requirement for an efficient execution of VLab models is the possibility to use cloud-hosted data as inputs; in particular, the possibility to access the entire set of Copernicus Sentinel data in EOSC would be of great value.

### 5.2.4 Future improvements

The main objective for next year is to consolidate the present integration, in particular concerning VLab. This will require to experiment the use of non-cloud provided services for the serverless and queue messaging functionalities, e.g. deploying Kubernetes-based software. Although not optimal, this solution might help filling the gap.

As far as new functionalities, the development for the AAI integration use-case will also be experimented, but no production-level can be guaranteed at the moment.

---

<sup>47</sup> <https://youtu.be/SigmXfKryVY>

## 6 OPENCoastS

### 6.1 Service description

<b>Service/Tool name</b>	OPENCoastS
<b>Service/Tool url</b>	<a href="https://opencoasts.ncg.ingrid.pt/">https://opencoasts.ncg.ingrid.pt/</a>
<b>Service/Tool information page</b>	<a href="http://opencoasts.lnec.pt/index_en.php">http://opencoasts.lnec.pt/index_en.php</a>
<b>Description</b>	The OPENCoastS service assembles on-demand circulation forecast systems for selected coastal areas and keeps them running operationally for a period defined by the user. This service generates daily forecasts of water levels and vertically averaged velocities over the region of interest for 48 hours, based on numerical simulations of the relevant physical processes.
<b>Value proposition</b>	A user-friendly web portal allows users to setup a forecast system of their coastal region of interest.
<b>Customer of the service/tool</b>	Coastal Research Communities, Coastal end-users, Coastal authorities
<b>User of the service/tool</b>	researchers, people in charge of emergency and daily planning at coastal infrastructures
<b>User Documentation</b>	Manual: <a href="http://opencoasts.lnec.pt/OPENCoastS_manual.htm">http://opencoasts.lnec.pt/OPENCoastS_manual.htm</a> 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: <a href="https://doi.org/10.1016/j.envsoft.2019.104585">10.1016/j.envsoft.2019.104585</a>
<b>Technical Documentation</b>	not public at this stage
<b>Product team</b>	LNEC, LIP, CNRS/LIENS, UNICAN
<b>License</b>	Apache License Version 2.0
<b>Source code</b>	not public at this stage
<b>Testing</b>	a suite of testing deployments, covering all options, was defined and is conducted every time a new release is performed

The OPENCoastS service builds on-demand circulation forecast systems for user-selected sections of the North Atlantic coast and maintain them running operationally for the time frame defined by the user. This service generates daily forecasts of water levels and 2D velocities over the spatial region of interest for periods of 48 hours, based on numerical simulations of the 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 or 3D baroclinic circulation in the coastal area of interest. The 3D baroclinic circulation option was added in the last few months.

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 and 7.2 [R1, R2], on the several papers published in 2019: OPENCoastS core paper (published in open access), OPENCoastS grid repository paper and OPENCoastS 3D version papers [R3-R5].

### **6.1.1 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, a backend where models and mapping services run and a storage tier for preservation (more information can be retrieved from EOSC-hub deliverable 7.1, 7.2 and Oliveira et al., 2020).

In the first 10 months of the second year of the project, effort was devoted to integrating OPENCoastS with the EGI High-Throughput Compute service and EGI Workload Manager (DIRAC). This is operational using Storm with WebDAV for data transfer between the cloud and GRID, and is already available at the staging site<sup>48</sup>. After detailed testing by the OPENCoastS team, it will be included in the production site<sup>49</sup>.

---

<sup>48</sup> <https://opencoasts-alt.ncg.ingrid.pt>

<sup>49</sup> <https://opencoasts.ncg.ingrid.pt/home/>

### 6.1.2 Main users

The OPENCoastS service is available freely to anyone who plays a role in coastal areas, from coastal authorities to the general public. Presently, it has been applied to estuaries and coastal regions in 6 continents with local and regional scale applications. Current users also span 6 continents.

The OPENCoastS platform is nowadays used by coastal managers, public institutions, research groups and private companies. Users are mainly from the research and academic communities, although the number of coastal managers has been growing, in particular after the last two trainings. Below is a map of the spatial distribution of service uptake order, starting from the national service at the beginning of the project:



**Fig. 21 - OPENCoastS service update geographic distribution (number indicates the first registration from each country, the pink box is the initial stage of the service, before EOSC-hub)**

## 6.2 Services exploitation

### 6.2.1 Usage statistics

OPENCoastS has over 200 users from 31 countries in 6 continents.

Regarding the resource's usage, there are two sites supporting OPENCoastS: INCD (managed by LIP) and IFCA. The resources used for each site since February 2018 are:

- INCD:
  - CPU hours: 483591,86
  - RAM MB hours: 990393998,57
  - Storage GB hours: 3083281,54
- IFCA:

- CPU hours: 454559.24
- RAM MB hours: 843470431.86
- Storage GB hours: 5079815.68

### 6.2.2 Users' feedbacks

Feedback from users are provided mainly at the training events, although the service has its own rating and feedback form.

#### Legend

excellent	6
very good	5
good	4
average	3
fair	2
poor	1

Below is a summary of the evaluations at the 4 training events:

#### 1) IMUM 2018

	average	stand. Dev.
objective	5,5	0,5
overall	5,2	0,4
tutorial	5,2	0,8

#### 2) E-course 2018

	average	stand. Dev.
<b>Overall Evaluation</b>	5,1	0,7
<b>T1 - Coastal processes</b>	5,3	0,9
<b>T2 - Understanding and using Model</b>	4,9	0,8
<b>T3 - Generation of triangular finite element grids for</b>	5,3	0,7
<b>T4 - Forecast systems: an overview</b>	5,2	0,8
<b>T5 - The OPENCoastS service</b>	5,2	0,8
<b>T6 - E-infrastructures and</b>	4,6	1,0

#### 3) Silusba 2019

	average	stand.dev.
objective	4,7	0,7
overall	4,4	1,0
tutorial	4,7	1,0

#### 4) Medcoast 2019

	average	stand.dev.
overall tutorial	5,7	0,6
coastal forecasting	6,0	0,0
Generation of triangular meshes	5,7	0,6
OPENCoastS hands-on	5,7	0,6

### 6.2.3 Identified gaps

For this thematic service, DIRAC is exploiting the INCD site using StoRM with WebDAV for data transfer between the cloud and GRID.

During the integration, several issues were identified related to the usage of DIRAC which were conveyed both to the DIRAC support team and also to the sites. The use of DIRAC with the INCD site is operational, while the use of DIRAC with the IFCA site is being further trouble-shooted. As a contingency plan an alternative approach using direct job submission to the computing elements was also developed by LIP. The support for WebDAV was also added to the StoRM installation at the IFCA site to support OPENCoastS.

### 6.2.4 Future improvements

Integration with EUDAT services is now being planned and will be implemented in the next few months.

Inter-thematic-service collaboration will be evaluated in 2020, in particular with the ECAS service. The data processing services in ECAS are a valuable add-on to OPENCoastS outputs. Additionally, given the proximity between both users' communities, joint training maybe planned.

## 7 WeNMR

### 7.1 Service description

Service/Tool name	WeNMR
Service/Tool url	<a href="https://www.eosc-hub.eu/services/WeNMR%20suite%20for%20Structural%20Biology">https://www.eosc-hub.eu/services/WeNMR%20suite%20for%20Structural%20Biology</a>
Service/Tool information page	<a href="http://www.wenmr.eu">http://www.wenmr.eu</a>
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	<a href="http://www.wenmr.eu">http://www.wenmr.eu</a>
Technical Documentation	<a href="http://www.wenmr.eu">http://www.wenmr.eu</a>
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 and 7.2 [R1, R2], 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).

### 7.1.1 Architecture and integrated services

All services provided by the WeNMR suite are web-based and they share a unified look'n'feel, despite being built on a variety of different technological solutions. They all make use of the EOSC HTC resources to distribute jobs to the sites supporting the enmr.eu VO. Most of the portals make use of the DIRAC4EGI machinery 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.

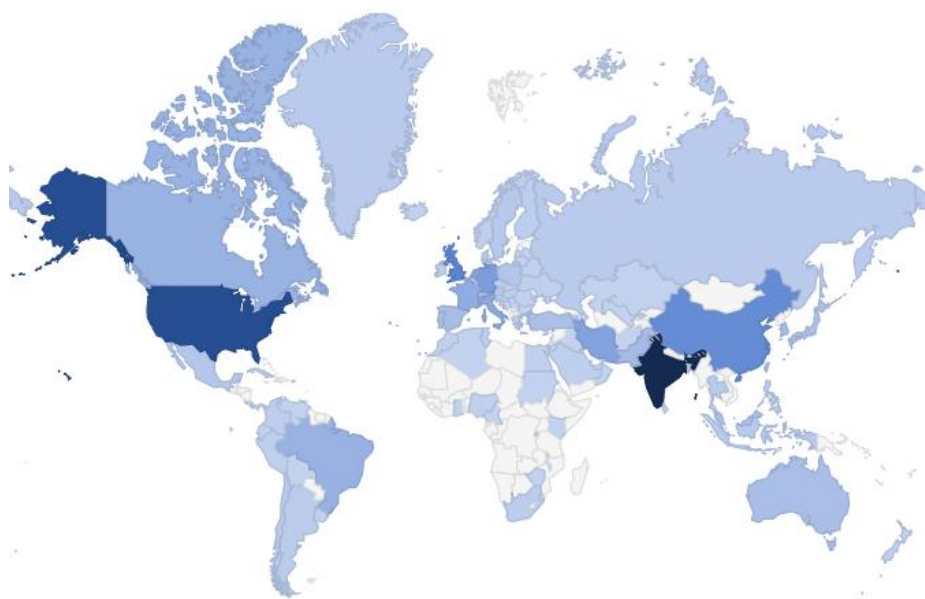
From the user experience point of view, the major part of the WeNMR portals integrate the EGI Check-in Single Sign On (SSO), but some of them provide a SSO through the West-Life SSO which connects to ARIA (the access management solution of the Structural Biology infrastructure INSTRUCT-ERIC). Please refer to *Deliverables 7.1 and 7.2* for more details about integration and technologies used.

### 7.1.2 Main users

Main users of the WeNMR are structural biologists of any degree of experience, with the aim of modelling and refining NMR structures, complexes of proteins and other biomolecules or fitting cryo-electron microscopy maps.

A user is a person making use of at least one of our thematic services. All portals except one (FANTEN) do require user registration. For FANTEN, users are identified by their IP address for collecting the various metrics.

The number of newly registered users per period has increased from 1,750 (starting of the project) to 2,273 in the first reporting period (M1-M8) and to 3,844 in the second reporting period (M9-M17) which represents an increase of 129.9% and 219.7% respectively. The number of countries reached by our thematic services has increased from 96 to 108 since the last reporting period (see world map below).



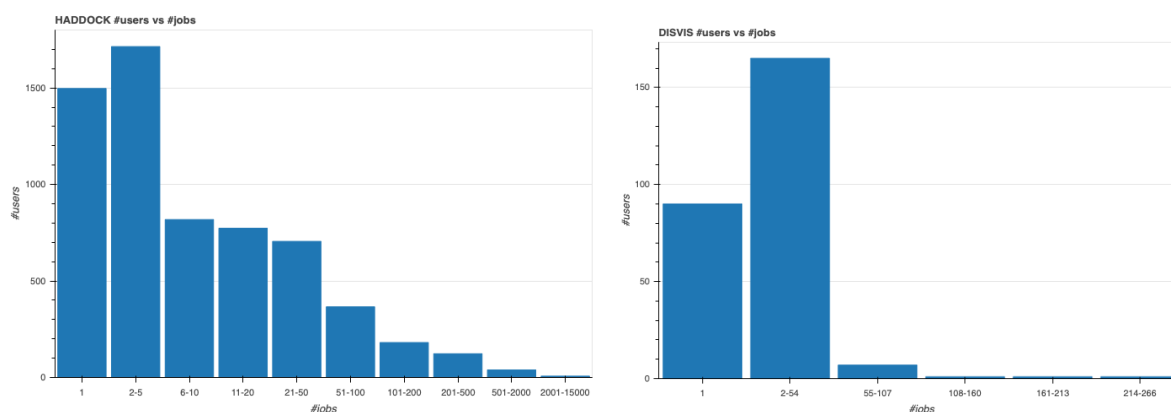
**Fig. 22 - World map of WeNMR thematic services users**

## 7.2 Services exploitation

### 7.2.1 Usage statistics

Number of user runs submitted to the thematic services portals are provided by the various portals keeping log files of submissions. For the second reporting period (month 9-17), 24,702 runs have been submitted by users (36,179 for first period - months 1-8), which represents an increase of 179% compared to the number of jobs reported as 3/4 of the 2017 submissions (13,800, baseline).

We have further analysed user submissions to check how many jobs are submitted per user in order to answer the question whether the services attract topical users or rather recurring ones. The results clearly point toward recurring usage of the WeNMR services. As an example, the distributions of number of job submissions per users are shown in the following plots for HADDOCK and DISVIS.



**Fig. 23 - Distributions of number of job submissions per user for the HADDOCK and DISVIS portals**

The number of grid/cloud job submissions for the second reporting period (months 9-17) is 10,343,738 (5,034,534 for first reporting period), which represents an increase of 153% compared to the baseline as 3/4 of the 2017 submissions (6,750,000). Note that not all portals are submitting to grid/cloud resources. Also, the number of jobs reported here consists of either direct submissions via UI or through the DIRAC4EGI service. The number of jobs for the enmr.eu VO in the EGI accounting portal is therefore be smaller since multiple submissions to DIRAC translate into one pilot job on the grid.

The number of grid/cloud job submissions represent a consumption in HS06 CPU Time/Wall Time hours of 20,684,686 for the grid and of 21,241 for the cloud. Those metrics were collected from the EGI accounting portal for the enmr.eu VO. The baseline is based on the 3/4 of the 2017 submissions (15,168,258). No jobs were submitted to the cloud in 2017.

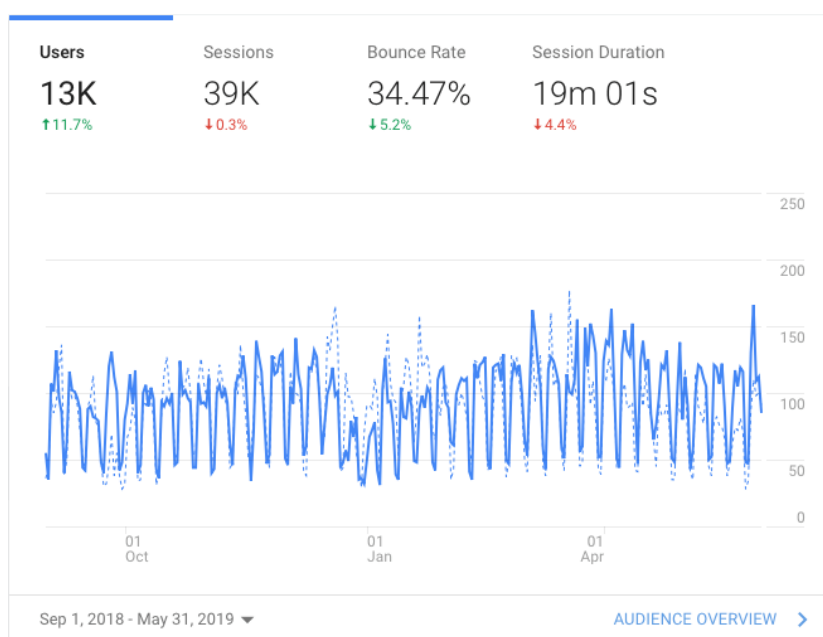
The number of citations of the services related publications is reported below:

- AMPS-NMR web portal publication: 6 citations since 2018 (<https://edu.nl/tgvjr>).
- FANTEN web portal publication: 12 citations since 2018 (<https://edu.nl/cqafu>).
- HADDOCK web portal two main publications:
  - 89 citations since 2018 (<https://edu.nl/vtruk>)
  - 133 citations since 2018 (<https://edu.nl/rk4cj>).
- DISVIS/POWERFIT web portals publication: 9 citations since 2018 (<https://edu.nl/yungk>).
- SpotON web portal: 15 citations since 2018 (<https://edu.nl/mw67y>).
- WeNMR J Grid Comp publications: 30 citations since 2018 (<https://edu.nl/w37yp>).

While most users are accessing the services directly, those are now also advertised in the EOSC-hub Marketplace. Statistics of access through the Marketplace have been collected by Google Analytics for views and orders (views / orders):

- AMPS-NMR: 97 / 5
- CS-ROSETTA3: 20 / 0
- DISVIS: 28 / 2
- FANTEN: 37 / 1
- HADDOCK: 40 / 6
- POWERFIT: 20 / 1
- SPOTON: 12 / 0

These numbers are still rather small compared to the direct access statistics of some portals over the same period (P2 - months 9-17). For example, the HADDOCK portal pages have >39K sessions over the same period.



**Fig. 24 - Google analytics access statistics of the HADDOCK portal over period 2 (months 10-18)**

### 7.2.2 Users' feedback

Users' feedback is collected as a continuous satisfaction feedback on a scale of 1 (worst) to 5 (best) on the result pages. An example of satisfaction feedback request can be found on the following example output page of the DISVIS service<sup>50</sup>. Online continuous stats for the Utrecht portals are available<sup>51</sup>.

Current user satisfaction is:

- AMPS-NMR: 4.2 (from 10 respondents)
- CS-ROSETTA3: 5.0 (from 4 respondents)
- DISVIS: 4.9 (from 15 respondents)
- HADDOCK: 4.9 (from 1310 respondents)
- HADDOCK2.4: 5.0 (from 23 respondents) - NEW!
- POWERFIT: 5.0 (from 2 respondents)
- SPOTON: 4.9 (from 27 respondents)

### 7.2.3 Identified gaps

EGI Check-in and EOSC MarketPlace integration was completed. The WeNMR thematic services were among the first ones to be fully integrated. The main gap at this time is a mechanism for users to simply upload their simulations results into EUDAT/Onedata repositories. Some portals do

<sup>50</sup> <http://milou.science.uu.nl/cgi/enmr/services/DISVIS/disvis/example>

<sup>51</sup> <https://milou.science.uu.nl/stats.php>

however support Onedata. This does however not seem a major issue from an end user perspective. Instead, based on user feedback we implemented some specific request, like the option to download all models within one archive in the new HADDOCK2.4 web portal.

#### **7.2.4 Future improvements**

WP7 funds for WeNMR Thematic Service have ended at M18, as planned in the initial DoW. We have migrated some of the Utrecht University contributed services to a new cluster to ensure operational and user growth in the coming years and are in the process of consolidating the infrastructure/deployment of the portals by systematic containerisation.

We will further continue our efforts in addressing specific user requests that would lead to an improvement of our services.

## 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. The remaining month between M12 and M15 have been used to complete the service developments and to integrate the services into the EOSC-hub marketplace.

### 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:

<b>Service/Tool name</b>	Geohazards Exploitation Platform (GEP)
<b>Service/Tool url</b>	<a href="https://geohazards-tep.eu/#!">https://geohazards-tep.eu/#!</a> <a href="https://geohazards-tep.eu/geobrowser/?id=eoschub-landslide-app">https://geohazards-tep.eu/geobrowser/?id=eoschub-landslide-app</a> <a href="https://geohazards-tep.eu/geobrowser/?id=eoschub-alpsmonitoring-app">https://geohazards-tep.eu/geobrowser/?id=eoschub-alpsmonitoring-app</a>
<b>Service/Tool information page</b>	<a href="https://geohazards-tep.eu/#!">https://geohazards-tep.eu/#!</a>
<b>Description</b>	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:

	<ul style="list-style-type: none"> <li>- Web portal</li> <li>- Data Agency including <ul style="list-style-type: none"> <li>- Catalogue service</li> <li>- Data Gateway</li> <li>- Production Center</li> <li>- Development environment</li> </ul> </li> </ul> <p>In the context of EOSC_Hub it exposes two main services:</p> <ul style="list-style-type: none"> <li>- A service providing an interferometric &amp; 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.</li> <li>- 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.</li> </ul>
<b>Value proposition</b>	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.
<b>Customer of the service/tool</b>	Research Communities
<b>User of the service/tool</b>	large research groups
<b>User Documentation</b>	<a href="https://terradue.github.io/doc-tep-geohazards/">https://terradue.github.io/doc-tep-geohazards/</a> <a href="https://geohazards-tep.eo.esa.int/#!blog">https://geohazards-tep.eo.esa.int/#!blog</a>
<b>Technical Documentation</b>	<a href="https://docs.terradue.com/t2-api/">https://docs.terradue.com/t2-api/</a>
<b>Product team</b>	Terradue Srl
<b>License</b>	-
<b>Source code</b>	<a href="https://github.com/geohazards-tep">https://github.com/geohazards-tep</a>
<b>Testing</b>	-

<b>Service/Tool name</b>	EODC Data Catalogue Service and EODC JupyterHub for global Copernicus data
--------------------------	---

<b>Service/Tool url</b>	<a href="https://csw.eodc.eu/">https://csw.eodc.eu/</a> <a href="https://jupyterhub.eodc.eu">https://jupyterhub.eodc.eu</a>
<b>Service/Tool information page</b>	<a href="https://marketplace.eosc-portal.eu/services/eodc-data-catalogue-service">https://marketplace.eosc-portal.eu/services/eodc-data-catalogue-service</a> <a href="https://marketplace.eosc-portal.eu/services/eodc-jupyterhub-for-global-copernicus-data">https://marketplace.eosc-portal.eu/services/eodc-jupyterhub-for-global-copernicus-data</a>
<b>Description</b>	<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>
<b>Value proposition</b>	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.
<b>Customer of the service/tool</b>	Research Communities
<b>User of the service/tool</b>	research groups
<b>User Documentation</b>	<a href="https://eomex.eodc.eu/manual">https://eomex.eodc.eu/manual</a>
<b>Technical Documentation</b>	<a href="https://eomex.eodc.eu/manual">https://eomex.eodc.eu/manual</a>
<b>Product team</b>	EODC
<b>License</b>	Different licenses for the individual components
<b>Source code</b>	<a href="http://git.eodc.eu">http://git.eodc.eu</a>
<b>Testing</b>	-

<b>Service/Tool name</b>	Sentinel Hub
<b>Service/Tool url</b>	<a href="https://services.sentinel-hub.com/">https://services.sentinel-hub.com/</a>

<b>Service/Tool information page</b>	<a href="https://sentinel-hub.com">https://sentinel-hub.com</a>
<b>Description</b>	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.
<b>Value proposition</b>	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.
<b>Customer of the service/tool</b>	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.
<b>User of the service/tool</b>	Application developer Data scientists Machine learning experts (EO)
<b>User Documentation</b>	<a href="https://docs.sentinel-hub.com">https://docs.sentinel-hub.com</a> <a href="https://www.sentinel-hub.com/develop/documentation/api/ogc_api">https://www.sentinel-hub.com/develop/documentation/api/ogc_api</a>
<b>Technical Documentation</b>	<a href="https://docs.sentinel-hub.com">https://docs.sentinel-hub.com</a> <a href="https://www.sentinel-hub.com/develop/documentation/api/ogc_api">https://www.sentinel-hub.com/develop/documentation/api/ogc_api</a>
<b>Product team</b>	Sinergise
<b>License</b>	Proprietary
<b>Source code</b>	Proprietary
<b>Testing</b>	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.

<b>Service/Tool name</b>	ADAM (former MEA Platform)
<b>Service/Tool url</b>	<a href="https://explorer.adamplatform.eu">https://explorer.adamplatform.eu</a>
<b>Service/Tool information page</b>	<a href="https://adamplatform.eu">https://adamplatform.eu</a>

<b>Description</b>	The Advanced geospatial Data Management platform (ADAM) is a tool to access a large variety and volume of global environmental data.
<b>Value proposition</b>	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.
<b>Customer of the service/tool</b>	E.g. RI; Resource Provider; Research Communities
<b>User of the service/tool</b>	Earth scientists; Professionals;
<b>User Documentation</b>	<a href="https://adamplatform.eu">https://adamplatform.eu</a> 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 <a href="#">ADAM's YouTube channel</a> .
<b>Technical Documentation</b>	<a href="https://adamplatform.eu">https://adamplatform.eu</a>
<b>Product team</b>	MEEO

<b>Service/Tool name</b>	rasdaman datacubes
<b>Service/Tool url</b>	<a href="http://eoschub.rasdaman.com/rasdaman/ows">http://eoschub.rasdaman.com/rasdaman/ows</a>
<b>Service/Tool information page</b>	<a href="http://eoschub.rasdaman.com">http://eoschub.rasdaman.com</a>
<b>Description</b>	The rasdaman datacube engine provides fast and flexible access, analytics, and fusion on massive spatio-temporal CreoDIAS data assets.
<b>Value proposition</b>	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.
<b>Customer of the service/tool</b>	Value-adding remote sensing businesses; researchers; authorities
<b>User of the service/tool</b>	Earth data analysts
<b>User Documentation</b>	<a href="http://eoschub.rasdaman.com/eosc/index.html?page=demonstrations">http://eoschub.rasdaman.com/eosc/index.html?page=demonstrations</a>
<b>Technical Documentation</b>	<a href="http://www.earthserver.eu/webinars">http://www.earthserver.eu/webinars</a>
<b>Product team</b>	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/EOSC/EO+Pillar+VA>

### **8.1.1 Architecture and integrated services**

The architecture of the individual services remained unchanged between M12 and M15 and is already explained within D7.2 (Section 8, T7.7 EO-Pillar).

### **8.1.2 Main users**

The target users are the scientific as well as public and private users acting in the field of earth observation and beyond.

## **8.2 Services exploitation**

### **8.2.1 Usage statistics**

CREODIAS EO Data is now 15050 TB (downloading about 40 TB per day). They have currently 3628 users of CREODIAS service (service started 2018-06).

Further usage statistics have been reported within WP13, see <https://wiki.eosc-hub.eu/display/EOSC/EO+Pillar+VA>.

### **8.2.2 Users' feedbacks**

Individual user feedback is collected per provider.

### **8.2.3 Identified gaps**

The planned services could be established and a few further services like the data catalogue services have been added additionally.

### **8.2.4 Future improvements**

The improvements have been completed till M15. Minor adaptations/improvements based on user feedback are performed in WP13 in order to better satisfy the user demands.

## 9 DARIAH

### 9.1 Service description

Service/Tool name	DARIAH Science Gateway
Service/Tool url	<a href="https://dariah-sg.irb.hr">https://dariah-sg.irb.hr</a>
Service/Tool information page	<a href="https://dariah-sg.irb.hr">https://dariah-sg.irb.hr</a>
Description	The <b>DARIAH Science Gateway</b> 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: <a href="https://dariah-sg.irb.hr/help2">https://dariah-sg.irb.hr/help2</a> Simple Cloud access service: <a href="https://ltos-gateway.lpds.szta.hu/web/wizard/help">https://ltos-gateway.lpds.szta.hu/web/wizard/help</a> Gutenberg use case for Simple Cloud access service: <a href="https://bitbucket.org/davordavidovic/textanalysis/src/master/">https://bitbucket.org/davordavidovic/textanalysis/src/master/</a>
Technical Documentation	-
Product team	RBI, SZTAI
License	-
Source code	gUSE Grid and Cloud Science Gateway: <a href="https://sourceforge.net/projects/guse/">https://sourceforge.net/projects/guse/</a>
Testing	-

Service/Tool name	Invenio-base repository
Service/Tool url	<a href="https://dariah-portal.cloud.ba.infn.it/">https://dariah-portal.cloud.ba.infn.it/</a>
Service/Tool information page	<a href="https://dariah-portal.cloud.ba.infn.it/">https://dariah-portal.cloud.ba.infn.it/</a>
Description	The <b>Invenio-based repository</b> 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, having
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): <a href="https://www.indigo-datacloud.eu/deploy-open-data-repository-cloud-using-marathon">https://www.indigo-datacloud.eu/deploy-open-data-repository-cloud-using-marathon</a>
Technical Documentation	-
Product team	RBI, Indigo-DataCloud
License	GNU General Public Licence 2
Source code	DockerHub container: <a href="https://hub.docker.com/r/indigodatacloudapps/dariah-repository">https://hub.docker.com/r/indigodatacloudapps/dariah-repository</a> Source code: <a href="https://github.com/indigo-dc/dariah-repository">https://github.com/indigo-dc/dariah-repository</a> Ansible roles: <a href="https://github.com/indigo-dc/ansible-role-dariahrepo">https://github.com/indigo-dc/ansible-role-dariahrepo</a>
Testing	-

Service/Tool name	DARIAH Repository
Service/Tool url	<a href="https://projects.gwdg.de/projects/dariah-de-repository">https://projects.gwdg.de/projects/dariah-de-repository</a>
Service/Tool information page	<a href="https://dariah-sg.irb.hr">https://dariah-sg.irb.hr</a>

<b>Description</b>	<b>DARIAH repository</b> 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 amount of diverse data, both structured and unstructured.
<b>Value proposition</b>	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.
<b>Customer of the service/tool</b>	DARIAH, Digital arts and humanities communities
<b>User of the service/tool</b>	Individual researchers, research groups, institutions, research projects
<b>User Documentation</b>	<a href="https://info.gwdg.de/docs/doku.php?id=en:services:storage_services:gwdg_cd_star:start">https://info.gwdg.de/docs/doku.php?id=en:services:storage_services:gwdg_cd_star:start</a>
<b>Technical Documentation</b>	Rest API: <a href="https://www.gwdg.de/documents/20182/31188/gwdg-bericht-78.pdf">https://www.gwdg.de/documents/20182/31188/gwdg-bericht-78.pdf</a>
<b>Product team</b>	GWDG
<b>License</b>	-
<b>Source code</b>	<a href="https://github.com/clld/pycdstar">https://github.com/clld/pycdstar</a>
<b>Testing</b>	-

The DARIAH Thematic Services is a service provides user-friendly solutions (services 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 web-based 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 deliverable D7.1 and D7.2 [R1, R2].

The DARIAH Thematic service provides three independent services:

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

### 9.1.1 Architecture and integrated services

The 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 integrates and provides 5 services: DBO@Cloud, Semantic Search Engine, Parallel Semantic Search Engine, Simple Cloud Access and workflow management. In the first project year the gateway was migrated from the local site in MZA SZTAKI to INFN-Bari Cloud site, which is a part of EGI FedCloud infrastructure. The DARIAH SG is published in the EOSC-hub Service Catalogue as well as in EOSC Marketplace.

More detailed information on the DARIAH Science Gateway, its components, architecture and integration activities can be found in D7.2 [R2].

The Invenio-based repository service is a set of several integrated and connected components (for more information see Figure 10 in Deliverable D7.1 [R1]). The service consists of the following components:

- Indigo IAM identity and access management,
- FutureGateway as a web-based user-interface,
- Prepared data repository images from the DockerHub,
- Pre-defined TOSCA template (filled via FutureGateway),
- Orchestrator,
- Mesos cluster,
- Marathon framework for holding and managing a long running jobs,
- Invenio-based repository composed of several sub-components packed into 5 docker images.

Detailed information on the architecture of the Invenio-based repository in the Cloud can be found in Deliverable D7.2 [R2]. At the time of that document writing, the integration process was still undergoing with only Indigo-IAM service ready and a clear integration plan provided. In the meantime, the integration of the Invenio-based repository service with the EOSC-hub services is completed. The FutureGateway is used as a web-frontend, where the users configure their new repository by providing the basic options, such as hardware requirements. Then, the service automatically allocates the required resources in the EGI FedCloud infrastructure and starts a new instance of a repository. The overall system resource requirements, background Invenio-based repository services requirements and basic configuration of the repository is provided via a TOSCA template. On the allocated resources an instance of Mesos cluster is started in which Marathon container orchestrator is started. The Marathon is used to deploy, manage and operate the docker containers, of which the Invenio-based digital repository is composed. The prepared docker images of the Invenio-based repository sub-components are automatically retrieved from the Indigo DockerHub and started inside Marathon container framework. The authentication to the service is done via the Indigo-IAM service. A new instance of Indigo-IAM service is launched for the DARIAH TS as an IAM-as-a-Service and is hosted at the INFN-BARI site.

The DARIAH Repository service integrated a number of services that are demanded by the arts and Humanities Communities with respect to the publication and long-term storage of research data. The DARIAH Repository consists of a number of sub-services:

- CRUD Service - a core storage service for the DARIAH Repository,
- Publish Service - a backend of the so-called Publikator for preparing and managing data objects for import,
- PID Service - assigning persistent identifiers to the imported objects: a DataCite DOI for citation, and an EPIC Handle PID for administrative use,
- OAI-PMH Service - for harvesting all metadata from the collections stored in the DARIAH-DE Repository.

A detailed description of the DARIAH Repository architecture can be found in deliverable D7.2. The service can be accessed through the DARIAH AAI, which integrates EOSC identities and is thus integrated into the EOSC infrastructure.

### 9.1.2 Main users

The target users of the *DARIAH Science Gateway* are the scholars and researchers coming from digital arts and humanities domain, who require, in their work, to browse, search and download records from digital collections and open repositories as well as to perform simple data analysis.

The *Invenio-based repository* focuses on meeting the needs of individual researchers, small research groups and (research) projects that do not have adequate expertise nor in-depth know-how and experience in deploying the digital repositories or does not have financial support to acquire and maintain the resources required to host a repository. The main users of the service are the members and small research groups involved in the DARIAH community. The service is targeting two different types of users:

- *repository manager* – a user that, via the FutureGateway, configures, initialize and deploys a new repository instance in the cloud, this is an administrator of the repository instance who has an access to Marathon dashboard from where he/she can monitor the resource allocation, scale the resources and overview the separate Invenio components,
- *end-user* – a user that, once a new repository instance is deployed, has an access to the repository. Based on the access rules (defined by the repository manager) the end-user can browse, search, upload and download the records from the repository.

The *DARIAH Repository* is a service for the Arts and Humanities to store research data sustainably and securely, to add metadata to it and to publish it. Furthermore, collections and data objects receive unique and permanently valid persistent identifiers (DOI and EPIC PIDs) for long term identification and citation. The DARIAH Repository service is targeting individual researchers, scholars and institutions coming from the DARIAH community that have large digital research data collections that have to be published, preserved for long term and uniquely identified (e.g. DOI).

## 9.2 Services exploitation

### 9.2.1 Usage statistics

The usage statistics are collected via the Google Analytics service and the internal/local accounting systems. The aggregated statistics are reported for M1-M17 in the WP13 Access provisioning work packed and the usage reports are presented in the following table.

Metric name	Baseline	How measure is done	M1-M17
Number of registered users	36	Number of registered users of the sub-services (users with valid DARIAH IdP account)	76
Number of visits	249	Number of visits of the sub-services collected via Google Analytics	250
Number of supported services/applications on the gateway	5	Number of end-user applications and use-cases based on the DARIAH TS services. Number of new applications integrated into DARIAH SG (number of new portlets). Number of external services accessing DARIAH TS services via APIs.	10
Usage DARIAH Repository: number of objects/DOIs		Number of published data sets/DOIs f are collected via the service	574
Usage DARIAH		Number of collections available via	70

Repository: number of collections		the data repository	
Number and names of the countries reached	10	Based on the IPs, measured using Google Analytics	10 (DE, HU, CRO, USA, IT, China, FR, NL, CAD, AT)
Satisfaction	Not applicable	from WP4	Data not reported

The DARIAH Science Gateway service is a long running service hosted at the RECAS BARI Cloud site in virtual environment occupying 1 CPU, 2 GB main memory and 200 GB disk space. The Gateway is a web-oriented service which requires only a small portion of resources required for hosting the gUSE/WS-Pgrade portal, while the sub-services provided via the Gateway utilize their local resources (e.g. PSSE, DBO@Cloud).

The Invenio-based repository service requires one virtual server for hosting the FutureGateway portal which provide web-based user interface enabling setting-up and configuring a new instance of Invenio-based repository on the cloud infrastructure. Currently, the RECAS-BARI Cloud site provides 40 CPUs, 50 GB memory and 1 TB storage space for hosting the virtual instances of Invenio-based repositories configured and initiated via the Services front-end, i.e. FutureGateway portal. The DARIAH-IAM services is also hosted on one virtual server at the RECAS-BARI Cloud site.

### 9.2.2 Users' feedbacks

The users' feedbacks for DARIAH Science Gateway and Invenio-based repository services has to been collected so far. We plan to collect the feedbacks and users' satisfaction via the online questionnaires that will be launched in the last year of the project. The design, collecting and analysis of the user feedback is a part of the work to be done in the Work Package WP11 and will be collocated with the training events and demonstrations that are also planned with the WP11 activities.

### 9.2.3 Identified gaps

No gaps have been identified for the DARIAH Science Gateway, Invenio-based repository in the cloud and DARIAH Repository services compared to the initial plan (M7.1).

### 9.2.4 Future improvements

No further improvements of the DARIAH Science Gateway and Invenio-based repository services are foreseen by the end of the project since the DARIAH Thematic Service (Task T7.8) has officially ended by June 2019 (as was initially planned in the project work plan). However, DARIAH TS is one of the Thematic Services that has been identified as a primary target for the integration of EOSC-hub and OpenAIRE Advance services. In that regard, the integration of the DARIAH Repository with

OpenAire Advance has been further planned by the DARIAH TS. A concise integration plan has been developed that include:

- Specification and implementation of integration workflows,
- Implementation of metadata transformation processes,
- Extension of the OAI-PMH Service to cover all metadata,
- Testing and documentation.

The implementation of the plan has started in the second project year and is expected to be finalized by the end of the project.

## 10 LifeWatch

### 10.1 Service description

<b>Service/Tool name</b>	Plant Classification
<b>Service/Tool url</b>	<a href="http://deep.ifca.es/plants/">http://deep.ifca.es/plants/</a>
<b>Service/Tool information page</b>	<a href="https://github.com/IgnacioHeredia/plant_classification">https://github.com/IgnacioHeredia/plant_classification</a>
<b>Description</b>	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.
<b>Value proposition</b>	This tool can definitely open the field to active contributions of non-expert users including citizen scientists.
<b>Customer of the service/tool</b>	<a href="http://deep.ifca.es/static/images/logoIFCA.png">http://deep.ifca.es/static/images/logoIFCA.png</a>
<b>User of the service/tool</b>	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
<b>User Documentation</b>	<a href="https://github.com/IgnacioHeredia/plant_classification">https://github.com/IgnacioHeredia/plant_classification</a>
<b>Technical Documentation</b>	<a href="https://github.com/IgnacioHeredia/plant_classification">https://github.com/IgnacioHeredia/plant_classification</a>
<b>Product team</b>	LifeWatch, IFCA-CSIC
<b>License</b>	-
<b>Source code</b>	<a href="https://github.com/IgnacioHeredia/plant_classification">https://github.com/IgnacioHeredia/plant_classification</a>
<b>Testing</b>	-

<b>Service/Tool name</b>	Glacier Lagoons of Sierra Nevada
<b>Service/Tool url</b>	<a href="https://lagunasdesierranevada.es/">https://lagunasdesierranevada.es/</a>
<b>Service/Tool information page</b>	<a href="https://lagunasdesierranevada.es/">https://lagunasdesierranevada.es/</a>
<b>Description</b>	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.
<b>Value proposition</b>	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
<b>Customer of the service/tool</b>	<a href="http://secretariageneral.ugr.es/pages/ivc/descarga/_img/vertical/ugrmarca01color_2/">http://secretariageneral.ugr.es/pages/ivc/descarga/_img/vertical/ugrmarca01color_2/</a>
<b>User of the service/tool</b>	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.
<b>User Documentation</b>	-
<b>Technical Documentation</b>	-
<b>Product team</b>	LifeWatch, University of Granada
<b>License</b>	<a href="https://lagunasdesierranevada.es/condiciones-de-uso/">https://lagunasdesierranevada.es/condiciones-de-uso/</a>
<b>Source code</b>	-
<b>Testing</b>	-

<b>Service/Tool name</b>	GBIF Spain Spatial Portal
<b>Service/Tool url</b>	<a href="https://espacial.gbif.es/">https://espacial.gbif.es/</a>
<b>Service/Tool information page</b>	<a href="https://espacial.gbif.es/">https://espacial.gbif.es/</a>
<b>Description</b>	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.

<b>Value proposition</b>	Map service, to visualise and analyse relationships between species, location and environment.
<b>Customer of the service/tool</b>	<a href="https://www.gbif.es/wp-content/uploads/2017/05/gbif-logo.svg">https://www.gbif.es/wp-content/uploads/2017/05/gbif-logo.svg</a>
<b>User of the service/tool</b>	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.
<b>User Documentation</b>	<a href="https://www.gbif.es/en/portal-nacional-de-datos/">https://www.gbif.es/en/portal-nacional-de-datos/</a>
<b>Technical Documentation</b>	<a href="https://www.gbif.es/en/portal-nacional-de-datos/">https://www.gbif.es/en/portal-nacional-de-datos/</a>
<b>Product team</b>	GBIF Spain, LifeWatch
<b>License</b>	<a href="https://www.gbif.es/en/politica-de-privacidad/">https://www.gbif.es/en/politica-de-privacidad/</a>
<b>Source code</b>	-
<b>Testing</b>	-

<b>Service/Tool name</b>	GBIF Spain Species
<b>Service/Tool url</b>	<a href="https://especies.gbif.es">https://especies.gbif.es</a>
<b>Service/Tool information page</b>	<a href="https://especies.gbif.es">https://especies.gbif.es</a>
<b>Description</b>	GBIF.es data access under biogeographic context provides access with advanced facets to GBIF biodiversity data under a biogeographic context. Aggregates data on species description, common names, taxonomy, image gallery, sequence data and bibliography.
<b>Value proposition</b>	Aggregates data on species description, common names, taxonomy, image gallery, sequence data and bibliography.
<b>Customer of the service/tool</b>	<a href="https://www.gbif.es/wp-content/uploads/2017/05/gbif-logo.svg">https://www.gbif.es/wp-content/uploads/2017/05/gbif-logo.svg</a>
<b>User of the service/tool</b>	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 only focused on research on Spanish species.
<b>User Documentation</b>	<a href="https://www.gbif.es/en/portal-nacional-de-datos/">https://www.gbif.es/en/portal-nacional-de-datos/</a>
<b>Technical Documentation</b>	<a href="https://www.gbif.es/en/portal-nacional-de-datos/">https://www.gbif.es/en/portal-nacional-de-datos/</a>
<b>Product team</b>	GBIF Spain, LifeWatch
<b>License</b>	<a href="https://www.gbif.es/en/politica-de-privacidad/">https://www.gbif.es/en/politica-de-privacidad/</a>
<b>Source code</b>	-

Testing	-
---------	---

Service/Tool name	GBIF Spain Species List
Service/Tool url	<a href="https://listas.gbif.es/public/speciesLists">https://listas.gbif.es/public/speciesLists</a>
Service/Tool information page	<a href="https://listas.gbif.es/public/speciesLists">https://listas.gbif.es/public/speciesLists</a>
Description	GBIF.es data access under biogeographic context provides access with advanced facets to GBIF biodiversity data under a biogeographic context. This service permits to upload your own list of species to use in the GBIF.ES portal or use a list uploaded by other GBIF.ES users or data providers.
Value proposition	This service permits to upload your own list of species to use in the GBIF.ES portal or use a list uploaded by other GBIF.ES users or data providers.
Customer of the service/tool	<a href="https://www.gbif.es/wp-content/uploads/2017/05/gbif-logo.svg">https://www.gbif.es/wp-content/uploads/2017/05/gbif-logo.svg</a>
User of the service/tool	Research use. This tool allows users to upload lists of species and work with that list within the Atlas. It is also possible to access lists made by others.
User Documentation	<a href="https://www.gbif.es/en/portal-nacional-de-datos/">https://www.gbif.es/en/portal-nacional-de-datos/</a>
Technical Documentation	<a href="https://www.gbif.es/en/portal-nacional-de-datos/">https://www.gbif.es/en/portal-nacional-de-datos/</a>
Product team	GBIF Spain, LifeWatch
License	<a href="https://www.gbif.es/en/politica-de-privacidad/">https://www.gbif.es/en/politica-de-privacidad/</a>
Source code	-
Testing	-

<b>Service/Tool name</b>	GBIF Spain Collections
<b>Service/Tool url</b>	<a href="https://colecciones.gbif.es">https://colecciones.gbif.es</a>
<b>Service/Tool information page</b>	<a href="https://colecciones.gbif.es">https://colecciones.gbif.es</a>
<b>Description</b>	GBIF.es data access under biogeographic context provides access with advanced facets to GBIF biodiversity data under a biogeographic context. Institutions, collections and datasets list, description and linked to occurrences.
<b>Value proposition</b>	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
<b>Customer of the service/tool</b>	<a href="https://www.gbif.es/wp-content/uploads/2017/05/gbif-logo.svg">https://www.gbif.es/wp-content/uploads/2017/05/gbif-logo.svg</a>
<b>User of the service/tool</b>	The access to these collections is open, this service is aimed both for people who want to learn and for research. Serves as a record.
<b>User Documentation</b>	<a href="https://www.gbif.es/en/portal-nacional-de-datos/">https://www.gbif.es/en/portal-nacional-de-datos/</a>
<b>Technical Documentation</b>	<a href="https://www.gbif.es/en/portal-nacional-de-datos/">https://www.gbif.es/en/portal-nacional-de-datos/</a>
<b>Product team</b>	GBIF Spain, LifeWatch
<b>License</b>	<a href="https://www.gbif.es/en/politica-de-privacidad/">https://www.gbif.es/en/politica-de-privacidad/</a>
<b>Source code</b>	-
<b>Testing</b>	-

<b>Service/Tool name</b>	GBIF Spain Regions
<b>Service/Tool url</b>	<a href="https://regiones.gbif.es">https://regiones.gbif.es</a>
<b>Service/Tool information page</b>	<a href="https://regiones.gbif.es">https://regiones.gbif.es</a>
<b>Description</b>	GBIF.es data access under biogeographic context provides access with advanced facets to GBIF biodiversity data under a biogeographic context. Map tool to browse states, territories, local government areas and biogeographic regions.
<b>Value proposition</b>	Map tool to browse states, territories, local government areas and biogeographic regions.
<b>Customer of the service/tool</b>	<a href="https://www.gbif.es/wp-content/uploads/2017/05/gbif-logo.svg">https://www.gbif.es/wp-content/uploads/2017/05/gbif-logo.svg</a>
<b>User of the service/tool</b>	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.
<b>User Documentation</b>	<a href="https://www.gbif.es/en/portal-nacional-de-datos/">https://www.gbif.es/en/portal-nacional-de-datos/</a>
<b>Technical Documentation</b>	<a href="https://www.gbif.es/en/portal-nacional-de-datos/">https://www.gbif.es/en/portal-nacional-de-datos/</a>
<b>Product team</b>	GBIF Spain, LifeWatch
<b>License</b>	<a href="https://www.gbif.es/en/politica-de-privacidad/">https://www.gbif.es/en/politica-de-privacidad/</a>
<b>Source code</b>	-
<b>Testing</b>	-

<b>Service/Tool name</b>	GBIF Spain Occurrences
<b>Service/Tool url</b>	<a href="https://registros.gbif.es">https://registros.gbif.es</a>
<b>Service/Tool information page</b>	<a href="https://registros.gbif.es">https://registros.gbif.es</a>
<b>Description</b>	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.
<b>Value proposition</b>	List of occurrence data, georeferenced records and data set search statistics. Allows access to the information of each record and the quality tests performed
<b>Customer of the service/tool</b>	<a href="https://www.gbif.es/wp-content/uploads/2017/05/gbif-logo.svg">https://www.gbif.es/wp-content/uploads/2017/05/gbif-logo.svg</a>

<b>User of the service/tool</b>	Research use. Access to occurrence biodiversity data published by Spanish providers
<b>User Documentation</b>	<a href="https://www.gbif.es/en/portal-nacional-de-datos/">https://www.gbif.es/en/portal-nacional-de-datos/</a>
<b>Technical Documentation</b>	<a href="https://www.gbif.es/en/portal-nacional-de-datos/">https://www.gbif.es/en/portal-nacional-de-datos/</a>
<b>Product team</b>	GBIF Spain, LifeWatch
<b>License</b>	<a href="https://www.gbif.es/en/politica-de-privacidad/">https://www.gbif.es/en/politica-de-privacidad/</a>
<b>Source code</b>	-
<b>Testing</b>	-

<b>Service/Tool name</b>	GBIF Spain Images
<b>Service/Tool url</b>	<a href="https://imagenes.gbif.es">https://imagenes.gbif.es</a>
<b>Service/Tool information page</b>	<a href="https://imagenes.gbif.es">https://imagenes.gbif.es</a>
<b>Description</b>	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
<b>Value proposition</b>	Image storage and web services application. Images can be measured, calibrated, zoomed or downloaded
<b>Customer of the service/tool</b>	<a href="https://www.gbif.es/wp-content/uploads/2017/05/gbif-logo.svg">https://www.gbif.es/wp-content/uploads/2017/05/gbif-logo.svg</a>
<b>User of the service/tool</b>	Research and academic use. Visualize images found in the data portal for a taxon.
<b>User Documentation</b>	<a href="https://www.gbif.es/en/portal-nacional-de-datos/">https://www.gbif.es/en/portal-nacional-de-datos/</a>
<b>Technical Documentation</b>	<a href="https://www.gbif.es/en/portal-nacional-de-datos/">https://www.gbif.es/en/portal-nacional-de-datos/</a>
<b>Product team</b>	GBIF Spain, LifeWatch
<b>License</b>	<a href="https://www.gbif.es/en/politica-de-privacidad/">https://www.gbif.es/en/politica-de-privacidad/</a>
<b>Source code</b>	-
<b>Testing</b>	-

<b>Service/Tool name</b>	GBIF Spain eLearning
<b>Service/Tool url</b>	<a href="https://elearning.gbif.es">https://elearning.gbif.es</a>
<b>Service/Tool information page</b>	<a href="https://elearning.gbif.es">https://elearning.gbif.es</a>
<b>Description</b>	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
<b>Value proposition</b>	The GBIF.ES e-learning platform offers online workshops on data quality, data publication, data use, etc. open to whole GBIF community
<b>Customer of the service/tool</b>	<a href="https://www.gbif.es/wp-content/uploads/2017/05/gbif-logo.svg">https://www.gbif.es/wp-content/uploads/2017/05/gbif-logo.svg</a>
<b>User of the service/tool</b>	Platform offers online workshops open to whole GBIF community.
<b>User Documentation</b>	<a href="https://www.gbif.es/en/portal-nacional-de-datos/">https://www.gbif.es/en/portal-nacional-de-datos/</a>
<b>Technical Documentation</b>	<a href="https://www.gbif.es/en/portal-nacional-de-datos/">https://www.gbif.es/en/portal-nacional-de-datos/</a>
<b>Product team</b>	GBIF Spain, LifeWatch
<b>License</b>	<a href="https://www.gbif.es/en/politica-de-privacidad/">https://www.gbif.es/en/politica-de-privacidad/</a>
<b>Source code</b>	-
<b>Testing</b>	-

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, it is the University of Sevilla (USE) 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: Institute of Physics of Cantabria – Spanish National Research Council (IFCA-CSIC), University of Granada (UGR) and Global Biodiversity Information Facility Spain (GBIF.ES). 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.

Ultimately, and as part of our commitments in the WP7, the following services will be 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 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. 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 data access under biogeographic context provides access with advanced facets to GBIF biodiversity data under a biogeographic context. There are 9 different services:

- 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
- SpeciesList: This service permits to upload your own list of species to use in the GBIF.ES portal or use a list uploaded by other GBIF.ES users or data providers.

- **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.
- **Logger:** Updated information related to data downloaded from collections available through <https://datos.gbif.es>
- **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**

It has been also developed by IFCA; 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 then provides different tools to manage and work with the products.

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, 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

#### **10.1.1 Architecture and integrated services**

With the exception of Remote Monitoring and Smart Sensing, all the thematic services intended to be integrated are available on the website of their corresponding providers. The Plant Classification is the only already available service on the Marketplace. However, the version on the platform is deprecated, much of the information is missing and has a low integration level (it only has a link that redirects to the LifeWatch home page). Thus, this version is going to be replaced.

Regarding the integration with EOSC common services, there is only one service that is going to be integrated using them: the Remote Sensing service. It uses 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, none requirement has been placed.

### 10.1.2 Main users

**Glacier Lagoons of Sierra:** 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. The participants provide experimental data and research facilities. As far as their scientific use is concerned, they will allow researchers to know how the Lagoons of Sierra Nevada and their surroundings evolve over time. It also helps in decision making.

**Plant Classification:** 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. The citizen public are nature lovers. As the API is open, its use is also aimed at developers and researchers.

**Remote Monitoring and Smart Sensing:** This service is oriented to the use of satellite imagery in research and management. It is mainly aimed at researchers in the field of ecology, computer vision, geography and any study related to remote monitoring of natural environment parameters. As this tool can compare and predict situations, it also serves as a decision-making tool. This service is oriented to the use in research. It is mainly aimed at researchers in the field of ecology, computer vision, geography and any study related to remote monitoring of natural environment parameters. As this tool can compare and predict situations, it can also serve as a decision-making tool.

**GBIF.ES:** It integrates many services, each of which has its own audience. All are easy to access and sign in. Most of these services are intended for being used by the research community and citizens. The services are connected between them and some even allow importing lists or records from other services and download contents. The eLearning platform offers online training for the other services.

## 10.2 Services exploitation

### 10.2.1 Usage statistics

Although there is only one service in EOSC-hub marketplace, there is usage information in this period of 3 of them in their native websites:

<b>Metric</b>	<b>Plant Classification*</b>	<b>Glacier Lagoons*</b>	<b>GBIF.ES*</b>
# of users: Number of user's accounts	-	51	157
# of users: number of unique users (without accounts)	428 352 532 818	1808	137
Usage: Amount of data stored	-	42Gb	4.8 Tb
Number and names of the countries reached	-	37**	-

\*The data usage recollected has a different period depending on the service.

- The Plant Classification data is from August, September to November 2019 (metrics from EOSC-hub channels and provider website).
- Glacier Lagoons data is from July to October 2019 (metrics from provider website).
- GBIF.ES data is from May to October 2019 (metrics from provider website).

\*\*Spain (1557), United States (75), France (37), China (11), Peru (11), Ecuador (10), Colombia (9), Germany (9), Mexico (9), United Kingdom (8), Italy (8), Belgium (6), Japan (6), Argentina (5), Israel (5), Portugal (5), Brazil (4), Chile (4), Poland (4), Bolivia (3), India (3), Netherlands (3), Czechia (2), Malta (2), Venezuela (2), Andorra (1), Austria (1), Benin (1), Canada (1), Denmark (1), Dominican Republic (1), Finland (1), Iceland (1), Nicaragua (1), Norway (1), Puerto Rico (1), Sweden (1)

### 10.2.2 Users' feedbacks

At the moment, there are not any means by which the users are allowed to give feedback on their service usage experience. We are open to recommendations for measuring feedback in future integrated services.

### 10.2.3 Identified gaps

During this period, and as abovementioned, the suitability of the initially considered thematic services has been needed to be evaluated. After the analysis, some changes have been established regarding the initial proposal. Some of the thematic services have been discarded (PAIRQRS and CINDA) whereas new ones have been added (Glacier Lagoons of Sierra Nevada, additional services from GBIF.ES).

### 10.2.4 Future improvements

By December 2019, all thematic services are being integrated into the Marketplace. The templates are uploaded, and some bugs are now being fixed. Monthly metrics will continue to be collected.

As future improvements, we are trying to:

- Increase the level of integration; some services will, as far as possible, be integrated with common services such as EUDAT or INDIGO.
- Add use cases and examples to increase and facilitate the use of thematic services.

- Develop a feedback method for thematic services.

## 11 Conclusions

All thematic services part of WP7 have been actively integrating with both services available into the EOSC-hub marketplace and to the marketplace itself by making their services part of it.

Besides the different advancements in the integration and in the services release, all services are already showing a good uptake and good feedbacks from users.

This report sets the baseline for the exploitation and thematic services usage which will be updated in future deliverables.

## 12References

<b>No</b>	<b>Description/Link</b>
<b>R1</b>	D7.1 “First Thematic Service software release”, <a href="https://documents.egi.eu/document/3411">https://documents.egi.eu/document/3411</a>
<b>R2</b>	D7.2 “First report on Thematic Service architecture and software integration”, <a href="https://documents.egi.eu/document/3412">https://documents.egi.eu/document/3412</a>
<b>R3</b>	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: <a href="https://doi.org/10.1016/j.envsoft.2019.104585">10.1016/j.envsoft.2019.104585</a>
<b>R4</b>	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
<b>R5</b>	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.