D7.4 - Transnational usage of EOSC-Pillar services

Lead Partner:	CNRS
Authors	Pablo de Andres (Fraunhofer IWM), Lisana Berberi (KIT), Vincent Breton (CNRS), Leonardo Candela (CNR-ISTI), Alain Franc (Inrae), Jean-Marc Frigerio (Inrae), Luciano Gaido (INFN), Pietro Mandreoli (CNR), Dominique Lechaudel (CNRS), Marco Lorini (GARR), Matteo Di Fazio (GARR), Alessandro Orro (CNR), Jérôme Pansanel (CNRS), Claudio Pisa (GARR), Sorina Pop (CNRS), Thomas Porquet (Couperin), Geneviève Romier (CNRS), Marco Antonio Tangaro (CNR), Enrico Vianello (INFN), Jos van Wezel (KIT).
Version:	V3
Status:	Submitted
Dissemination Level:	PU
Document Link:	https://repository.eosc-pillar.eu/index.php/s/6k2WpnTWcJCAYsZ

Deliverable Abstract

This document presents the transnational usage of the “Ready to use services”. It reports on the technical and organisational, either qualitative or quantitative developments of the different services made in the framework of the project in order to adapt the services to the user needs. It also presents the impact of the EOSC-Pillar project for the ready to use services: the evolutions and success stories indirectly related to the project and that would probably not occur or in a longer time without EOSC-Pillar.

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No. 857650.

COPYRIGHT NOTICE

DELIVERY SLIP

Date	*Name*	*Partner/Activity*	*Date*
From:	Geneviève Romier	CNRS	
Moderated by:			
Reviewed by:			
Approved by:			

DOCUMENT LOG

Issue	*Date*	*Comment*	*Author*
v.1	16/03/2022	Table of content	Geneviève Romier (CNRS)
		First draft	Geneviève Romier (CNRS)
		T7.4.3 & T7.4.4	Leonardo Candela (CNR-ISTI)
		T7.4.1 & FG-iRODS France Grilles	Jérôme Pansanel (CNRS)
		Added services descriptions and usage for B2SAFE(CINECA, CINES, KIT) and VM(KIT)	Jos van Wezel (KIT)
		T7.4.2	Pietro Mandreoli Marco Antonio Tangaro
		Added service description and usage for GARR	Claudio Pisa (GARR)
		Added T7.4.6	Pablo de Andres (Fraunhofer IWM)
		T7.4.5	Jean-Marc Frigerio & Alain Franc(Inrae)
		added VIP section (3.3.3)	Sorina Pop (CNRS)
		inkind services GARR	Marco Lorini
		added Readmetrics section	Dominique Lechaudel & Thomas Porquet (Couperin)
		Covid-19 use case description in Appendix 1	Alessandro Orro (CNR)
		Executive summary, introduction, conclusion	Geneviève Romier (CNRS)
v.3		comments and review along the deliverable writing process	all

TERMINOLOGY

https://eosc-portal.eu/glossary

Terminology/Acronym	*Definition*
iRODS	Integrated Rule-Oriented Data System
TRL	Technology readiness level
Virtual Research Environment	A web-based working environment providing its designated community with the services and data needed to accomplish certain tasks.
VRE	Virtual Research Environment

Contents

1 Introduction 6
2 Ready to use services 7
2.1 Services described in the proposal 7
2.2 Service coming from T6.10 call 9
2.3 In-kind services 10
3 Transnational usage, adoption by EOSC-Pillar use cases, success stories and issues encountered 13
3.1 Services described in the proposal 13
3.2 Service coming from T6.10 call (since end of 2021) 28
3.3 In-kind services 30
4 Conclusion 34
5 References 37
Appendix I. Covid-19 Use Case (CNR) 39
Objectives 39
Implementation 39

Executive summary

This document presents the activities performed in WP7.4: "Ready to use services: deploying services already available for several scientific communities & support their integration in the EOSC framework". It reports the initial status and the evolution of each service in the different categories:

- the services that are part of the proposal in WP7.4 (The use of acceleration resources, Galaxy as a service platform: Laniakea@Recas service and contribution to Galaxy-E, Virtual Laboratories/Virtual environments, Virtual Research Environment-based service for research data publishing, Converging to common deployment tools to facilitate code circulation and mobility between different infrastructures (HPC, HTC) - pico2 service, Materials modelling MarketPlace and Open Innovation Environment;
- one of the services added during the project in the framework of the WP6.10 call: (Readmetrics);
- and the services proposed as in kind services by several EOSC-Pillar partners: Indigo-IAM by INFN-CNAF, FG-iRODS and VIP (Virtual Imaging Platform) by France Grilles, HPC resources by CINECA, Cloud computing and resources by GARR, B2SAFE storage services based on iRODS by CINES, KIT and CINECA and VM service by KIT.

The report provides first a short presentation of what the "ready to use services" are. The description is intended to be short to avoid redundancy with other WP7 deliverables and because this is not the objective of the present deliverable.

This document presents secondly the usage of the services at transnational level, the technical and organisational, either qualitative or quantitative developments of the different services made in the framework of the project in order to adapt the services to the user needs and to the EOSC integration and also the dissemination activities conducted by the services' teams in order to report on their evolutions. The Covid-19 use case added to the "ready to use services" during the lifetime of EOSC-Pillar is described in the Appendix 1 while its use of services is included in the services supporting it. This document also presents the impact of the project on the ready to use services: the evolutions and success stories indirectly related to the project and that would probably have occurred over a longer timespan without EOSC-Pillar.

The conclusion reports the global view of the results and benefits from EOSC-Pillar activities, the remaining issues and the recommendations to services providers for their future integration and evolution in EOSC: the services' providers are encouraged to study and adopt the methods and tools, to follow the guidelines, the recommendations and the documentation proposed by EOSC-Pillar and to stay tuned with EOSC-Future and Horizon Europe program to continue to receive support.

1 Introduction

This document reports on the transnational usage of the ready to use services. These services are the services that are part of the proposal in WP7.4 (subtasks T7.4.1 to subtask T7.4.6), one of the services added during the project in the framework of the WP6.10 call and the services proposed as in kind services by several partners. All those services are not at the same level of requirements in the EOSC-Pillar project: the work to be done on the WP7.4 subtasks is defined in the proposal and the teams in charge of them are members of a project partner while the objective of the service added from the T6.10 call is to join EOSC and those provided as in kind by the partners do not have the same level of required involvement. For this reason we differentiate them clearly in this report.

The document is organised in two main sections. The first one presents briefly the services as initially described and gives their technology readiness level evolution during the project lifetime. The second section presents the usage of the services at transnational level, the technical and organisational, either qualitative or quantitative developments of the different services made in the framework of the project in order to adapt the services to the user needs. If there were issues that came up during the process, solutions to solve them are explained. Publications and dissemination actions in order to foster services adoption are also presented.

The conclusion presents a more global view of the lessons learnt, remaining issues and recommendations.

Moreover, the document contains an appendix where the experience of the COVID-19 use case is described.

2 Ready to use services

The "Ready to use services" task is defined as "deploying services already available for several scientific communities & support their integration in the EOSC framework." Each service is shortly described here, from the point of view of their usage in the EOSC-Pillar project. Additional information about the description, architecture outlined, technical specifications, deployment and operation is included in D7.2 deliverable and not included here in order to avoid repetition.

2.1 Services described in the proposal

Six services of European-wide interest belong to this category and are described below.

2.1.1 T7.4.1: The use of acceleration resources (GPU, FPGA) in cloud computing (partner CNRS - IPHC)

The Cloud computing service provides a broad set of compute, storage, networking and container-based modules for scientific research. This service, available through multiple interfaces (API, SDK, command line and web dashboard), allows researchers to quickly access a wide range of servers sized for scientific computing, as well as containers with GPUs for AI workload. Based on open-source technologies and open standards (OpenStack[1], Kubernetes[2]), it offers users an easy-to-use innovative digital environment.

2.1.1.1 TRL

This service, provided by the SCIGNE platform[3] hosted by IPHC, was already in production before the beginning of the project. The TRL has evolved during the lifetime of the EOSC-Pillar project from 8 to 9.

2.1.2 T7.4.2: Galaxy as a service platform: Laniakea@ReCaS service & contribution to Galaxy-E

The "Galaxy as a service" platform is dedicated to research groups needing powerful computational infrastructure to run complex analyses over large datasets within the familiar Galaxy environment, e.g. those using a modelling platforms running on Tier 2 or HTC infrastructure. This notably covers WP6.6 use-case, for health related communities, as Galaxy-E initiative is dedicated to biodiversity researchers.

2.1.2.1 Laniakea@ReCaS service (partners (INFN, CNR, INSERM)

The Laniakea@ReCaS[4] service, based on Laniakea software stack, provides the possibility to automate the creation of Galaxy-based virtualised environments through an easy to set-

[1] https://www.openstack.org/
[2] https://kubernetes.io/
[3] https://scigne.fr/en/services/on-demand-servers-and-containers/
[4] https://laniakea-elixir-it.github.io/

up procedure, providing an on-demand workspace ready to be used by life scientists and bioinformaticians. At the end of the process, the user gains access to a private, production-grade, fully customizable, Galaxy virtual instance. Laniakea features the deployment of standalone or cluster-backed Galaxy instances, shared reference data volumes, encrypted data volumes and rapid development of novel Galaxy flavours for specific tasks.

2.1.2.1.1 TRL

The service was released as production ready and registered in EOSC marketplace before the beginning of the project. The current TRL is 9.

2.1.2.2 Contribution to Galaxy-E (partner CNRS)

Galaxy-E[5] is a web platform of the Galaxy project, to get, process, analyse and visualize ecological data. The PAMPA tool suite was already available in the context of Galaxy-E.

2.1.2.2.1 TRL

At the beginning of the EOSC-Pillar project, the web platform and the PAMPA tool suite were usable in production. The TRL is 9.

2.1.3 T7.4.3: Virtual Laboratories/Virtual Research Environments & T7.4.4: Virtual Research Environment-based service for research data publishing (partner CNR - ISTI)

The two subtasks are linked and in order to allow a better understanding without repeating text, are presented below in the same section.

T7.4.3 is defined as “Virtual Laboratories/Virtual Research Environments supporting collaboration and co-operation among scientists working in institutions located in different Member States”. The VRE will offer a transparent and high-level access to federated resources by guaranteeing that established policies are met.

T7.4.4 is defined as “Virtual Research Environment-based service for research data publishing”. This service provides a community-tailored, web-based working environment facilitating the publication of “research objects”. Such a working environment offers (i) a workspace VRE members can use to collaboratively organise the items to publish, (ii) a configurable catalogue VRE members can use to publish (and discover) items according to VRE-defined typologies of “research objects” and the metadata profiles accompanying each of such objects. The content of the catalogue resulting from each VRE is made available by standard protocols (e.g. OAI-PMH, DCAT) thus facilitating reuse.

2.1.3.1 TRL

Both the services are provided by D4Science [12], an IT infrastructure specifically conceived to support the development and operation of Virtual Research Environments by the as-a-

[5] https://ecology.usegalaxy.eu/

service provisioning mode [13]. During its lifetime, the D4Science infrastructure has been selected from several communities of practice to be the platform to develop and operate their VREs [14]. Because of this, the overall service offered by D4Science is at TRL9.

In the context of EOSC-Pillar, a dedicated gateway (https://eosc-pillar.d4science.org/) was deployed and made available since the early phases of the project.

2.1.4 T7.4.5: Converging to common deployment tools to facilitate code circulation and mobility between different infrastructures (HPC, HTC) - pico2 service (partner Inrae - Biogeco, Inria - Plafrim, CNRS - GRICAD)

Pico2 service provides a seamless access to fully operational deployment on servers, docker containers and code sources.

2.1.4.1 TRL

Pico2 was in development at the beginning of the EOSC-Pillar project (TRL6) and the service is at TRL 7-8 at the end of the project.

2.1.5 T7.4.6: Materials modelling MarketPlace and Open Innovation Environment (partner Fraunhofer)

The MarketPlace project[6] and the OIE platform[7] (developed in the OYSTER project), aim to on-board services from two important domains of materials science, namely materials modelling and materials characterisation. Both systems connect experts in their field and enable collaboration by allowing an easier access to data.

2.1.5.1 TRL

OIE and Materials Modelling Marketplace platforms aim to reach TRL 7 at the end of EOSC-Pillar (currently – September 2022 - are at 5-6).

2.2 Service coming from T6.10 call

An open call for communities was conducted by Task 6.10 of the EOSC-Pillar project. Eventually (November 2021), Readmetrics was selected as an additional “ready to use” service.

[6] https://materials-marketplace.eu/
[7] https://oyster-oie.eu/

2.2.1 Readmetrics (partners Couperin, CNRS)

Readmetrics is the name of a tool set[8], based on the already existing and field-tested ezPAARSE[9] and ezMESURE[10] Free and open-source tools, and is a new turnkey solution for:

- monitoring transformative agreements from the usage of electronic resources perspective, including before and beyond;
- analysing usage across disciplines and consortia;
- assessing value of publishing expenditure;
- enabling comparative citation on analysis.

Within the context of EOSC-Pillar, a public Readmetrics instance (behind an Indigo-IAM authentication layer) was installed, serving as a sandbox for potential users to encourage them to install their own dedicated instances in production.

2.2.1.1 TRL

Readmetrics is currently at TRL 8. The new instance dedicated to EOSC-Pillar will be at TRL 8 before the end of the EOSC-Pillar project.

2.3 In-kind services

The “in-kind” services are services proposed “as-is” by a project partner in order to complete the “ready to use services” portfolio in term of type of services or resources. They are all at least at TRL 8 and already used in production by different types of users. They are available for EOSC-Pillar use cases only during the lifetime of the project and conditions may apply to use them. For other usages (for instance after the project or outside the framework of EOSC-Pillar), use cases have to contact the services’ owners.

2.3.1 Indigo-IAM (INFN-CNAF)

The IAM service provides a layer where identities, enrolment, group membership, attributes and policies to access distributed resources and services can be managed in a homogeneous and interoperable way. It supports the EduGAIN federated authentication mechanisms.

The IAM service provides user identity and policy information to services so that consistent authorization decisions can be enforced across distributed services.

Identity and Access Management is provided through multiple methods (SAML, OpenID Connect and X.509) by leveraging the credentials provided by the existing Identity Federations (i.e. IDEM, EDUGAIN, etc.). Distributed authorization policies and Token

[8] https://ezpaarse-project.github.io/readmetrics/
[9] https://www.ezpaarse.org/
[10] https://ezmesure.couperin.org/

Translation Service guarantee selective access to the resources as well as data protection and privacy.

2.3.1.1 TRL

The Indigo-IAM instance dedicated to EOSC-Pillar is at TRL 8.

2.3.2 FG-iRODS (France Grilles)[11]

This service provides geographically distributed storage management system based on the iRODS middleware[12]. At the beginning of the EOSC-Pillar project, the service was already in production for the French research community. It was made available to the project through the France Grilles service offering.

2.3.2.1 TRL

The service is currently at TRL 8.

2.3.3 VIP - Virtual Imaging Platform (France Grilles)

The Virtual Imaging Platform (VIP[13]) is a web portal for medical simulation and image data analysis. It leverages resources available in the "biomed" Virtual Organisation of the European Grid Infrastructure to offer an open service to academic researchers worldwide. .

2.3.3.1 TRL

The service is provided at TRL 8 since several years.

2.3.4 HPC resources (CINECA)

This HPC service is based on a Marconi100 server and at production level before the beginning of EOSC-Pillar.

2.3.4.1 TRL

The service is provided at TRL 9.

2.3.5 Cloud computing & resources (GARR)

The GARR Cloud Platform is a federated and multi-tenant environment spanning different geographical regions. It is based on OpenStack, an open-source software suite aimed at implementing cloud services in datacenters.

The GARR cloud is federated through the IDEM/eduGAIN Authentication and Authorization Infrastructure. Moreover, members of the Italian scientific community can federate their

[11] http://www.france-grilles.fr/catalogue-de-services/fg-irods/
[12] https://irods.org/
[13] https://www.creatis.insa-lyon.fr/vip/

computational and storage resources with GARR, which manages the cloud core services, and disseminates its know-how on cloud technologies.

The persistency layer provides enhanced performance and data redundancy through the Ceph technology. From the networking point of view, the GARR Cloud infrastructures are connected to the facilities of the international research community through the high-performance network links of the GÉANT network.

The service was at production level before the beginning of EOSC-Pillar.

2.3.5.1 TRL

The service is provided at TRL 8.

2.3.6 B2SAFE storage services based on iRODS (CINES, KIT and CINECA)

The B2SAFE service is run by CINES, CINECA, KIT and several other institutions that are members of the EUDAT CDI. The service, which is based on iRODS, and its resource offering are part of the EUDAT CDI services portfolio. It was at production level before the beginning of the EOSC-Pillar project. CINECA and KIT offer the B2SAFE service also in INFRAEOSC-07 project DICE under Virtual Access (VA) conditions.

2.3.6.1 TRL

The services are provided at TRL 8 at CINES and at TRL 9 at KIT and CINECA.

2.3.7 VM (KIT)

The VM service at KIT is based on VMware and runs 24/7 on a high availability cluster. Allocation of the number of virtual CPUs, memory and storage is flexible and can accommodate environments for testing, and running infrastructure enabling services such as monitoring or accounting but also to run thematic service. The service was at production level before the beginning of EOSC-Pillar.

The VM service has not been requested during the EOSC-Pillar project.

2.3.7.1 *TRL*

The service is provided at TRL 9.

3 Transnational usage, adoption by EOSC-Pillar use cases, success stories and issues encountered

This section presents the usage of the services at transnational level, the technical and organisational evolutions, as well as qualitative or quantitative developments of the different services made in the framework of the project in order to adapt the services to the user needs. It also presents the impact of the project for the ready to use services: the evolutions and success stories indirectly related to the project and that would probably have occurred in a longer time frame (or not at all) without EOSC-Pillar. If there were issues that also came up during the process, and solutions to solve them are explained. Publications and dissemination actions in order to foster services adoption are also presented.

3.1 Services described in the proposal

3.1.1 T7.4.1: The use of acceleration resources (GPU, FPGA) in cloud computing

Inside the EOSC-Pillar community, the use cases 6.2 (Agile FAIR Data for Environment and Earth System Communities), 6.3 (Integration of data repositories into EOSC based on communities' approaches) & 6.5 (FAIR principles in data life-cycles for Humanities - Archeology Lidar use case) are accessing the service. In addition, the ready to use service 7.4.5 - pico2 - uses this service since June 2020. In a global manner, the service is actually used by ten active users and has provided more than 600.000 CPU hours since the beginning of the project. Additional hardware resources have been put in production and a container registry based the Harbor Project open source software[14].has been set up in order to anticipate on a need expressed by French and European users

The successful use of GPU and containers as a service requires a solid technical expertise in the field of digital infrastructures for scientific computing. If the scientific team of the use case does not have these skills, it is important to have the possibility to be accompanied by an expert straddling the scientific and the technical domains. It would be interesting to include a service offering such expert support in the EOSC catalogue, and to be able to associate them with this type of high skill services, to facilitate the use of advanced tools.

To be able to register the EOSC catalogue, the privacy, access and acceptable usage policies of the Cloud infrastructure have been published on https://scigne.fr/resources/policies/. The documentation for users and website for the users have been completed and translated to English (https://scigne.fr/en/user-support/online-documentation/openstack-user-guide/). The technical operation measures have also been written (to guarantee data security) and a new page to show services and resources

[14] https://csanhub.org

availability to the user has been put online. Interaction with the EOSC-Pillar project allowed to increase the overall quality of the service and raise the TRL to 9.

The service has been registered in the EOSC catalogue and in the EOSC MarketPlace (https://marketplace.eosc-portal.eu/services/scigne-cloud-compute).

Outside the EOSC-Pillar project

Before the EOSC-Pillar project, the authentication mechanism for this service was based on EGI Check-in, in turn based on OpenID. The OpenID interface to EGI Checkin does not permit to know where the users come from at a site level. The EGI Checkin service administrators have to be contacted to get the information. In the first quarter of 2022, users came from 9 countries in addition to France.

The visibility of this service in EOSC-Pillar facilitates its adoption in the EGI-ACE project. A more diverse user community is now using the service.

The requests of the EOSC-Pillar use-cases were a good driver to enrich the catalogue of service: for example, a new container catalogue for distributing research software was developed, aiming to provide a simple way to share research software between e-Infrastructures.

Publications and conferences

The deployment of GPU servers and containers with the help of FG-Cloud [1] and a demonstration showing the automation of the applications' deployment on GPUs with containers in a multi-cloud environment [2] were presented at the JCAD 2021[15] conference in France. JCAD events are dedicated to users, technical experts of HPC, HTC and cloud infrastructures and related services.

3.1.2 T7.4.2: Galaxy as a service platform: Laniakea@ReCaS service & contribution to Galaxy-E

3.1.2.1 Laniakea@ReCaS service

Inside the EOSC-Pillar project the Laniakea@ReCaS service is adopted by the use case 6.6 (Exploring reference data through existing computing services for the bioinformatics community). This use case has two main objectives, to ensure the reproducibility and coherency of the data analysis and conform to data protection regulations concerning health personal data.

To achieve these objectives, the use case adopted the Laniakea service and through this system has compared the analysis between a private deployment of Galaxy and a public instance manually configured. The test was performed developing a specific "Somatic Variant Detection" Galaxy flavour and running a variant calling workflow, using the same input.

[15] https://jcad2021.sciencesconf.org/

The test was recorded and presented as a demo during one of the "Face to Face meeting" organized within the EOSC-Pillar project, dedicated to WP6, (https://www.youtube.com/watch?v=SK9C_MWz6Yk&t=68s), showing that using the same workflow and the same input data, an identical output was obtained.

During the implementation of this use case scenarios, Laniakea has been updated, following the best practice decided in the subtask 6.6.1 "share and harmonize Galaxy deployment practices". To do so, the Laniakea core, the one involved in the configuration of Galaxy servers, has been completely revised to employ the official Galaxy Project Ansible roles allowing for faster alignment of Laniakea flavours to the latest Galaxy releases.

In parallel, a continuous integration system based on Jenkins (https://www.jenkins.io/) was developed to improve the long-term sustainability and maintainability allowing the automatic Galaxy deployment, automatic image creation, Galaxy flavour testing and update.

After a robust test phase, a new version of Laniakea will be released (v3.0.0) and the Laniakea@ReCaS service will be consequently updated: the new version provides more stability, simplified maintainability and offers EOSC and ELIXIR users new applications in addition to beyond Galaxy.

Outside the EOSC-Pillar project

The service is currently used by 15 research groups and organizations, for a total of 18 Galaxy instances, serving 113 users.

A selection of 10 use cases have been described in detail in a dedicated paper *"Laniakea@ReCaS: Exploring the Potential of Customisable Galaxy on-Demand Instances as a Cloud-Based Service"* published on BMC-Bioinformatics [3].

Publications and conferences

A poster [4] dedicated to Laniakea@ReCaS service and a talk [5] on the Laniakea software platform update were presented at the Galaxy Community Conference 2021.

Laniakea has been presented to the workshop "Web Applications for Life Sciences" the 8th July 2022, with a talk named "The Laniakea Dashboard and storage encryption components: a foundation for developing on-demand cloud services for Life Science information on submission".

The Laniakea@ReCaS service has been described in a dedicated publication on BMC Bioinformatics [3] "Laniakea@ReCaS: exploring the potential of customisable Galaxy on-demand instances as a cloud-based service.".

A poster [6] that presents a bioinformatics use case has been presented by Gilles Mathieu & Yosra Sanaa to French days on computing and data JCAD 2021 conference[16].

[16] https://jcad2021.sciencesconf.org/data/Poster_Cas_dusage_EOSC_Pillar_Yosra_SANAA.pdf

3.1.2.2 Contribution to Galaxy-E

During the EOSC-Pillar project, the PAMPA tool suite has been completed and tested on Galaxy. PAMPA allows to compute common biodiversity metrics from species abundance data and analyse it through generalized linear (mixed) models (GLM and GLMM). This workflow made up of 5 tools allows to process temporal series data that include at least year, location and species sampled along with abundance value and, finally, generate article-ready data products. The results of this development is a scientific workflow dedicated to ecology in Galaxy. Different types of data were analysed with PAMPA in order to test the workflow. The PAMPA suite is included in https://ecology.usegalaxy.eu/

A documentation has been provided and a tutorial developed[17].

Publications and conferences

The results have been disseminated in different contexts:

- ecoinfoFAIR2020: French national action on FAIR development in Ecology: "How create Galaxy tools from R scripts and why ?" [7]
- French days on computing and data "JCAD 2020" & "Bioinformatics days for French Museum 2020": "Galaxy-E : une instance de Galaxy dédiée à l'analyse de données en Ecologie" (JCAD 2020 / bioinfodays)[8]
- CODATA / GO FAIR "International FAIR convergence 2020 symposium": "Galaxy-E : a Galaxy instance dedicated to data analysis in Ecology"
- Biomac ecology lab, Amsterdam : "How to integrate an EBV workflow in Galaxy and why?"
- French ecology and conservation lab (UMR CESCO): "EBV workflow for developers" & "EBV workflows for users" presentation.

3.1.3 T7.4.3: Virtual Laboratories/Virtual Research Environments & T7.4.4: Virtual Research Environment-based service for research data publishing

These subtasks are linked and in order to allow a better understanding without repeating text, they are presented together in the present section.

Before EOSC-Pillar, the service was operated by D4Science (www.d4science.org) and supported several communities. In particular, in July 2019 the D4Science infrastructure was supporting the activity of more than 10K users (10,507 active users) accessing via 14 dedicated gateways and more than 100 Virtual Research Environments.[18]

The service was fully operational and enacted the creation of the existing VREs as well as the forthcoming ones.

[17] https://training.galaxyproject.org/training-material//topics/ecology/tutorials/PAMPA-toolsuite-tutorial/tutorial.html

[18] A continuously updated list of D4Science supported gateways and VREs is available at https://services.d4science.org/thematic-gateways

The EOSC-Pillar project represented a scenario where the service could be further exploited and developed to serve the needs of the project use cases and communities.

The service as a whole was monitored by D4Science [12], [13]. Any VRE consists of a set of services that might be operated by third-party providers. Depending on the level of integration D4Science might monitor the operation of these third-party services too [14].

D4Science is not offering a "status" web page where service availability can be known by users, yet (i) the availability of VREs is very high (over 99% in the last 5 years), and (ii) known downtimes are announced in advance.

The community served by D4Science is very heterogeneous across "space" and "scope" (topics of interest). It is international and very often paired with EU projects aiming at specific use cases.

Community

Documentation (often how-tos) is available both in the form of a Wiki enriched by the same users of the platform, as well as from a dedicated site: https://dev.d4science.org/

The AAI solution is in place. It was conceived to support the federation of several identity providers.

The status and usage of the service at the time of writing this report (September 2022) is characterised as follows:

- The service is fully operational and made available by the **project gateway** https://eosc-pillar.d4science.org/ (a screenshot of the gateway is in Fig. 1)
- **12 Virtual Research Environments** have been created (browse the available VREs by https://eosc-pillar.d4science.org/explore)
- **500+ Registered Users** on the gateway;
- **13,250+ Working Sessions** initiated by the Gateway (since July 2020), with 500 circa working sessions per month in average;

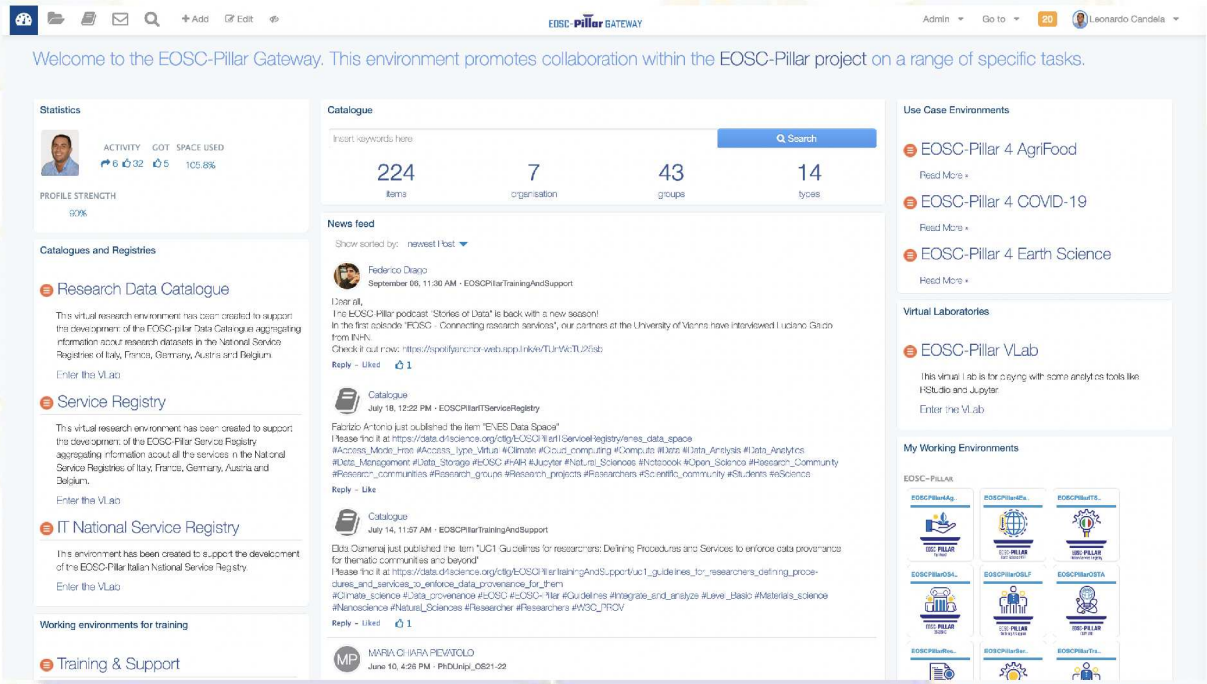
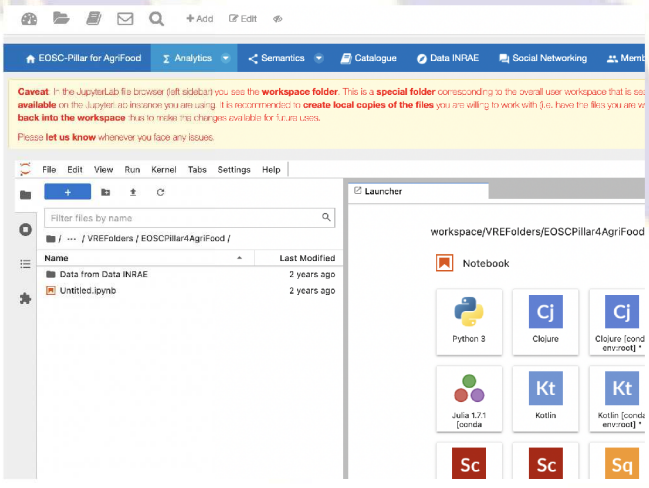
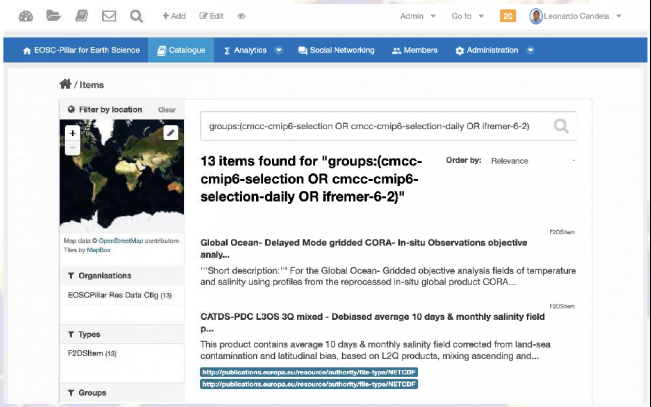
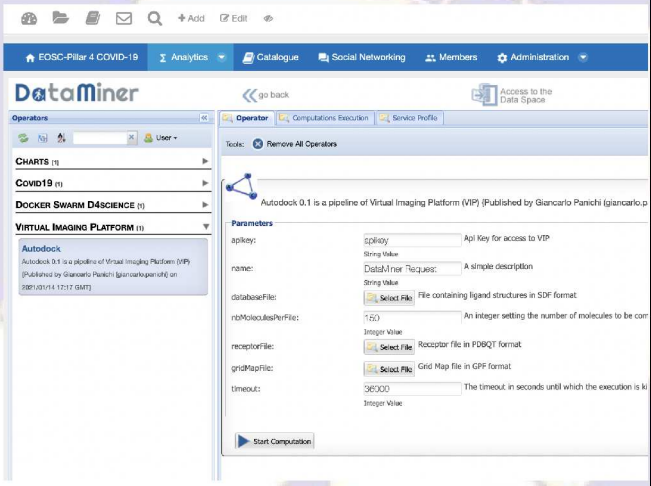
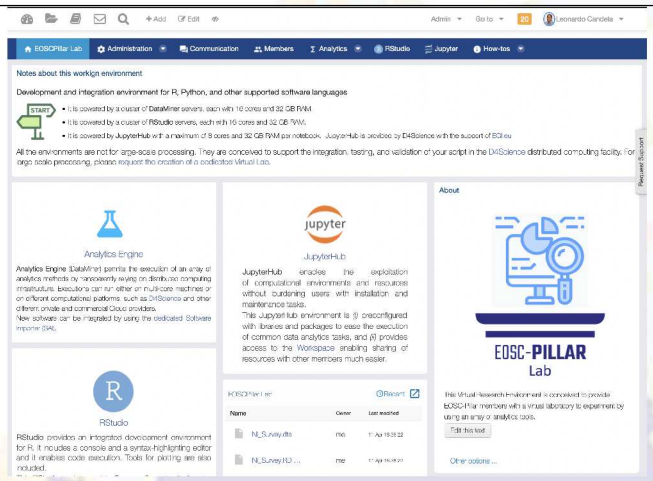
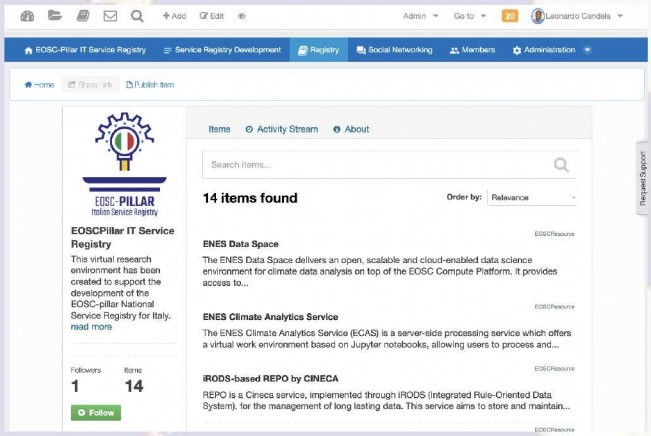
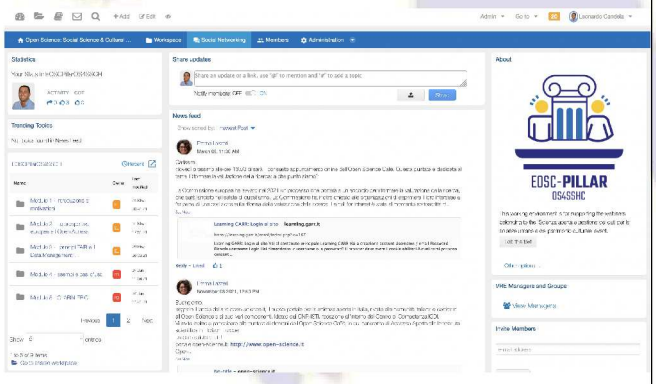


Figure 1 EOSC-Pillar Gateway Dashboard

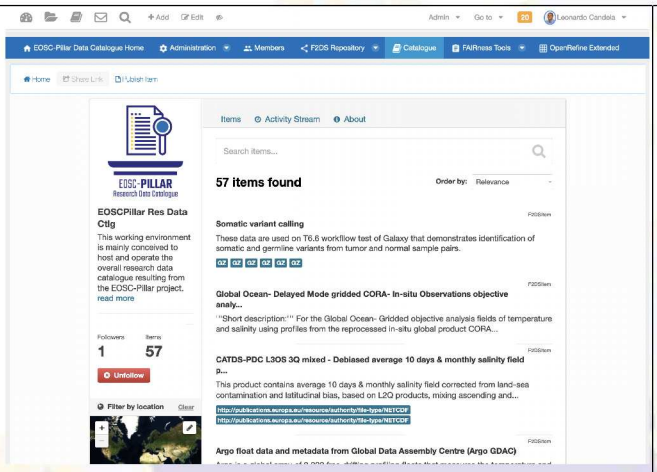
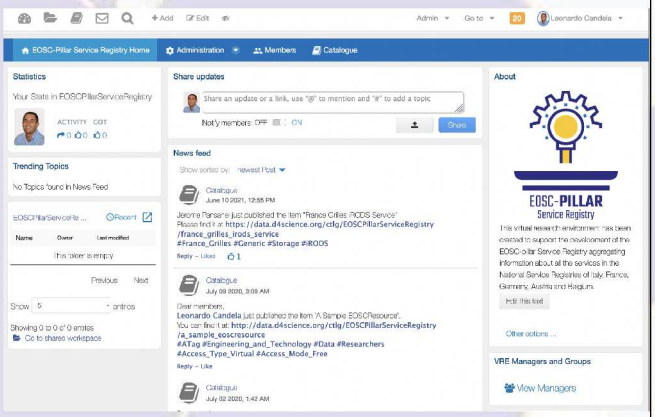
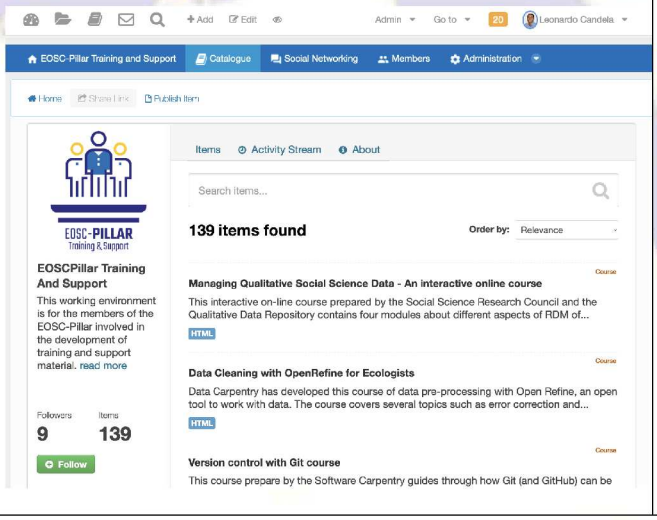
For each VRE, the following table presents a short description, and the current number of registered users:

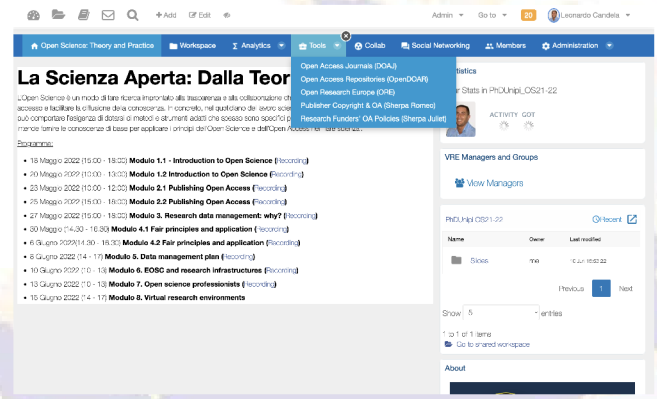
VRE Name	VRE Description	Users
eoscpillar4agrifood (https://eosc-pillar.d4science.org/web/eoscpillar4agrifood)	A Virtual Research Environment developed in the context of the Use Case 6.3. A detailed description of the use case was in Sec. 2.3 of D6.2 [12]. The VRE offers several services and is integrated with the INRAE Repository, e.g. data from this repository can be transferred into the VRE workspace and then analysed by Jupyter. A screenshot is below. 	19

eoscpillar4earthscience (https://eosc-pillar.d4science.org/web/eoscpillar4earthscience/)	A Virtual Research Environment developed in the context of the Use Case 6.2. A detailed description of the use case was in Sec. 2.2 of D6.2 [12]. A screenshot of the catalogue of datasets offered by the VRE is below. 	31
eoscpillar_covid-19 (https://eosc-pillar.d4science.org/web/eoscpillar_covid-19/)	A Virtual Research Environment developed in the context of the COVID Use Case. This working environment integrates also the VIP platform (cf. Sec. 2.3.3 and 3.3). A screenshot showcasing how VIP processes (Autodock in this case) can be launched from the VRE is below. Details on the case are in the appendix Covid-19 Use Case . 	12
eoscpillar_lab (https://eosc-pillar.d4science.org/web/eoscpillar_lab)	An open access Virtual Research Environment deployed to provide project members with a working environment for data analytics (e.g. exploited to crunch WP3 datasets). A screenshot of the VRE is below.	30

		
eoscpillaritserviceregistry (https://eosc-pillar.d4science.org/web/eoscpillaritserviceregistry)	A Virtual Research Environment supporting the development of the Italian Registry of Services (WP4). A screenshot of the resulting registry is reported below. 	13
eoscpillaros4ssch (https://eosc-pillar.d4science.org/web/eoscpillaros4ssch)	A Virtual Research Environment supporting one of the training events organised by WP5 (Scienza aperta e gestione dei dati per le scienze umane e del patrimonio culturale). 	47
eoscpillaroslf (https://eosc-	A Virtual Research Environment supporting one of the training events organised by WP5 (FAIR	27

pillar.d4science.org/web/eoscpillarosIf)	Data Stewardship in Life Science). 	
eoscpillarosta (https://eosc-pillar.d4science.org/web/eoscpillarosta)	A Virtual Research Environment supporting one of the training events organised by WP5 (Praticare l'Open Science nelle Scienze della Terra e dell'Ambiente) 	134
eoscpillarresdatactlg (https://eosc-pillar.d4science.org/web/eoscpillarresdatactlg)	A Virtual Research Environment realising a working environment for exploiting the services and data stemming from the F2DS activity (WP5)	39

		
[eoscpillarserviceregistry](https://eosc-pillar.d4science.org/web/eoscpillarserviceregistry) (https://eosc-pillar.d4science.org/web/eoscpillarserviceregistry)	A Virtual Research Environment for the experimentation of the Service Registry technology (WP4) for the development of the project Service Catalogue 	8
[eoscpillartrainingandsupport](https://eosc-pillar.d4science.org/web/eoscpillartrainingandsupport) (https://eosc-pillar.d4science.org/web/eoscpillartrainingandsupport)	A Virtual Research Environment supporting the development of the EOSC-Pillar Training Catalogue (WP5). 	150

phdunipi_os21-22 (https://eosc-pillar.d4science.org/web/phdunipi_os21-22)	A Virtual Research Environment supporting one of the training events organised by WP5 (The course on Open Science for University of Pisa PhD Students) 	14

Publications and conferences

D4Science is an established service provider supporting several communities and use cases. An overview of its offering for supporting the implementation of Open Science practices was described in various publications [13, 14,15]. A recent publication describes how its service has been exploited to support the development of the EOSC-Pillar RDM Training Catalogue [16].

3.1.4 T7.4.5: Converging to common deployment tools to facilitate code circulation and mobility between different infrastructures (HPC, HTC) - pico2 service

pico2 in EOSC-Pillar is a follow-up of pico2 in the EOSCPilot [19]project. Within EOSCPilot, pico2 (Pilot for Connecting Computing Centres) was an initiative aiming at facilitating data and code flows between HPC and HTC infrastructures at regional, national and European level, being agnostic as far as scientific communities were concerned. The centres involved were national (IDRIS, CC-IN2P3, France Grilles in France and DESY in Germany) or regional in France (GRICAD Grenoble, MCIA and Plafrim Bordeaux, GRIF CEA Orsay) and the French NREN, RENATER. Several areas of work were addressed during the EOSCpilot project:

- Data sharing through an IRODS zone federation,
- Dedicated network connection via L3VPN,
- Evaluation of different code carrying solutions: packaging and container systems.

Among them, only the last item, for sharing and circulating codes, has been carried on in EOSC-Pillar, with the following French participants: GRICAD Grenoble, INRIA infrastructure

[19] https://eoscpilot.eu/

PlaFRIM, and BioGeCo at INRAE. pico2 is focused on deploying tools on HPC infrastructures while easing their use for non-HPC specialists (biologists of the metabarcoding community) through access to computing facilities via Galaxy servers and/or Jupyter hubs.

During the course of EOSC-Pillar, a new feature has been developed upon request of the use case 6.3 (INRAE Dataverse), namely the development of a connection between datasets available on INRAE Dataverse (now data.gouv.fr) (public datasets with DOI) and software available on public GitLab. The scientific objective was to provide to a user a seamless access to fully operational deployment on servers (even HPC ones), docker containers and code sources through GitLab, and at the same time access to data through Dataverse.

Architecture outlined, technical specifications and capabilities:

BioGeCo[20] has developed a user-friendly library for intensive computing in metabarcoding, with a python based interface called yap!, and a series of C programs for computing distances, as well as provided datasets for tests (small datasets for demos as well as real size large datasets for testing access to HPC). The datasets are publicly available on INRAE's Dataverse system, the different software components (disseq, jelly_diskm,yap!) are publicly available on INRIA's GitLab, as well as on Software Heritage.

GRICAD[21] has provided a dedicated VM for these tools, which is connected with DAHU, their HPC Cluster. BioGeCo has deployed a Galaxy server on it, which will be publicly available in 2023, soon after the project closure. Currently, for those who have credentials at GRICAD, this Galaxy server can work as an interface between the VM and DAHU for HPC, enabling those who are not familiar with OAR scheduler to launch HPC programs. This has been successfully tested with the MPI distributed version of disseq (written in C) on GRICAD DAHU with an access via the dedicated VM.

jelly_diskm is written in python with a distributed version with MPI python. It has been successfully tested on BioGeCo server PGTVM4, and next step is to make it available on GRICAD VM with access to DAHU.

INRIA has offered a service on a dedicated machine of their infrastructure PlaFRIM[22] (a dedicated node). We have deployed a Galaxy server on this machine as well, which will be used for HPC on the nodes of PlaFRIM. However, due to INRIA policy of access to the platform PlaFRIM, which is a research tool for the community of mathematics and computer sciences in Bordeaux, independent from EOSC-Pillar, it was not possible to open it to public access beyond this community, and test the connections with GRICAD. However, it has been used successfully as a development tool.

[20] https://www6.bordeaux-aquitaine.inrae.fr/biogeco
[21] https://gricad.univ-grenoble-alpes.fr/
[22] https://www.plafrim.fr/

The next expected step for work on GRICAD infrastructure is to open the service to the public. However, BioGeCo team has not the competences to comply with all security rules, be they on the access to the infrastructure or to the web service necessary for interfacing a Galaxy server. Thanks to the project a tight cooperation has been established with GRICAD on these issues, this cooperation will continue in 2023 after EOSC-Pillar, to fix them.

Interface with WP6.3 Dataverse instance is operational for connecting data and computing (download tested, upload in progress).

All these tools working on GRICAD dedicated VM and BioGeCo server PGTVM4 have been encapsulated in a Docker container, tests are in progress. If successful, this will permit the use of the encapsulated tools from a laptop with Windows OS, which is crucial to reach the community of biologists.

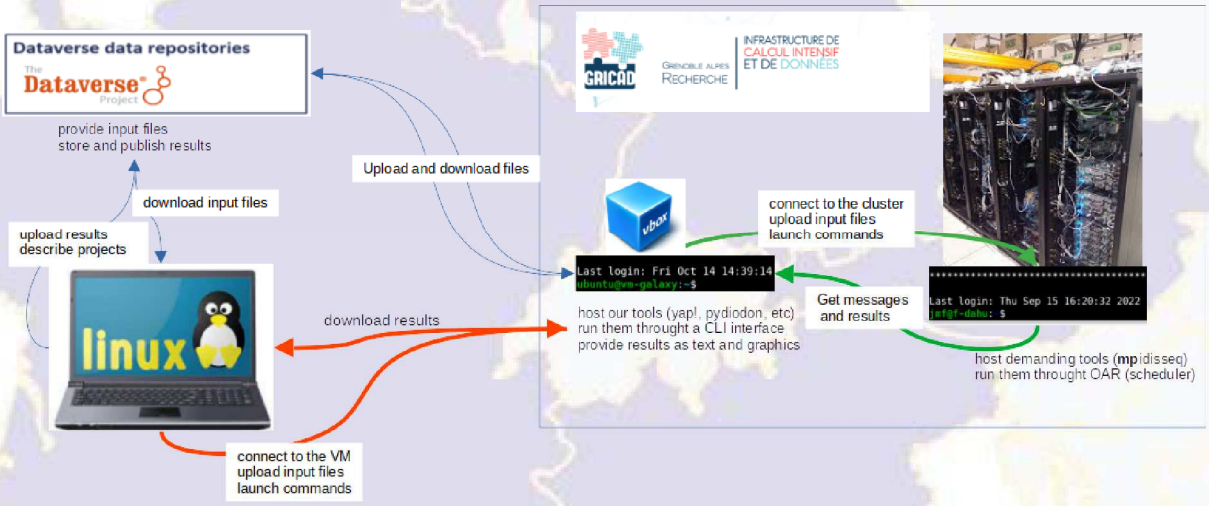


Figure 2: diagram of the connections between the different components of pico2

Outcomes - successful deployments and operations

The main achievement in pico2 is twofold:

- possibility to launch a distributed memory program on DAHU (HPC cluster) from a galaxy server without writing a script in OAR

- seamless communication between data and calculations through a connection between INRAE Dataverse (now data.gouv.fr) and the VM deployed on GRICAD. This has been demonstrated during a demo for members of WP6.3.

More precisely, the following steps have been successfully deployed:

- public access of the tools on a public git repository (https://gitlab.inria.fr/biodiversiton/disseq and https://gitlab.inria.fr/biodiversiton/yap).
- deployment of the tools on two regional centres: PlaFrim at Bordeaux and Gricad at Grenoble.
- deployment of a Galaxy server on each of these centres, as a node or a virtual machine.
- distribution of the MPI version of the code on the computing nodes of GRICAD, either as a Command-line interface from the VM (operational), or through a Galaxy server (with GRICAD credentials)
- connection between data (in INRAE's Dataverse) and codes deployment (either from Git or from Software Heritage) publicly accessible.
- integration of these tools in a docker container.

As mentioned above, current efforts are on letting the Galaxy server publicly available while complying with the security specifications of the computing centres hosting the executable.

Last step is, on top of solving these issues, to make these achievements visible for the community of metabarcoding, which can benefit from access to HPC but is not very familiar with it. This has been initiated in the course of COST project DNAqua.net and with a close collaboration with a team involved in metabarcoding for testing the seamless availability of datasets and software.

3.1.5 T7.4.6: OIE and Marketplace projects.

The development of both platforms is still ongoing, which means new features are continuously being added.

The most direct contributions from EOSC-Pillar are the connection of EOSC-Pillar services to the MarketPlace (specifically to the Dataverse repository in T6.3.3) and the accessibility to the MarketPlace and OIE via AAI from the EOSC-Pillar. However, EOSC-Pillar has also greatly contributed in the definition and prioritisation of requirements through the service maturity assessment carried out in WP7.

Some code[23] has already been published, as well as user documentation[24].

While currently only project members have access, the aim is to progressively open up the platforms, allowing for instance access to selected members of EOSC-Pillar before opening up registration completely.

Most issues linked to the EOSC-Pillar project have to do with the timeline. Since the platform development is still ongoing, the maturity of the service is not yet as high as other services.

The success of the activities related to EOSC-Pillar will come in the form of a higher exposure and accessibility which translate in new users for the services.

[23] https://github.com/materials-marketplace/
[24] https://materials-marketplace.readthedocs.io/

The MarketPlace platform is available (for now only for Consortium members) on https://materials-marketplace.eu/ .

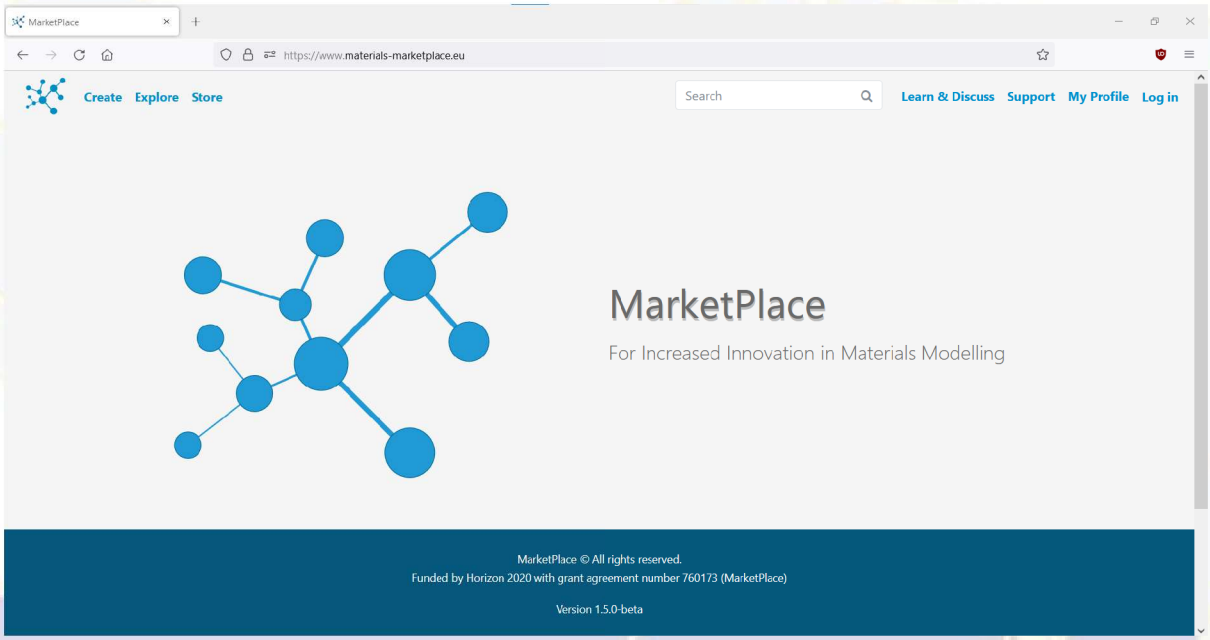


Figure 3: MarketPlace landing page

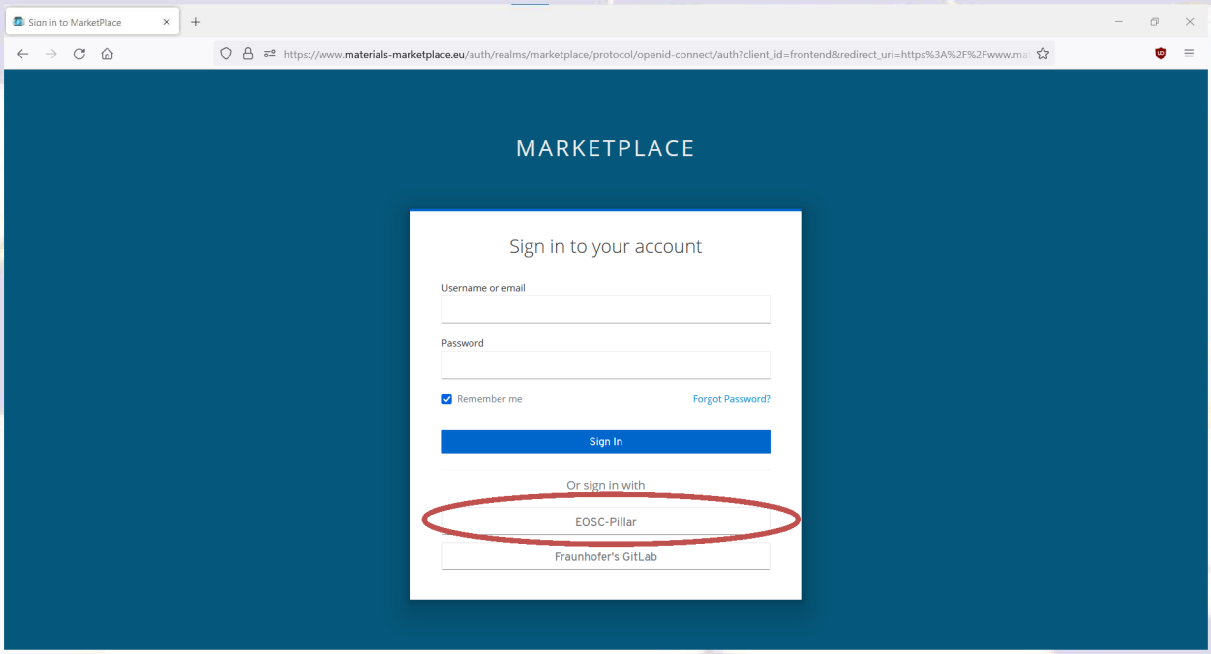


Figure 4: MarketPlace login, with Indigo IAM integration

Similarly, the OIE is available, with restricted access, on https://oyster-oie.eu/.

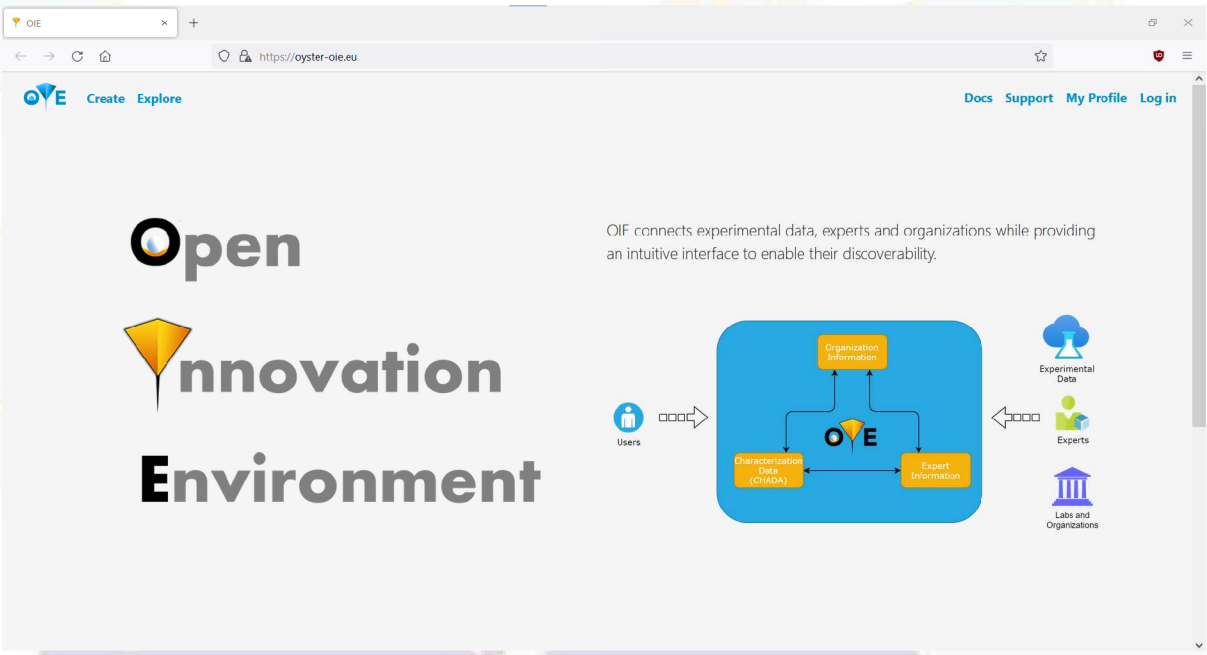


Figure 5: OIE Landing page

Given the status of the platforms, not many public dissemination activities have been done. There is a publication [17] presenting the benefit of the MarketPlace and its sister project (VIMMP) to materials modelling, as well as a conference talk [18] demonstrating one feature of the MarketPlace, the materials relation database.

3.2 Service coming from T6.10 call (since end of 2021)

3.2.1 Readmetrics

The Readmetrics service is operational in production-mode since a couple of years, serving 130 institutions from the French Couperin.org consortium (and targeting 250+ institutions in this specific context).

It lives on https://ezmesure.couperin.org/ and is hosted on a cluster operated by Inist-CNRS since the beginning. We evaluate that we have reached a TRL9 in this HER French context with this specific instance.

The goal with participating in the T6.10 call was to deploy a platform dedicated to Readmetrics EOSC for institutions external to the “core” user base so that the service could be exposed and used by a broader range of institutions and consortia that share the need to collect, aggregate and analyse usage data to electronic resources.

A dedicated EOSC instance of the ReadMetrics platform and services was deployed (https://eosc.readmetrics.org/login) during summer 2022 on a dedicated cluster. This instance allows users to test the service without having to install it locally (which can be time and resource consuming).

The work of deploying this new ReadMetrics installation for the outside world started with deploying a dedicated Indigo-IAM instance (replacing shibboleth) and to connect it to the ezMESURE authentication layer. The opportunity to discover Indigo-IAM was made possible by the connection to the EOSC-Pillar project.

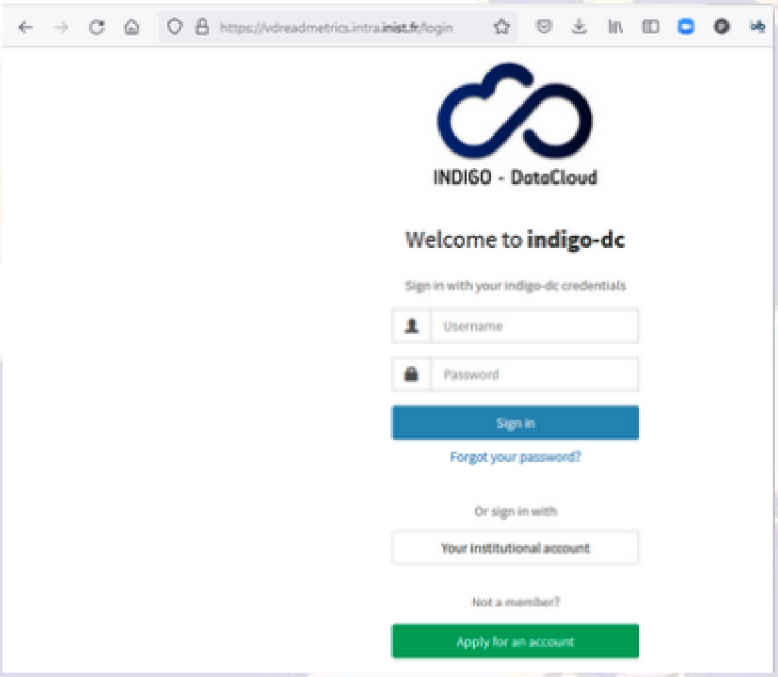


Figure 6: Sign in picture

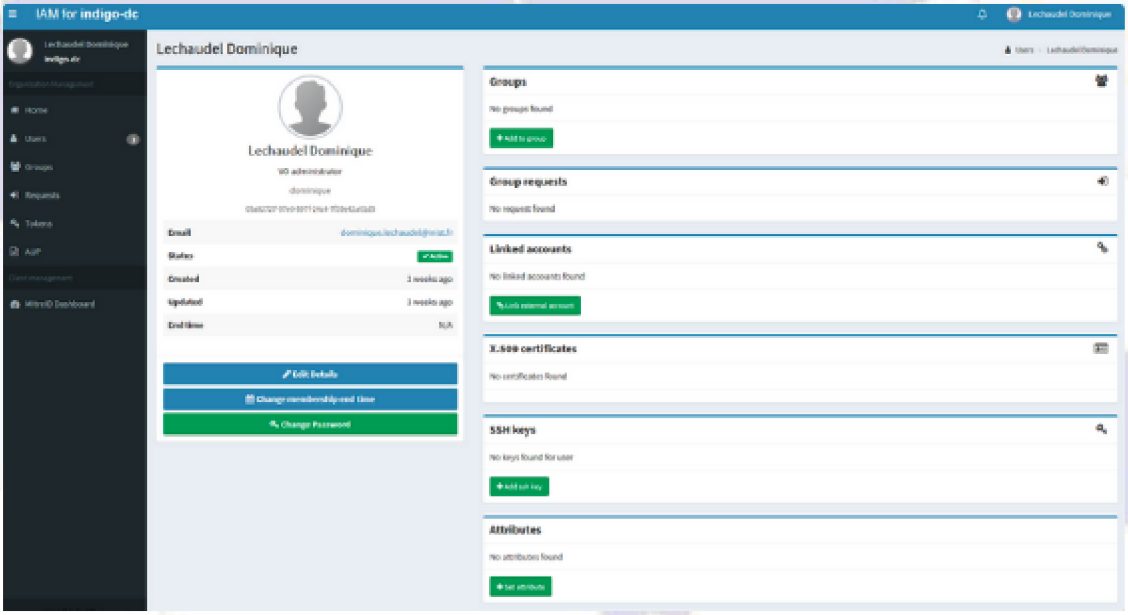


Figure 7: Indigo-dc personal page example

This installation suffered some initial delay due to a lack of available human resources in the ReadMetrics team and is still ongoing at time of writing: nevertheless, the final goal remains unchanged and we hope a ReadMetrics instance can be exposed in the marketplace as soon as the service interface with the Indigo IAM layer will be complete.

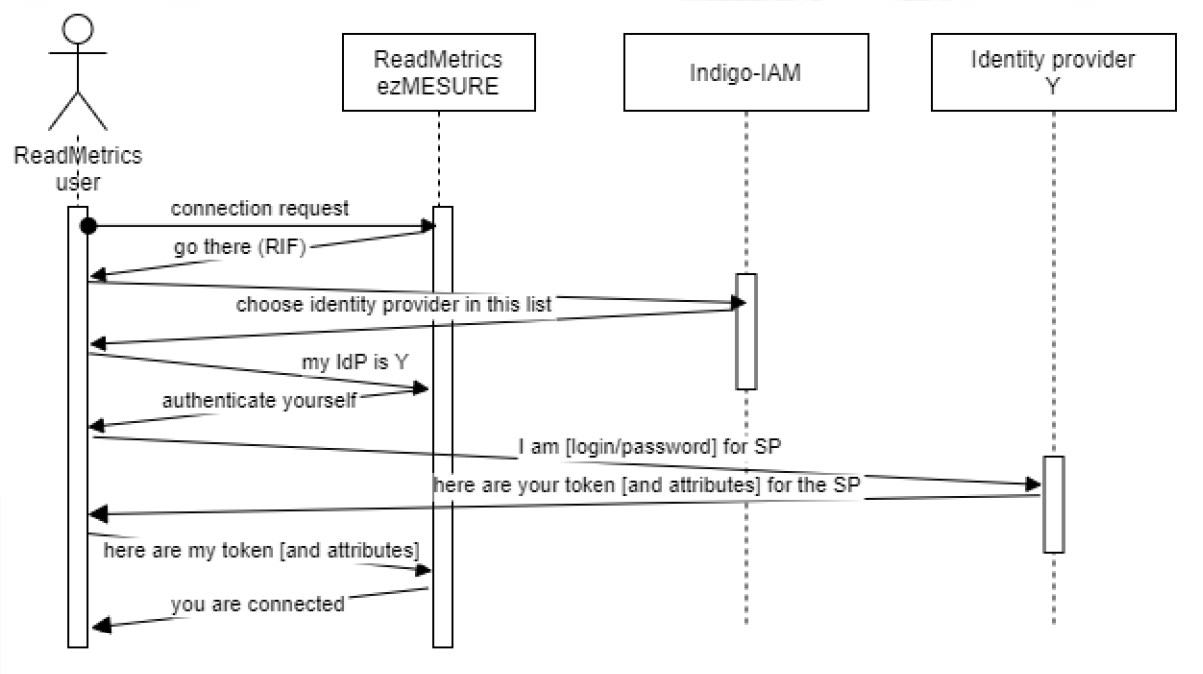


Figure 8: Readmetrics authentication process

3.3 In-kind services

3.3.1 Indigo-IAM (INFN-CNAF)

An instance of INDIGO-IAM service dedicated to the EOSC-Pillar project has been configured and deployed at INFN-CNAF. Since the first deployed version (v1.6.0), some features and fixes has been applied through the time through the version updates. With the upgrade to v1.7.0, the main features/fixes added were:

- an improved token exchange flexibility
- linking ssh keys to an account has been allowed
- added the support for including custom HTML content at the bottom of the login page
- prevent the deletion of dynamically registered clients under certain conditions
- intermediate group membership enforcing
- support for X.509 managed proxies
- added more flexibility in populating registration fields from SAML authentication assertion attributes
- fixed bug on submitting multiple group requests for the same group

Then, the upgrade to v1.7.1 and then to v1.7.2 added other several bug fixes and dependencies migrations. At this time, there are 39 registered users.The plan is to update soon to the next upcoming release v1.8.0 which mainly provides more support to AARC guidelines[25], introduces support for JWT-based client-auth, fixes and restyles administrators' client user interface and several other bug fixes and improvements.

Publications and conferences

Indigo-IAM is an established software used to provide services in distributed and heterogeneous contexts. A lot of material about Indigo-IAM is available on the web, for example: general documentation [20], a webinar [21] by Andrea Ceccanti for ESCAPE project and a conference article[22]: "Beyond X.509: token-based authentication and authorization for HEP".

3.3.2 FG-iRODS France Grilles

At the beginning of EOSC-Pillar, the service was already in production for the French research community. It was made available through the France Grilles service offering.

The service is already monitored at the provider level and at a federation or external level. The user availability level of the service is provided at the provider level.

The accounting of the service is available internally. Storage usage details are available at the user level.

The main development in the framework of the project was to connect to an AAI framework using the OpenID connect protocol. This is still a work in progress, as it was dependant of the new authentication system that has been released in June 2022. The work will be achieved early in 2023.

Initially, 27 users (all of them French researchers) from 6 communities are using the service. Many of them were from the Phenome research infrastructure[26], but also other communities, like astroparticle physics and bioinformatics. There used to be no English documentation available for the users. User credential and X509 based authorization is used for AAI.

End of 2022, the evolutions made in the framework of EOSC-Pillar are:

- AAI integration: connection to the Indigo-IAM service (still in progress, as it is based on a version of iRODS that has just been released).
- English documentation: the policies and documentation to access the service have been translated in English and published.
- A new storage server has been added to offer a better user experience to the EOSC-Pillar project members.

[25] https://aarc-project.eu/architecture/
[26] https://www.phenome-emphasis.fr/phenome_eng/

The user availability level of the service is now provided at a federation or external level. Three use cases are using the service (WP6.2, WP6.3 and WP6.5) and there are 5 users from EOSC-Pillar. Their accounting data are made available directly from the iRODS system.

The main ongoing work is to join the EOSC-Pillar IAM.

The most important added value from EOSC-Pillar context for the service was the increase of the service visibility and the work accomplished to be able to use a new authentication and authorisation framework based on OpenID.

3.3.3 VIP (Virtual Imaging Platform) France Grilles

VIP has been used by the COVID19 use case "Modelling and virtual screening of existing drugs against SARS-CoV2 target proteins" (cf. Appendix 1) for integrating the autodock application in a interoperable manner with the rest of the EOSC-Pillar infrastructure. Autodock was integrated in VIP as a workflow allowing to automatically distribute the load (based on the number of molecules) in multiple parallel jobs executing on the resources available to VIP users through the biomed Virtual Organization within the EGI e-infrastructure. To ease the deployment of the application and ensure a certain level of reproducibility, a Singularity container was built. Users can launch autodock executions on VIP from the EOSC-Pillar VRE (see section 3.1.3). The submission is handled transparently at the VRE level by using calls to the VIP REST API.

3.3.4 HPC resources (CINECA)

At some point, an expression of interest for HPC resources usage was raised by UC6.2: however, a closer inspection identified the Cloud resources as the best match to the computing needs expressed.

3.3.5 Cloud computing and resources (GARR)

Inside the EOSC-pillar project, GARR's Cloud computing and resources service provides cloud resources for Laniakea@ReCaS, F2DS and F2OS services.

In the context of Laniakea, GARR hosts the CVMFS mirror server which improves connectivity to the service. The resources of the service are provided in the cloud region of Palermo (garr-pa1) and are equivalent to: 32 vCPU, 64GB of RAM and 15TB of storage.

For the F2DS project GARR provides a Kubernetes cluster as a staging environment. This cluster is composed of a master node and three worker nodes and the resources are reserved in the Catania region (garr-ct1) and amount to: 32 vCPUs, 64GB of RAM and 500GB of storage.

A Kubernetes cluster is also provided for the F2OS project as a development environment and is always composed of a master node and three worker nodes. The cluster is built in the Palermo region (garr-pa1) with the following quantities of resources: 32 vCPU, 64 GB of RAM and 600 GB of storage. Furthermore, GARR provides IaaS resources for the F2OS

(Ontoportal) service on the OpenStack platform. For the F2OS service, GARR reserves: 30 vCPUs, 50GB of RAM and 2TB of storage.

3.3.6 B2SAFE storage services based on iRODS (CINES, KIT and CINECA)

Cineca and KIT have not received any request by the use cases for using their B2SAFE iRODS instance.

3.3.7 VM (KIT)

KIT has not received any request by the use cases for the use of the VM service.

4 Conclusion

An important objective of Transnational Access (TNA) is to provide efficient, free and smooth access to services and resources offered from a specific provider to external researchers and users outside its national community.

In addition to immediate scientific outcomes, users from European nations other than where the facilities are based contribute to the development of the European research infrastructure of facilities by actively collaborating with them in terms of new techniques, new instruments and new updates. The TNA allows all European scientists to support different needs and/or to realise experiments to address scientifically as well as socially important topics, including the Societal Challenges of Horizon Europe.

During the last years national policies evolved in the European countries and the barriers that prevent the use of EOSC resources outside the researchers' countries or communities are now lower. Furthermore, obstacles due to the legal and ethical rules have been resolved because of the TNA. For instance, on the one hand, offering a resource or a service from a service provider outside their original country may cause a legal issue and on the other hand, for the users gaining access to these kind of resources the TNA comes in as a concept and a solution and the same time (ethical aspect can be confidentiality for the user). So this bureaucracy for both facilities and users is addressed based on a "programme" agreement to support TNA.

In order to facilitate the integration of services in EOSC, EOSC-Pillar teams provided "Guidelines and Recommendations for the Technical Integration of Resources and Services in the EOSC" [19]. In addition, two main support activities were carried out. Support has been provided to the EOSC-Pillar services either from "ready to use" services or from "use cases" to integrate EOSC: the support on integration with the EOSC-core services, specifically Indigo-IAM as one of the federated EOSC-AAI solutions to have a seamless access on distributed applications and resources and the support on on-boarding a EOSC-Pillar resource or service into EOSC catalogue and Marketplace. Most of the services presented in this deliverable are now registered and findable in the EOSC catalogue and in the EOSC Marketplace.

The Indigo-IAM EOSC-Pillar instance is then a keystone to facilitate smooth access to services and resources thanks to the Single Sign On mechanism it offers and its support of the EduGAIN authentication federation. Adding this instance brings major advantages to a service: easily expanding the user base (from the federation) while removing some complexity in managing user account creation. Many providers took the opportunity to adapt their service with the support of the EOSC-Pillar teams. The possibility to add such a functionality in a service depends also on the maturity of the software used to provide the service. For example, FG-iRODS had to wait on an iRODS version compatible with OpenID before its successful integration in the EOSC-Pillar Single Sign On federation.

It is important to note that the services have also benefited from their maturity level and operational readiness assessment conducted in the framework of EOSC-Pillar WP7 [cf. D7.3 deliverable: https://doi.org/10.5281/zenodo.5838071]. It contributed to a better knowledge of the level required to be fully integrated to EOSC and leveraged their ability to the transnational access. For example, their documentation has been translated in English and their policies have been better defined and published by several services.

In addition, interoperability and connectivity between use cases and services was improved. We can cite of course the support of use cases that expressed the need of the services but also the support of another ready to use service (e.g. the acceleration resources supported Pico2). Another example is the Covid19 use case that was included in the "ready to use services" portfolio during the pandemic. It was possible to integrate the application needed for the analysis in the VIP platform allowing through it to use the biomed distributed resources and give access to VIP in a VRE especially created in D4Science.

Despite these very positive evolutions, we note remaining difficulties.

- The policy of several services restricts their access to their project members (for instance OIE and Marketplace projects), others restricts the access to a well-defined set of users and are protected by high level security measures (for instance PlaFRIM HPC infrastructure). Such policies are not a barrier for the services to integrate EOSC provided that they are publicly available. But the technical integration in EOSC of resources protected by high level security resources such as those of HPC centres does not depend on ourselves. Some solutions are under development but are not currently supported.
- One factor common to multiple services in terms of tracking the origin of users is the use of Single Sign-On (SSO) mechanisms. Adding these providers (such as the above presented Indigo IAM) brings many benefits to a platform, easily expanding the user base while removing some complexity in managing user account creation. However, trusting a provider to authenticate external users entails partly giving away the control (and thus knowledge) about incoming users. A platform can know which identity provider the users originate from, as well as some additional information the provider may share, but in many instances, the country of origin is not one of them. It is then impossible to quantify this transnational access.

Recommendations:

In order to guarantee high level quality and mature enough services in the EOSC infrastructure level, we strongly recommend to disseminate in the services providers' community:

- The three videos produced in the framework of EOSC-Pillar, presenting in an effective and nice way three main topics that potential user communities may be interested in:
  - Why should you trust EOSC Portal resources (https://youtu.be/F4oST9VdaxI)
  - Join the development of EOSC (https://youtu.be/UYMXZWXMSYI)

    - The EOSC federated AAI: How can it help you. (https://youtu.be/71pv7zZRB3k)
- Guidelines and Recommendations for the Technical Integration of Resources and Services in the EOSC[27]: this document provides technical guidance and recommendations to support the integration of services with the so-called EOSC core services and the onboarding of services and resources into an EOSC catalogue, which is the primary tool to connect demand (researchers) and supply (resource providers) side.
- Maturity models: the methodology adopted to assess the maturity level of a service and its operational readiness in EOSC-Pillar uses a checklist template (adapted from the EOSC-Nordic template). It allows to check at what degree a service delivery is compliant to all (or to a category) the defined requirements to EOSC. This methodology, the checklist and a set of recommendations for future integration work are described in the EOSC-Pillar "Report on the validation statistics, operational infrastructure services and recommendations for future integration work" [23].
- Single Sign On services such as Indigo-IAM (as explained above) because of their important benefits to a platform, easily expanding the user base while removing some complexity in managing user account creation.
- The need of experts skilled in both scientific and technical domains linked to EOSC: Successful use of advanced services requires a solid technical expertise in the field of digital infrastructures for scientific computing. If the scientific team of the use case does not have these skills, it is important to have the possibility to be accompanied by an expert straddling the scientific and the technical domains. It would be interesting to include a service offering such expert support in the EOSC catalogue, and to be able to associate them with this type of high skill services, to facilitate the use of advanced tools.

Despite the end of the support due to the end of the EOSC-Pillar project, the basis are here: methods and tools, guidelines, recommendations and documentation. Stay tuned with EOSC-Future and Horizon Europe program to continue to receive support. Services are encouraged to watch the videos and read the documentation, assess their maturity, follow the guidelines, consider Single Sign On services and prepare their active integration in the EOSC.

[27] https://zenodo.org/record/5648215

5 References

No	Description/Link
[1]	Conference talk: Jérôme Pansanel, Vincent Negre, Ayoub Nachite, Marie Weiss. Déploiement de serveurs et de conteneurs GPU avec le service FG-Cloud. JCAD2021 https://jcad2021.sciencesconf.org/data/Conteneurs_GPU_Ayoub_NACHITE.pdf
[2]	Demonstration: Rémi Cailletaud, Yannis Govinda, Jérôme Pansanel. Automatiser ses déploiements applicatifs GPUs avec des conteneurs dans un environnement multi-cloud. JCAD2021 pdf – video – demonstration
[3]	Tangaro, M.A., Mandreoli, P., Chiara, M. et al. Laniakea@ReCaS: exploring the potential of customisable Galaxy on-demand instances as a cloud-based service. BMC Bioinformatics 22 (Suppl 15), 544 (2021). doi.org/10.1186/s12859-021-04401-3
[4]	Poster: Laniakea@ReCaS service, Galaxy Community Conference 2021. doi.org/10.7490/f1000research.1118635.1
[5]	Talk: the Laniakea software platform update were presented at the Galaxy Community Conference 2021. doi.org/10.7490/f1000research.1117073.1
[6]	poster: Gilles Mathieu, Yosra Sanaa. Une utilisation concrète d'EOSC : Cas d'usage en bioinformatique dans le cadre du projet EOSC-Pillar. JCAD 2021
[7]	Talk: Coline Royaux. How create Galaxy tools from R scripts and why ? ecoinfoFAIR2020: French national action on FAIR development in Ecology
[8]	Talk: Yvan Le-Bras . Galaxy-E : une instance de Galaxy dédiée à l'analyse de données en Ecologie. JCAD 2020 / Bioinformatics days for French Museum 2020
[9]	Talk: Coline Royaux. Galaxy-E : a Galaxy instance dedicated to data analysis in Ecology. CODATA / GO FAIR
[10]	Talk: Coline Royaux, Yvan Le-Bras. How to integrate an EBV workflow in Galaxy and why? Biomac ecology lab, Amsterdam
[11]	Talk: EBV workflow for developers" & "EBV workflows for users. French ecology and conservation lab
[12]	AA.VV. Demonstrators and success stories from the use cases. EOSC-Pillar D6.2. March 2022
[13]	Assante, M., Candela, L., Castelli, D., et al. (2019) Enacting open science by D4Science. Future Gener. Comput. Syst. 101: 555-563 10.1016/j.future.2019.05.063

[14]	Assante M, Candela L, Castelli D, et al. The gCube system: delivering virtual research environments as-a-service. Futur Gener Comput Syst. 2019;95:445-453. doi:10.1016/j.future.2018.10.035
[15]	Assante, M, Candela, L, Castelli, D, et al. Virtual research environments co-creation: The D4Science experience. Concurrency Computat Pract Exper. 2022;e6925. doi:10.1002/cpe.6925
[16]	Garcia, P.O., Berberi, L., Candela, L., Van Nieuwerburgh, I., Lazzeri, E., Czuray, M. (2022). Developing the EOSC-Pillar RDM Training and Support Catalogue. In: Silvello, et al. Linking Theory and Practice of Digital Libraries. TPDL 2022. Lecture Notes in Computer Science, vol 13541. Springer, Cham. https://doi.org/10.1007/978-3-031-16802-4_22
[17]	Goldbeck, Gerhard, Simperler, Alexandra, & Mogni, Gabriele. (2022). Digital Marketplaces and their value for the Materials Modelling Ecosystem (Version 1). Zenodo. https://doi.org/10.5281/zenodo.6329760
[18]	Conference Talk: "An ontological approach for classification of continuum and mesoscopic models". Huschle, Tobias at the Materials Science and Engineering (MSE) Congress 2022, dgm.de/mse/2022/.
[19]	Gaido, Luciano, Chiavarini, Beatrice, Pisa, Claudio, Galeazzi, Fulvio, & Verlato, Marco. (2021). EOSC-Pillar D7.1 - Guidelines and Recommendations for the Technical Integration of Resources and Services in the EOSC. Zenodo. https://doi.org/10.5281/zenodo.5648215
[20]	Indigo-IAM general documentation: https://indigo-iam.github.io/v/v1.8.0/docs/
[21]	Andrea Ceccanti, webinar for ESCAPE project: "Authentication and Authorization" https://drive.google.com/drive/folders/1iqWmTFKWM6jTG5qcNR4GlzgbhRcMwc1G
[22]	Conference article: "Beyond X.509: token-based authentication and authorization for HEP". Andrea Ceccanti, Enrico Vianello, Marco Caberletti and Francesco Giacomini. EPJ Web Conf., 214 (2019) 09002 DOI: https://doi.org/10.1051/epjconf/201921409002
[23]	EOSC-Pillar D7.3: Report on the validation statistics, operational infrastructure services and recommendations for future integration work

Appendix I. Covid-19 Use Case (CNR)

Objectives

EOSC-Pillar use case on COVID-19 was conceived to develop a solution for rapid screening of chemical compounds against the known COVID-19 targets leveraging the service offering of the project (in terms of services and computing capacity) and the Open Science principles. In this way a speed-up can be achieved in the search for new potential drugs when new pathogenic variants will be discovered in the future.

The envisaged solution was meant to be integrated into classic drug discovery pipelines that allow the screening of well characterized and approved drugs.

Implementation

In order to support the implementation of the use case a specific Virtual Research Environment was deployed and made available[28]. A screenshot of the working environment is reported below.

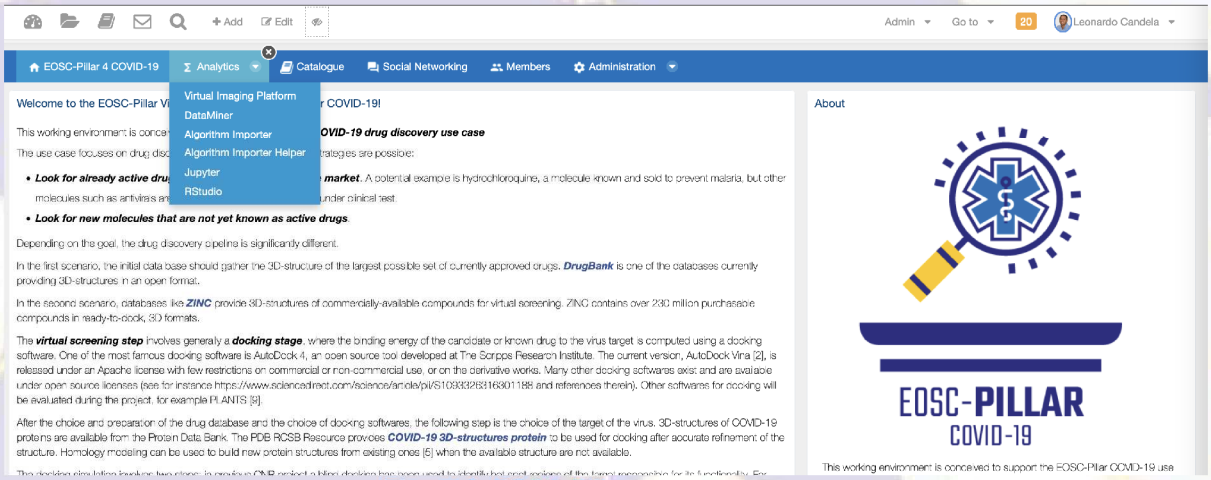


Figure 9. EOSC-Pillar VRE Screenshot

Among the functionality this environment offers, there is a rich portfolio of analytics solutions including: (i) DataMiner [13], a computing platform enabling its users to integrate analytics methods and execute them via a distributed computing infrastructure; (ii) JupyterLab, the platform for implementing and executing notebooks, and (iii) RStudio, the R-based development environment.

The software for Virtual Screening was designed and developed to perform thousands of docking simulations of a set of targets against a set of compounds and to assemble the single results to produce a final table of results integrating annotation of the poses. The pipeline is described in the following figure. Each computational step is represented as blue box (normal programs) or yellow box (parallel programs).

[28] EOSC-Pillar COVID-19 Virtual Research Environment https://eosc-pillar.d4science.org/web/eoscpillar_covid-19/home

Figure 10. EOSC-Pillar Screening Pipeline

The model of the receptors and a compound library are provided by users as inputs to the software. The receptor is a model of the target protein of interest that needs to be enabled or disabled (depending on the application and specific disease) with a potential drug. The compound library is a set of potential drugs. Both receptor and compound are represented in PDBQT format that is a well-known and useful way to represent molecules for docking application.

The first step is the computation of a set of druggable pockets of the receptor. In drug discovery, pocket is a term used to indicate a 3D region of the receptor that, due to specific atoms and bonds, is potentially able to bind a small molecule.

The next step is the docking simulation (based on Autodock Vina software) of each target pocket against each molecule of the compound library. The result of the docking, represented as a protein-ligand complex, is then processed (by using the Plip software) to extract the interaction profile and binding properties.

The last step gathers all results and builds a final report of all docking hits.

The complete version of the above pipeline was implemented as a standalone command line program in a docker image, ready to be used.

Three methods have been integrated into the DataMiner platform.

The Vinadocker method[29] (Figure 11) executes the first part of the pipeline (the docking part) and accepts as input parameter receptors and ligands taken from the VRE workspace.

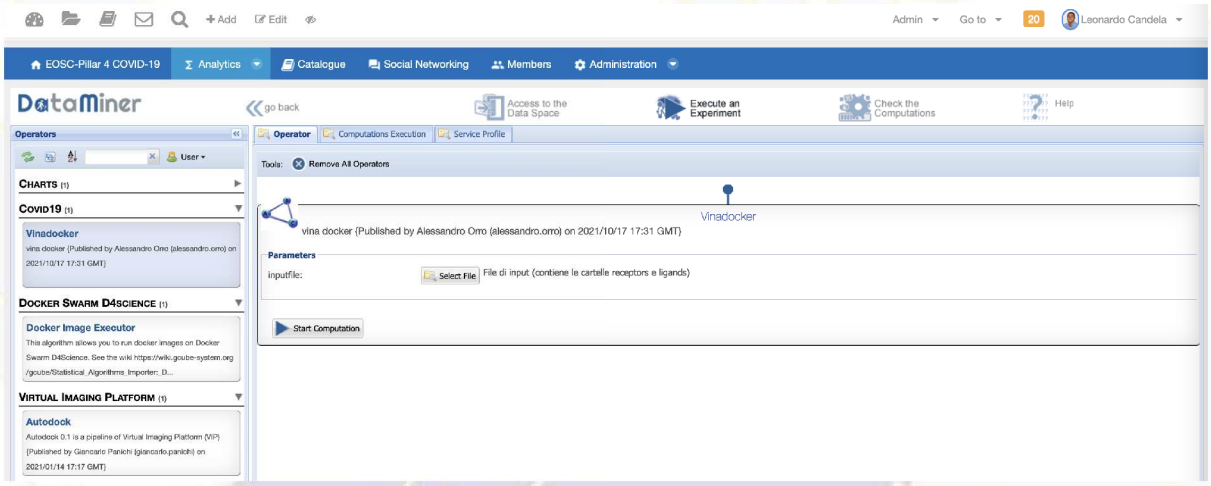


Figure 11. Vinadocker Method User Interface

The Docker Image Executor[30] (Figure 12) executes a Docker image with some input parameters from the workspace. This method was used to execute the entire pipeline above.

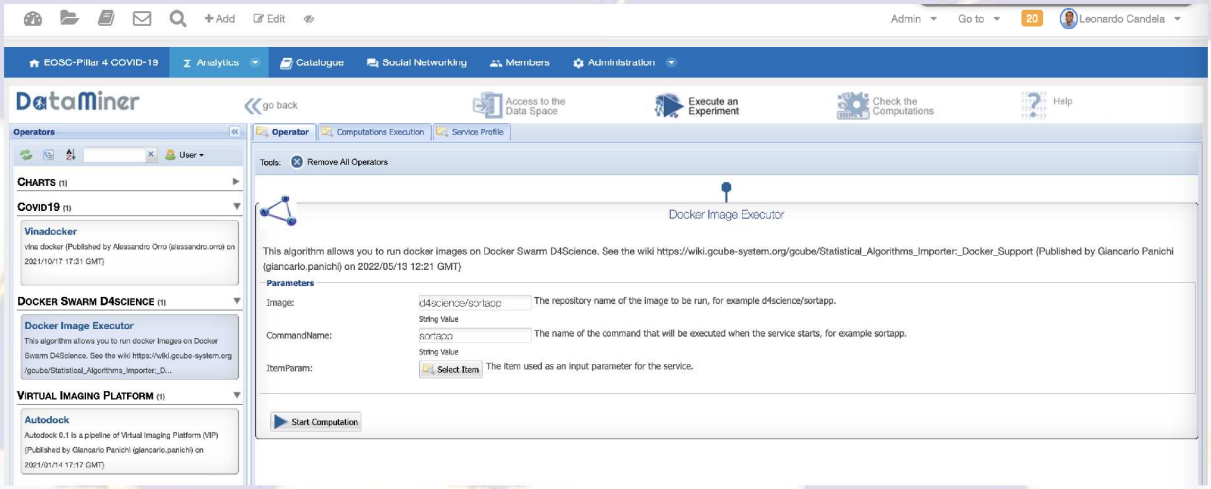


Figure 12. Docker Image Executor Method User Interface

The Autodock method[31] (Figure 13) executing the Autodock pipeline integrated into the Virtual Imaging Platform (Sec. 2.3.3). In particular, it accepts the ligands, the receptors and

[29] Vinadocker method https://eosc-pillar.d4science.org/group/eoscpillar_covid-19/dataminer?OperatorId=org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.transducerers.VINADOCKER

[30] Docker Image Executor method https://eosc-pillar.d4science.org/group/eoscpillar_covid-19/dataminer?OperatorId=org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.transducerers.DOCKER_IMAGE_EXECUTOR

[31] Autodock method https://eosc-pillar.d4science.org/group/eoscpillar_covid-19/dataminer?OperatorId=org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.transducerers.AUTODOCK

the grid map from the workspace plus other parameters (e.g. the API Key to invoke VIP and a couple of numbers tuning the execution).

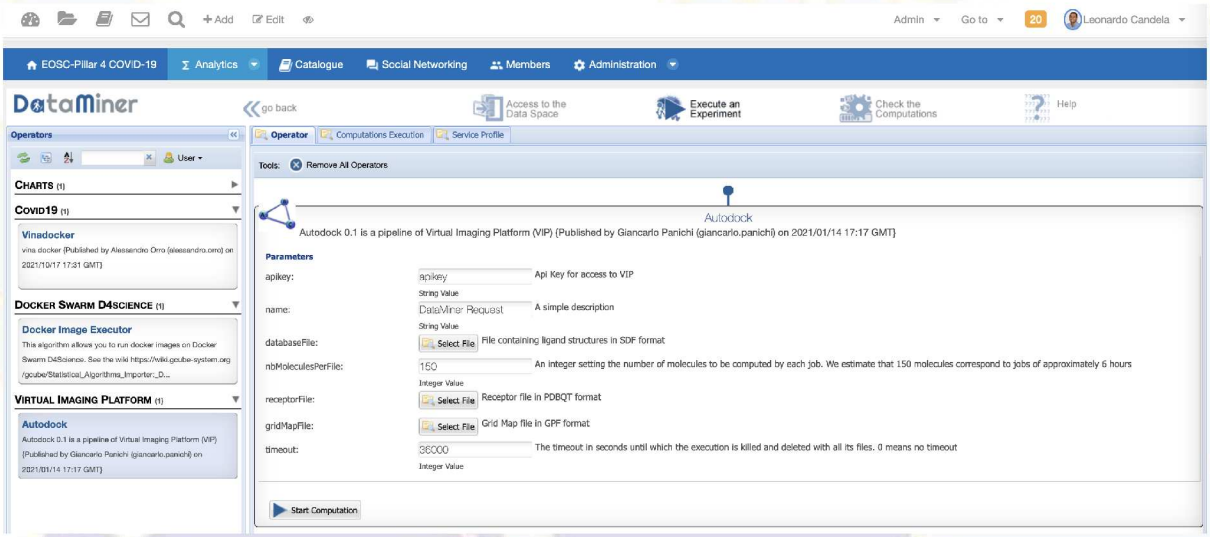


Figure 13. Autodock Method User Interface

Experiments

The pipeline was used to perform screening against the SARS-CoV-2 common targets, in particular the Spike Protein that is the main mechanism that the virus uses to infect target cells. Variants of Spike, deposited in the VRE workspace, are SARS-CoV-2 S-Delta variant B.1.617.2 (7V7N) and SARS-CoV-2 S Omicron Spike B.1.1.529 (7QO7).

The collaboration between CNR-ITB and University of Brescia (prof Arnaldo Caruso, Head of the Department of Diagnostic Laboratories at the “Spedali Civili” of Brescia) established also thanks to this EOSC-Pillar activity will lead to the set-up of new experiments involving novel Covid19 variants found during clinical activities.

The report of docking results against the two spike models has been inserted into the repository. The ligand dataset is a selection of DrugBank and Zinc library and contains 1600 molecules. The top molecules from the energetic ranking of docking simulation have energy < -8.8 kcal/mole (low tail of energy distribution plot, see Figure 14 below).

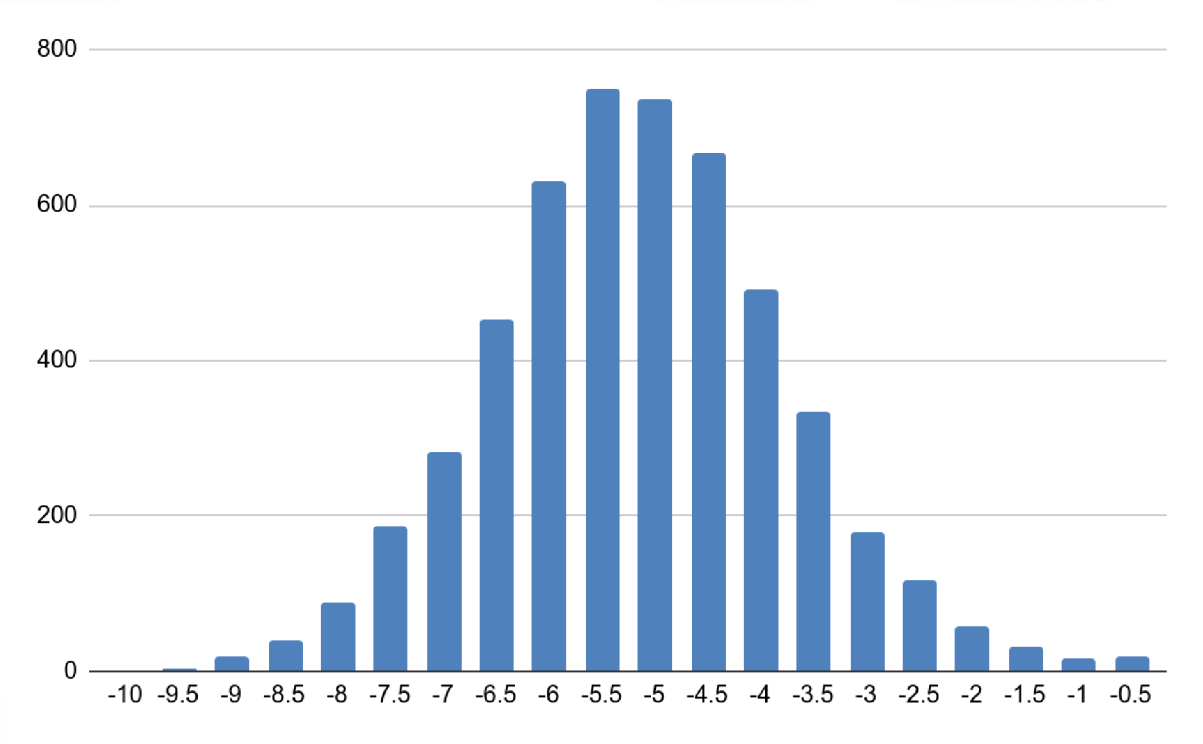


Figure 14: Histogram distribution of biding affinity

About 13000 simulations have been performed including both docking and binding site analyses. In the range affinity < -7.5 kcal/mole (the left tail computed as average minus standard deviation), 183 hits (poses) has been found that correspond to 153 molecules.

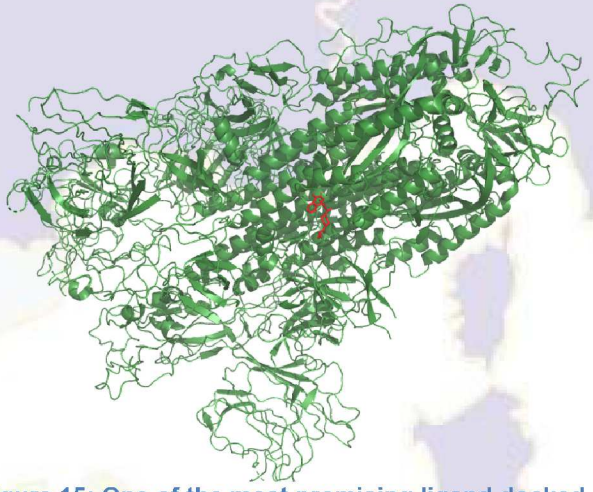


Figure 15: One of the most promising ligand docked in the Spike poket

These drug candidates have been found with the proposed virtual screening pipeline and are currently subjected to validation process by computational approaches (Molecular

Dynamics, binding affinity analysis, structural stability) at the CNR-ITB Bioinformatics Laboratory. Experimental validation by bioassay will be performed by UniBS.

Moreover two other important collaborations have been activated: The Computer Science Laboratory of Cambridge University (Lab of prof Pietro Liò) provided expertise in Artificial Intelligence applied to docking ligand optimization and de novo drug design; the Department of Pharmacy of University of Genova (prof.ssa Paola Fossa) has been involved for the ADME Tox analysis and ligand optimization by chemical synthesis.

In the future, these hits will be subjected to following analysis:

1. Molecular Dynamics to evaluate the complex stability and energy
2. Kinetic analysis of the complex to determinate the association (k_{on}) and dissociation (k_{off}) rates
3. Standard assays will be carried out to measure the effects of each compound on cell cytotoxicity and antiviral activity by CC50

From the characterization of the resulting promising hits, we will perform:

1. Ligand optimization and Denovo Drug Design by Artificial Intelligence methods
2. ADME tox analysis
3. Synthesis of new compounds

University of Brescia is involved in intense search and sequencing of new variants in their clinical setup that in future will be subjected to the same process that involves the computational analysis with the described EOSC-Pillar Pipeline and the following experimental validation. The computational pipeline will be further developed and improved in order to adapt to virus evolution (evaluating the sequence variability) and to new biological and medical knowledge.