

## Contributing to uptake of AI in research: advanced AI for scientists - the AI4EOSC platform

*Marcin Plociennik (PSNC - impact and collaboration), Álvaro López García (IFCA-CSIC - project coordinator)*

AI4EOSC  
Documentation of the  
services and software



Funded by  
the European Union

# AI4 | eosc Artificial Intelligence for the EOSC

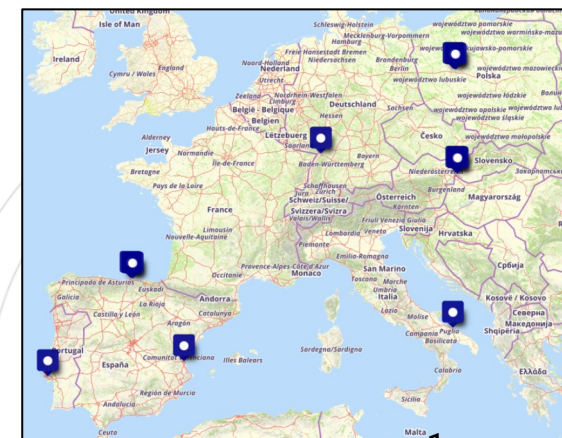
- Evolution of the DEEP Hybrid DataCloud platform
- Runs September 1st 2022 – August 2025 (36 months)

Overall Goal: Deliver an **EOSC AI platform with advanced features** for distributed, federated, composite learning, metadata provenance, MLOps, event-driven data processing, and provision of AI/ML/DL services

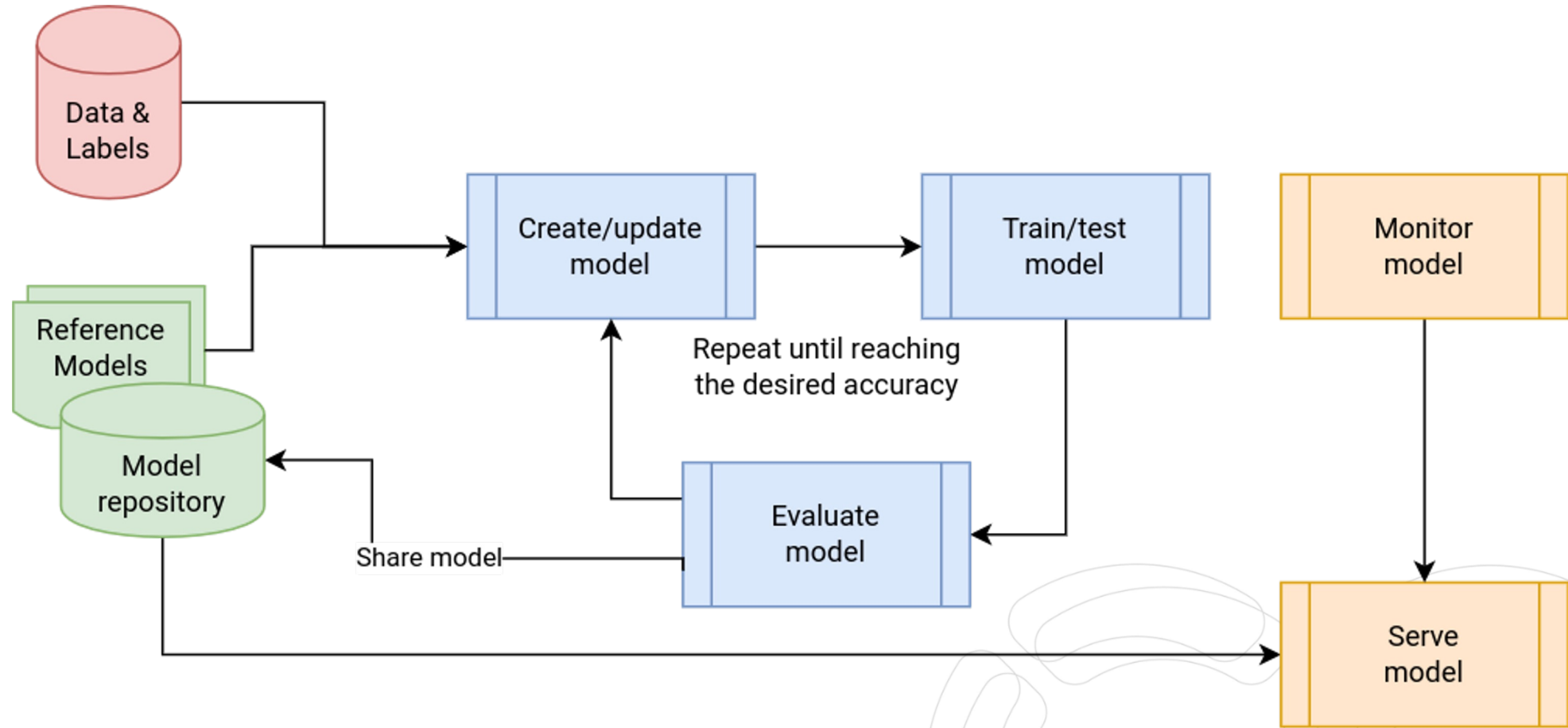
- Funding: 5M€
- 3 workshops on AI, image processing, federated learning
- 1 external users open call
- 8 peer reviewed publications in high impact journals
- 2 peer reviewed publications in high impact conferences
- Collaboration with several EU funded and INFRAEOSC



1. Feature rich **services and platform** to build and deploy **custom AI applications in the EOSC**
2. Support for **building AI systems on distributed datasets**, with a particular focus on **federated learning**
3. Services to **compose AI tool workflows**, enabling the development of complex data-driven AI applications
4. **AI Exchange Hub** in the context of the EOSC, enhancing and increasing the application offer currently available
5. **Extend the service offer and the capabilities** being offered through the **EOSC**, with focus on AI

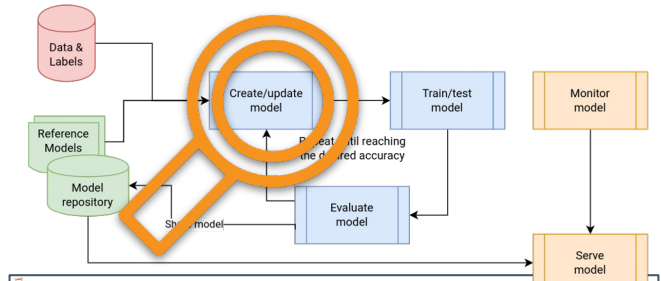


# The Machine Learning Lifecycle



# AI4 | eosc Services for AI/ML development

## Dev tools: sandbox and online IDE



AI4 | eosc Alvaro Lopez Garcia

Deployments

Name	Status	Container name	GPUs	Creation time	Actions
test-vscode	Running	deephub/dep-oc-generic-dev-latest	0	2023-06-20 09:14:29	
test-jupyter	Running	deephub/dep-oc-plants-classification-latest	0	2023-06-20 09:10:18	

[New deployment](#)

AI4 | eosc Alvaro Lopez Garcia

Module

Modules / Plants species classifier

### Plants species classifier

Classify plant images among 10K species from the iNaturalist dataset.

Build status: build passing License: Apache 2.0 Created: 2019-01-01

The deep learning revolution has brought significant advances in a number of fields [1], primarily linked to image and speech recognition. The standardization of image classification tasks like the [ImageNet Large Scale Visual Recognition Challenge](#) [2] has resulted in a reliable way to compare top performing architectures.

The use of deep learning for plant classification is not novel [3, 4] but has mainly focused in leaves and has been restricted to a limited amount of species, therefore making it of limited use for large-scale biodiversity monitoring purposes.

This Docker container contains a trained Convolutional Neural network optimized for plant identification using images. The architecture used is an Xception [5] network using Keras on top of Tensorflow. A detailed article about this network and the results obtained with it can be found in [6].

The PREDICT method expects an RGB image as input (or the url of an RGB image) and will return a JSON with the top 5 predictions.

The original training dataset was the great collection of images which are available in [PlantNet](#) under a Creative-Common AttributionShareAlike 2.0 license. It consists of around 250K images belonging to more than 6K plant species of Western Europe. These species are distributed in 1500 genera and 200 families.

A new iteration of the application has been trained using plant images from [iNaturalist](#). This dataset has around 4.4M observations with 7M images from 58K worldwide species. We have restricted our training to the 10K most popular species.

This service is based in the [Image Classification with Tensorflow](#) model.

References

- [1]: Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. [Deep learning](#). Nature, 521(7553):436-444, May 2015.
- [2]: Olga Russakovsky et al. [ImageNet Large Scale Visual Recognition Challenge](#). International Journal of Computer Vision (IJCV), 115(3):211-252, 2015.
- [3]: Sue Han Lee, Chee Seng Chan, Paul Wilkin, and Paolo Remagnino. [Deep-plant: Plant identification with convolutional neural networks](#), 2015.
- [4]: Mads Dyrmann, Henrik Karstoft, and Henrik Skov Mithiby. [Plant species classification using deep convolutional neural network](#). Biosystems Engineering, 151:72-80, 2016.
- [5]: Chollet, Francois. [Xception: Deep learning with depthwise separable convolutions](#) arXiv preprint (2017): 1610-02357.
- [6]: Heredia, Ignacio. [Large-scale plant classification with deep neural networks](#). Proceedings of the Computing Frontiers Conference. ACM, 2017.

Categories: tensorflow, docker, deep learning, tramable, inference, pre-trained, image classification, api v2

Additional Resources: Get the code (GitHub, Dockerhub), Get the data (Dataset)

[Train module](#)

code-server

Editing evolved

Next Up: Deploy code-server for your team

Recent: /

Walkthroughs: Get Started with..., Discover the best customizations to make VS Code for the Web pros...

Computing the predictions

It's time to see how your net performs on unseen TEST data. We will divide the tutorial in 3 parts depending on whether you want to:

- Predict leafpaths contained in a dataset split.txt file
- Predict local images
- Predict all images

Note: Parameters in UPPERCASE letters must be changed by the user if needed.

```

import os
import json
import numpy as np
import matplotlib.pyplot as plt

from tensorflow.keras.models import load_model
from imgclas_data_utils import load_image, load_data_splits, load_class_names
from imgclas_test_utils import predict
from imgclas import paths, plot_utils, utils

# User parameters to set
TIMESTAMP = '2023-11-08 19:34:59' # timestamp of the model
MODEL_NAME = 'final_model.h5' # model to use to make the prediction
TOP_K = 5 # number of top classes predictions to save

# Set the timestamp
paths.timestamp = TIMESTAMP

# Load the data
class_names = load_class_names(paths.get_ts_splits_dir())

# Load training configuration
conf_path = os.path.join(paths.get_conf_dir(), 'conf.json')
with open(conf_path) as f:
    conf = json.load(f)

# Load the model
model = load_model(os.path.join(paths.get_checkpoints_dir(), MODEL_NAME), custom_objects=utils.get_custom_objects)

# Predicting a datasplit.txt file
Here we are going to compute the predictions and save them into a .json file to retrieve them later.

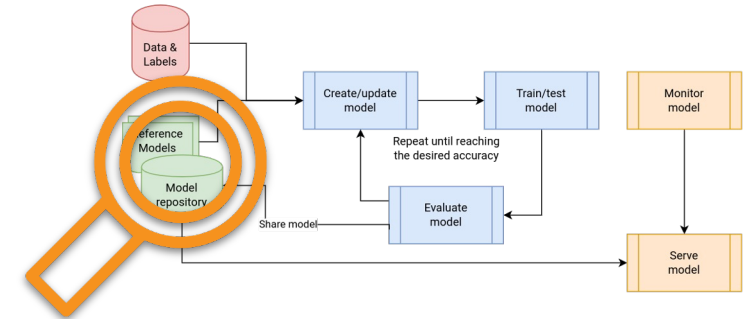
SPLIT_NAME = 'test_R0B' # data split to use
conf['general']['images_directory'] = '~/media/ignacio/datos/datasets/sonil/cas/datasets/R0B' # custom the
# Load the data
X, y = load_data_splits(paths.get_ts_splits_dir(),
                        in_dir=conf['general']['images_directory'],
                        split_name=SPLIT_NAME)

# Predict
pred_prob = predict(model, X, conf, top_k=TOP_K, filemode='local')

# Save the predictions
pred_dict = {} # filenames: list(),
              pred_lab: pred_lab.tolist(),
              pred_prob: pred_prob.tolist()
if y is not None:
    pred_dict['true_lab'] = y.tolist()
    
```



## Model marketplace and dashboard



<https://dashboard.cloud.ai4eosc.eu/>

[https://open-science-cloud.ec.europa.eu/resources/services/eosc.ifca-csic.ai4eosc\\_platform](https://open-science-cloud.ec.europa.eu/resources/services/eosc.ifca-csic.ai4eosc_platform)



# AI4 | eosc Services for AI/ML development

## Deployment: standalone service

AI4 | eosc  
Dashboard  
Marketplace  
Deployments  
Useful links  
Identity and Access  
AI4EOOSC documentation  
Project page  
Storage  
Status  
Experiment tracking

Module: OBSEA Fish Detection

Build status: Build Status  
License: MIT  
Created: 2023-11-29

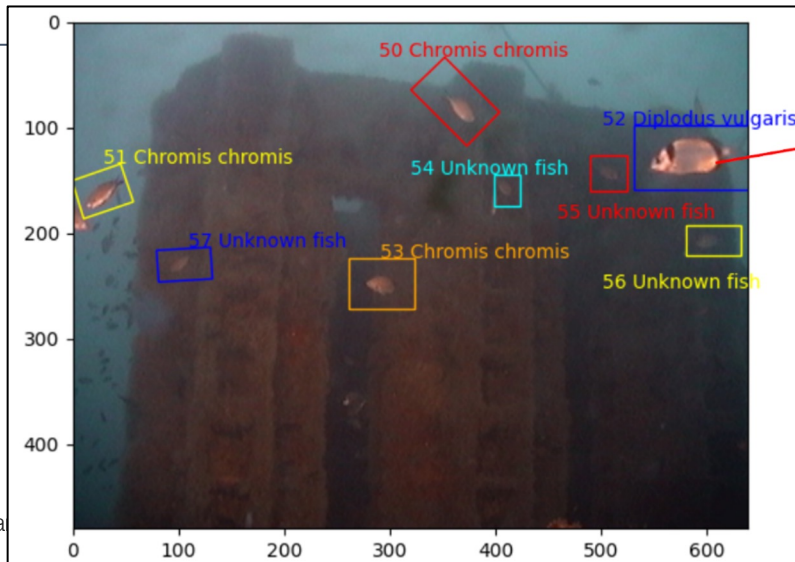
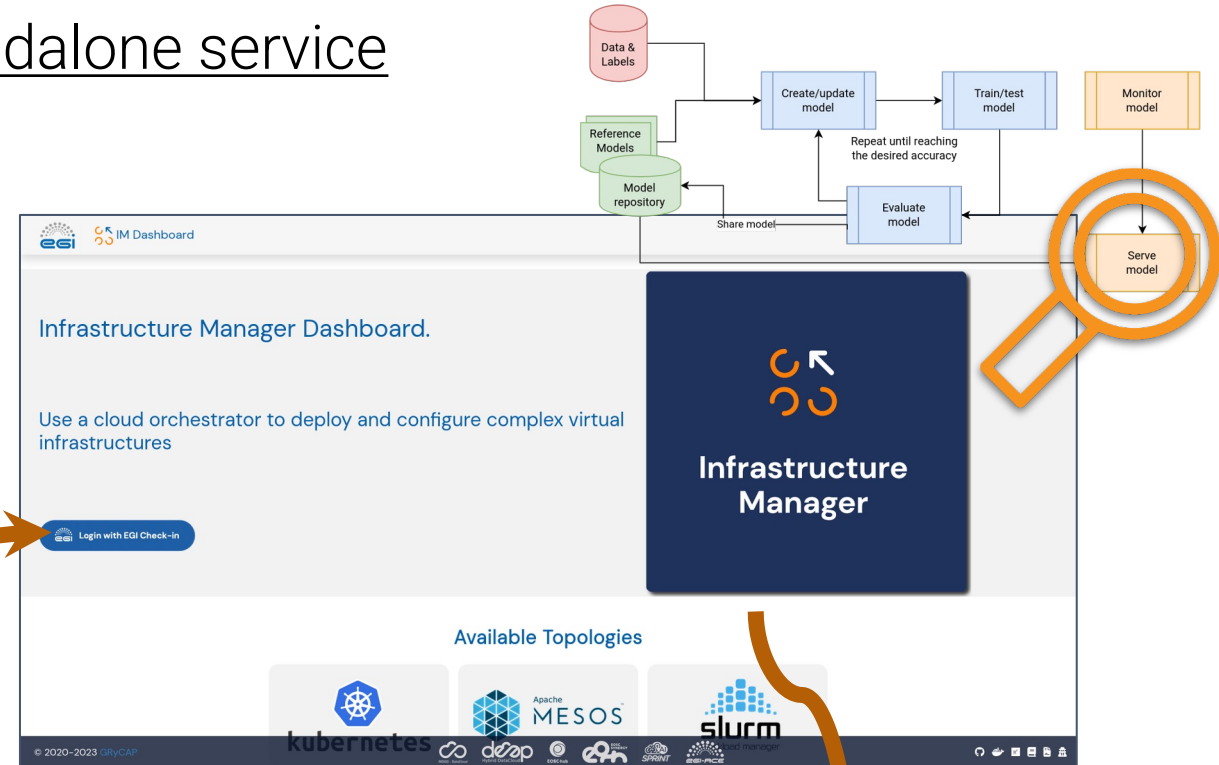
AI-based fish detection and classification algorithm based on YOLOv8. The model has been finetuned to detect and classify fish at the OBSEA underwater observatory.

obsea\_fish\_detection is an application using the DEEPaaS API.  
Write additional information for your users (how to predict, how to retrain, dataset description, training description, etc).

Categories: docker, apisv2, yolov8, vo.imagine-ai.eu

Additional Resources: Get the code (Github, Dockerhub), Get the data (Dataset)

Deploy via the IM



Swagger /swagger.json

### DEEP as a Service API endpoint 1.3.0

Hybrid DataCloud

This is a REST API that is focused on providing access to machine learning models. By using the DEEPaaS API users can easily run a REST API in front of their model, thus accessing its functionality via HTTP calls.

Currently you are browsing the Swagger UI for this API, a tool that allows you to visualize and interact with the API and the underlying model.

- Project website
- Project documentation
- Model marketplace

API documentation

versions

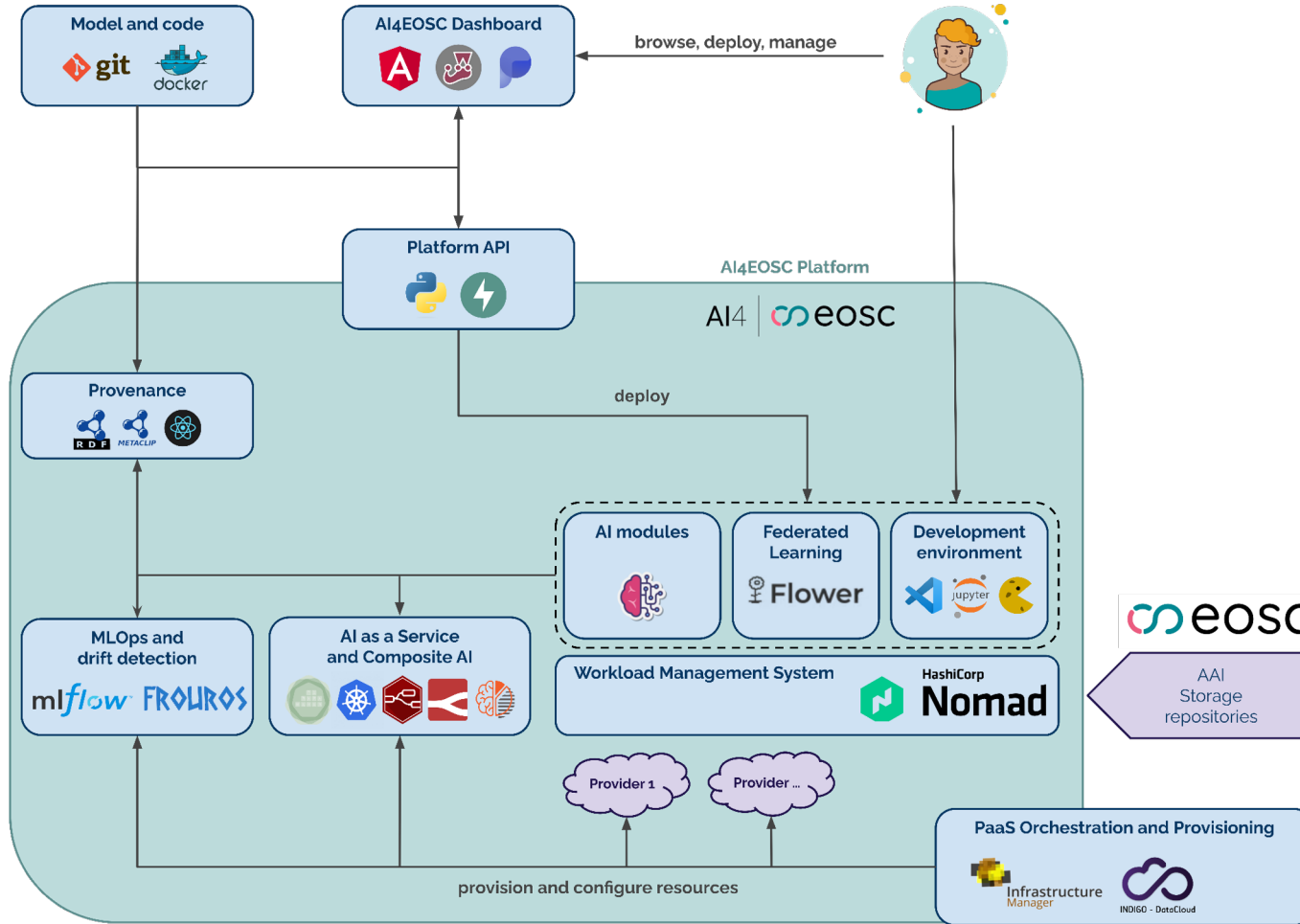
- GET /v2/ Get V2 API version information
- GET / Get available API versions

debug

- GET /v2/debug/ Return debug information if enabled by API.

models

- GET /v2/models/ Return loaded models and its information
- GET /v2/models/obj\_detect\_pytorch/ Return model's metadata



Interactive C4 diagrams available [here](#).



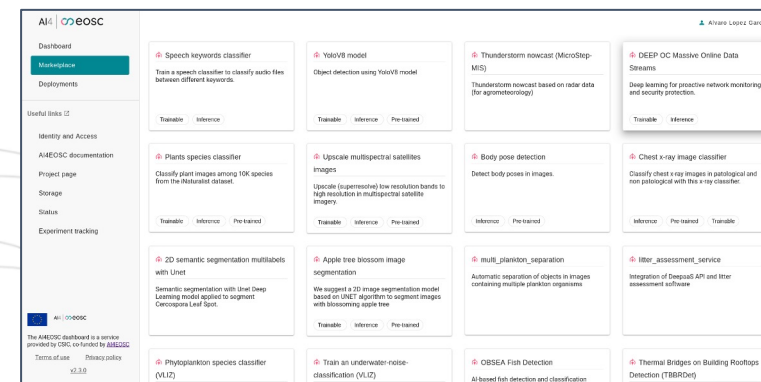
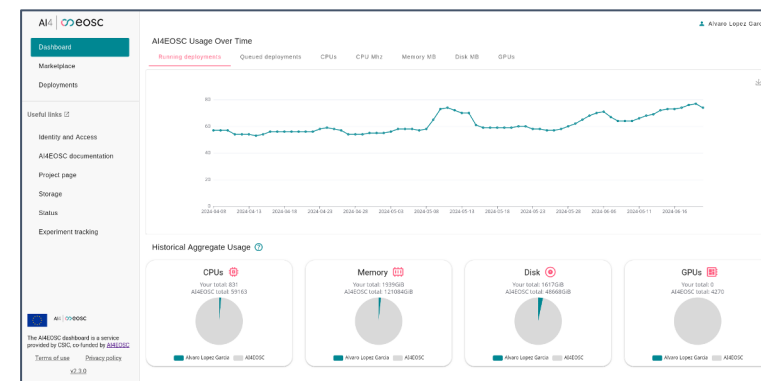
*AI4EOSC empowers scientific research by...*

- Providing users with advanced AI tools:
  - **Model retraining** (iterative learning, fine tuning)
  - **Federated learning** (including client authentication)
  - **Parallel training in multiple GPUs** (distributed training - data parallelism)
  - **Model monitoring**: MLOps, drift detection
  - **Model inference**
- Providing a simple and **intuitive IDE for developing AI/ML/DL models** (VSCode or JupyterLab)
- Allowing **seamless and transparent access to computational resources** to accelerate model development
- **Deploying your models** in production in a **serverless** environment

**More info:**

<https://ai4eosc.eu/>  
<https://docs.ai4os.eu/>

- Increased number of services in the EOSC Exchange
  - AI/ML generic (AI4EOSC platform) or community specific (e.g. iImagine <https://www.imagine-ai.eu/> and <https://dashboard.cloud.imagine-ai.eu/>)
- Transforming development of AI models for science in the EOSC
  - FAIR, reproducibility of pipelines, model provenance
  - Best practices for development AI and ML models
- Improvement of robustness of AI systems → Trustworthy AI
  - MLOps infrastructure and services for AI scientists
  - Drift detection tools and services to assess data/model/inference validity
  - Provenance of models (reproducibility), model metadata
  - FAIR-ification of ML assets
- (Main, relevant) Outcomes so far
  - AI platform for the EOSC (AI4EOSC platform)
    - Integrated popular IDE environments (JupyterLab, Visual Studio Code)
    - Advanced ML features: federated learning, homomorphic encryption, privacy tools
    - Composite AI, cross discipline pipelines, serverless platforms
  - Software stack (AI4OS) to build customized AI platforms
  - Best-practices documents for the AI community of practice in the EOSC





-> *towards strategy/pathways for the integration, and maintenance, of services developed by projects within EOSC*

... *usual issue* → *declared TRL vs real TRL*

- Need for quality standards
- External, independent validation
- Very good documentation from the day one is essential: for the providers, developers, users + trainings on top

*Preparation phase: towards achieving the adequate TRL*

- Provide the high quality software and services → procedures and services for delivering it in place (SQaaS, )
- Validation and testing by the use cases within the project
- Integration with other initiatives and frameworks (e.g. Flower.AI, NVIDIA MONAI )
- Exploitation strategy → technical liaison with projects (EOSC ecosystem, AI-ondemand ecosystem, RI, etc), scientific community, private, innovator sector (through EOSC DIH)

*Validation phase*

- Call for external use cases to validate the services
- Other initiatives to reuse the software stack (AI4OS) → provide feedback or contribution to enhance

*Exploitation phase*

- Usage of the platform within the Scientific Communities, innovators, other projects
- Different models of collaboration (depending on business model) → need for infrastructure and resources
- AI community of practice for the EOSC

- Define clear methodology and processes for integration
- Be pragmatic: architecture must be presented in an easy to understand, clearly defined and easy to update documents and notations. Use community standards (e.g. C4 model)
- Adequate quality of the software and services should be provided: similar standards should be followed
- Define a common standard to document things in order to have a clear understanding of the federation and integration paths e.g.:
  - OpenAPI specification,
  - C4 model,
- Foster reusability, promote visibility of results



- AI4EOSC is interested in **exploring the exploitation of EOSC Node compute capacity** as underlying infrastructure to deliver the advanced AI platform
- AI4EOSC is ready to implement integration with EOSC Node core capabilities (AAI, catalogs, accounting, monitoring)
- AI4EOSC vision on contributions to EOSC core:
  - Comprehensive provenance tracking system for ML assets (planned Q4 of 2024)
    - Service oriented, standards based, extensive for other scientific products
  - Proposing common standards and approaches:
    - Common an easily understandable model to define the architecture (e.g. C4Model)
    - Open API specifications and interoperability guidelines
- Discussion is needed to understand which of the other general AI related services and capabilities could be included in the EOSC Core

- Metadata and ontologies
  - **Provenance tracking system for AI:** Service oriented, semantic standards and domain specific ontologies
- FAIR Metrics
  - RDA FAIR Data Maturity, **extension of FAIR EVA** to support ML model provenance
- User environments
  - **Enhanced AI platform for the EOSC** (powered by the AI4OS software stack) providing EU researchers and data scientists with a comprehensive set of tools to share, develop and deploy AI models following the open science and FAIR principles
  - **Integrations with EOSC repositories** for automatic downloading of datasets (Zenodo, Figshare, etc.)
  - **Automatic deposit of AI assets** (Zenodo + GitHub)

- User provider environments
  - **AI4OS software stack:** The AI4OS software stack allows providers to build customized platforms (similar to the AI4EOSC platform) that can be tailor-made for the specific needs of a community
  - **Service mesh approach as a model for EOSC**, allowing more rich distributed platforms
- EOSC interoperability
  - **Guidelines for AI for EOSC Interoperability Framework**
  - **Open Architecture definition** following common notation model (C4 model)
- Skills development
  - **Training materials** and courses on AI for EOSC
  - Collaboration with EOSC sister projects (INFRAEOSC) on **AI and EOSC skills development**



- Finish and settle collaborations with INFRAEOSC (and other) projects



- AI + Open Science + FAIR data intersection


- **Uptake of AI for science** and research with a solid foundations and scientific background
- Start taking into account **quality aspects of the data** (QAFAIRD)

- Increased trust in AI-based research

- **Reputation of AI systems**
- **FaccT for AI systems:** Fairness, Accountability, and Transparency

- AI must be seen as an interdisciplinary collaboration

- Data experts, domain scientists, data scientists, research software engineers must work together
- **AI community of practice** for the EOSC

AI4 |  eosc Thank you for your attention!

## The AI4EOSC platform

Marcin Plociennik

[marcinp@man.poznan.pl](mailto:marcinp@man.poznan.pl)

(impact and dissemination)

Álvaro López García

[aloga@ifca.unican.es](mailto:aloga@ifca.unican.es)

(project coordinator)



<https://dashboard.cloud.ai4eosc.eu>

u



Funded by  
the European Union

AI4EOSC | EOSC projects meeting (2024/06/20)

---

eosc | cancer

# **EOSC4Cancer** contributions to session #5.2:

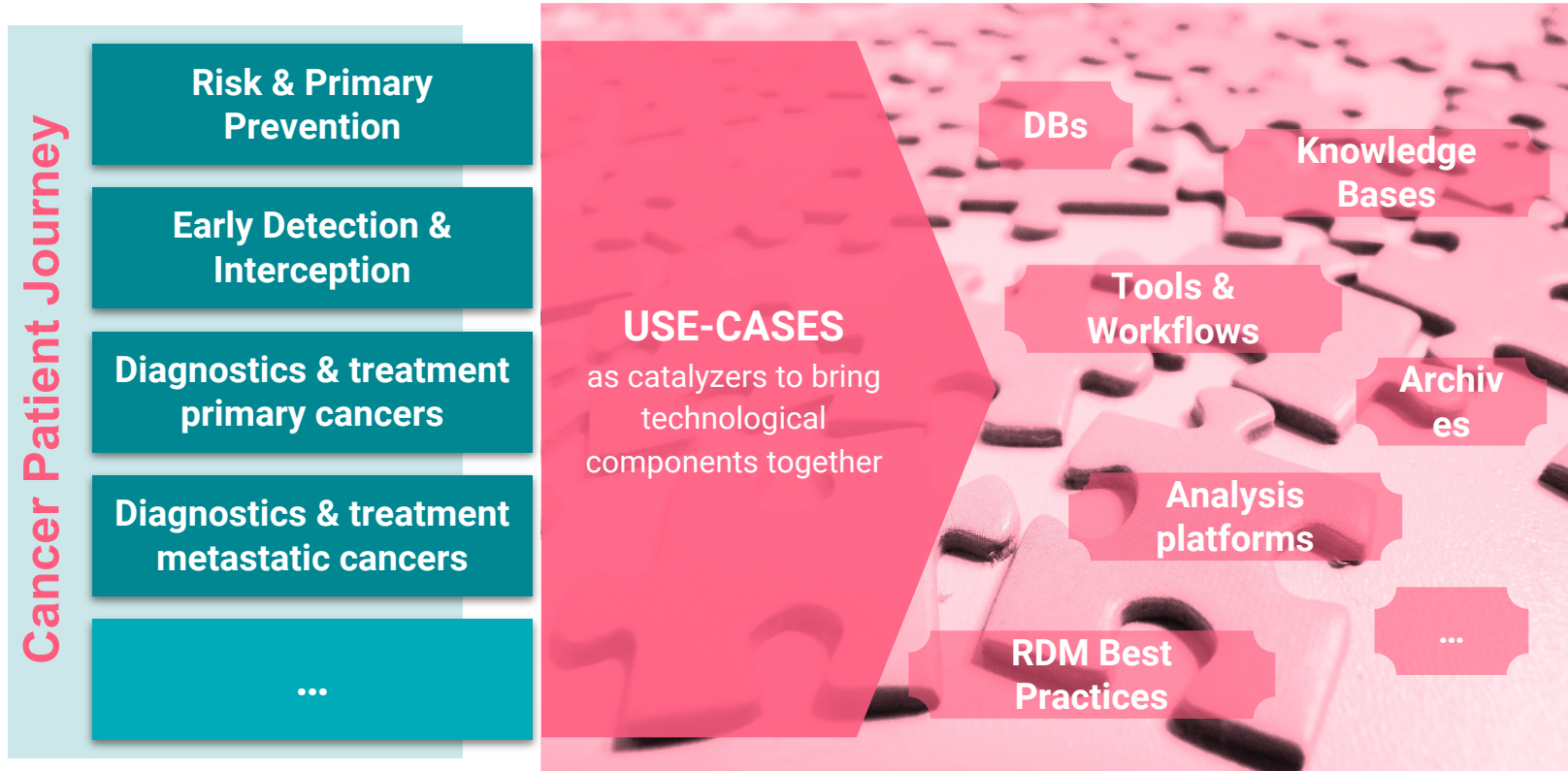
*Adding value through the EOSC Federation: Users and Resource environments in thematic communities*

**2024 Coordination meeting of EOSC-related projects funded under Horizon Europe**

June 20th, 2024

Salvador Capella-Gutierrez (BSC) · Veli Stroetmann (Empirica)

# EOSC4Cancer: the cancer research point of view



# EOSC4Cancer: the cancer research point of view

## Colorectal Cancer use-cases

Cancer risk identification and prevention by linking environmental data to **cancer registry** data

Data-driven optimisation of **cancer screening** programs

Data-driven treatment selection for localised tumours with **multiple patient-derived data types**

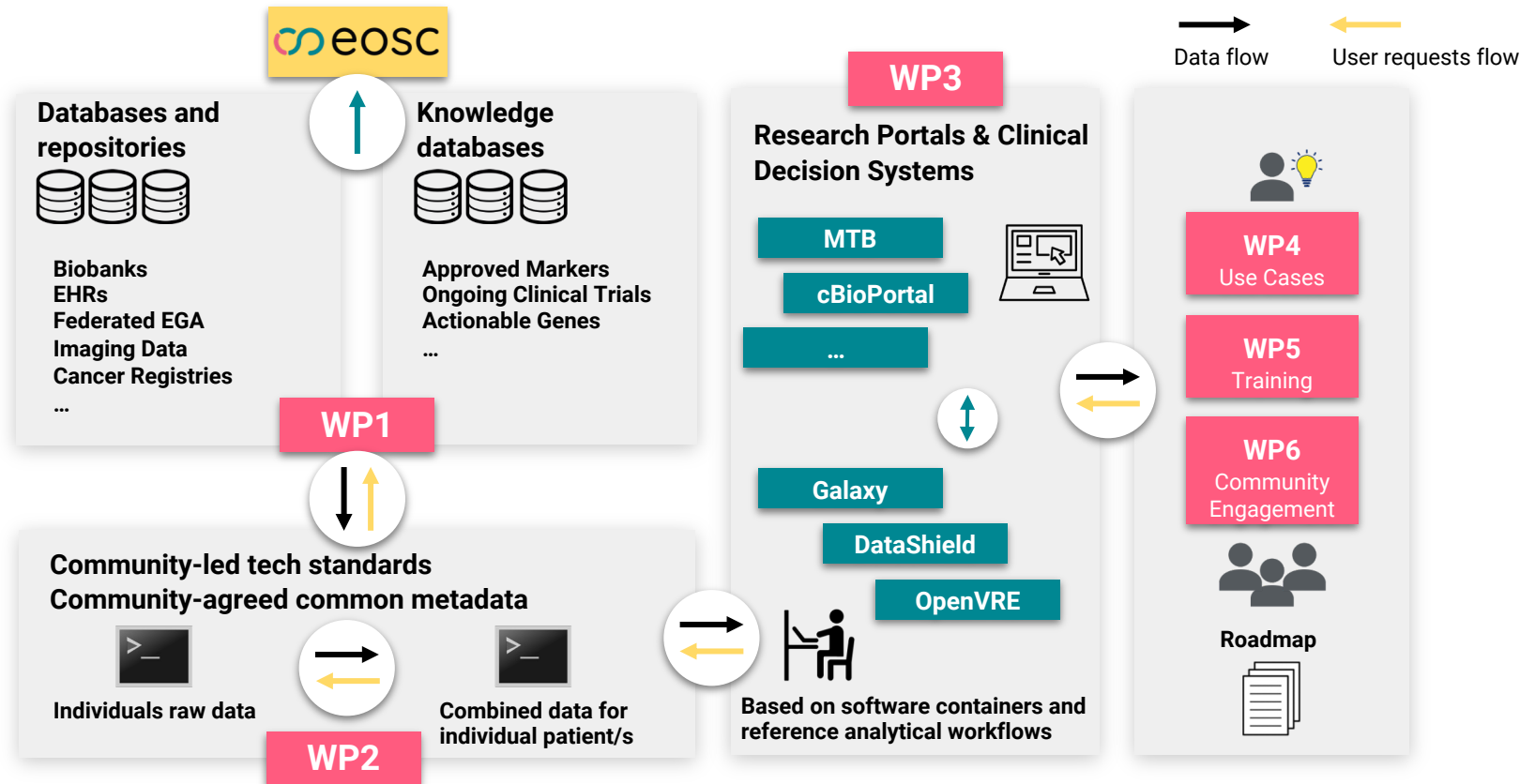
Data-driven treatment selection for localised tumour: improving the treatment of colorectal cancer by the inclusion of **circulating DNA information**

Connecting omics data from multiple sources to a **Clinical Decision Support System** (CDSS) for precision treatment of metastatic CRC

Provide **TECHNOLOGICAL SOLUTIONS** + identify **GAPS** and **EMERGING NEEDS** to drive further developments



# Technical overview



## KEY OUTCOMES ACHIEVED (first half)

### → Towards EU cancer research catalogue

- EOSC4Cancer Cancer Catalog, providing standardised descriptions of cohorts and other data resources, and the collaboration networks of experts surrounding them.

### → Synthetic data

- First synthetic dataset made available (longitudinal genomic + clinical data)

### → Integration of imaging data

- Started deployment of cBioPortal at BBMRI with slide image integration
- Work on XNAT and cBioPortal for image integration

### → Reference cBioPortal instance hosted by Health-RI

### → Development of a model for biomarker-driven clinical trials

### → Initiated the development of an AI-powered NLP tool for Cancer Patient-to-Precision Clinical Trials Matching

### → Establishment of a European training support network for Cancer Data Research

### → Initial framework for the Cancer Roadmap defined

### → Engagement with KEY stakeholders

## KEY OUTCOMES ACHIEVED (first half)

### → Towards EU cancer research catalogue

- EOSC4Cancer Cancer Catalog, providing standardised descriptions of cohorts and other data resources, and the collaboration networks of experts surrounding them.

### → Synthetic data

- First synthetic dataset made available (longitudinal

### → Inter

- S

→ Development of a model for **biomarker-driven clinical trials**

→ Initiated the development of an **AI-powered NLP tool** for Cancer Patient-to-Precision Clinical Trials Matching

→ Establishment of a **European training support**

## UPCOMING OUTCOMES

→ **Demonstrator of data mobilization** across analysis platforms (from Galaxy to cBioPortal)

→ **SOPs for the exchange of different data types** (Genomics, Clinical data, Images, Cancer registries, Screening programs, Exposome data)

→ **Harmonization models, procedures, and recommendations**

→ Demonstrator to trigger **Galaxy workflows from within cBioPortal**

→ **RDMkit Cancer View** and **EOSC4Cancer course portfolio**

→ **Demonstrator of federated data flow**

# Engagement with KEY stakeholders



Data Generation



Standards, Services,  
Software & Portals



Deployment of  
Infrastructure



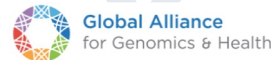
Demonstration  
through use-cases



**canSERV**  
providing cutting edge  
cancer research services  
across europe



European  
Genomic Data  
Infrastructure



---

eosc | cancer

**How can EOSC help us demonstrate and promote the benefits of open science and related collaborative practices for thematic users and resource environments?**



Community-led tech standards  
Community-agreed common metadata



Individuals raw data



Combined data for  
individual patient/s

WP2

# Examples: Towards EU cancer research catalogue

<http://data-catalogue.molgenisccloud.org/catalogue/ssr-catalogue/EOSC4Cancer>

## Metadata included:

Resource summary, Population, collection events, linkage, access conditions, mappings to standards, links to other 'networks', data access conditions (DUO)

## Enhancements made:




Cancer landing page







Cross linking to other EU programs (e.g. exposome)

## Standards:

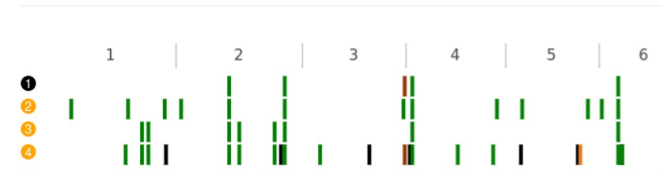
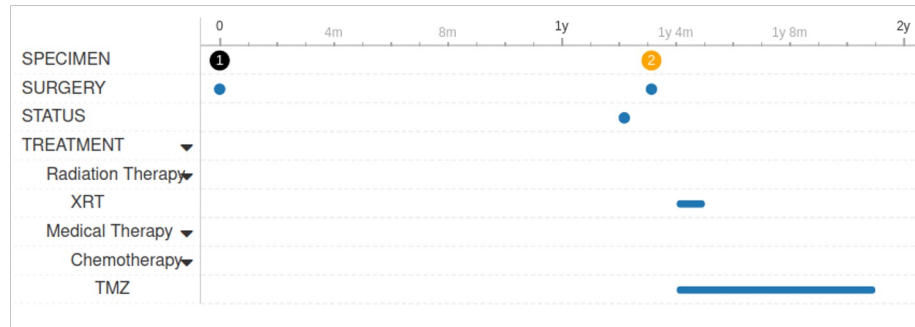
DCAT, FDP, DUO, Beacon, RDF, GraphQL, CSV, OI DC/LSAAI

The image displays three overlapping screenshots of the EOSC cancer research catalogue interface. The top screenshot shows the 'EUROPEAN HEALTH RESEARCH DATA AND SAMPLE CATALOGUE' landing page, which includes a table of thematic catalogues. The middle screenshot shows a search results page for 'MYFIT STUDY' with various filters. The bottom screenshot shows a search results page for 'EUChild' with various filters.

THEMATIC CATALOGUES		
Catalogues focused on a particular theme, developed by a collaboration of projects, networks and/or organisations:		
 EHCN	European Human Exposome Network	<a href="#">→</a>
 EOSC4Cancer	European Open Science Cloud for Cancer	<a href="#">→</a>
 EUChild	The EU Child Cohort Network	<a href="#">→</a>

PROJECT CATALOGUES		
Catalogues maintained by individual research projects or consortia, such as EC-RIA:		
 ATHLETE	ATHLETE	Advancing Tools for Human Early Lifecourse Exposome Research and Translation <a href="#">→</a>
 EXPANSE	EXPANSE	EXPANSE <a href="#">→</a>
 FORCE	FORCE	Infrastructure for Rare Cancers in the Netherlands <a href="#">→</a>
 HEAP	HEAP	Human exposome assessment platform <a href="#">→</a>
 Immune-Image	Immune-Image	Immune-Image: Specific Imaging of Immune Cell Dynamics Using Novel Tracer Strategies <a href="#">→</a>
 IPEC	IPEC	International Psychosis Epidemiology Consortium <a href="#">→</a>

## Longitudinal Data



80 Mutations (page 1 of 8)

Samples	Gene	Protein Change
1 2 3 4	IDH1	R132C
1 2 3 4	TP53	R248Q

- Inclusion of additional data views (longitudinal data)
- Inclusion of additional data types (radiological imaging visualization).
- Development of guidelines to facilitate interoperability across different cBioPortal instances.

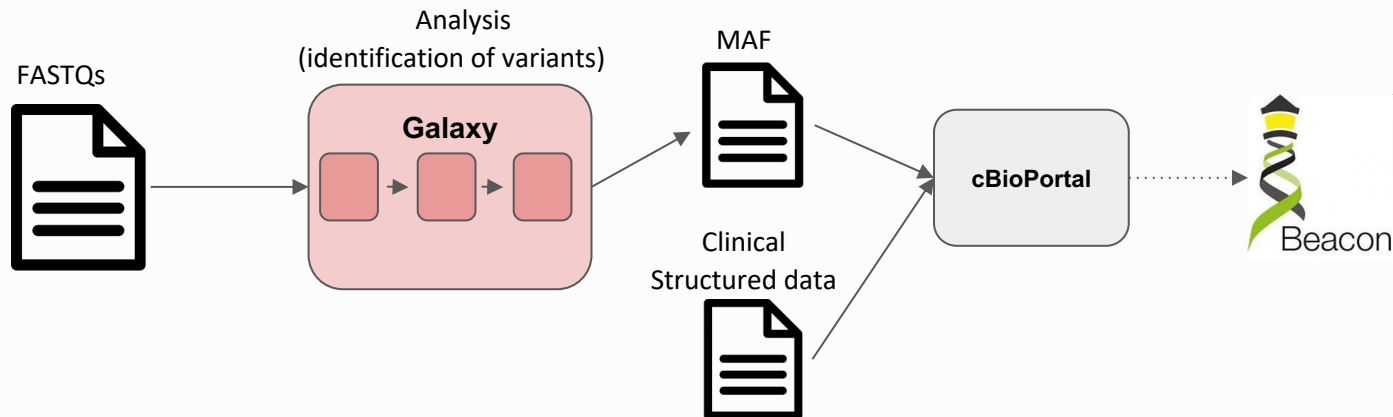




# Examples: Data mobilization across analysis platform (Cancer use-case connecting EOSC4Cancer & GDI)

## Description of the demonstrator

Workflow from FastQ to data analysis and visualisation in cBioPortal. Then, the data in cBioPortal, if it is well formatted, will be queried directly by Beacon (optional).





# Examples: Data mobilization across analysis platform (Cancer use-case connecting EOSC4Cancer & GDI)



Data mobilization across analysis  
platform (Cancer use-case)



# EOSC4Cancer: the cancer research point of view

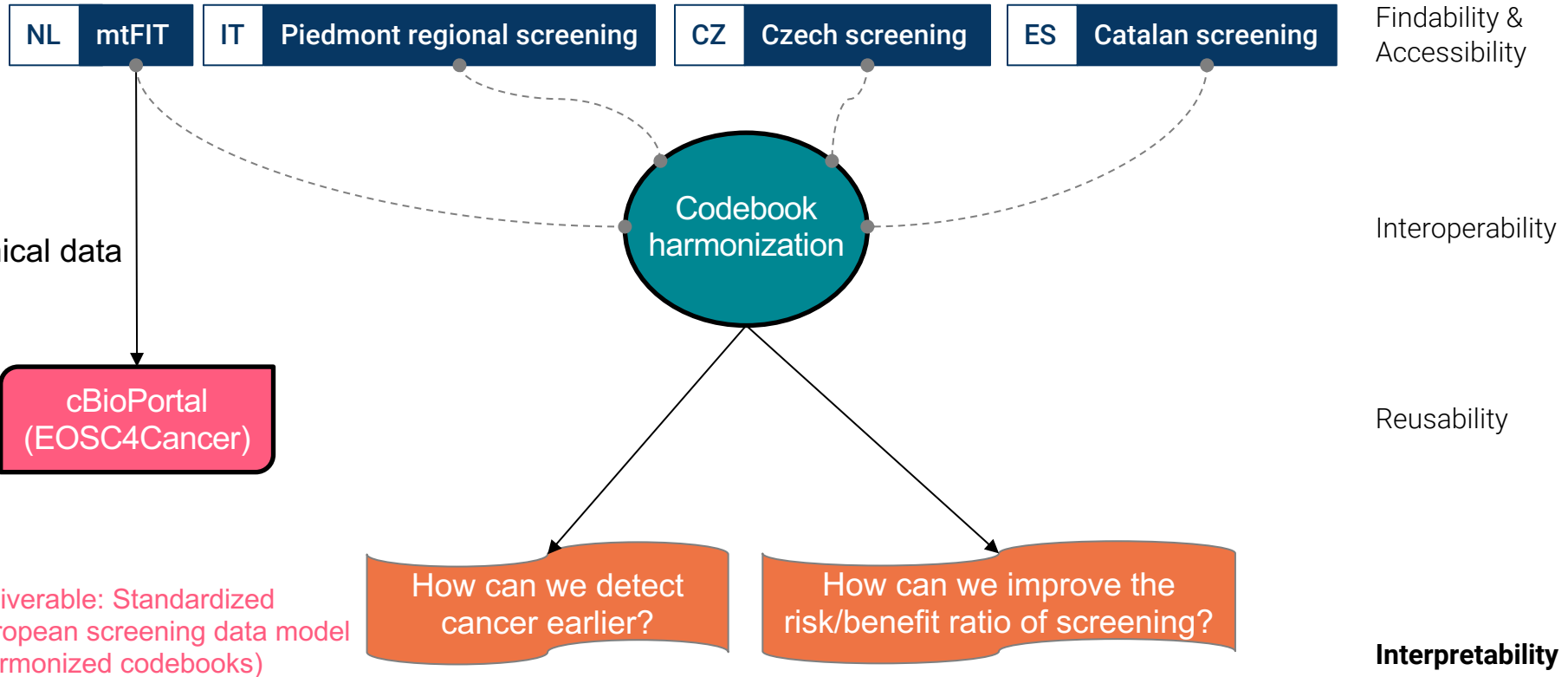
## Colorectal Cancer use-cases

Data-driven optimisation of **cancer screening** programs

Connecting omics data from multiple sources to a **Clinical Decision Support System** (CDSS) for precision treatment of metastatic CRC

Provide **TECHNOLOGICAL SOLUTIONS** + identify **GAPS** and **EMERGING NEEDS** to drive further developments

## 4.2: Data driven optimisation of cancer screening programs





# RDMkit

Data management | About | Contribute | GitHub | Search RDMkit

### Data management

- Data life cycle
- Your role
- Your domain
  - Bioimaging data
  - Biomolecular simulation data
  - Epitranscriptome data
  - Health data**
  - Human data
  - Human pathogen genomics
  - Intrinsically disordered proteins
  - Machine learning
  - Marine metagenomics
  - Microbial biotechnology
  - Plant sciences
  - Proteomics
  - Rare disease data
  - Single-cell sequencing
  - Structural bioinformatics
  - Toxicology data

Your domain

## Health data

### Introduction

Human 'health data' is a broad concept encapsulating diverse data types and modalities, including omics and clinical data. Clinical data (routinely collected or originating from clinical studies) includes but is not limited to images, healthcare administrative data (e.g. demographics), free text and patient-generated data from questionnaires or real-world wearables/mobile devices. This page describes data management considerations and solutions for two widely collected data types used in health data research studies: data about the patient from questionnaires and electronic health records (EHRs) generated from interactions with the healthcare system. Future versions of this page will include additional health data types.

All scientific research involving data processing concerning identifiable people in the European Union is subject to the General Data Protection Regulation (GDPR) and may require ethics approval. This page will not repeat the GDPR, ethics and data anonymization information given elsewhere in RDMkit, namely on the [GDPR compliance](#), [Ethical aspects](#), [Data sensitivity](#) and [Human data](#) pages, which should be familiar to scientists working with health data. The content on this page is also distinct from the [Rare disease data page](#), which considers collecting and processing data specific to rare diseases. Country-specific RDM resources, including existing national solutions or RDM advice specific to national policies/funders/infrastructures are on the [National Resources](#) pages. The information presented on this page is disease and country-agnostic.

### Patient-generated health data from questionnaires

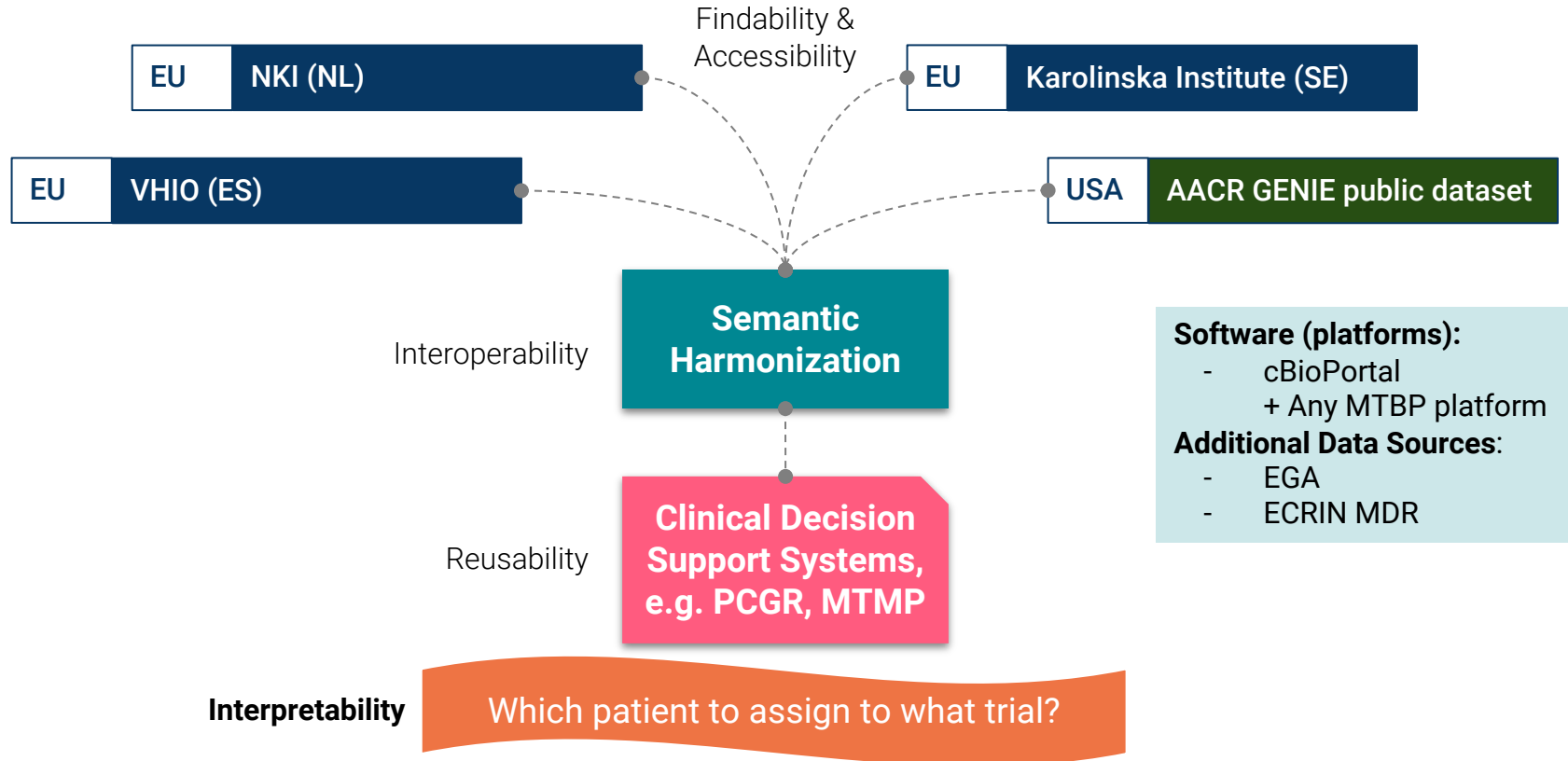
#### Description

Participants' health data which can be collected via surveys, interviews, and monitoring, are called Clinical Outcome Measures (COMs). In surveys, the questionnaire is either completed by the participants or their representative (e.g. family member or caregiver). In interviews, a healthcare professional asks/explains the questionnaire to the participants and records their responses. Monitoring includes measuring a participant's behaviour or activity in both a clinical and non-clinical setup. Based on this, COMs can be classified into four broad categories: Patient Reported Outcome Measure (PROM) is a measurement that comes directly from the patient (i.e. study participant) about the status of a patient's health condition without amendment or interpretation of the patient's response by a clinician or anyone else. Observer-Reported Outcome Measure (ObsROM) is a measurement based on a report of observable signs, events or behaviours related to a patient's health condition by someone (e.g. family member or caregiver) other than the patient or a health professional. Clinician-Reported Outcome Measure

### On this page

- Introduction**
- Patient-generated health data from questionnaires
- Electronic Health Record (EHR) data
- Related pages
- More information

# 4.5: Clinical Decision Support System





---

# Thank you!

**Horizon Europe projects coordination meeting  
in the context of the EOSC European Partnership**

Session 5.2: Adding value through the EOSC  
Federation: Users and Resource environments in  
thematic communities

# AquaINFRA

**Restoring our oceans, seas,  
coastal and inland waters**

Lise Schrøder and Henning Sten Hansen

Aalborg University Copenhagen



*This project has received funding from the European Commission's Horizon Europe Research and Innovation programme under grant agreement No 101094434. The European Commission is not responsible for any use that may be made of the information it contains.*

*/aquainfraeu*

*/aquainfraeu*

*/aquainfra*

## Agenda

- *Introduction to AqualNFRA*
- *Potentials of EOSC regarding demonstration and promotion of the benefits of open science and related collaborative practices for thematic users and resource environments*
- *Opportunities of the connection to the EOSC Federation to enhance user/resources environments already in place in thematic communities?*
- *Services of features that could be accessible for the Federation that could have the largest impact in thematic environments?*
- *Major difficulties or overheads (technical, managerial etc.) entail the eventual connection to the EOSC Federation for thematic user environments?*





## Overall objectives of AqualNFRA

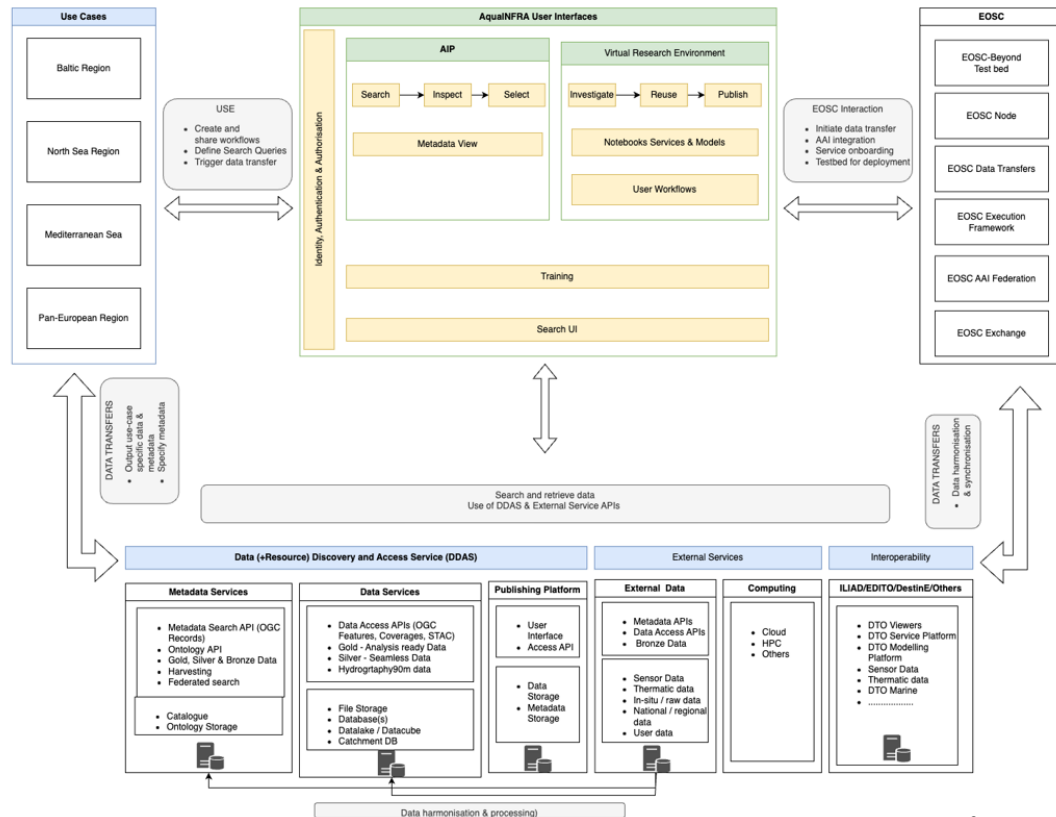
*The overall objective of the project is to develop a virtual environment equipped with FAIR multi-disciplinary data and services to support marine and freshwater scientists and stakeholders restoring healthy oceans, seas, coastal and inland waters*

*The aim of the AqualNFRA virtual environment is to enable the target stakeholders to store, share, access, analyse and process research data and other research digital objects from their own discipline, across research infrastructures, disciplines and national borders leveraging on EOSC and the other existing operational dataspace (e.g., EMODnet, Copernicus Marine Service, Digital Twins, etcetera)*



# The AqualNFRA setting

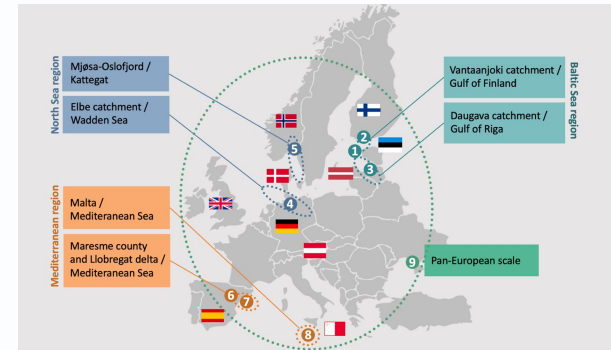
- The technical framework is designed to enhance cross-domain research, enabling seamless discovery and access to marine and freshwater data via interdisciplinary discovery services
- The Data Discovery and Access Service (DDAS) will act as a bridge connecting diverse data catalogues, APIs, and Digital Twins
- The AqualNFRA architecture will utilise international standards such as those from ISO TC211, the Open Geospatial Consortium (OGC), and the EU INSPIRE Directive



# Regional case study areas and technical use cases across the hydrosphere within an EOSC setting

across the

1. *Effectiveness of measures, their social and economic effects*
2. *Creation of seamless water quality data from land to sea utilising high-frequency water quality monitoring with FerryBox observations and combining sensor data with satellite observations.*
3. *Changes in nutrient loading and the Gulf of Riga optical properties: effectiveness, costs and socio-economic effects of environmental measures.*
4. *Monitoring the aquatic carbon cycle and quantifying the human impact (eutrophication).*
5. *Land- Sea interactions and quantifying the human impact (eutrophication)*
6. *Marine ecological state by the land catchment interactions*
7. *Future scenarios for inland water quantity and coastal impacts*
8. *Marine ecological state by the land groundwater interactions*
9. *Freshwater biodiversity*





# Potentials of EOSC regarding demonstration and promotion of the benefits of open science and related collaborative practices for thematic users and resource environments

*Marine and freshwater scientists will be equipped with a user environment for data harmonisation, processing, quality assessment, interaction and sharing, emphasising the pan-European seamless connectivity across the marine and freshwater realms*

- *The AqualNFRA Data Discovery & Access Service enables search for marine data, freshwater data, and socio-economic data seamlessly across borders and data repositories*

*EOSC as a central entrance point might enable collaboration and sharing across this very interdisciplinary research field.*

The screenshot displays the AqualNFRA DDAS (Data Discovery and Access Service) interface. At the top, it features the eOSC logo and navigation links for 'Baltic Sea', 'North Sea', 'Mediterranean', and 'Pan-European'. The main header reads 'AqualNFRA DDAS' and 'AqualNFRA Data Discovery and Access Service'. Below this, there are three primary service buttons: 'Discovery' (DDAS Data Discovery Service), 'Vector Access' (DDAS Vector Data Access Services), and 'Raster Access' (DDAS Raster Data Access Services). A 'Learn more' button is also present.

Two inset windows show the 'Collections in this service' section. The left inset shows a table of various data collections with columns for Name, Type, and Description. The right inset shows a similar table with a 'Coverage' column, listing specific datasets like 'CORINE Land Cover Plus 2018 dataset'.

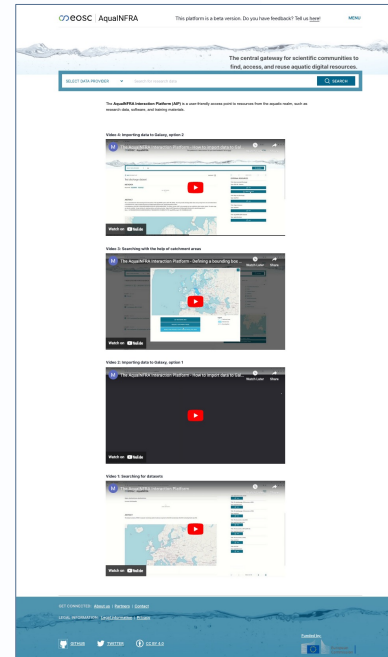
At the bottom, the footer identifies the service as being provided by the Finnish Geospatial Research Institute.

# Potentials of EOSC regarding demonstration and promotion of the benefits of open science and related collaborative practices for thematic users and resource environments

*An open and reproducible science capacity is being established in the marine, coastal, and inland water research communities, from bachelor's students to experienced researchers*

- *demo versions supporting the internal capacity building are available via the AqualNFRA Interaction Platform (AIP)*
- *a suite of educational materials and resources are being co-developed during the project promoting a sustainable culture of open science within the marine and freshwater research communities.*

*Capacity building within the consortium as part of being an EOSC project as well as the promotion of the AqualNFRA training platform via EOSC*



## Opportunities of the connection to the EOSC Federation to enhance user/resources environments already in place in thematic communities?

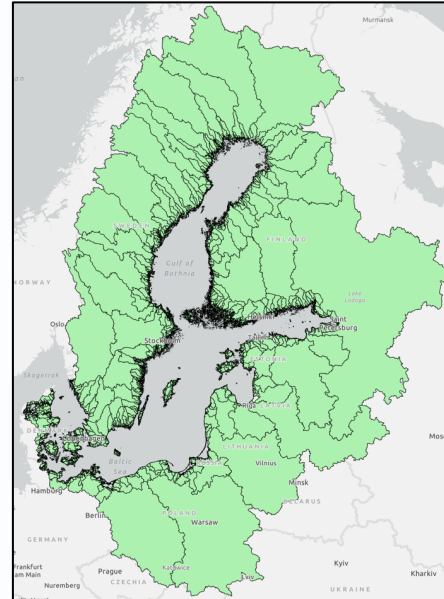
- *To provide an EOSC service that allows users to seamlessly and easily find and access data from the diverse set of catalogues*
- *To develop use cases for improved water quality and healthy aquatic environments to demonstrate the value of sharing open and FAIR research data and provide feedback to the EOSC Partnership*
- *To synergise with the oceans, seas, coastal and inland waters initiatives and the EOSC Partnership to avoid overlaps and reuse existing resources*



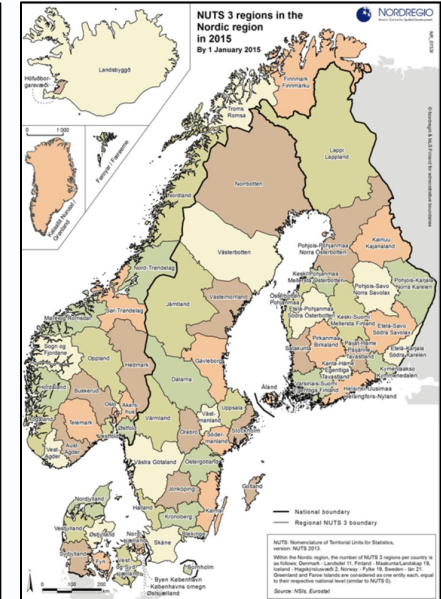
## Services of features that could be accessible for the Federation that could have the largest impact in thematic environments?

*AqualNFRA aims at providing a robust, user-friendly, and innovative environment that enables and propels forward the research in marine and freshwater sciences.*

- *A cornerstone of AqualNFRA is data harmonisation, essential for integrating diverse datasets from disparate sources, ensuring consistency and compatibility across different systems.*
- *The process will involve standardising data formats, metadata, and vocabularies to align with FAIR principles.*
- *The harmonised data, once processed, will feed into the AqualNFRA Data Space.*



<https://maps.helcom.fi/website/mapservice/>



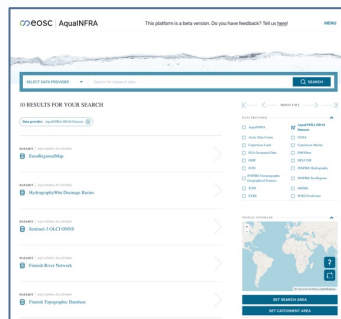
<https://nordregio.org/maps/nuts-3-regions-in-the-nordic-region-in-2015/>



# Services of features that could be accessible for the Federation that could have the largest impact in thematic environments?

*Through leveraging high-performance computing, adhering to and advancing standards, and promoting open science and FAIR data principles, AqualNFRA aims to be a transformative force in aquatic research infrastructures*

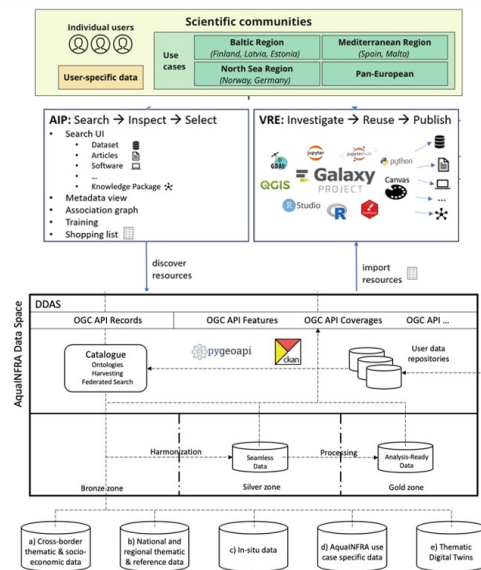
- The AqualNFRA Interaction platform (AIP) provides access to DDAS as well as the Virtual Research Environment (VRE)
- The VRE will provide computational tools and services necessary for sophisticated data analysis and modelling and support a range of programming languages, such as R and Python
- Notebook services offering customisable workflows to facilitate the analysis of raw or harmonised data will be included, allowing researchers to create tailored processes for their specific research needs.



<https://aquainfra.dev.52north.org/>

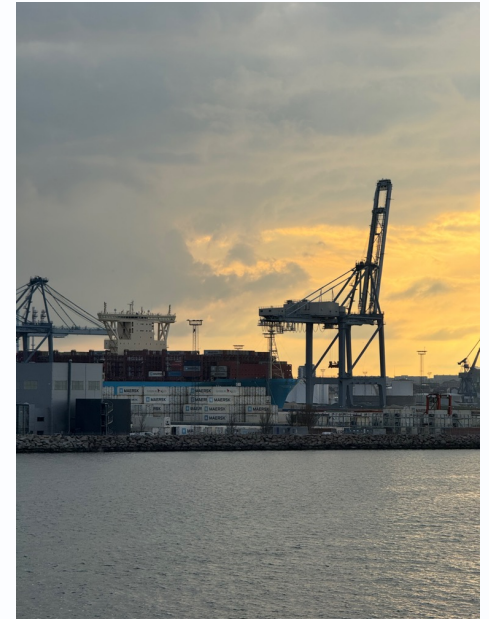


<https://aqua.usegalaxy.eu/>



## Major difficulties or overheads (technical, managerial etc.) that entail the eventual connection to the EOSC Federation for thematic user environments?

- *EOSC is new to most of the domain experts in AqualNFRA especially on the freshwater side*
- *Tools and workflows are typically tailored within national organisations and data are not necessarily in a searchable format*
- *Combining the fresh water and marine research environments and also include socioeconomic data in a seamless setting is in itself a challenge*
- *How to adapt the AqualNFRA architecture to connect to the new EOSC node setting and contribute to the EOSC Federation*
- *Uncertainty regarding the future of EOSC*
- *How to consider EOSC vs. DTO efforts/synergies*




## Partners



- AALBORG UNIVERSITET (AAU) DK
- MAANMITTAUSLAITOS (NLS FGI) FI
- THE BALTIC MARINE ENVIRONMENT PROTECTION COMMISSION (HELCOM)
- CSC-TIETEEN TIETOTEKNIIKAN KESKUS OY (CSC) FI
- SUOMEN YMPARISTOKESKUS (SYKE) FI
- SINTEF AS (SINTEF) NO
- NORSK INSTITUTT FOR VANNFORSKNING (NIVA) NO
- TARTU ULIKOOL (UTARTU) EE
- LATVIJAS HIDROEKOLOGIJAS INSTITŪTS (LIAE) LV
- HOCHSCHULE BOCHUM (HSBO) DE
- FORSCHUNGSVERBUND BERLIN EV (FVB-IGB) DE
- 52NORTH SPATIAL INFORMATION RESEARCH GMBH (52°North) DE
- HELMHOLTZ-ZENTRUM HEREON GMBH (HEREON) DE
- DEUTSCHES KLIMARECHENZENTRUM GMBH (DKRZ) DE
- CENTRO DE INVESTIGACION ECOLOGICA Y APLICACIONES FORESTALES (CREAF)
- AGENCIA ESTATAL CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS (CSIC)
- UNIVERSITA TA MALTA (UMALTA) MT
- UNIVERSITAET FUER BODENKULTUR WIEN (BOKU) AT
- KARLSRUHER INSTITUT FUER TECHNOLOGIE (KIT) DE
- TECHNISCHE UNIVERSITAET DRESDEN (TUD) DE
- BLUE LOBSTER IT LIMITED (BLIT) UK







Thank you for your attention!

Lise Schrøder and Henning Sten Hansen  
Aalborg University Copenhagen  
[lisesch@plan.aau.dk](mailto:lisesch@plan.aau.dk), [hsh@plan.aau.dk](mailto:hsh@plan.aau.dk)

# EOSC HE Coordinators

## June 2024

Breakout 5b - Adding value through the EOSC Federation: Users and Resource environments in thematic communities.



Funded by  
the European Union



# Agenda

- Introduction, breakout approach (Peter, Sara)
- Intro talk 1: Lise Schrøder, Aqua-INFRA.
- [Intro talk 2](#): Salvador Capella-Gutierrez, EOOSC4Cancer
  
- Discussion question 1:
  - How can EOOSC help us demonstrate and promote the benefits of open science and related collaborative practices for thematic users and resource environments?
- Discussion question 2:
  - What opportunities does the eventual connection to the EOOSC Federation bring to enhance user/resources environments already in place in thematic communities?
- Discussion question 3:
  - What services or features that could be accessible for the Federation could have the largest impact in thematic environments?
- Discussion question 4:
  - What major difficulties or overheads (technical, managerial etc) entail the eventual connection to the EOOSC Federation for thematic user environments?
  
- Summary, discussion readouts (Sara, Peter)

# Adding value through the EOSC Federation: Users and Resource environments in thematic communities

## Practical example from the EOSC4Cancer project

**Reusing catalogues** via common standards

**Open access platform (cBioPortal)** available to anyone

**Data driven harmonisation** of cancer screening programme: interoperability was put in practice in two screening programmes (NL - IT) working together by harmonising a codebook, which was then used also by IT and CZ

**Clinical decision support system:** creation of a (new) standard for semantic harmonisation for clinical decisions

# Discussion question 1: How can EOSC help us demonstrate and promote the benefits of open science and related collaborative practices for thematic users and resource environments?

The **human engagement factor** is perceived as a value add in EOSC. People wouldn't have had the chance to meet without EOSC, domain specific practices would have been shared without EOSC.

EOSC is not a portal, should be considered as **an umbrella framework** going much beyond data.

**Training** becomes strategic at this point: to inform researchers and scientists about the opportunity to use services, tools and methods already in place

Discussion question 2: What opportunities does the eventual connection to the EOSC Federation bring to enhance user/resources environments already in place in thematic communities?

EOSC has the potential of making available resources from other spaces (institutes, communities, nodes) not available before. **Shared authentication** is one of the opportunities, with a unique ID that is broad enough to include availability of a broader set of services (so not just limiting it to EduGAIN)

An **EOSC helpdesk** would be a great opportunity - if sustainability of such a service is ensured

**Scalability of VRE resources** is also seen as an opportunity, with SLAs that can vary according to the type of resources needed, the scope of using the services and the domain

## Discussion question 3: What services or features that could be accessible for the Federation could have the largest impact in thematic environments?

A **managed, reliable cloud service** with ensured uptime service would be the most desirable service for scientific user communities

A **modular SLA model** to be adapted for different services and different scientific communities, so to facilitate the understanding of what services are critical to be ensured at what conditions

A **credit system** could be effective if linked to an effective tracking of usage of resources (and here the AAI is necessary) and an **elastic resizing of resources** across the federation, to avoid stopping services when credits end

Discussion question 4: What major difficulties or overheads (technical, managerial etc) entail the eventual connection to the EOSC Federation for thematic user environments?

Data can be used in many different research environments, what's difficult is **making data findable by other domains** and to describe data in a way that can be integrated to other domains

Stronger (human and technological) connection between EOSC and **Digital Twins** should be made

**Time** is a challenge: putting EOSC in practice requires years!



# Stay up to date



By-covid.org

newsletter signup | news & events | project output



@BYCOVID\_eu



/company/by-covid



Join our **industry community** to discuss visionary ideas, challenges and solutions in COVID-19 and infectious diseases



Fill in our industry survey



Funded by  
the European Union

*Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.*



# INFRAEOSC Coordination Meeting 2024

## Sustainability of projects' outcomes

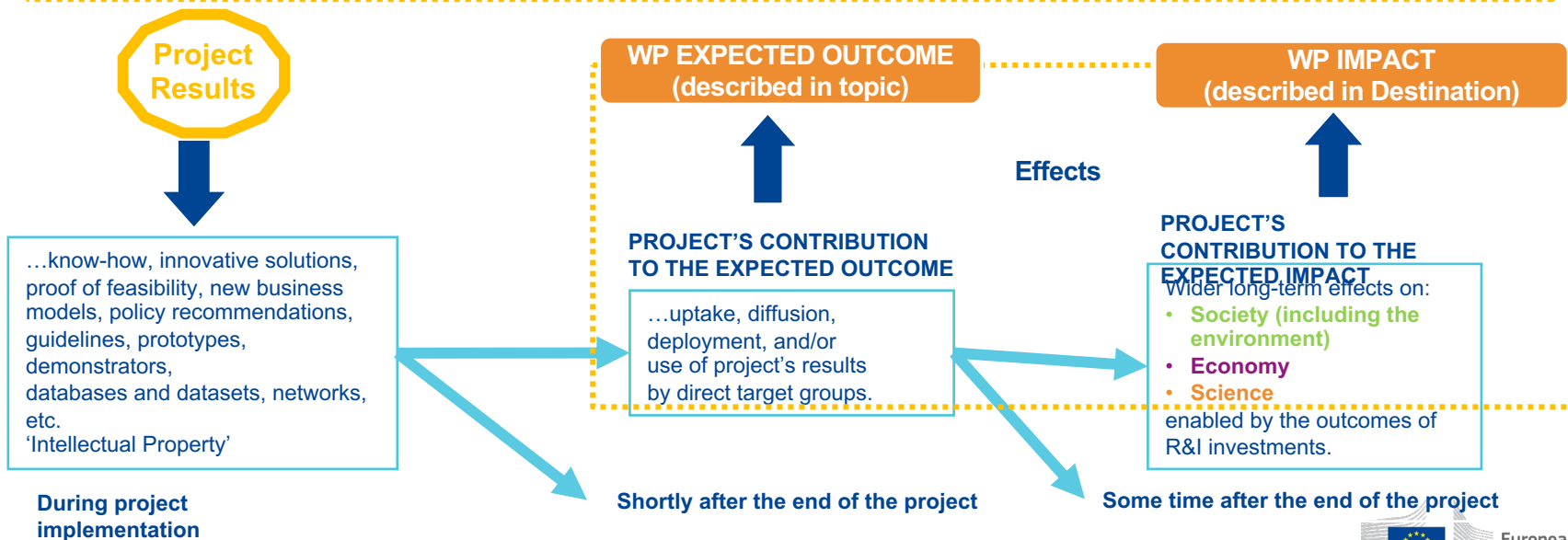
**Flavius PANA**  
European Research Executive Agency

# Project's pathway towards impact



**Project's pathway towards impact**

**Specific contribution that projects can make to the expected outcomes and impacts set out in the Work Programme**



# IPR - BACKGROUND AND RESULTS – ARTICLE 16

## Exploitation of results

- Beneficiaries which have received funding under the grant must — up to four years after the end of the action - use their best efforts to exploit their results directly or to have them exploited indirectly by another entity ...
- If, despite a beneficiary's best efforts, the results are not exploited within one year after the end of the action, the beneficiaries must (unless otherwise agreed in writing with the granting authority) use the **Horizon Results Platform** to find interested parties to exploit the results.

## About

### What is Horizon Results Booster?

**Horizon Results Booster (HRB)** is an initiative of the [European Commission](#) which aims to bring a continual stream of innovation to the market and maximise the impact of public funded research within the EU.

It supports projects eager to go beyond their Dissemination and Exploitation (D&E) obligations - steering research towards strong societal impact and concretising the value of Research and Innovation (R&I) activity for societal challenges.

To achieve this, HRB offers free consulting services to closed or ongoing research projects funded by FP7, Horizon 2020, or Horizon Europe programmes.

### What Services Are Available?

The Horizon Results Booster offers 3 types of services:

#### **1. Portfolio Dissemination & Exploitation Strategy**

- Identifying and creating the portfolio of Research & Innovation project results (module A),
- Creating the portfolio of results; design and execute a portfolio dissemination plan (module B)
- Improving existing exploitation strategy (module C)

#### **2. Tailor made support services to develop a business plan**

#### **3. Assistance, coaching and mentoring for go-to-market activities.**

### When Can You Apply?

It is possible to submit an HRB application as soon as the project's results are identified. HRB services **must be completed** at the latest by **7th of November 2024**.

### Who Is Providing the Services?

The services are provided by a consortium of specialized companies: META Group, Ecorys Europe, ICONS, Trust-IT Services, BDO, and PNO. For more information about Horizon Results Booster, visit the [FAQ section](#) or contact us at [booster@meta-group.com](mailto:booster@meta-group.com).

Follow HRB on [LinkedIn](#), [Twitter](#) and subscribe on [YouTube](#) to stay connected!

# Open Questions

- How is your project ensuring the sustainability of the project Outcomes?
- What are the challenges (if any) that you are facing?
- What possible solutions do you envisage ?