iMagine: AI-Powered Image Data Analysis in Aquatic Science

Elnaz Azmi*, Khadijeh Alibabaei, Valentin Kozlov, Álvaro López García, Dick Schaap, Gergely Sipos

*elnaz.azmi@kit.edu

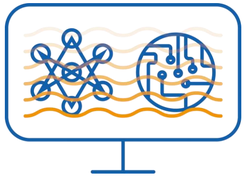
PASC25 | Brugg-Windisch, Switzerland | June 16-18, 2025

iMagine receives funding from the European Union's Horizon Europe research and innovation program under grant agreement No. 101058625

Funded by the European Union

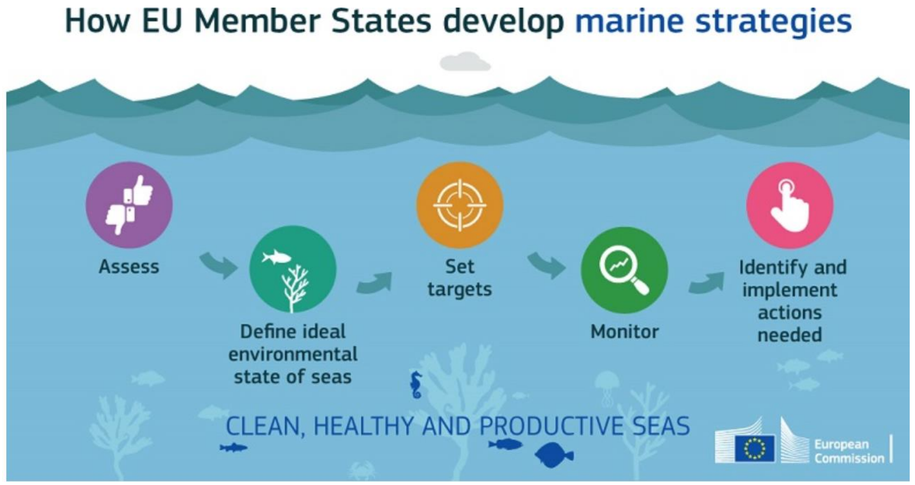
Outline

- Introduction
- Competence Center
- AI Platform
- Best Practices and Tools
- Case Study
- Summary



Sustainable Development

EU Directives and Initiatives for Aquatic Domain



Source: EU MSFD website

- Marine Strategy Framework Directive (MSFD)
- Water Framework Directive (WFD)
- European Green Deal
- Mission Restore our Ocean and Waters by 2030
- United Nation's 2030 Agenda for Sustainable Development
- United Nation's Decade of Ocean Science (2021-2030)

SUSTAINABLE DEVELOPMENT GOALS

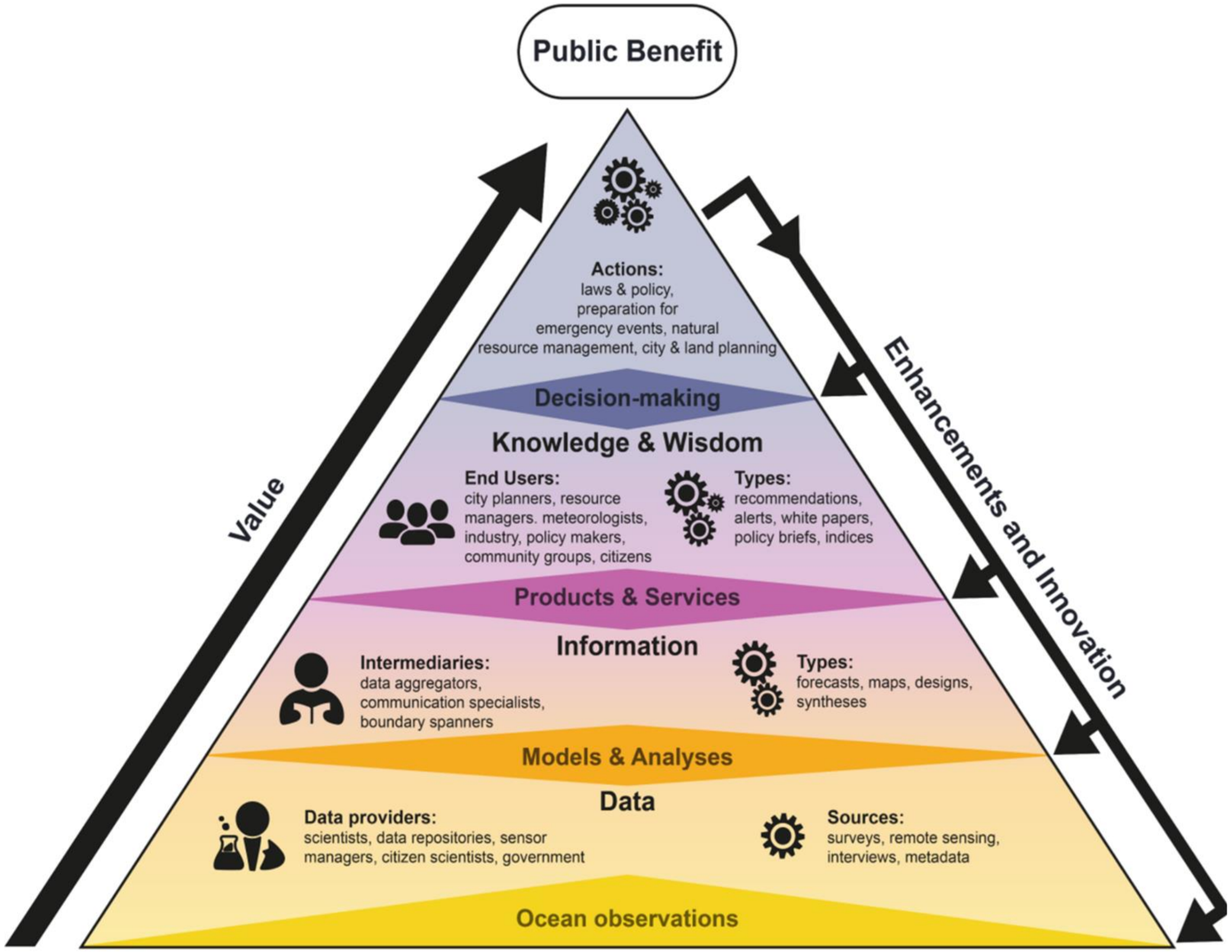
Green Deal

EU MISSIONS
RESTORE OUR OCEAN & WATERS

2021 2030 United Nations Decade of Ocean Science for Sustainable Development

14 LIFE BELOW WATER

Marine Value Chain

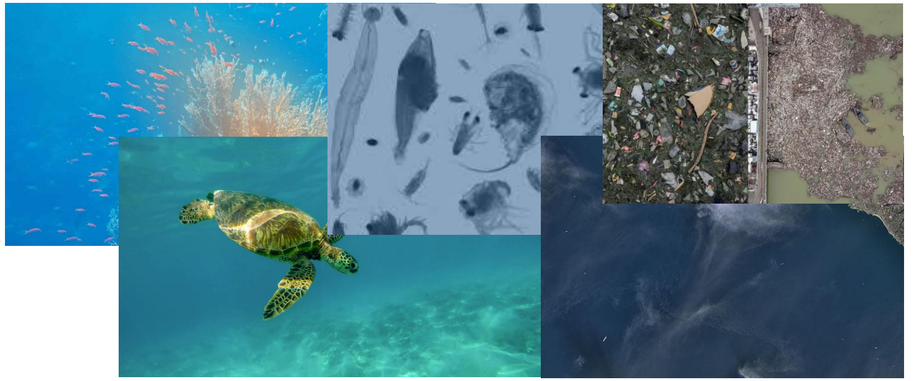


Source: European Marine Board

Statistics, Models, Policies, …

Object Detection & Identification

AI



Images

Imaging instruments

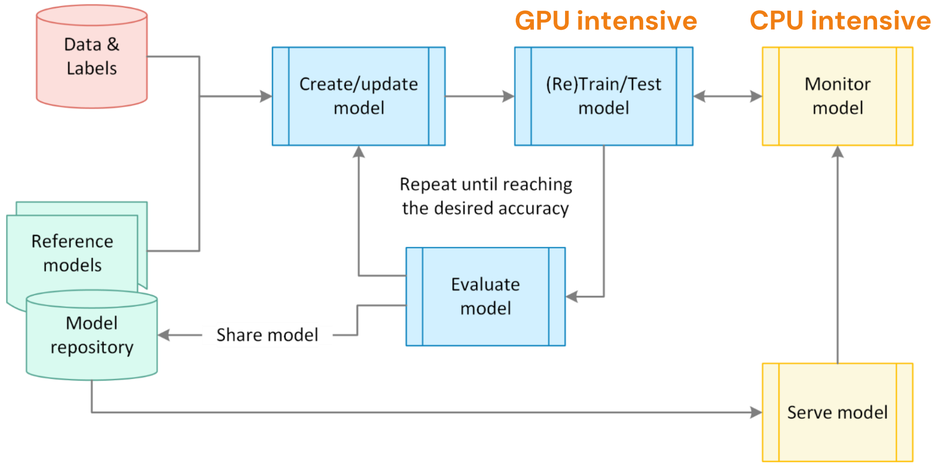
The iMagine Project

4 national cloud compute centers and
5 AI/ML technology development institutes supports
14 research institutes for **12** research infrastructures

The iMagine Approach

Thematic Services

Competence Centre



Use cases

- 8 internal
- 6 onboarded via open calls

Platform + Competence Centre

iMagine AI Platform:

Generic and scalable platform for developing and sharing AI/ML applications

iMagine Competence Centre:

Distributed support team for the full application lifecycle

Infrastructure

Federated clouds from EGI

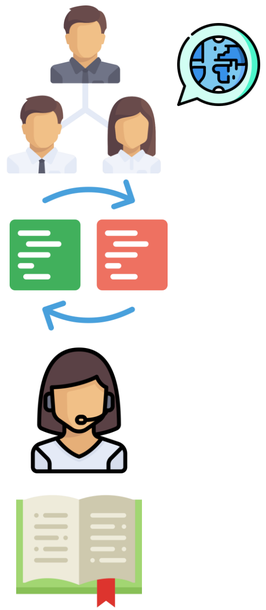
- 1500 TB-months
- 132,000 GPU-hours
- 6 Million CPU-hours

Competence Center:

Mission & Methodology

Mission: **Facilitate** the path from image **data** to imaging **applications** and **services** for scientists in aquatic sciences

Methodology:

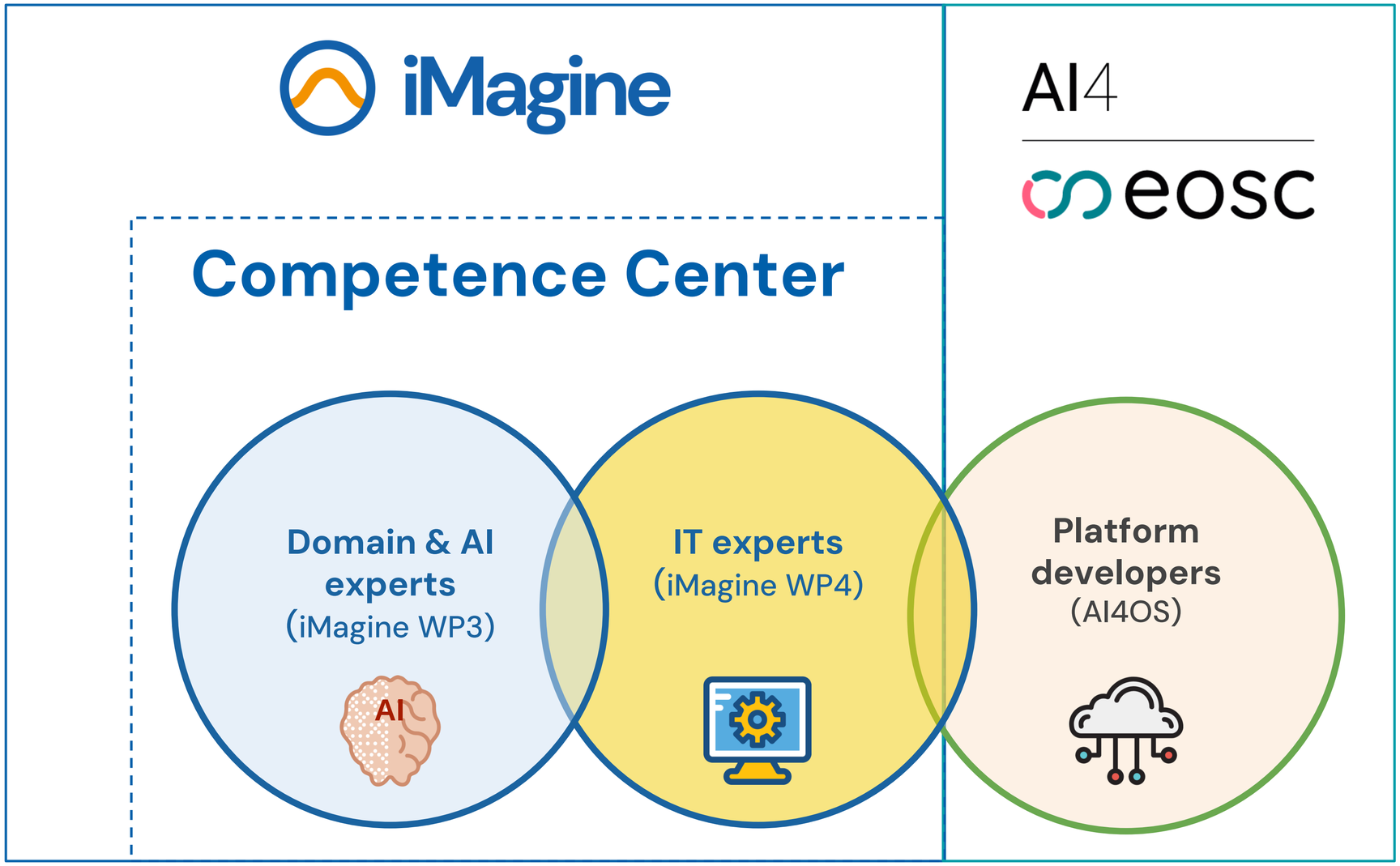


- Connect **experts** from different domains (science, AI, IT, …)
- Stimulate **knowledge exchange** between the parties
- Speak **common language**
- Provide **training** and **support**
  (data, AI, FAIR, platform, service management)
- Create **best practices**

Vectors and icons by SVGRepo

Competence Center:

Synergetic Approach



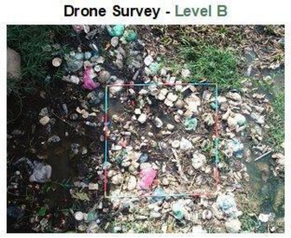
Vectors and icons by SVGRepo

Use Cases

5 mature + 3 prototypes

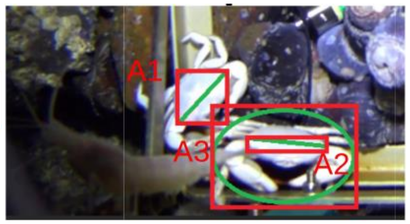
Marine Litter Assessment
(DFKI, MARIS, OGS)

Monitoring system for aquatic litter pollution



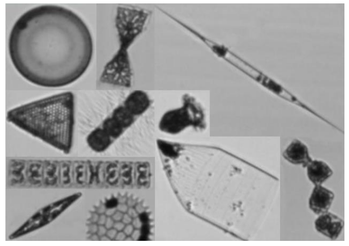
Marine Ecosystem Monitoring
(EMSO ERIC, UPC, IFREMER, MI)

Ecosystem monitoring by means of video imagery from cameras at EMSO sites



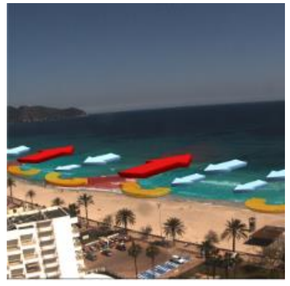
FlowCam Phytoplankton Identification
(VLIZ)

Taxonomic identification of phytoplankton using FlowCam images



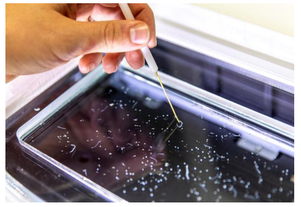
Beach Monitoring
(SOCIB)

Posidonia oceanica berms and rip-currents detection from beach monitoring systems



ZooScan – EcoTaxa Pipeline
(Sorbonne Uni.)

Taxonomic identification of zooplankton using ZooScan

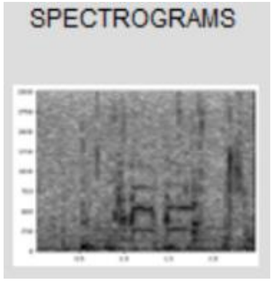


Oil Spill Detection
(CMCC, OrbitalEOS, UNITN)

Oil spill detection from satellite images

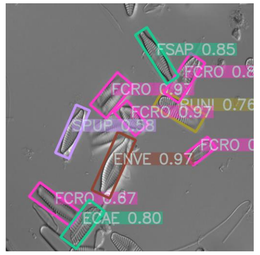
Underwater Noise Identification
(VLIZ)

Identification of sound events from acoustic recordings using spectrograms



Freshwater Diatoms Identification
(UL-LIEC)

Diatom-based bioidentification using automatic pattern recognition on microscope images



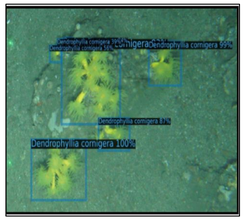
Use Cases

Onboarded via Open Calls

Improving Knowledge about Cold Water Coral Reef

(Spanish Oceanographic Institute)

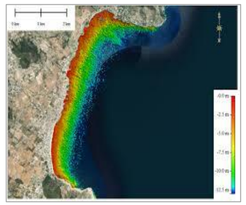
Use AI to precisely delineate areas of living coral and dead coral



Satellite-Derived Bathymetry

(Institute of Marine Sciences of Andalusia)

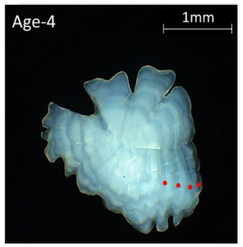
Nearshore bathymetry for coastal studies



AI for Image-Based Age Reading from Fish Otoliths

(Danish National Institute of Aquatic Resources)

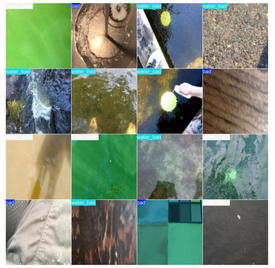
Develop an AI algorithm to identify the age of a fish from otolith images



Validating Images Collected with the EyeonWater App

(MARIS)

Detect and automatically reject false images, measuring life in water



Sea Wave and Coastal Inundation Detection Methodology (SWiM)

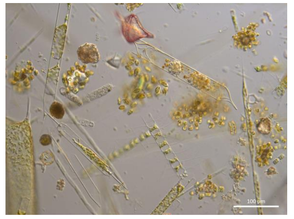
(National Laboratory for Civil Engineering)

A methodology for an automated wave runup and wave detection from images for the surveillance of coastline

DEcentrAlised Learning (DEAL)

(Plymouth Marine Laboratory)

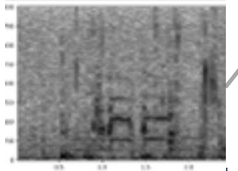
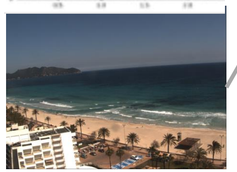
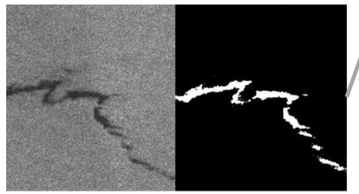
Use decentralized learning to increase the robustness of automated plankton and deep-water animal imagery processing



Competence Center:

From Data to Knowledge

UCx collected raw data

AI/ML development cycle

Data & Labels

Create/update model

(Re)Train/Test model

Monitor model

Repeat until reaching the desired accuracy

Reference models

Model repository

Share model

Evaluate model

Serve model

Continuous interaction with users, feedback & knowledge sharing

5 Image analysis services (powered by trained models)

iMagine Marketplace

OSCAR

20 Labeled training datasets

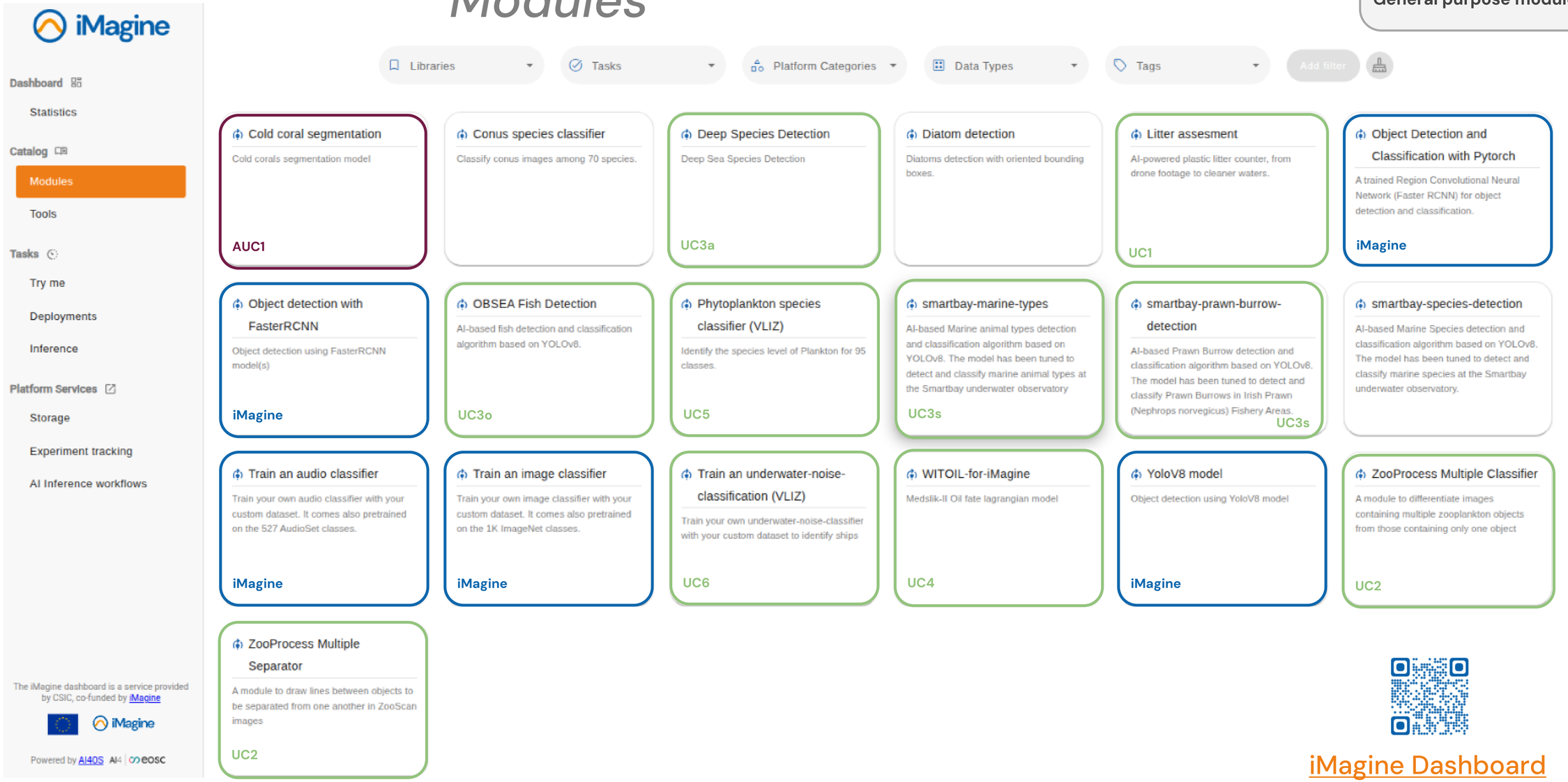
zenodo

Best practices

Dashboard

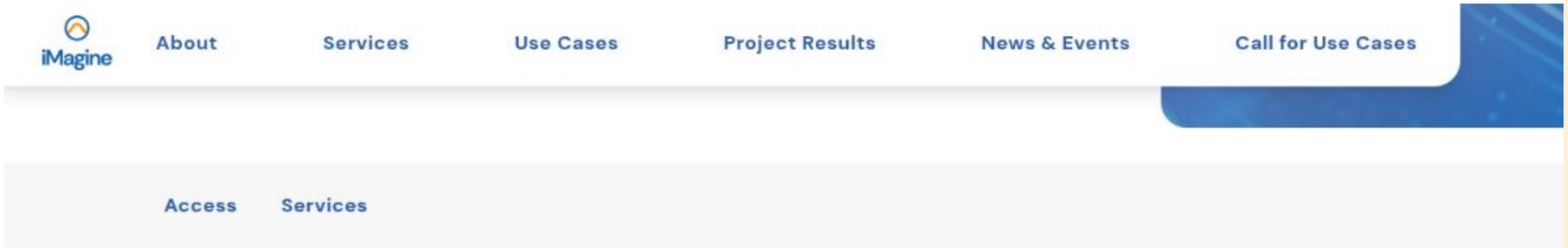
Modules

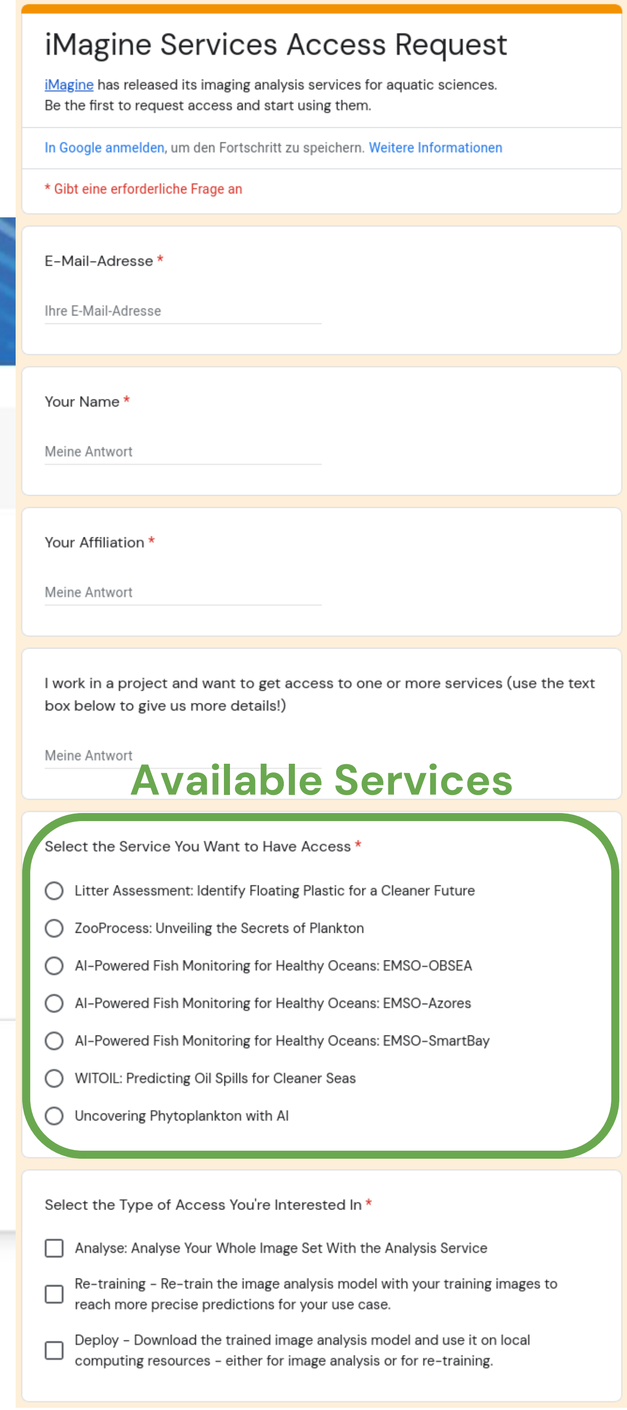
iMagine use case modules
Open call use case modules
General purpose modules



iMagine Dashboard

iMagine





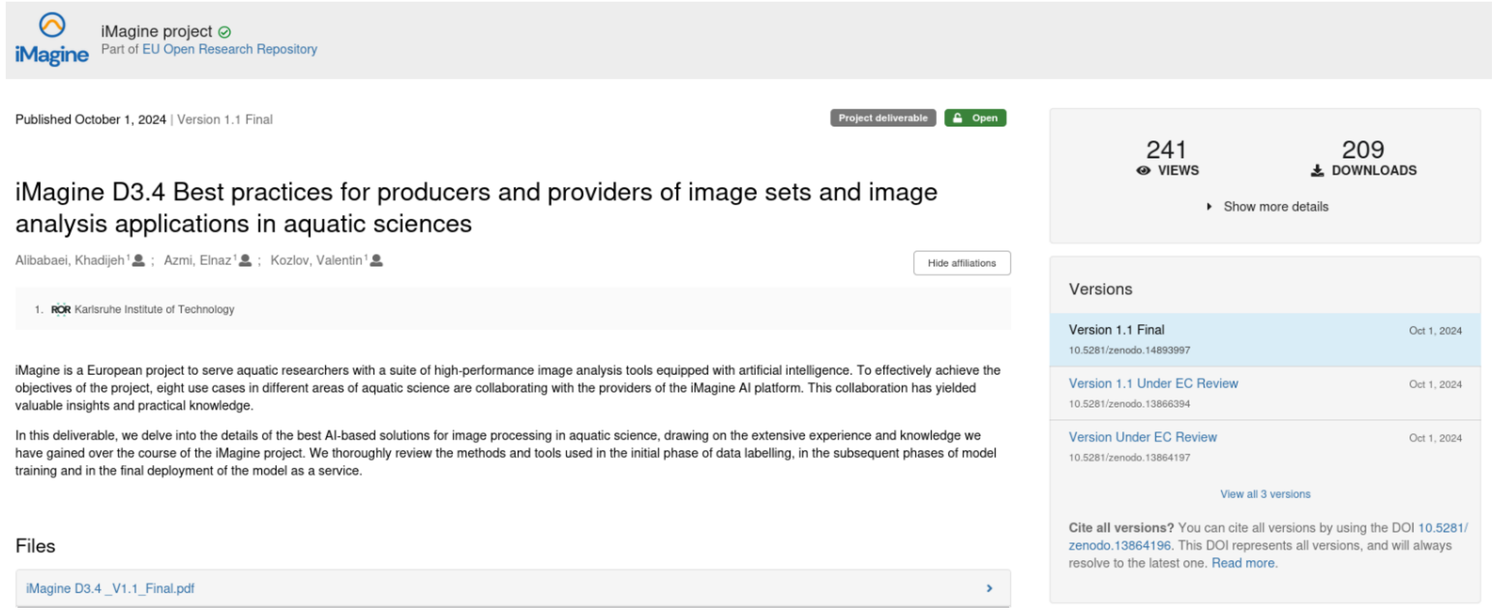
Thematic Services for Image Analysis

https://www.imagine-ai.eu/services/image-analysis-services-for-aquatic-sciences

Best Practices:

Publication

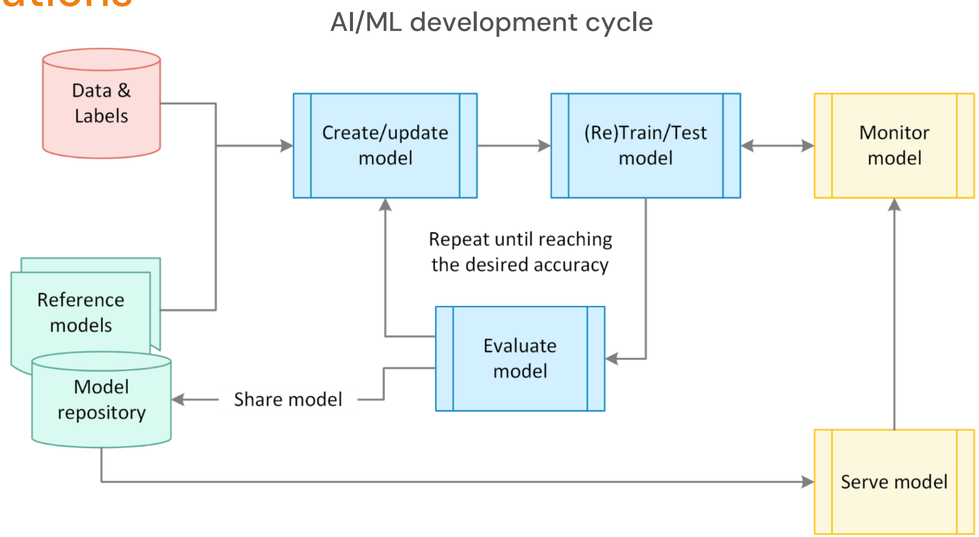
- Lessons learned during the whole AI development cycle
- Experiences from 8 aquatic image analysis use cases
- Available since October 2024 at Zenodo: https://doi.org/10.5281/zenodo.13864196
- Accepted for publication at the "Ecological Informatics" peer-reviewed journal



Best Practices:

Covered topics

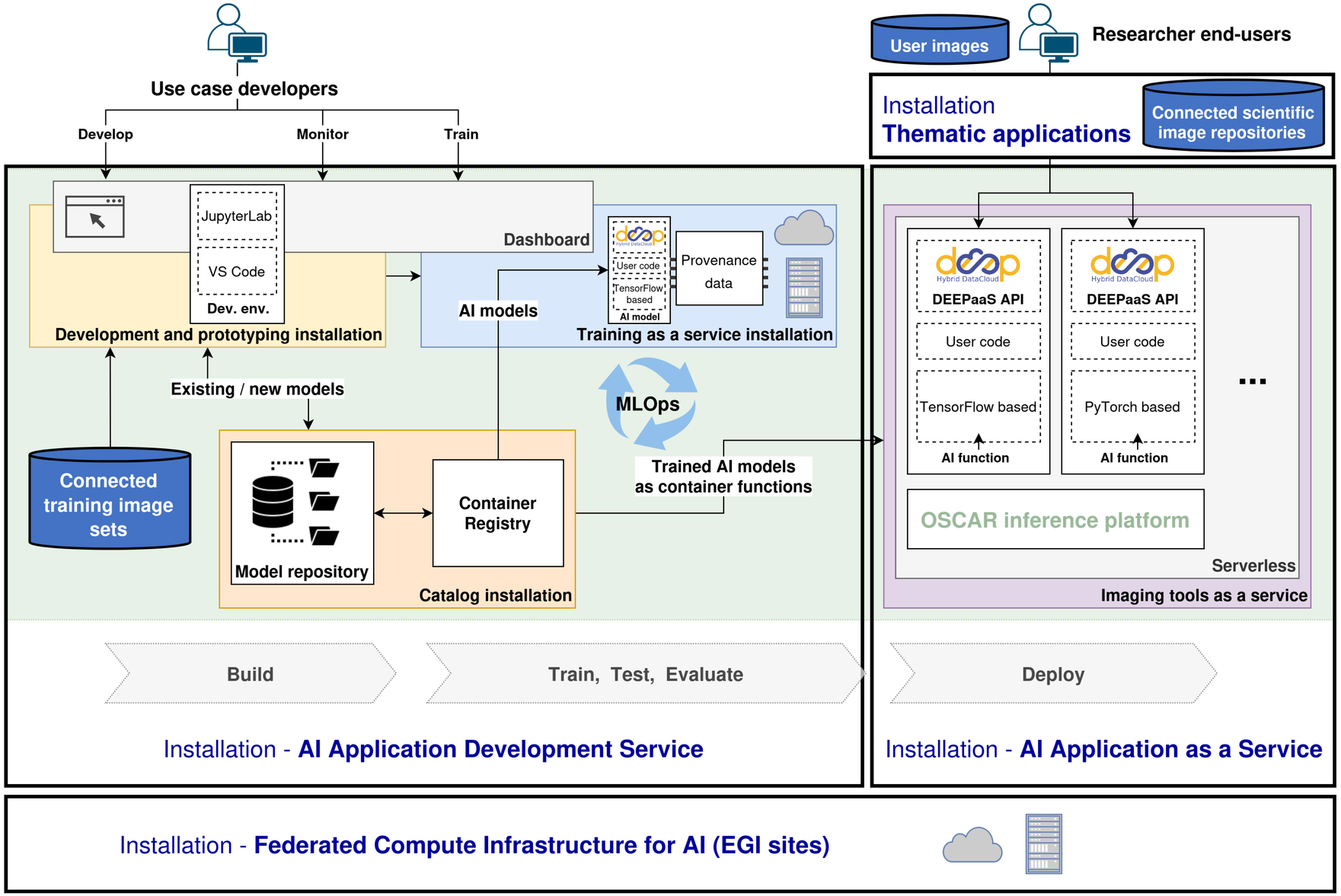
- Deep Learning
- Annotation Tools
- Data Repositories and Open-source Dataset for Marine Applications
- Preprocessing Techniques
- Performance Metrics and Evaluation Methods
- Tools for Monitoring Model Performance
- Data Biases and Fairness in Aquatic Science Models and Data
- Model Delivery
- AI Model Drift Tools
- Use Case Experiences



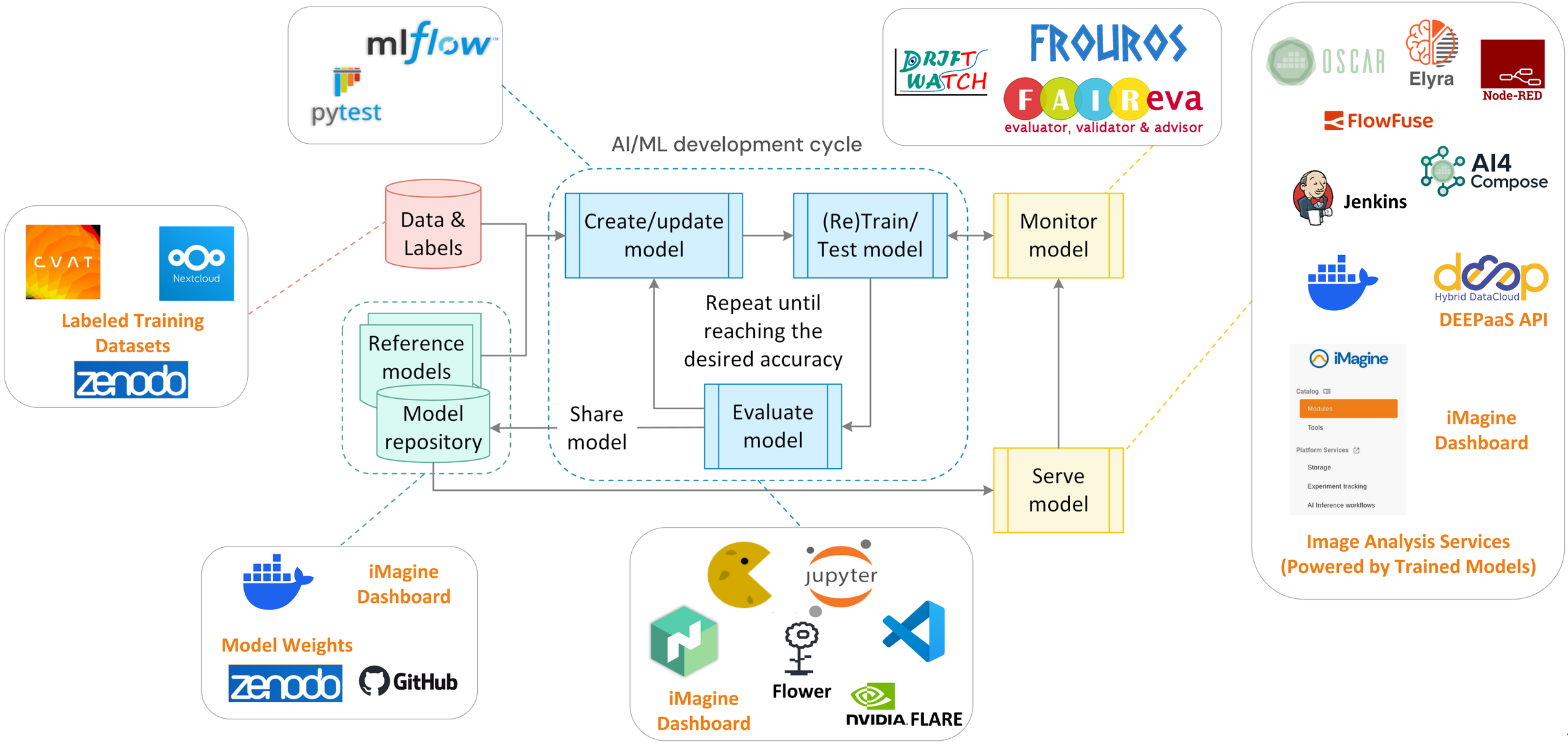
Diversity in AI development

Use cases	Labeling	Preprocessing	Data biases	Models	Performance metrics
UC1 - Marine litter assessment	prelabeled with **in-house** litter categories(JLIST)	resizing, **augmentation**, (horizontal flips, vertical flips, and random rotations)	**unbalanced** dataset, AI model checked, no need to address	**MobileNetV2**, SqueezeNet1.1., **DenseNet121**, ResNet50	**accuracy, precision, recall, F1-Score**
UC2 - ZooScanNet	labeling with EcoTaxa, segmentation with current version of ZooProcess (both **in-house** tools)	Before AI: **background subtraction** on the original image (with a blank scan, by ZooProcess) Classif: data **augmentation** for (rotation, resizing, cropping, flipping) done just-in-time during training and inference Segment: no augmentation	unbalanced dataset, combination of **resampling** (for classif and segment) and **weights** (for classif)	Classif: **MobileNet** Segment: MaskRCNN, **Mask2Former**	Classif: binary cross-entropy Segment: **accuracy, recall, precision, IoU**
UC3o - EMSO OBSEA	**labelImg**, roboflow, label studio, **BIIGLE**	**augmentation**	cover the **seasonality**, one-year dataset, unbalanced dataset, adding dataset of **other species**	Faster-RCNN, **YOLOv8**	mAP@50, **precision, recall**
UC3a - EMSO Azores	annotated by citizens	padding, bounding box corrections, **addressing redundancy**	unbalanced dataset, focused on **two specific species**	**YOLOv8**	**precision, recall**, mAP50-95
UC3s -Smartbay	annotated by **CVAT**	**augmentation** (roboflow noise and contrast)	different sediment types, larger training datasets and **more classes**	**YOLOv8**, DOVER VQA	mAP50-95
UC4 - Oil spill detection	proprietary method developed by Orbital EOS	weight-loss	false positives, almost 98% of pixels are not real oil spills	Bayesian optimization	Fractions Skills Score
UC5 - Flowcam plankton identification	**in-house** developed labeling tool	**augmentation**	unbalanced dataset, **resampling**	CNN based on an Xception architecture	**precision, recall, F1 score**
UC6 - Underwater Noise Identification		window size, **features extraction**,	data from **different locations** and dates during one year	Contrastive Language-Audio model based on the CLAP model	RMSE
UC7 - Shoreline extraction, Beach monitoring	**CVAT**, **in-house** tool written in R	label format adaptation, colour space transformation, **data scaling**, split rows, patchify, **augmentation**	unbalanced dataset, (**geographic locations** and/or camera FOVs)	U-Net, **YOLOv8**,9,11-seg, Bi-LSTM	**accuracy**, loss, **precision, recall, F1 score, IoU**
UC8 - Freshwater diatoms identification	**BIIGLE**, LabelBox	pre-training on synthetic datasets, padding, **bounding box corrections** (resizing, rotation), **features extraction**	unbalanced dataset	**YOLOv5**, EfficientNet, MAPLE, SAM2	**accuracy**, Expected Calibration Error, Negative Log-Likelihood, **Precision-Recall curve**

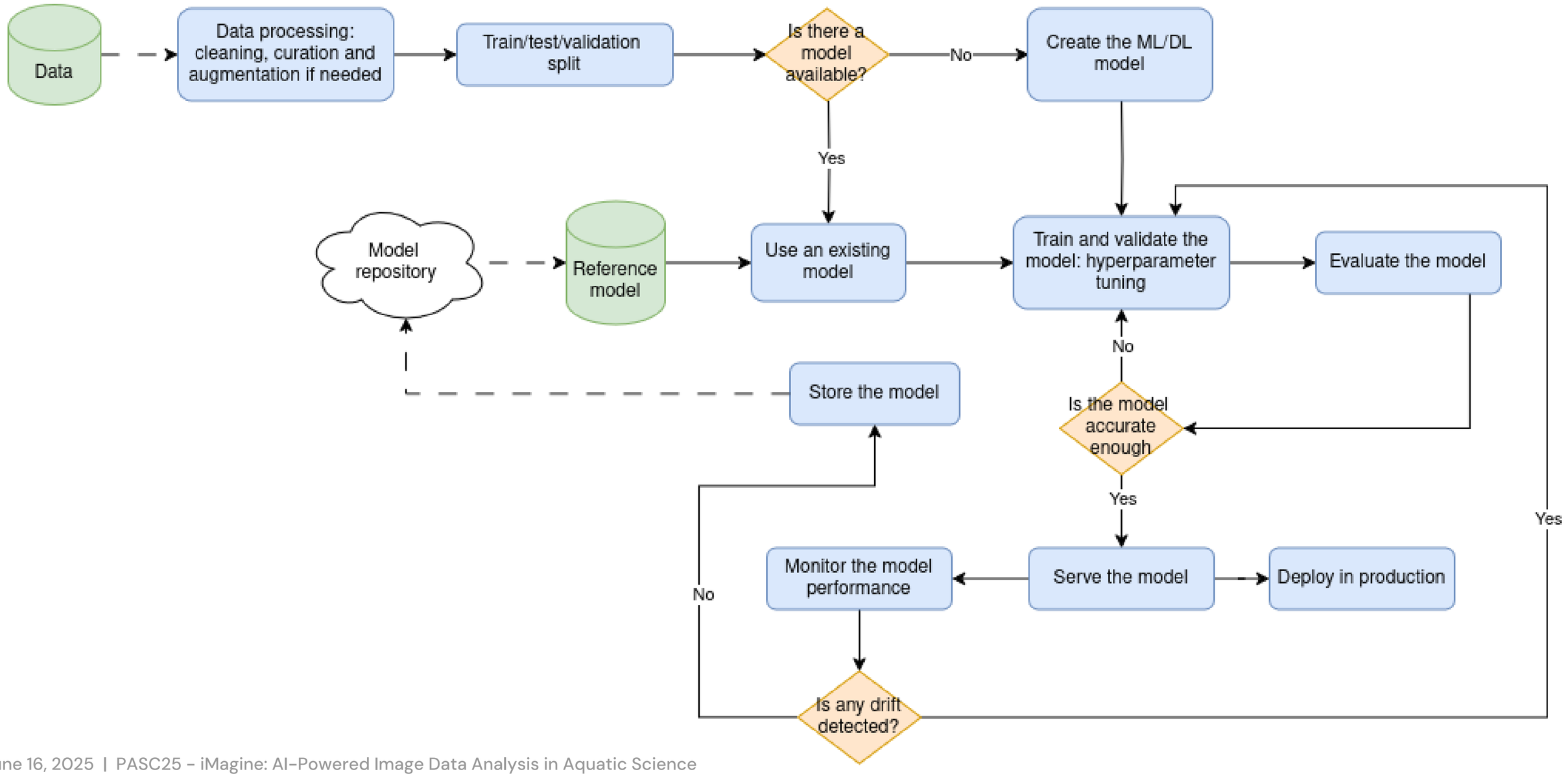
AI4OS Architecture



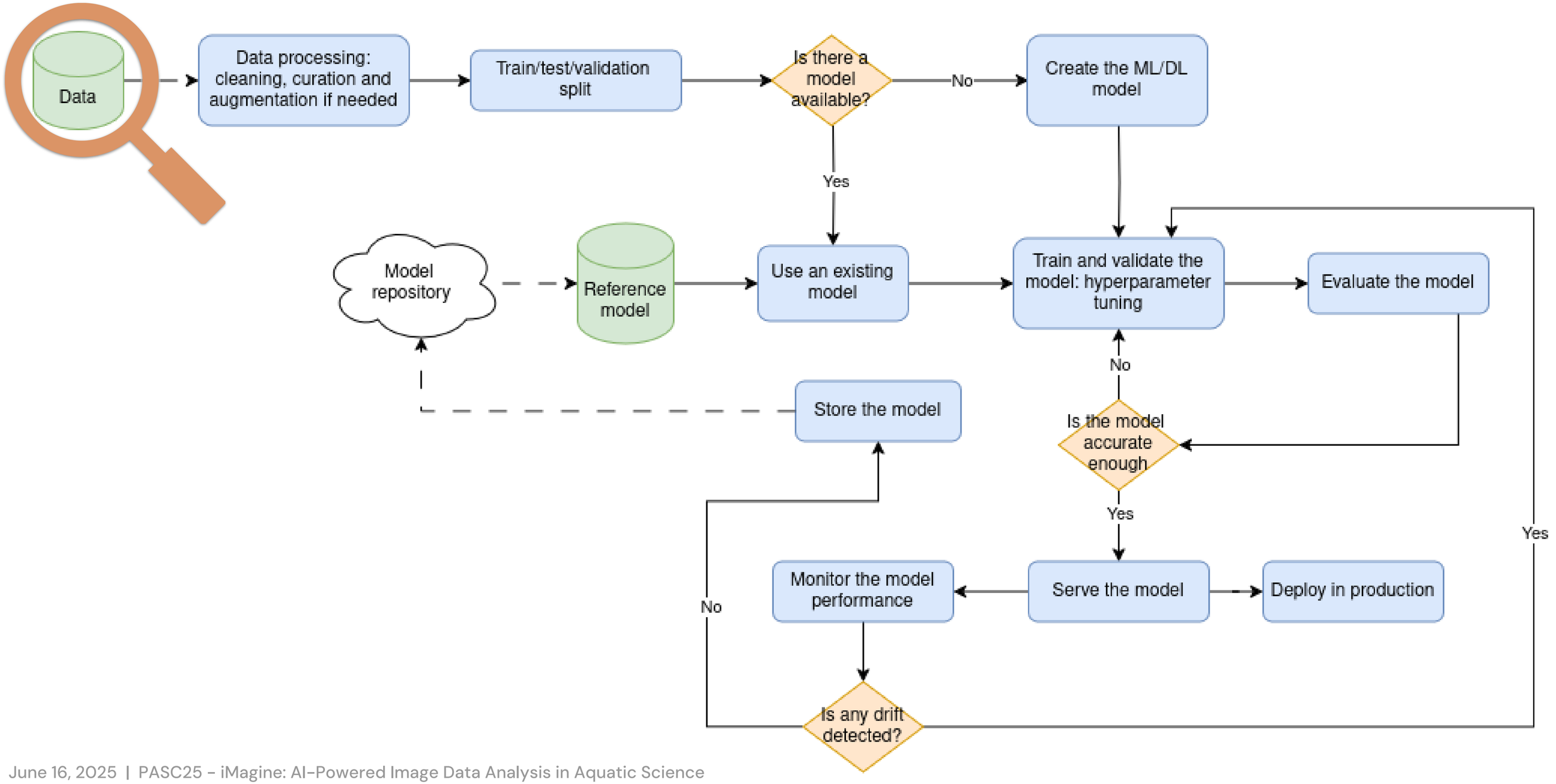
Tools and Infrastructures



AI/ML Application Development

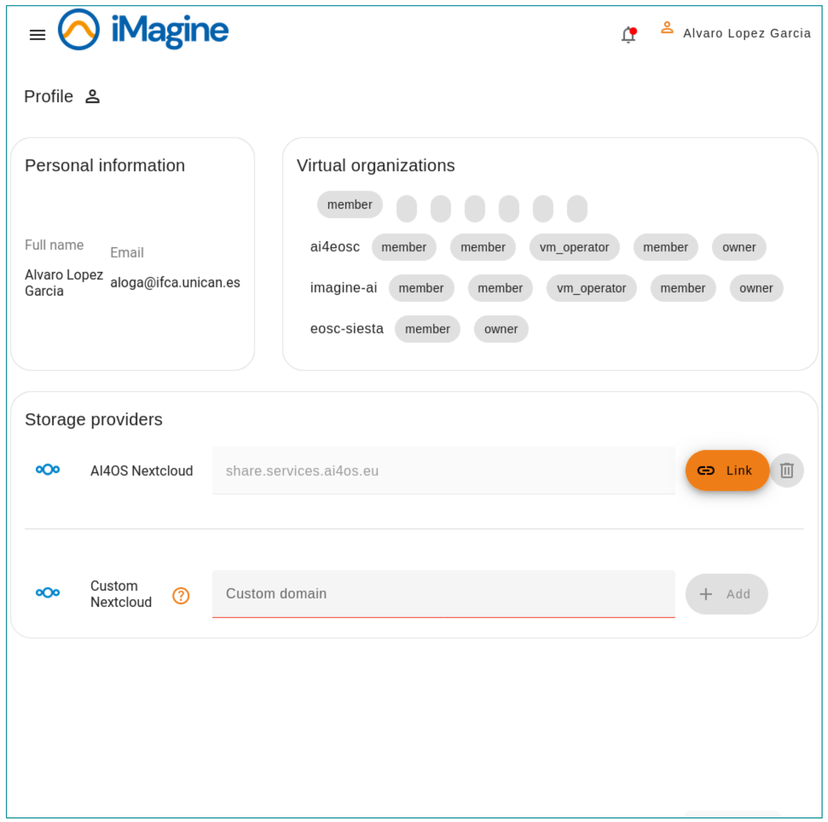


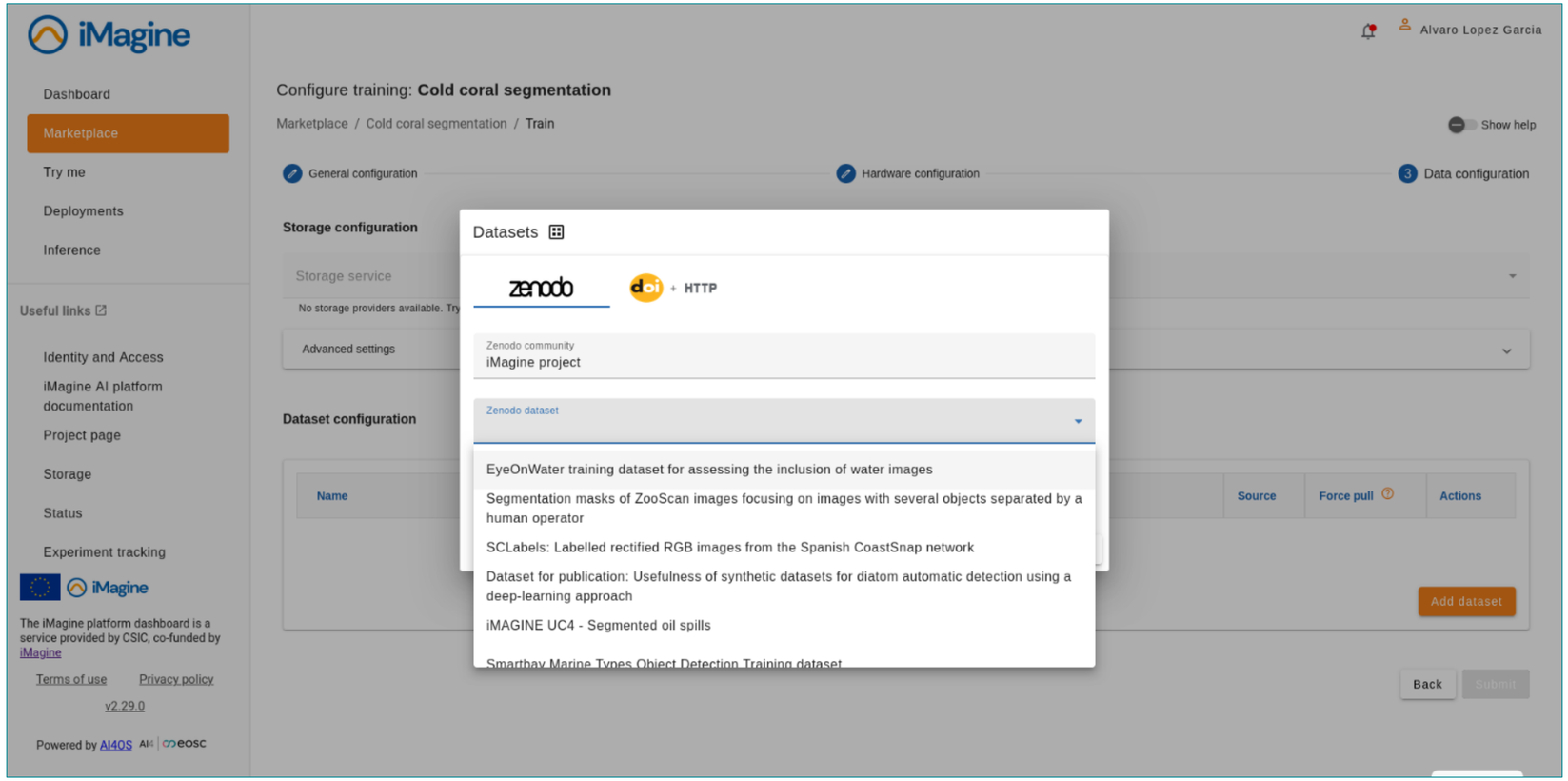
AI/ML Application Development



AI-Platform Features

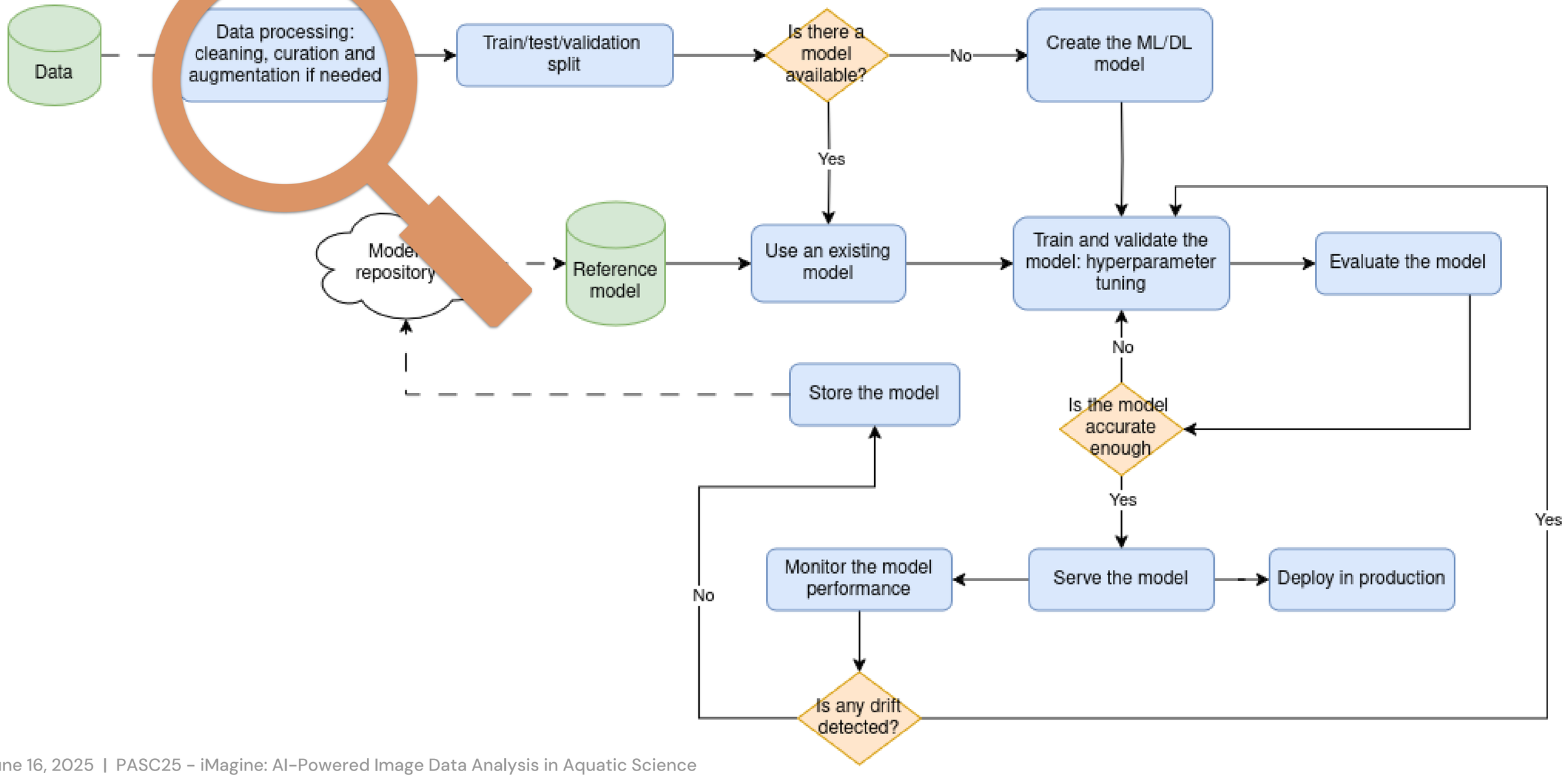
Storage Services & Data Synchronization





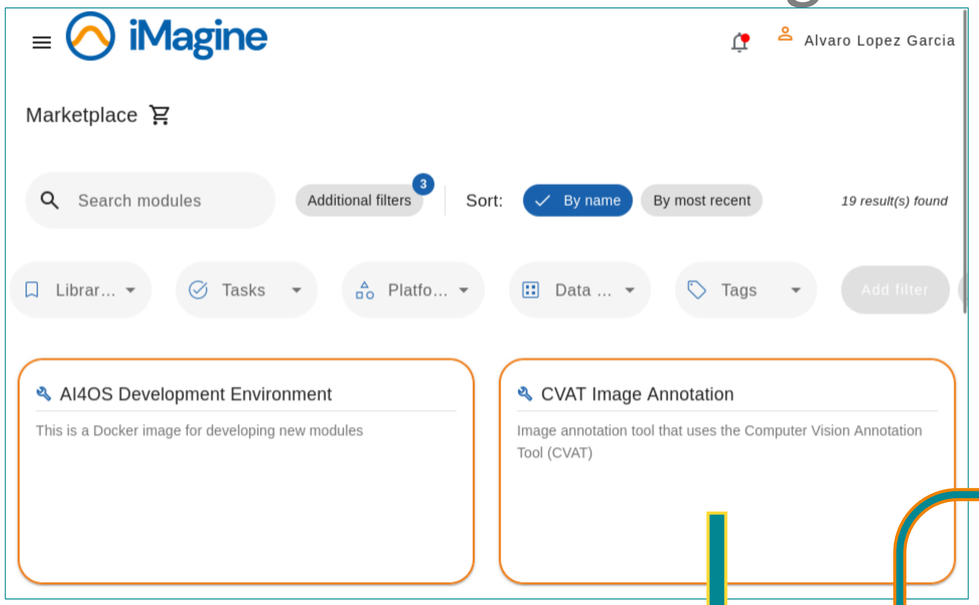
- The platform allows to link with external storage systems (currently NextCloud)
- Once linked, it allows to download and sync data from external sources (default Zenodo) via DOI
- FAIR evaluation of datasets performed via FAIR-EVA: https://fair.dev.imagine-ai.eu/en

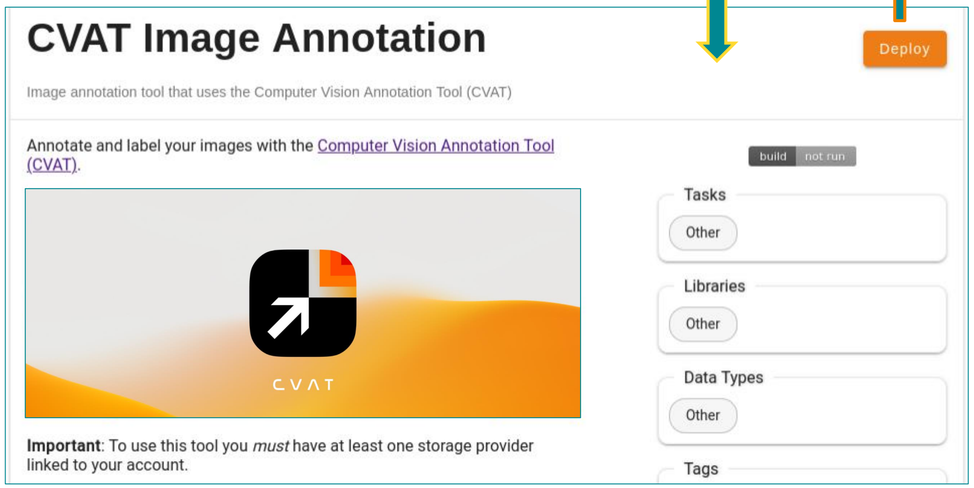
AI/ML Application Development

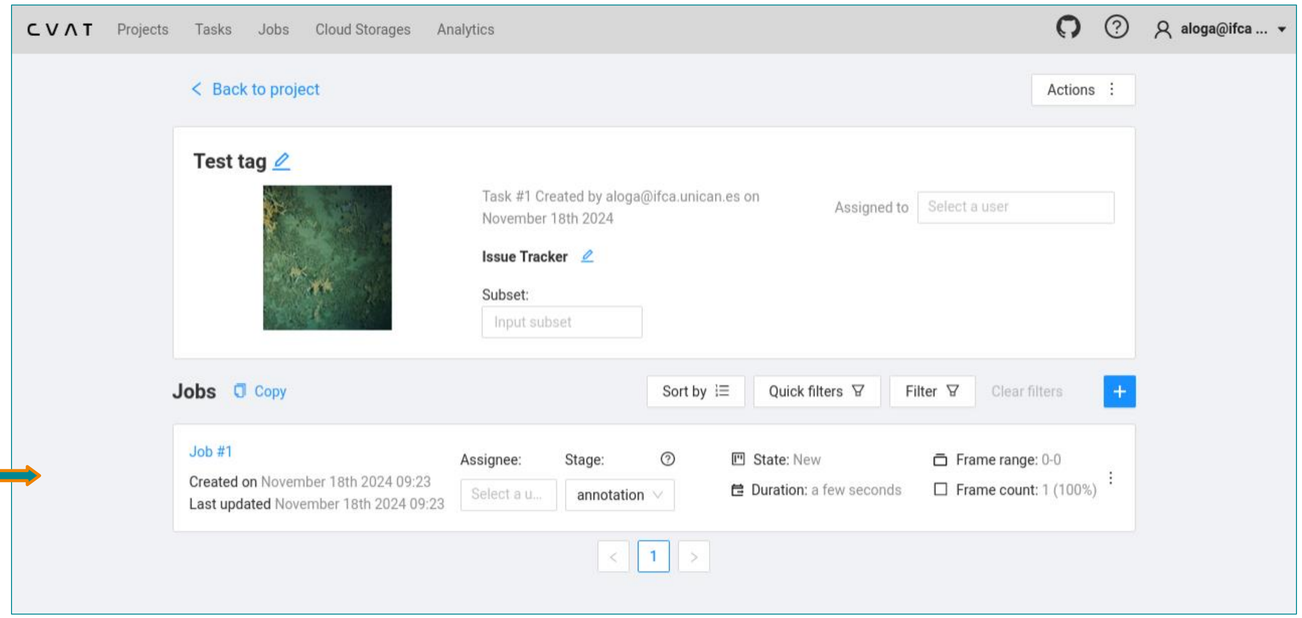


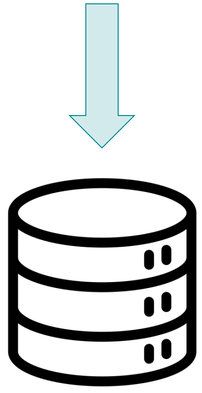
Data Preparation

Image Annotation Tool via Dashboard



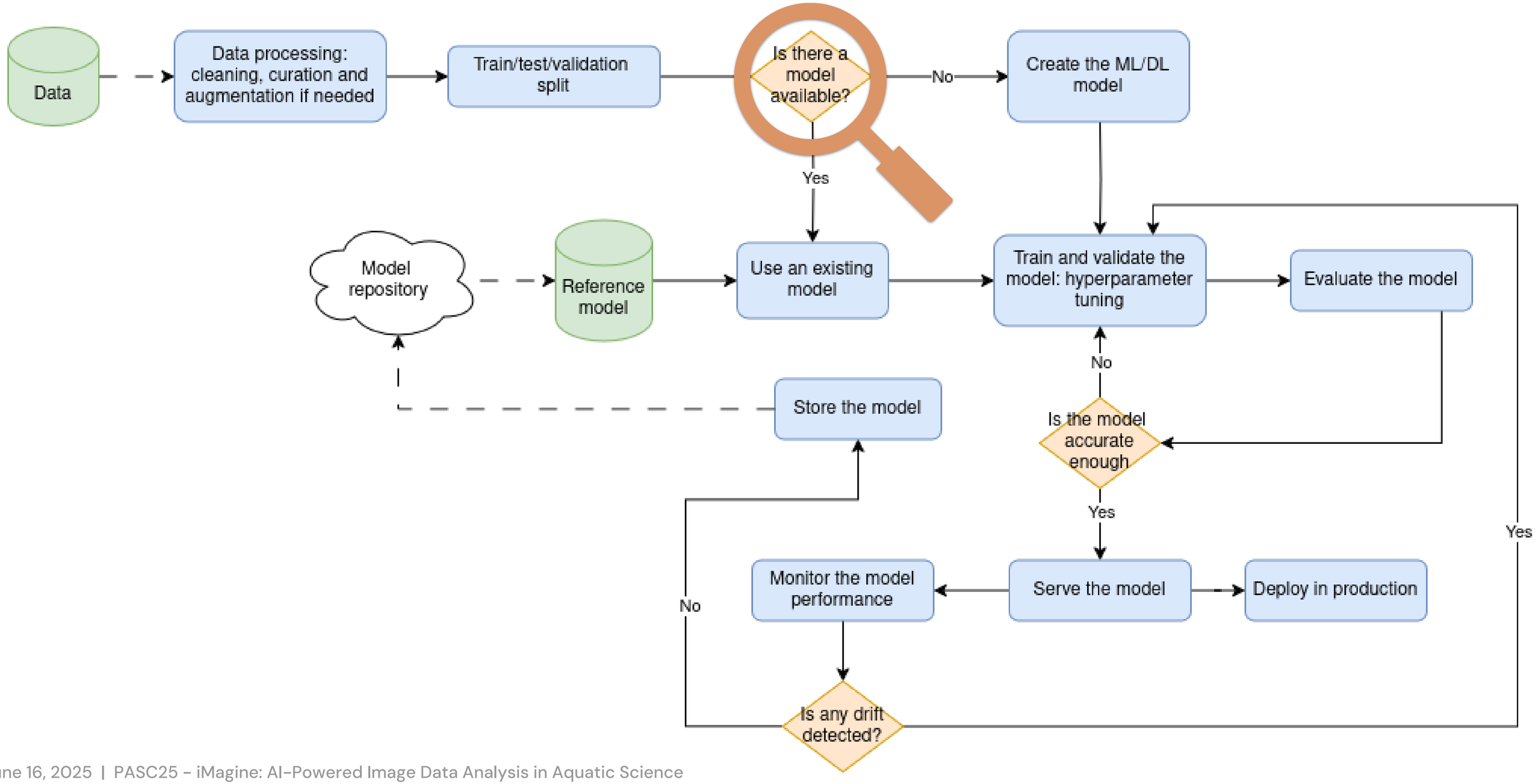




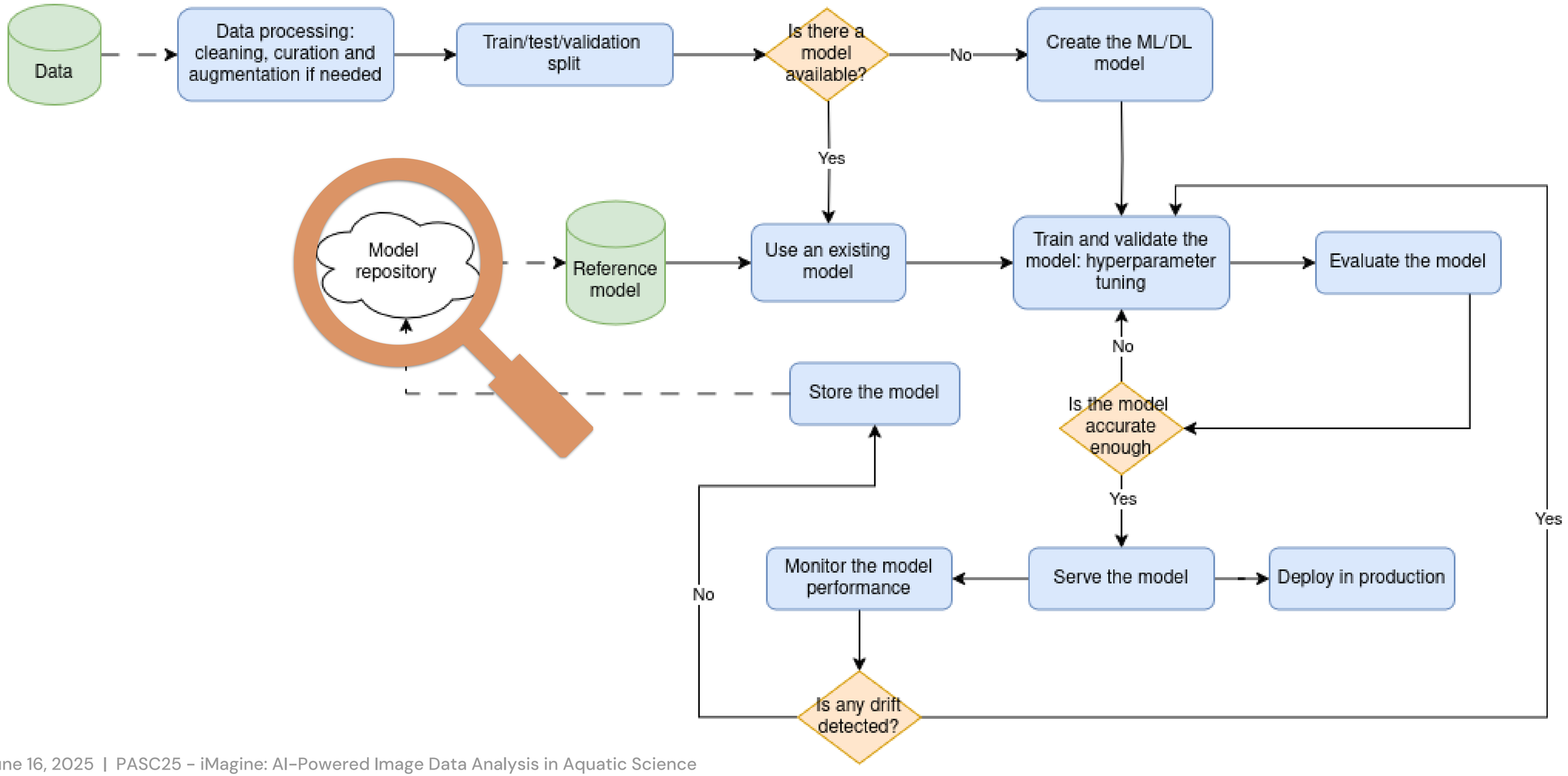


- Annotations synced with linked storage https://share.services.ai4os.eu
- No need to work offline and sync data with the platform

AI/ML Application Development

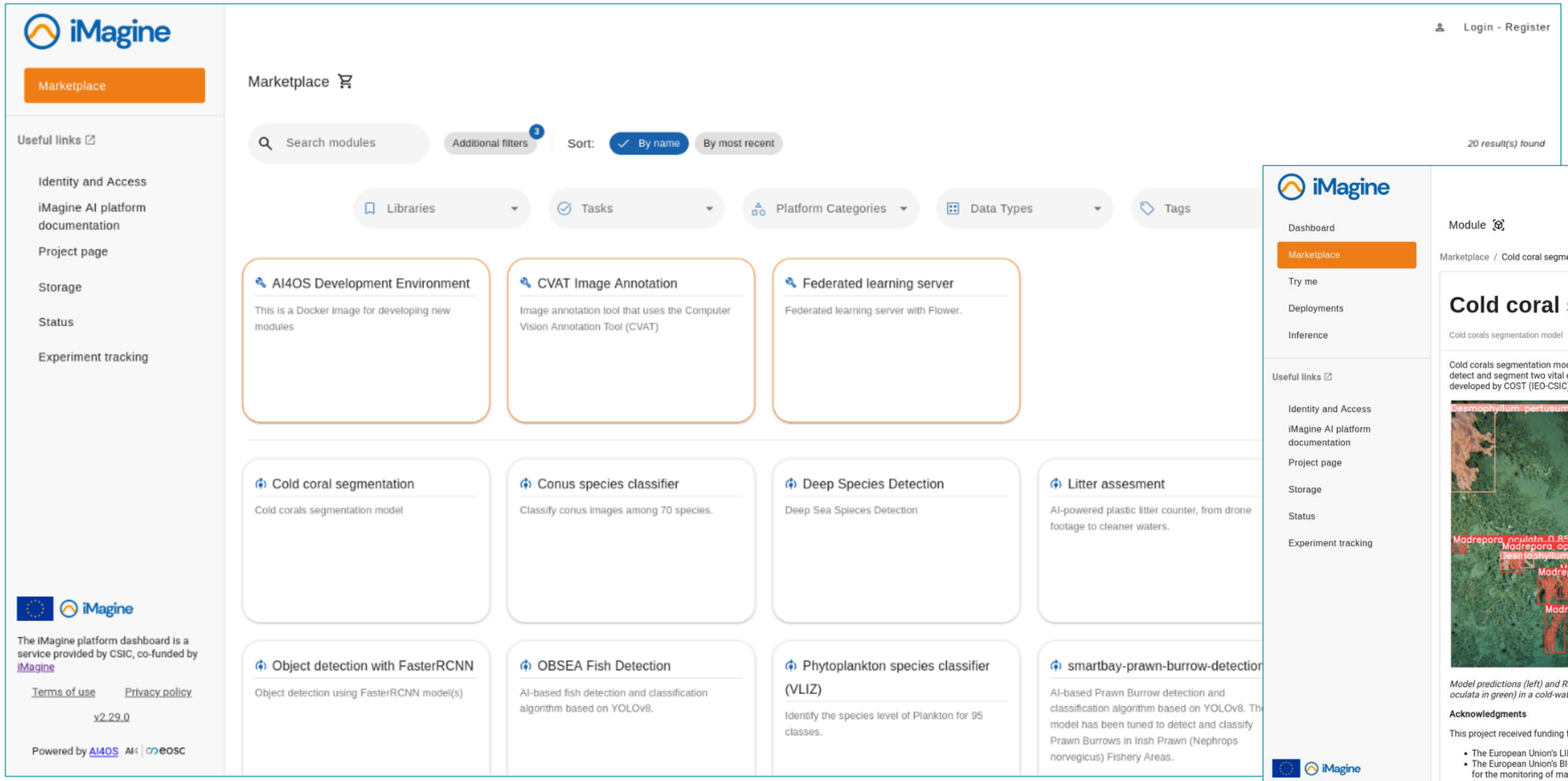


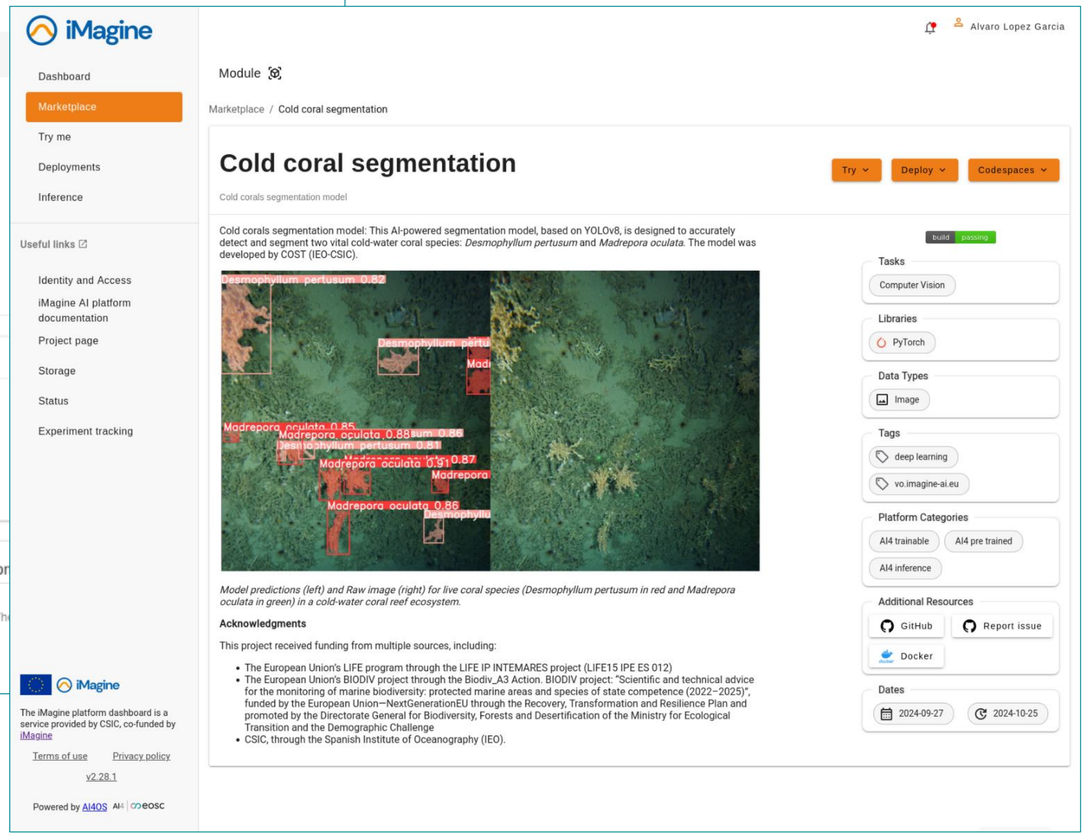
AI/ML Application Development



Models Available

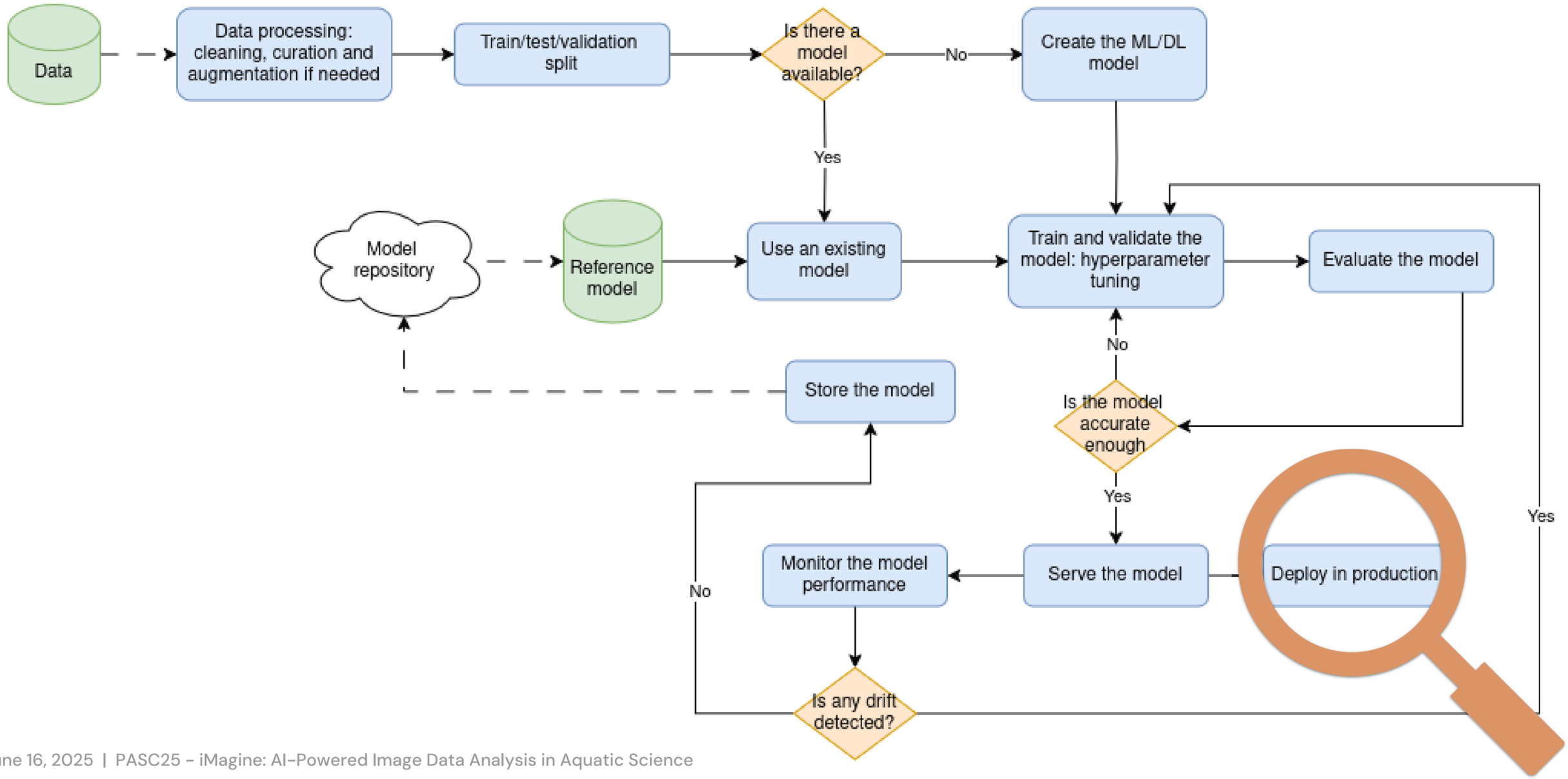
Dashboard & Marketplace





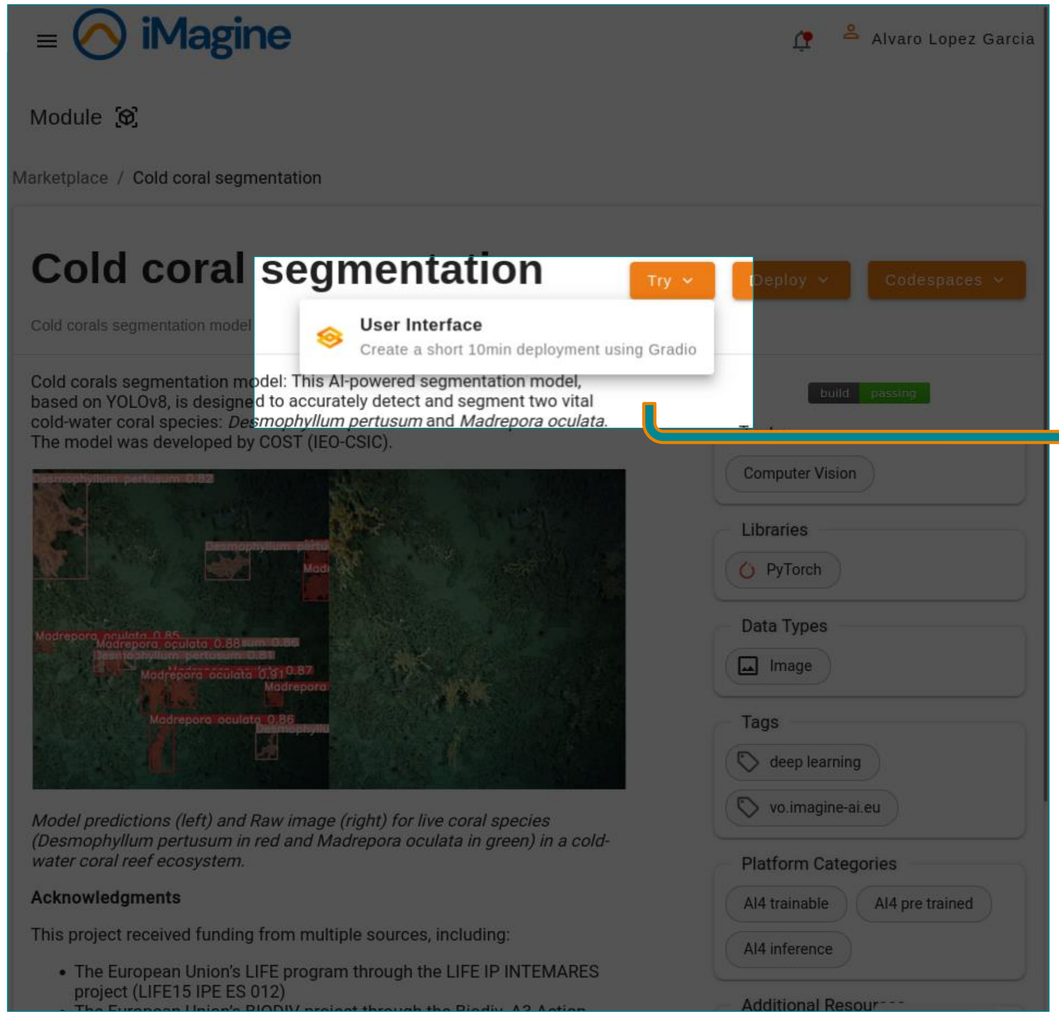
https://dashboard.cloud.imagine-ai.eu

AI/ML Application Development



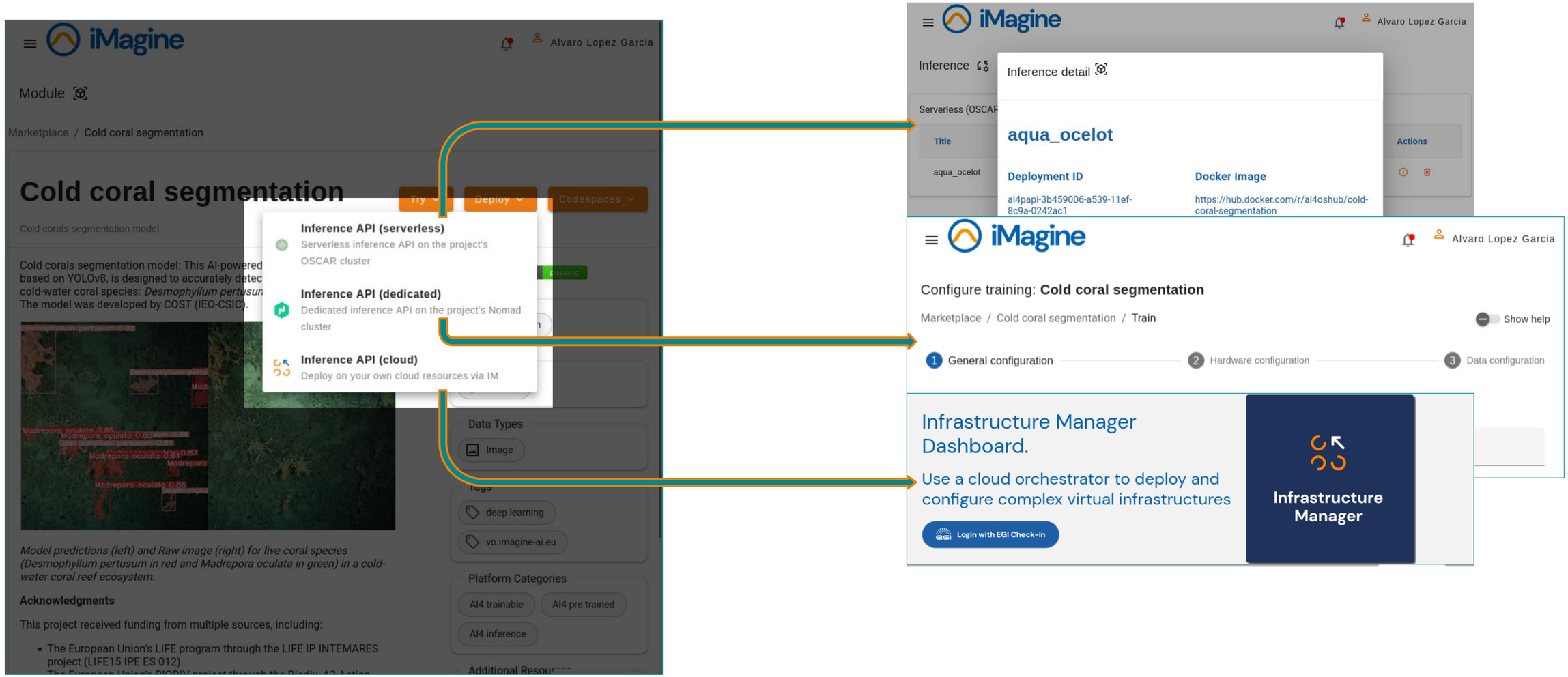
Model Deployment & Serving

Dashboard for Model Management

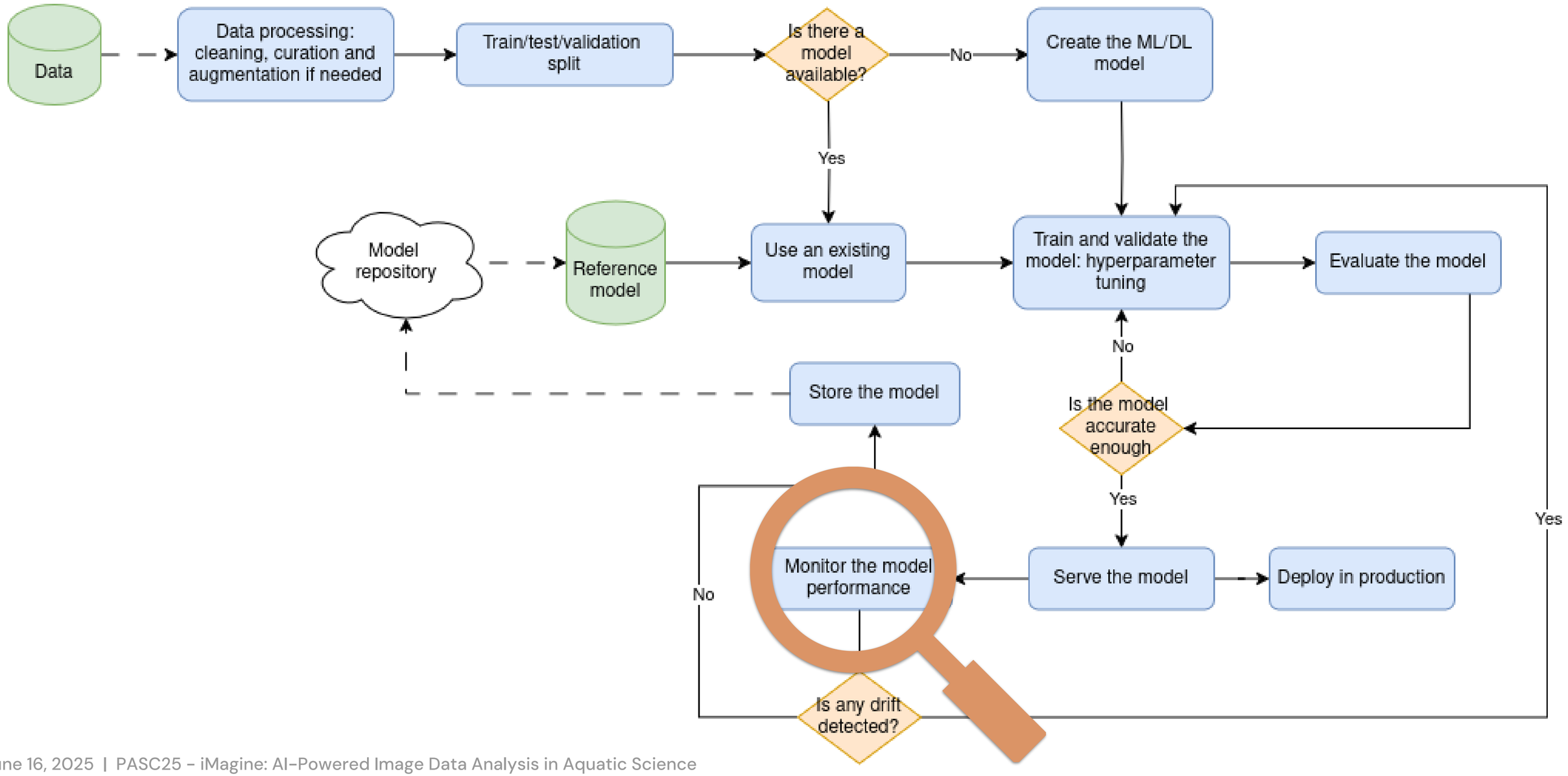


Model Deployment & Serving

Dashboard for Model Management

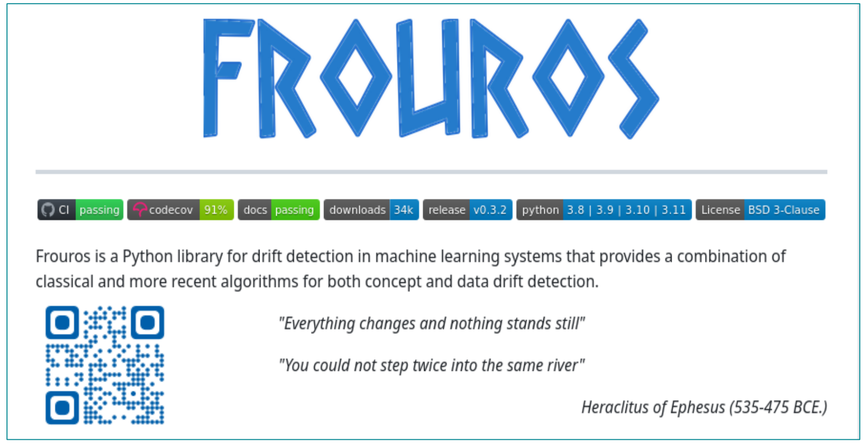


AI/ML Application Development



Monitoring the Model Performance

Drift Detection

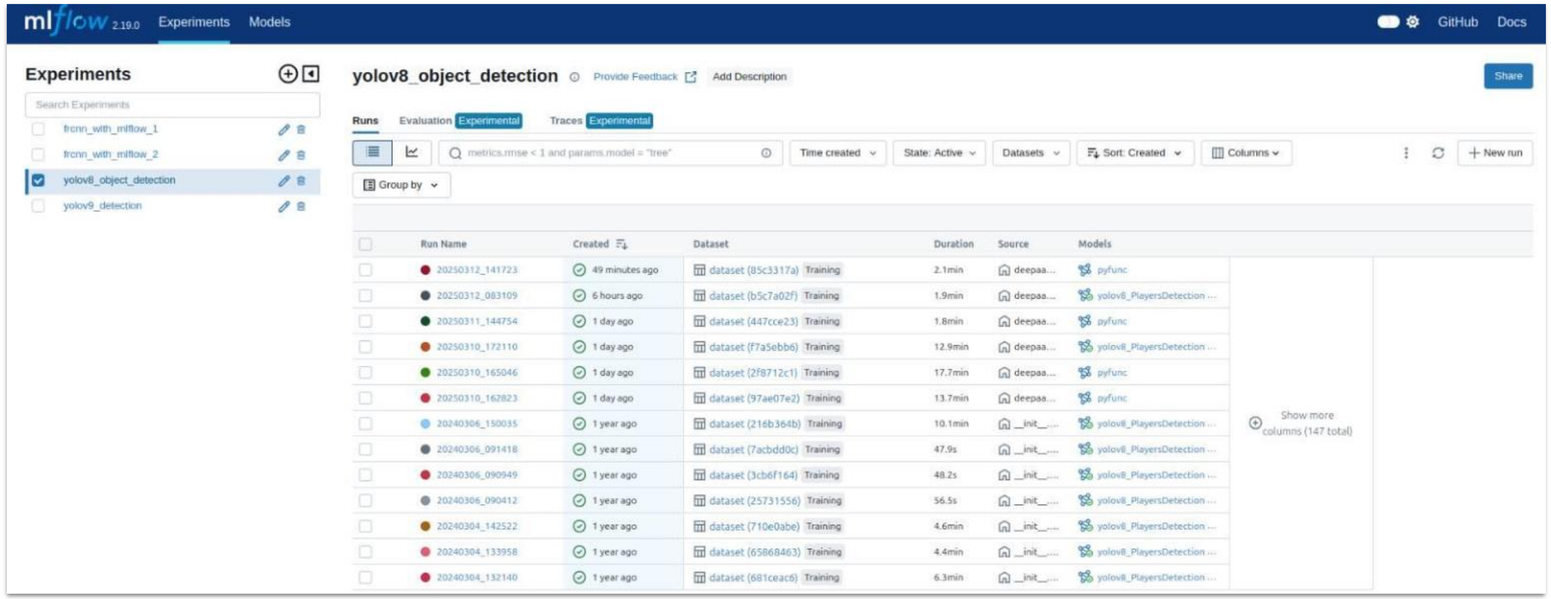


- Monitoring of models in production is not enough
  - Model learns from data, data is not stationary
  - Concept learnt by the model may change over time
- Data and concept drift detection is essential to build more robust models
- Frouros: A library for drift detection in ML
- Drift monitoring (DriftWatch): Monitor data drift change over time from multiple Drift detection libraries

Monitoring the Model Performance

MLflow Experiment Tracking Tool

- Package code into reproducible runs, and sharing and deploying models
- Track experiments to record and compare parameters and results
- Manage and deploy models from a variety of ML libraries to a variety of model serving and inference platforms

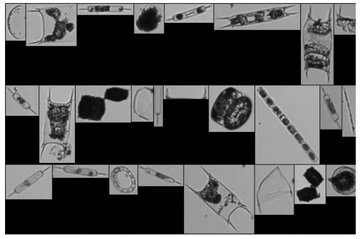


FlowCam Phytoplankton Identification

Case Study

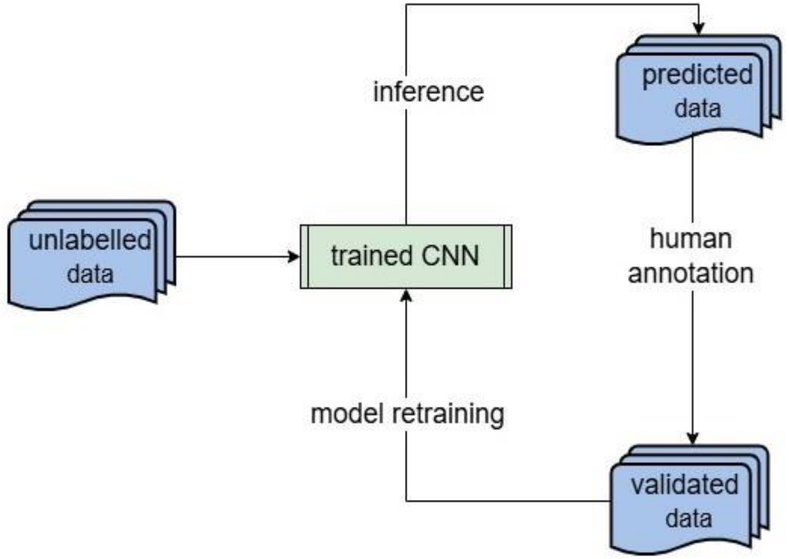
- Since 2017: monthly monitoring of phytoplankton in the Belgian North Sea

- Data acquisition using high-throughput imaging device: FlowCam



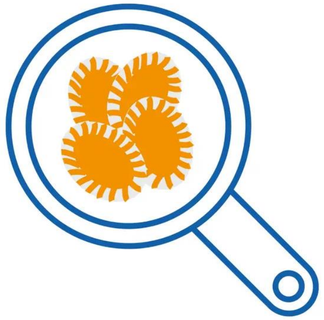
- 300 – 400,000 images to be analyzed annually + Infrastructure provisioning

- Semi-automated pipelines to process raw output data
- Trained CNNs + Manual validation check
- + 1.2 M validated FlowCam images



FlowCam Phytoplankton Identification

Objectives

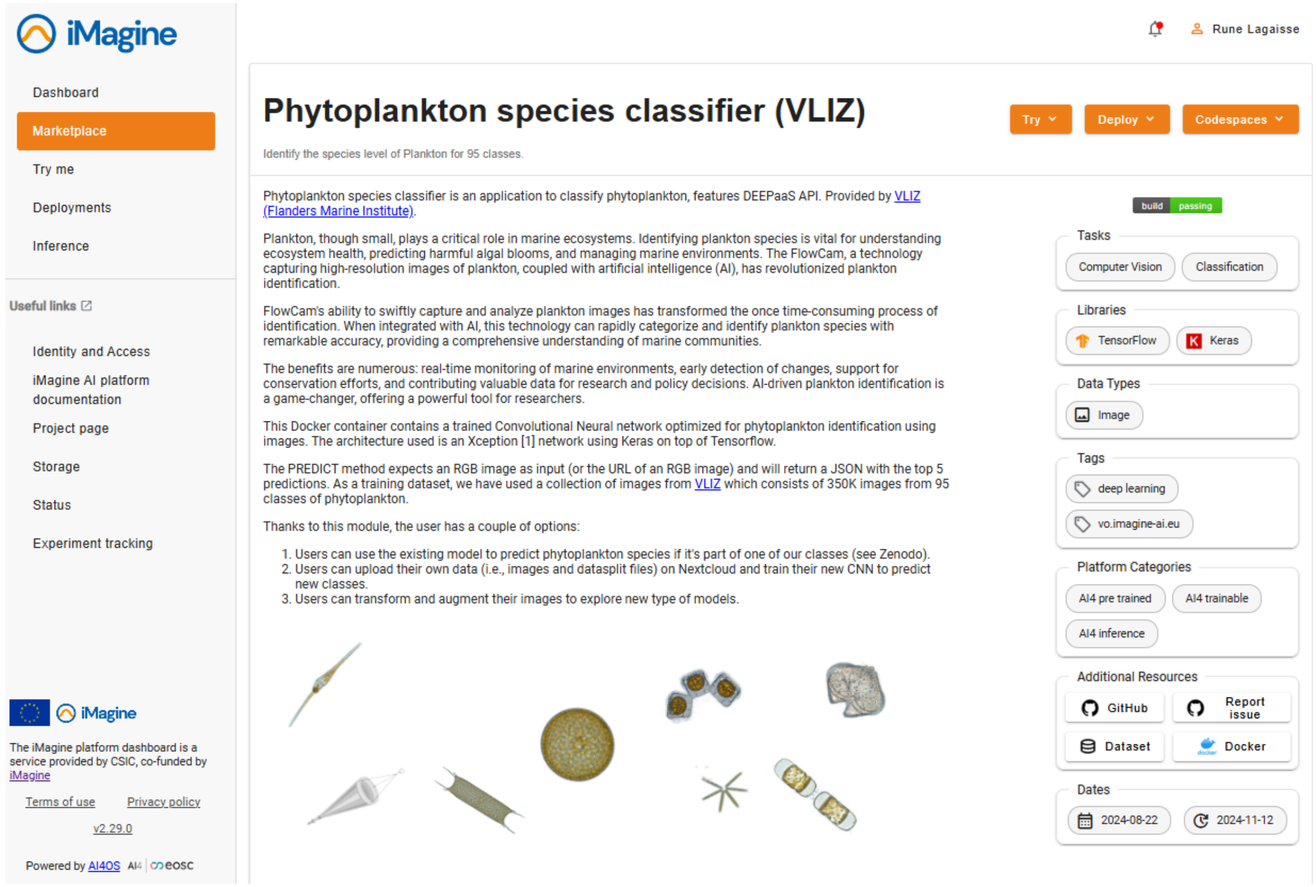


- Optimize existing pipeline from sensor to database
- Improve current metadata & data output formats towards compliance with community-based standards and vocabularies
- Improve the service to incorporate the context input and increase the classification accuracy
- Extend the training dataset by identification of additional particles currently grouped under a rest class
- Prepare the data and processing components for connection, synchronization and migration to enable access from the iMagine platform

Experience with the iMagine Platform

FlowCam Phytoplankton Identification Service

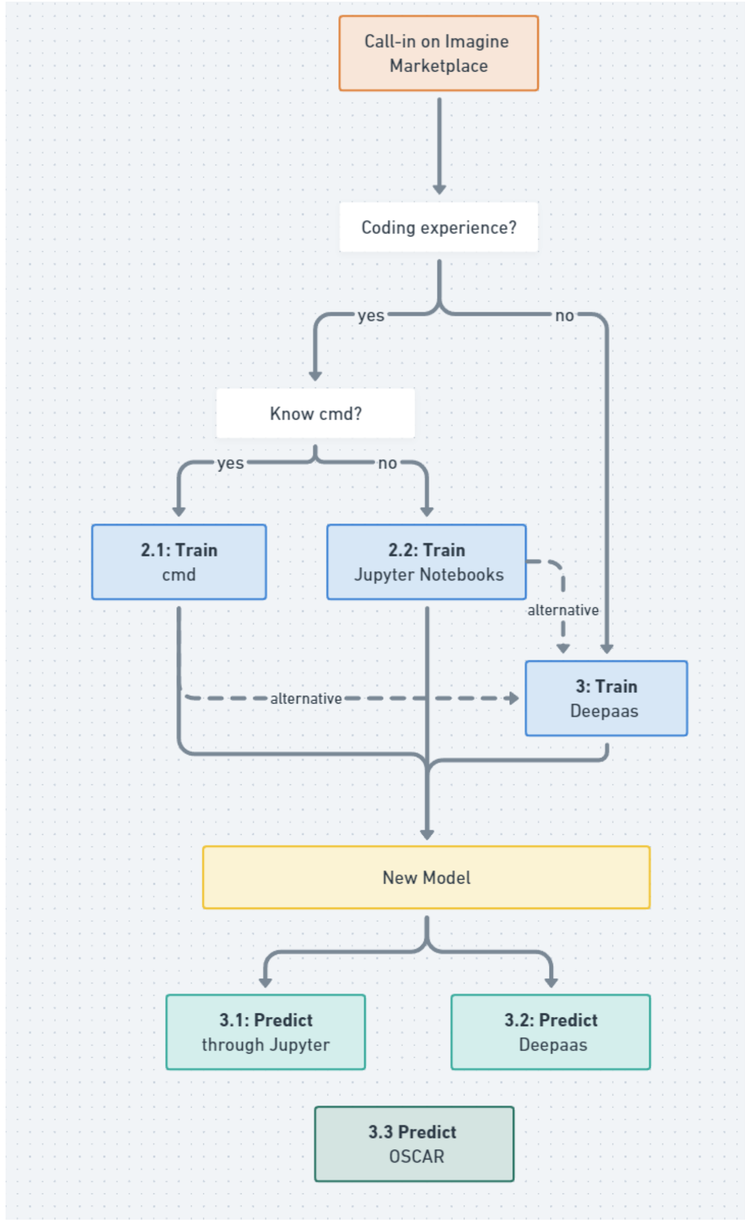
- User friendly
- Clear documentation and tutorials
- Efficient user support communication and follow-up
- Continuous improvement through regular updates
- Adequate storage space and computing resources for current needs



FlowCam Phytoplankton Identification

- Dashboard:
  https://dashboard.cloud.imagine-ai.eu/catalog/modules/phyto-plankton-classification

- Webinar:
  https://www.youtube.com/watch?v=GanTZ5nbxm0



Summary

- Empowering aquatic scientists with IT/AI infrastructure for healthy oceans
- Ready to use labeled datasets and AI-powered image analysis services
- Open-source AI platform and models
- Generated best practices for the aquatic science community
- Collaboration between researchers, developers, and infrastructure providers

Marketplace Best Practices imagine-ai.eu twitter.com zenodo.org

Thank you for your attention

imagine-ai-platform-support@mailman.egi.eu

imagine-ai.eu

Funded by the European Union

iMagine receives funding from the European Union's Horizon Europe research and innovation program under grant agreement No. 101058625