iMagine Competence Center: Facilitate collaboration between Scientists

Fahimeh Alibabaei, Elnaz Azmi, Valentin Kozlov

but also UPV, IFCA, EGI, …

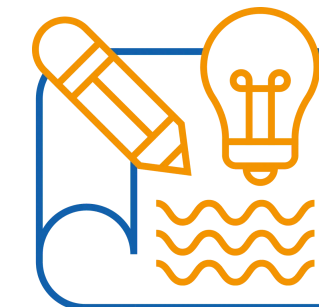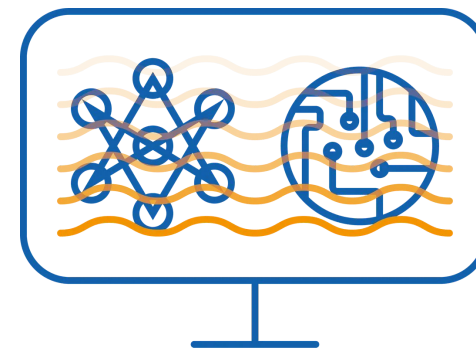Dissemination level: Public

02-06 June 2025

EGI2025, Santander, Spain

Outline

- iMagine Competence Center

- Use Cases

- Best Practices

- Summary

Competence Center:
Mission & Methodology

Mission: **Facilitate** the path from image **data** to imaging **applications** and **services** for scientists in aquatic sciences
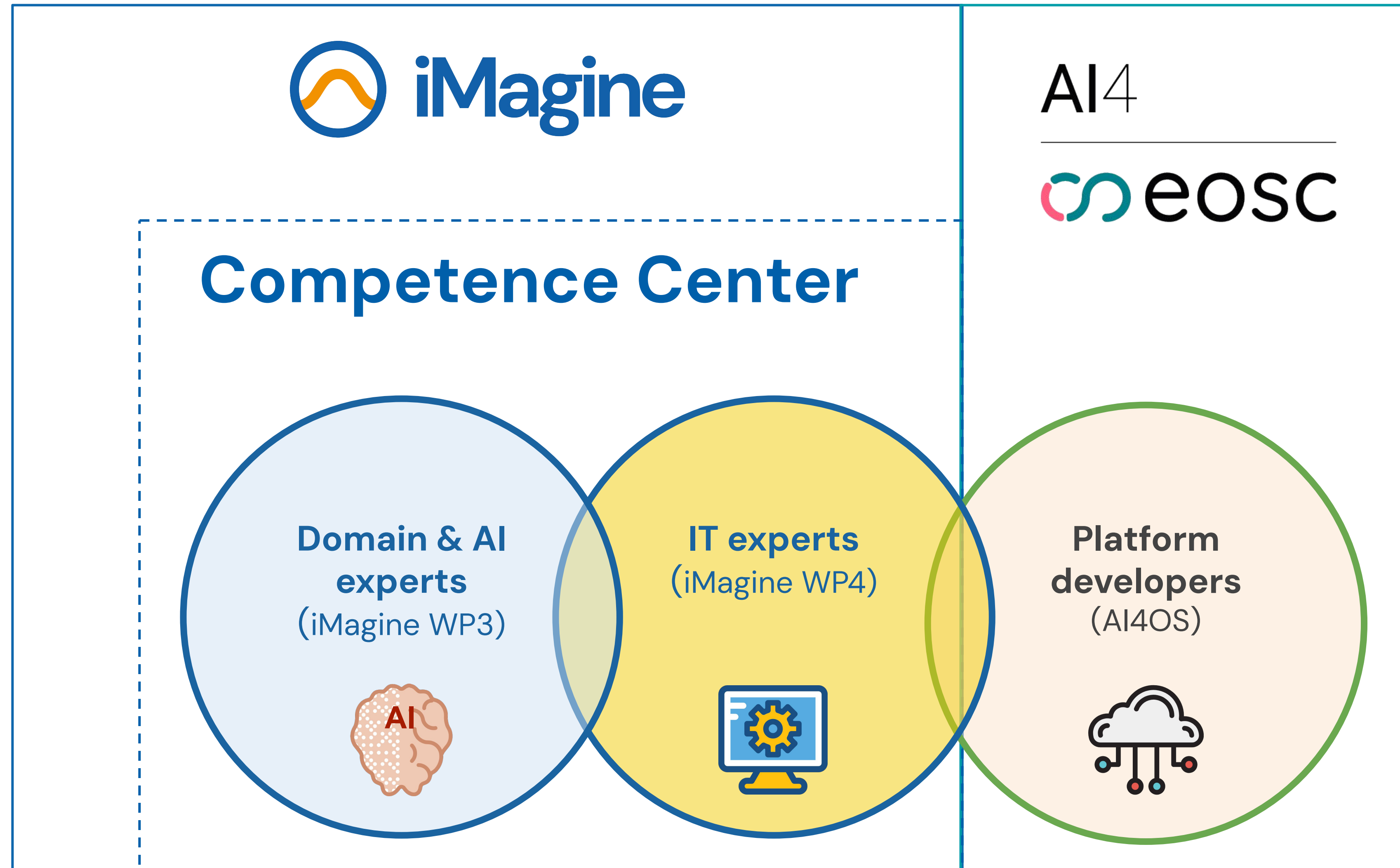
Methodology:
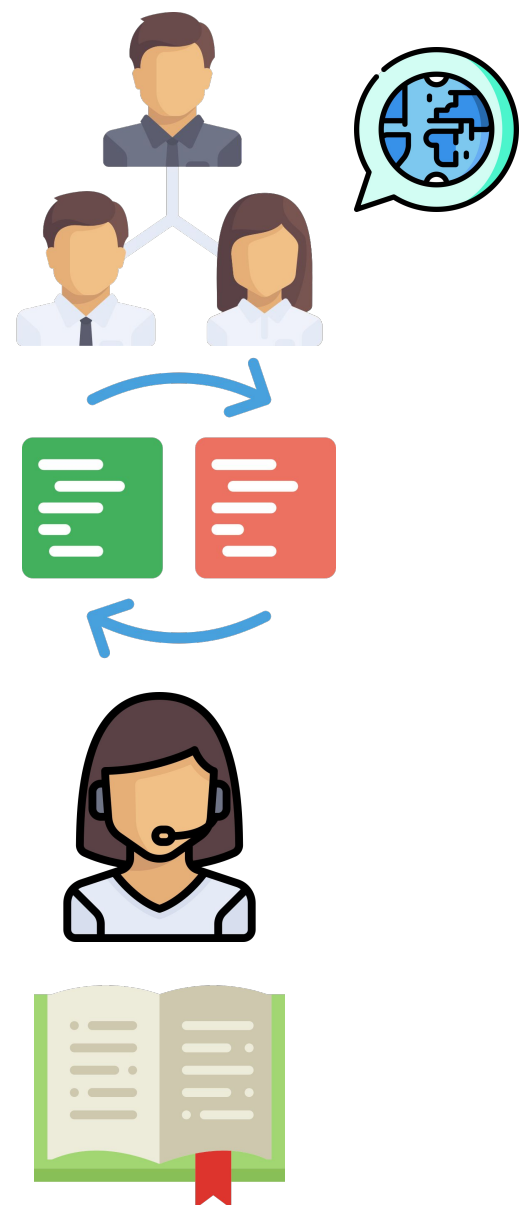
- Connect **experts** from different domains (science, AI, IT, …)

- Stimulate **knowledge exchange** between the parties

- Speak **common language**

- Provide **training** and **support**

  (data, AI, FAIR, platform, service management)

- Create **best practices**

Competence Center:

Synergetic approach

iMagine

AI4

eosc

Competence Center

Domain & AI experts
(iMagine WP3)

AI

IT experts
(iMagine WP4)

Platform developers
(AI4OS)

Competence Center:
Support

Knowledge exchange:
- Regular meetings with iMagine and Open Call use cases
- We try to understand each other & speak **same** language
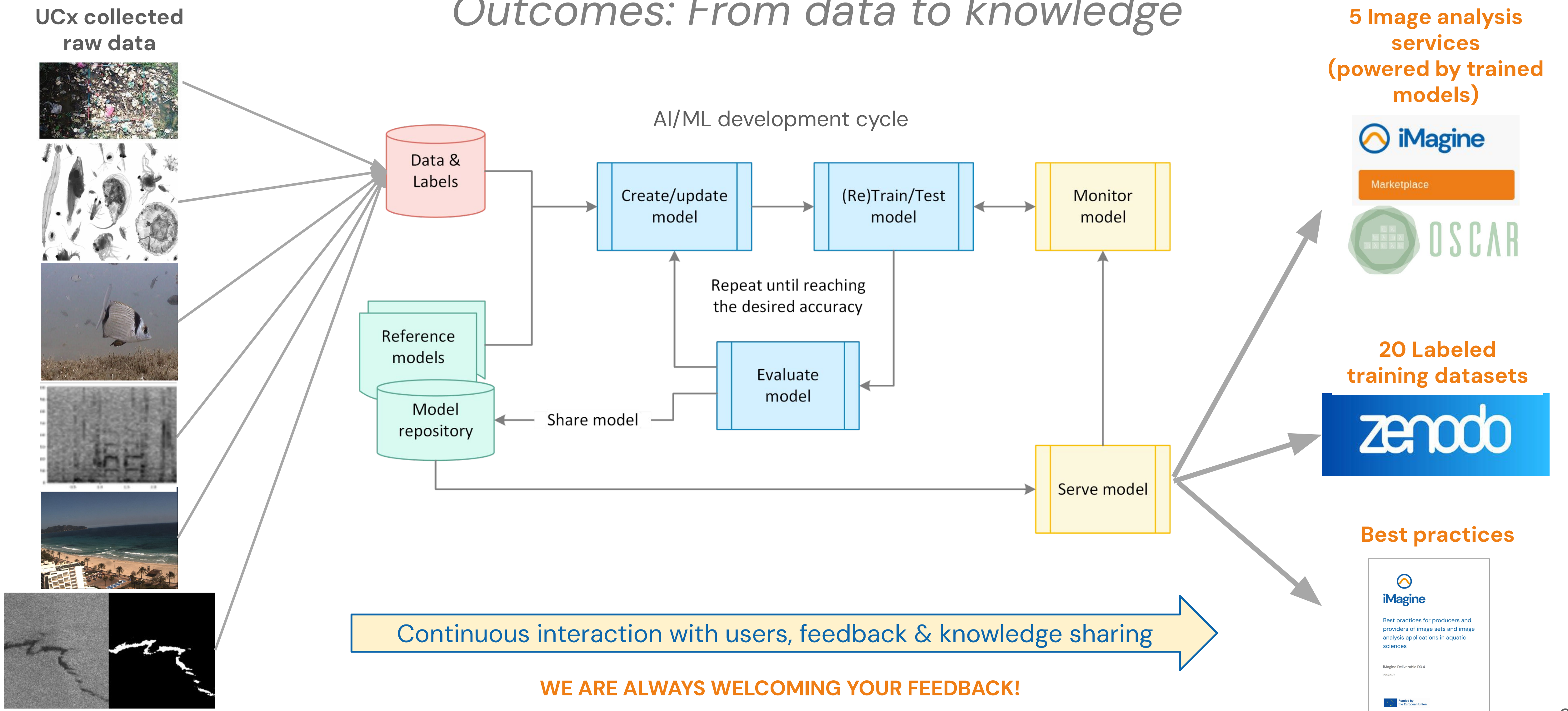
Training and **Support**:
- Competence center workshops ([2023](#), [2024](#), [2025](#))
- Dedicated [events](#) and [webinars](#)
- AI4EOSC platform [user workshop](#) (2023) and [webinars](#)
- [FitSM](#) course for imagine service providers (EGI)
- [Support email list](#), individual meetings and consultancy, e.g.
  - ▷ Marketplace publishing
  - ▷ Infrastructure (e.g. OSCAR, thanks to UPV!)

Best practices: [documentation](#), [deliverables](#), articles, [datasets](#)

Villefranche-sur-Mer 2023

Palma 2024

Brest 2025

5

Competence Center:

Outcomes: From data to knowledge

UCx collected raw data

5 Image analysis services (powered by trained models)

AI/ML development cycle

Data & Labels

Create/update model

(Re)Train/Test model

Monitor model

Repeat until reaching the desired accuracy

Reference models

Evaluate model

Model repository

Share model

Serve model

20 Labeled training datasets

Best practices

Continuous interaction with users, feedback & knowledge sharing

WE ARE ALWAYS WELCOMING YOUR FEEDBACK!

4th JUNE 2025 | EGI2025 – Addressing AI needs for processing of imaging data: the case of Aquatic Science and the iMagine project

6

Marketplace

iMagine Dashboard

iMagine use case modules

Open call use case modules

General purpose modules

How to Access

Try out the AI modules on the iMagine AI Platform for 10 minutes

- Check this tutorial first
- Log in on the iMagine AI dashboard with any account
- Select the module of interest
- Click on the TRY button on the right. A Gradio interface will open.

Request Access

Fill in the form here.

WE ARE OPEN FOR BUSINESS!

Thematic Services for Image Analysis

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

iMagine Services Access Request

iMagine has released its imaging analysis services for aquatic sciences. Be the first to request access and start using them.

In Google anmelden, um den Fortschritt zu speichern. Weitere Informationen

* Gibt eine erforderliche Frage an

E-Mail-Adresse *

Ihre E-Mail-Adresse

Your Name *

Meine Antwort

Your Affiliation *

Meine Antwort

I work in a project and want to get access to one or more services (use the text box below to give us more details!)

Meine Antwort

Select the Service You Want to Have Access *

○ Litter Assessment: Identify Floating Plastic for a Cleaner Future
○ ZooProcess: Unveiling the Secrets of Plankton
○ AI-Powered Fish Monitoring for Healthy Oceans: EMSO-OBSEA
○ AI-Powered Fish Monitoring for Healthy Oceans: EMSO-Azores
○ AI-Powered Fish Monitoring for Healthy Oceans: EMSO-SmartBay
○ WITOIL: Predicting Oil Spills for Cleaner Seas
○ Uncovering Phytoplankton with AI

Select the Type of Access You're Interested In *

☐ Analyse: Analyse Your Whole Image Set With the Analysis Service
☐ Re-training – Re-train the image analysis model with your training images to reach more precise predictions for your use case.
☐ Deploy – Download the trained image analysis model and use it on local computing resources – either for image analysis or for re-training.

4th JUNE 2025 | EGI2025 – Addressing AI needs for processing of imaging data: the case of Aquatic Science and the iMagine project

8

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

4th JUNE 2025 | EGI2025 – Addressing AI needs for processing of imaging data: the case of Aquatic Science and the iMagine project

9

Best Practices:

Covered topics

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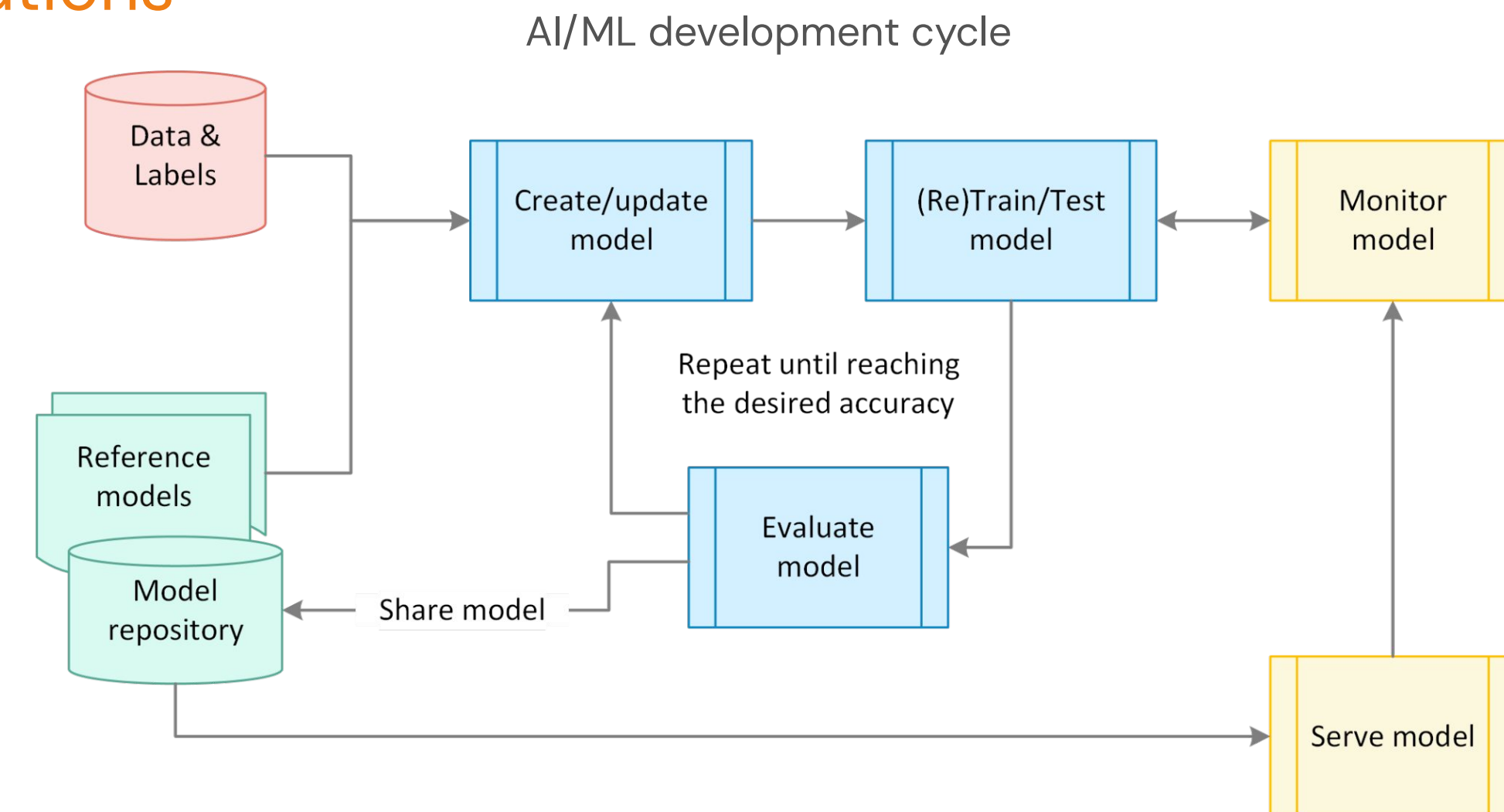- Performance Metrics and Evaluation Methods
- Tools for Monitoring Model Performance
- Data Biases and Fairness in Aquatic Science Models and Data
- Model Delivery
- AI Model Drift Tools
- Use Case Experiences

AI/ML development cycle

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**

Tools and Infrastructures
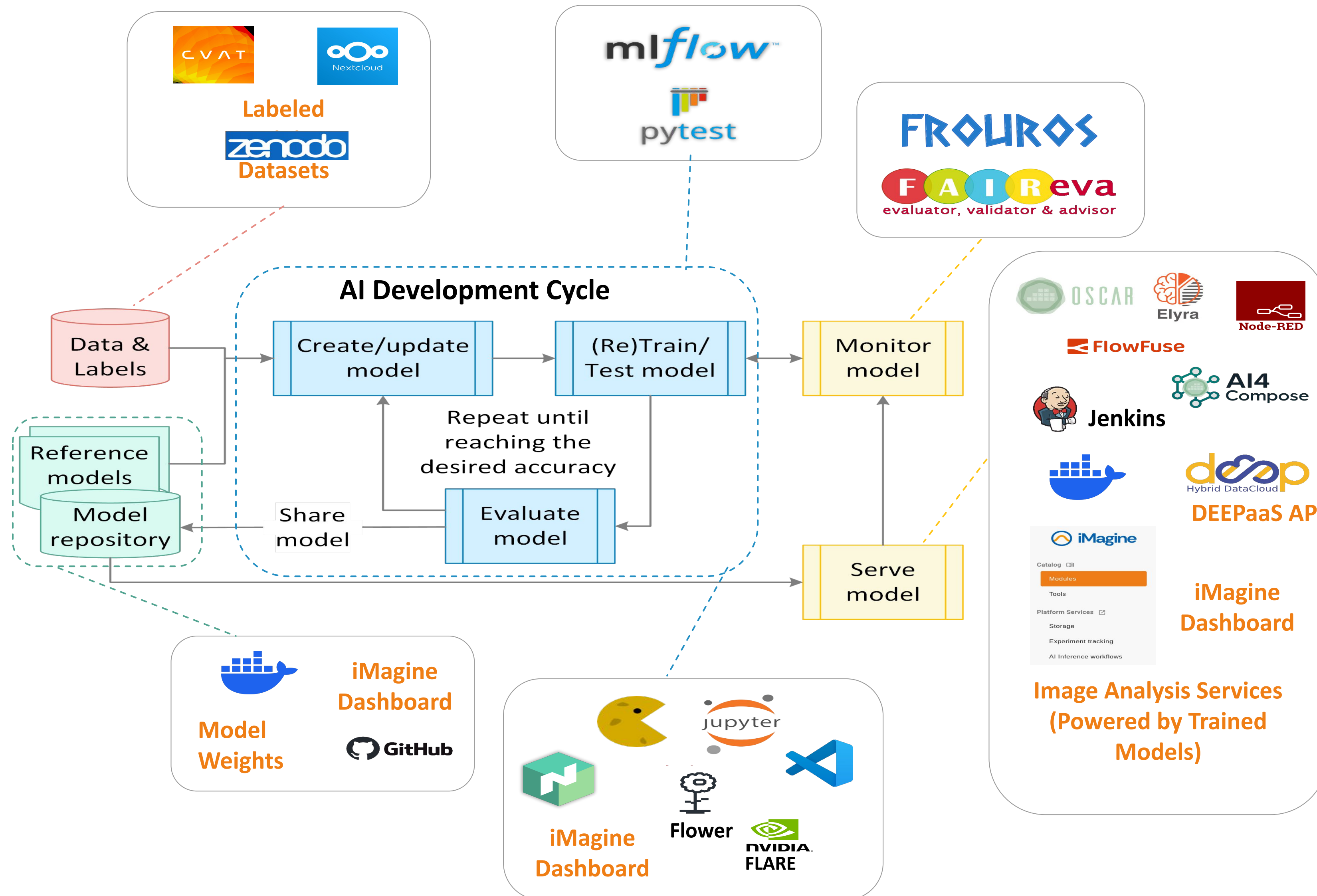
Summary

- The Competence Center of iMagine facilitates the transition from image data to imaging services for scientists in aquatic sciences through various means.
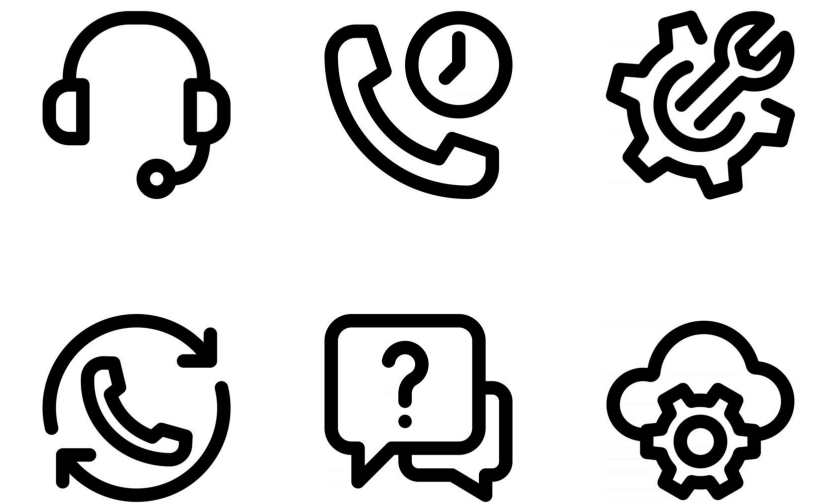
- **Key Outcomes of Collaboration within the Competence Center:**

  - Collaboration between researchers, developers, and infrastructure providers
  - AI–powered image analysis services
  - Support and training for scientists
  - Best practices for the aquatic science community

Marketplace   Best Practices   imagine-ai.eu

twitter.com   zenodo.org

![iMagine logo]

Thank you for your attention

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

imagine-ai.eu | 🐦 in

Best Practices:

Data and Labels

AI/ML development cycle

Annotation Tools

- ■ CVAT: user-friendly interface, robust annotation features

CVAT

15

4th JUNE 2025 | EGI2025 - Addressing AI needs for processing of imaging data: the case of Aquatic Science and the iMagine project

Use Cases Overview

5 mature + 3 prototypes

Aquatic Litter Drones
(DFKI, MARIS, OGS)

Zooscan – EcoTaxa pipeline
(Sorbonne Uni.)
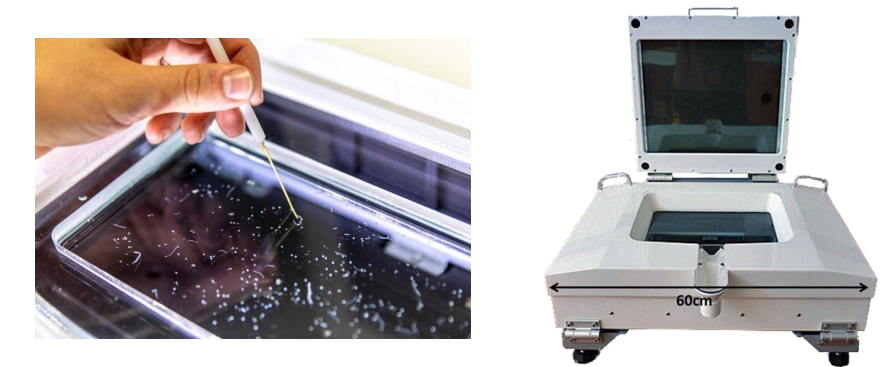
Marine Ecosystem Monitoring
(EMSO ERIC, UPC, IFREMER, MI)
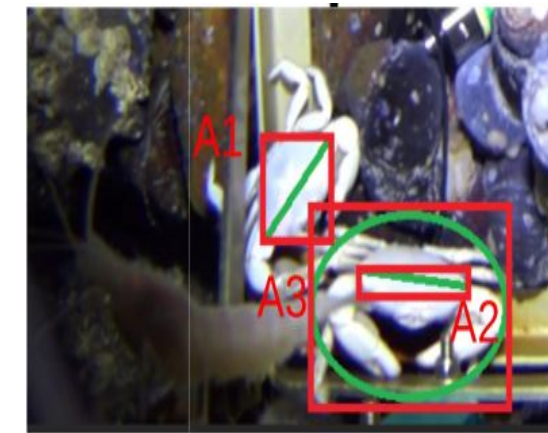
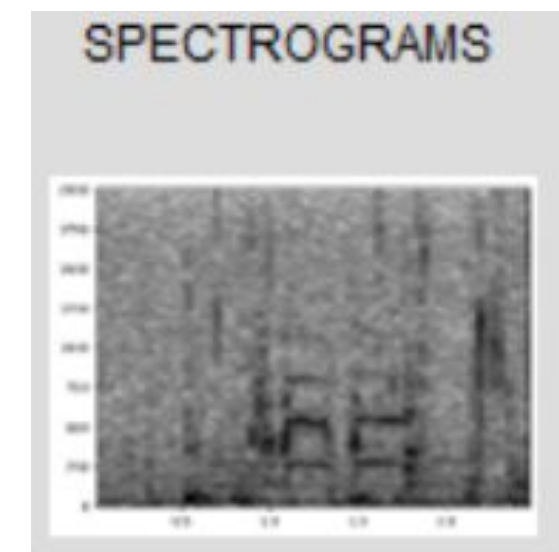**Oil Spill Detection**
(CMCC, OrbitalEOS, UNITN)
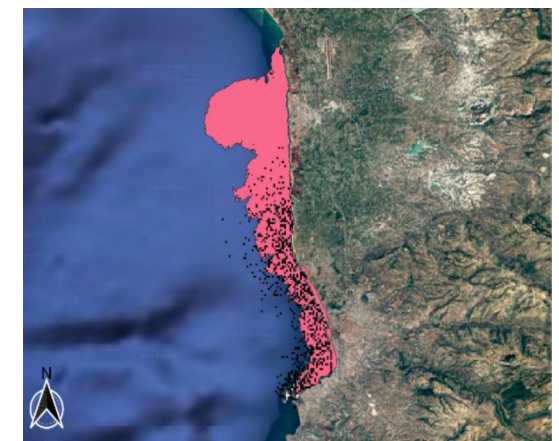
Flowcam phytoplankton identification (VLIZ)
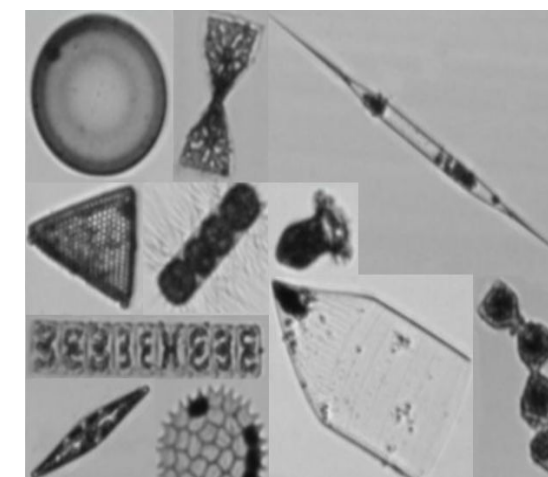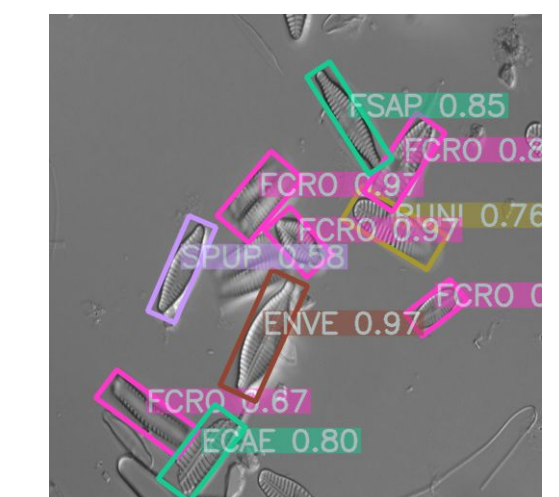
Underwater noise identification
(VLIZ)

Beach monitoring (SOCIB)

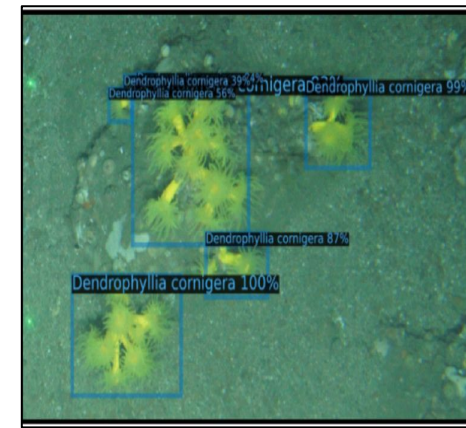**Freshwater diatoms identification** (UL–LIEC)

16

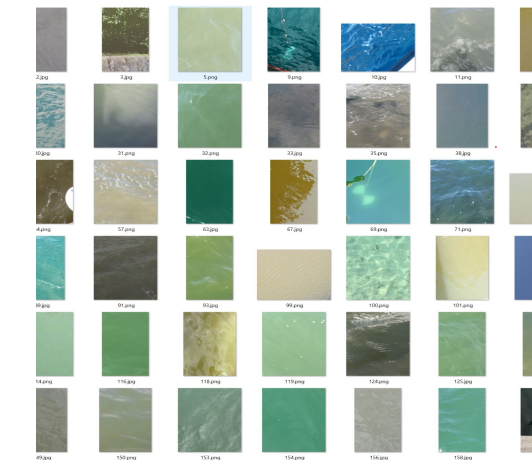# Open Call Use Cases

Onboarded via Open Calls

Improving knowledge about Cold Water Coral Reef
Spanish Oceanographic Institute (ES)

Validating images collected with the EyeonWater app
MARIS (NL)

Satellite–Derived Bathymetry
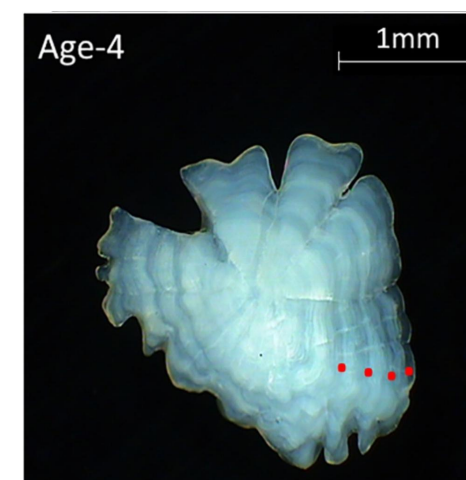Institute of Marine Sciences of Andalusia (ES)

DEcentrAlised Learning (DEAL)
Plymouth Marine Laboratory, UK

AI for image based age reading from fish otoliths
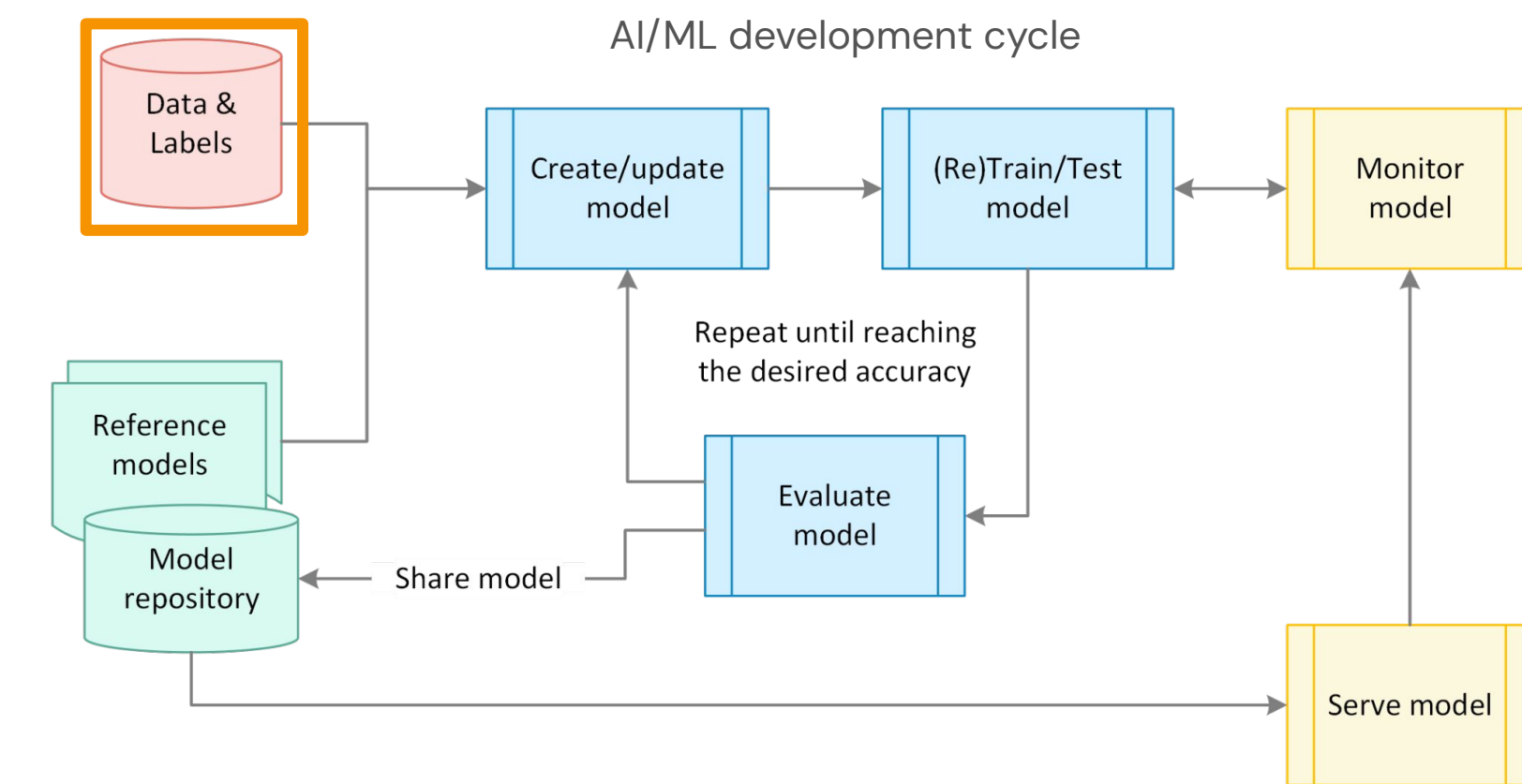Danish National Institute of Aquatic Resources (DK)

Sea Wave and coastal inundation detection Methodology (SWiM)
LNEC PT

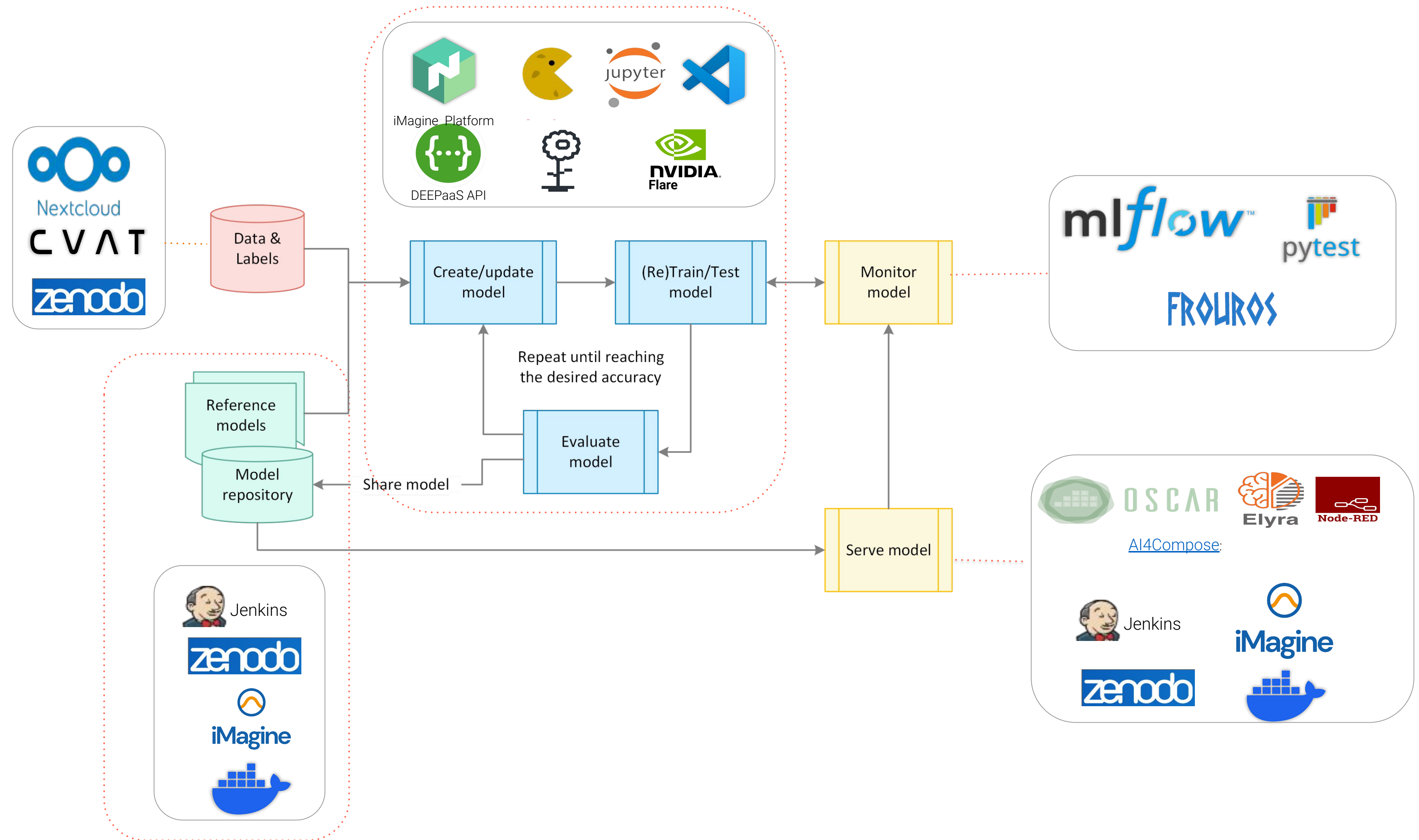# Best Practices:

Data Repository

AI/ML development cycle

Data Repositories and Open-source Dataset for Marine Applications

- **Zenodo**: robust solution, open-access platform to store and share datasets
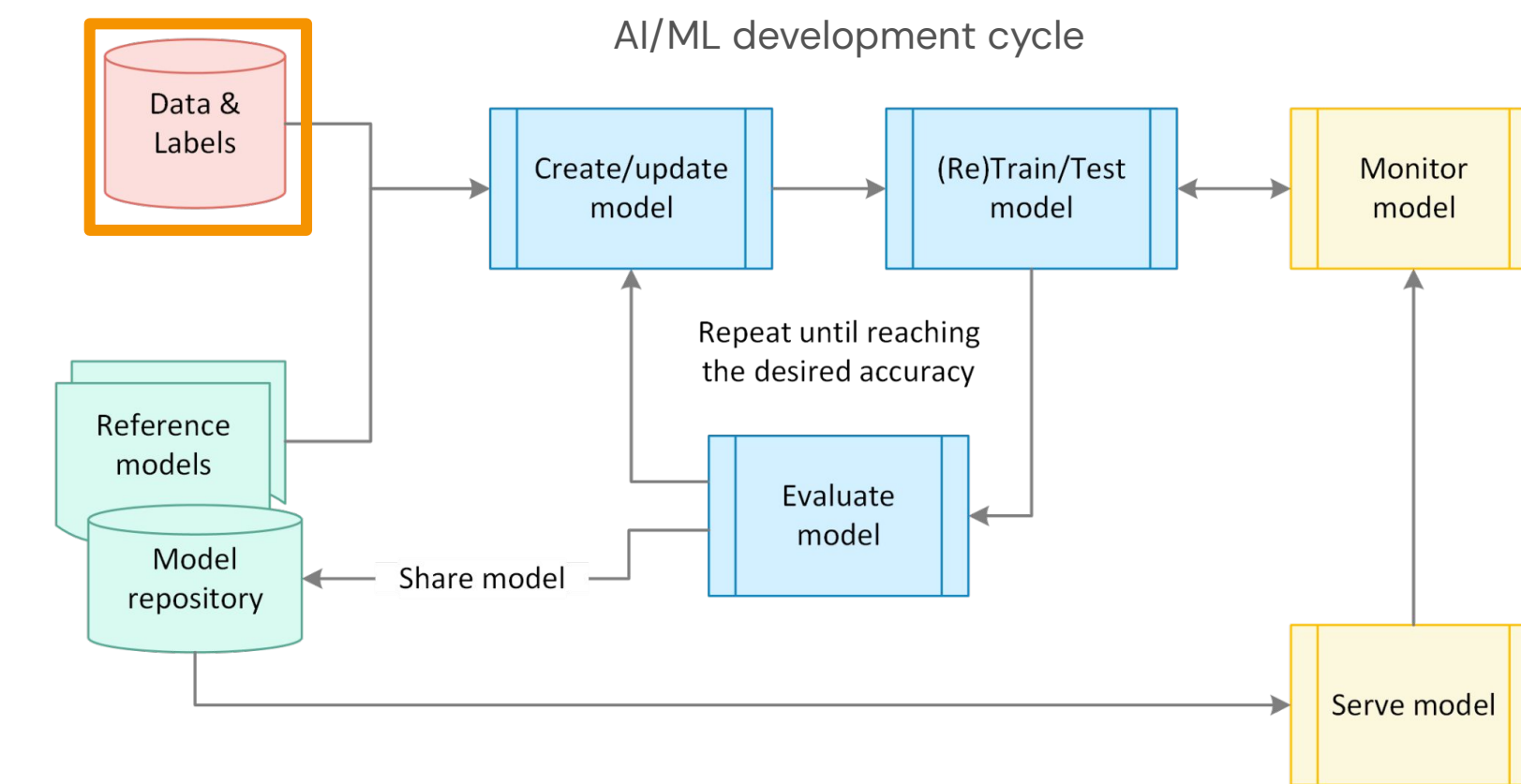- iMagine training datasets are in the **iMagine community at Zenodo**

zenodo

> iMagine and Zenodo **extend metadata** format for aquatic science

Tools and Infrastructures

Best Practices:

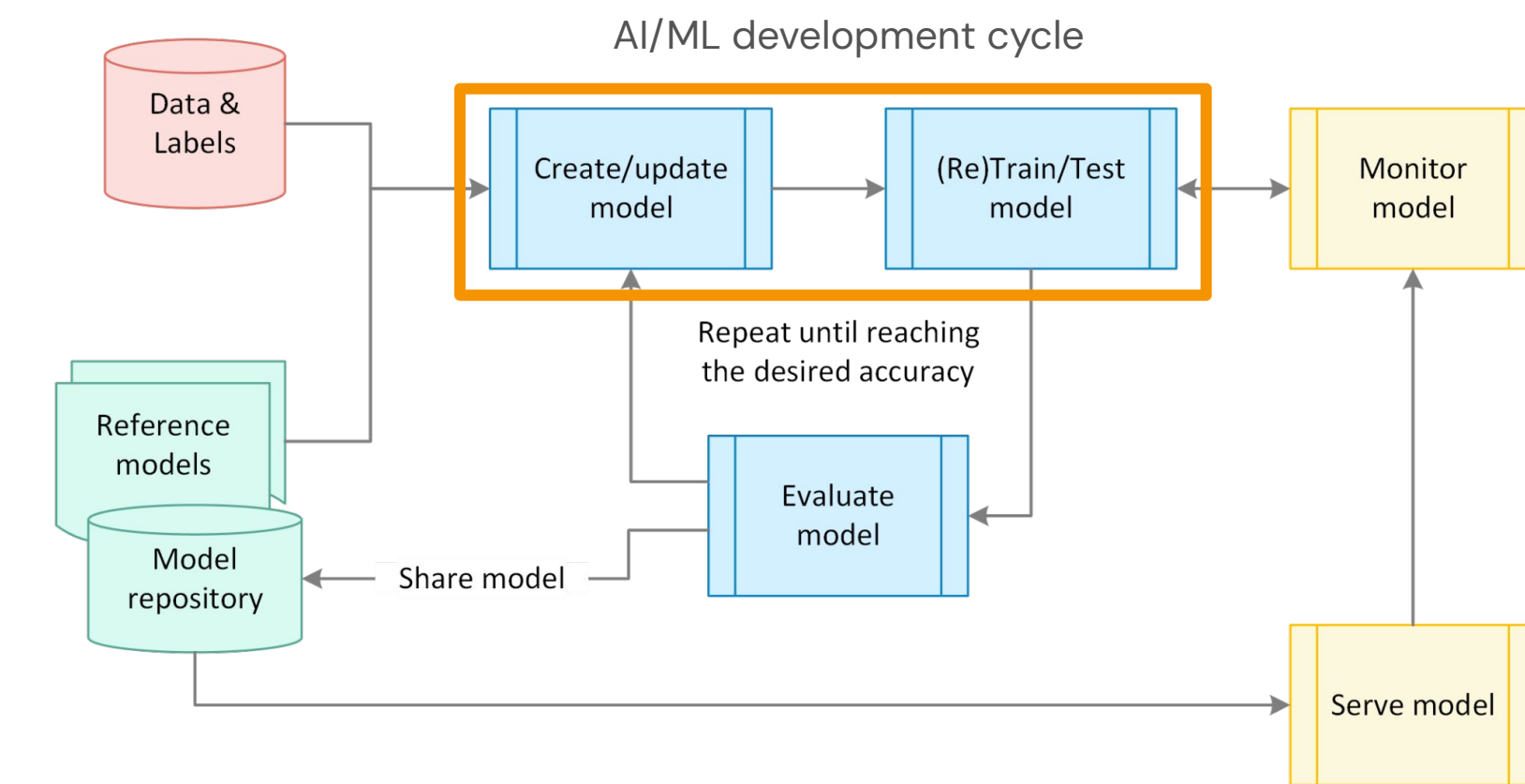*Data Bias*

AI/ML development cycle

Data Bias and Fairness in Aquatic Science Data

- Recommendations about handling **imbalanced datasets**
- **FAIR EVA** tool to improve the FAIRness of published training datasets concerning the metadata used in Zenodo

FAIReva

Best Practices:

Create/Update/Train/Test Model

AI/ML development cycle

- Deep Learning Models
 - Classification
    - Mobilenet
  - Object Detection
    - YOLO: most commonly used, showed strong performance
  - Segmentation
    - YOLO
    - Mask2Former: achieved better performance in panoptic segmentation than instance segmentation with Mask R–CNN

- Training and Testing on iMagine Marketplace

Marketplace

YOLO module

- Based on the Ultralytics' python package. Ultralytics[2] is a company that develops computer vision models

- **YOLO can perform:**
  - Classification: Identifying what object is in an image
  - Object Detection: Detecting and localizing multiple objects in an image with bounding boxes
  - Segmentation: Identifying the exact shape and boundaries of objects in an image (pixel-wise classification)

- **YOLO module** on the marketplace:
  - Integrated with DEEPaaS API (platform support)
  - Dockerized
  - Supports MLflow
  - Hyperparameter optimization with Hydra, Optuna & MLflow
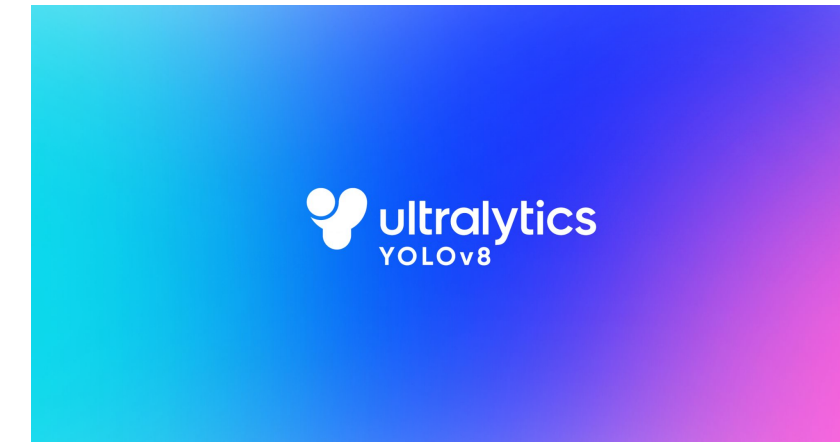
Once retrained on your data:
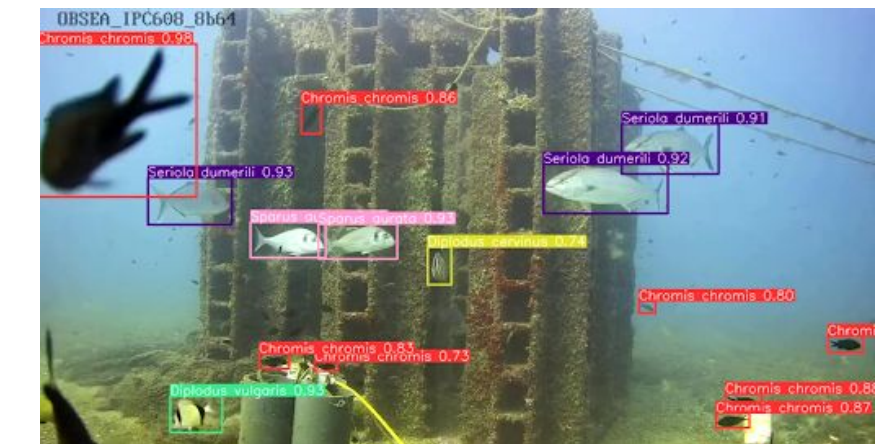
OBSEA module able to **detect** various fishes
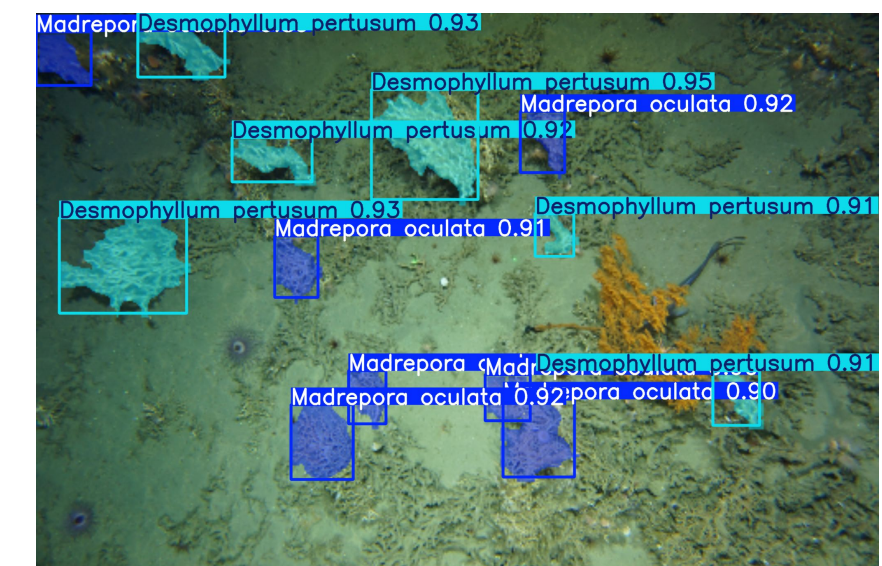
Cold water corals are properly **segmented**

2 https://github.com/ultralytics/ultralytics

To-Do: Validation of prototypes

WP3 point-of-view: tech. integration

- To Publish on the marketplace:
  - (very desired) Integrate with DEEPaaS API
  - Create application tests
  - Dockerize
  - Pass CI/CD

- Marketplace makes available:

  - Open source and docker access with rich metadata
  - Try for 10-minutes by everyone
  - Deployable for inference by VO-members
  - Deployable on 3d-party resources via IM

2. Marketplace download service delivery

1. Marketplace inference service delivery

3. Inference service delivery

⇒ TRL7 "System prototype demonstration in operational environment"

Metrics: Requested in VA

a) Number of unique users of the AI image processing service (unit: No. of users)

b) Number of images processed per year
   (= Images uploaded by your end-users for processing by your service)
   (unit: No. of images)

c) Number of images ingested
   (= annotated images, made available for download) (unit: No. of images)

d) Number of countries of users

e) Names of countries of users

Best Practices:

Evaluate Model

AI/ML development cycle
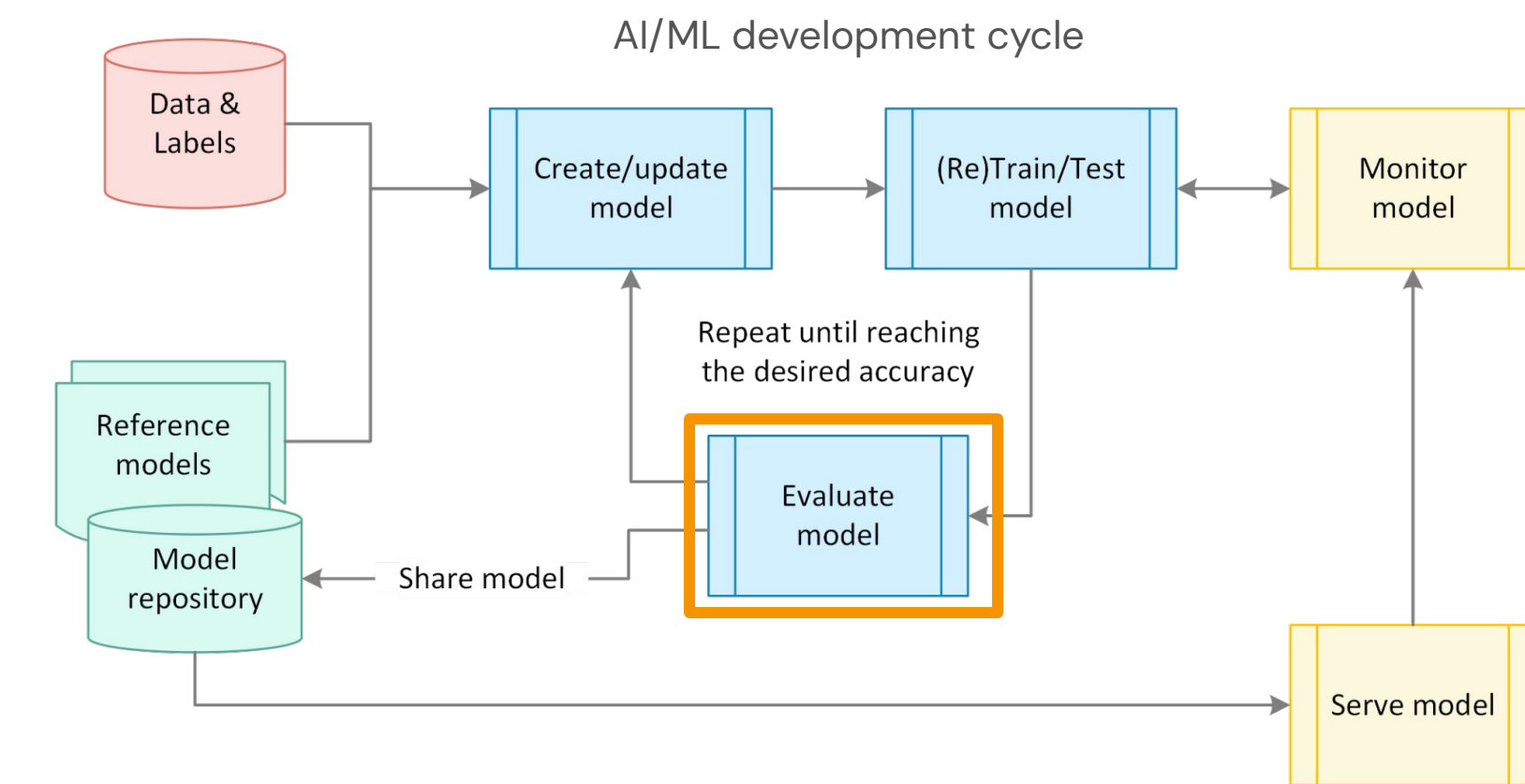
Performance Metrics and Evaluation Methods

- ○ **Accuracy**: correct predictions over total predictions
- ○ **Precision**: ratio of correctly predicted positive observations to the total predicted positives
- ○ **Recall**: ratio of correctly predicted positive observations to all observations in the actual class
- ○ **F1 score**: harmonic mean of precision and recall, useful for imbalanced dataset
- ○ **Intersection over Union** (IoU): overlap between two boxes

iMagine Use Cases

5 mature + 3 prototype

Beach monitoring
SOCIB (ES)
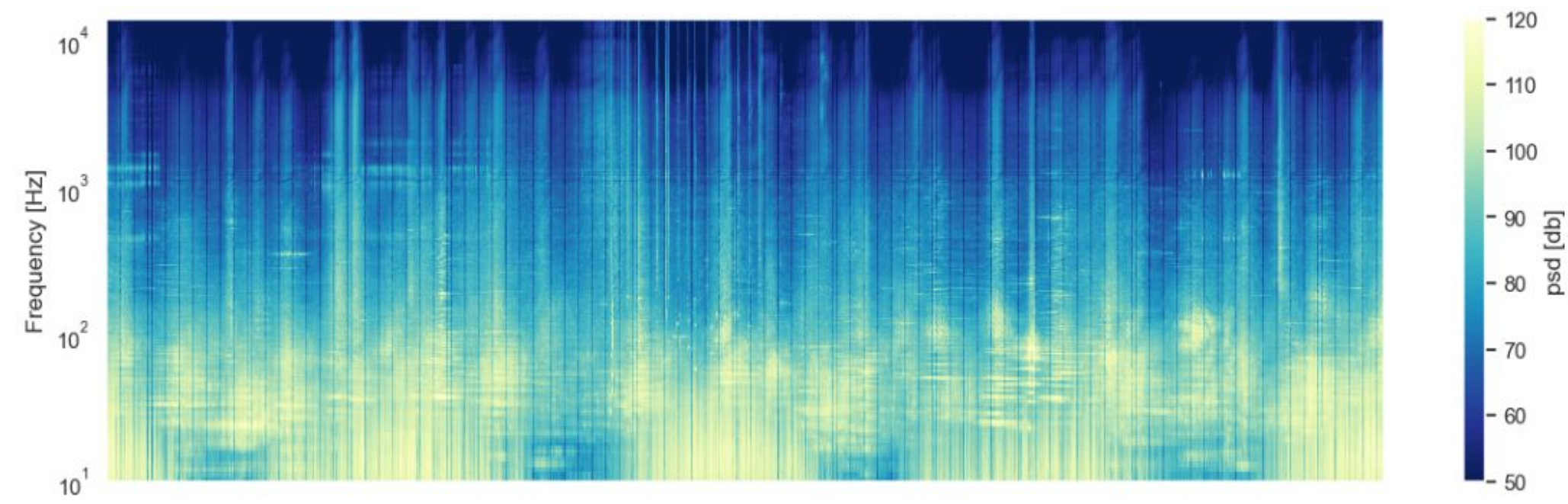
**Underwater noise
identification**
VLIZ (BE)
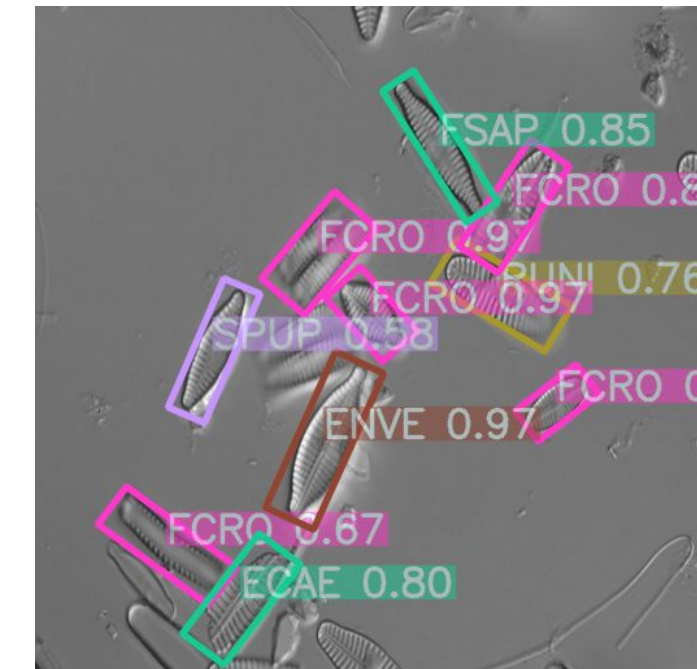
**Freshwater diatoms
identification**
Université de Lorraine–LIEC (FR)

Open Call* Use Cases

Being onboarded via Open Calls

Image Credit: Claire Widdicombe – PML · 100 μm

Image Source: https://mosaic.lnec.pt/
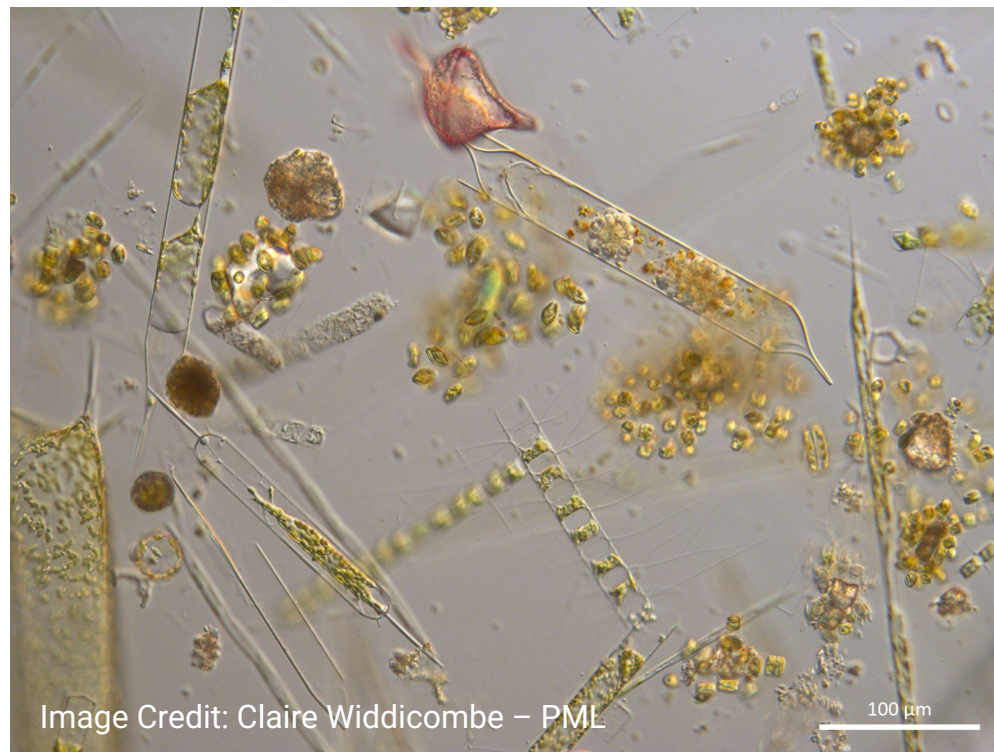
DEcentrAlised Learning (DEAL)

Plymouth Marine Laboratory, UK

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

Sea Wave and coastal inundation detection Methodology (SWiM)

LNEC (National Laboratory for Civil Engineering), PT

To enhance a methodology for an automated wave runup and wave detection from camera images for the surveillance of coastline

iMagine Use Cases

5 mature + 3 prototype
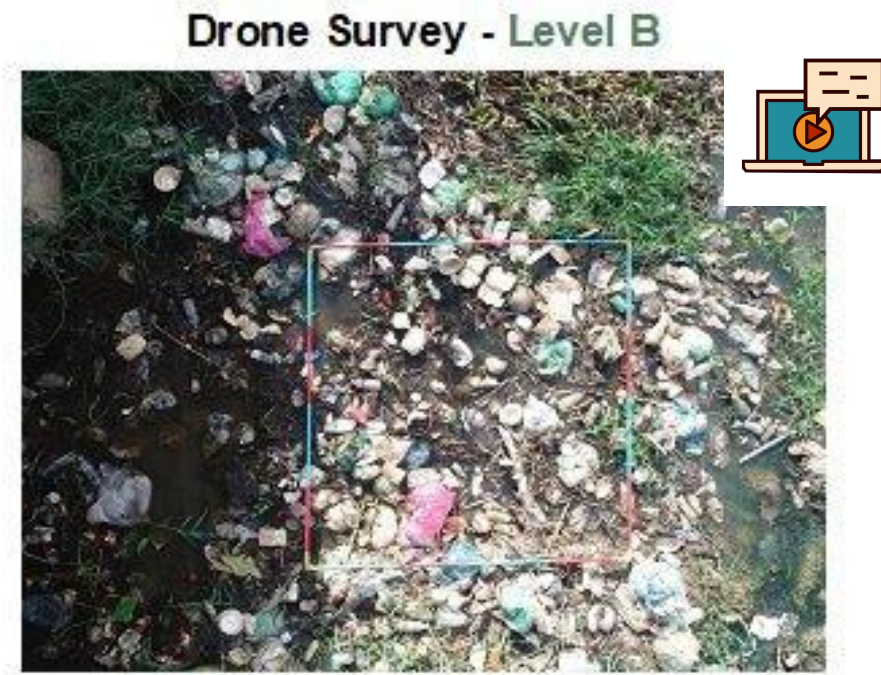
Aquatic Litter Drones

DFKI (DE), MARIS (NL),
OGS (IT)
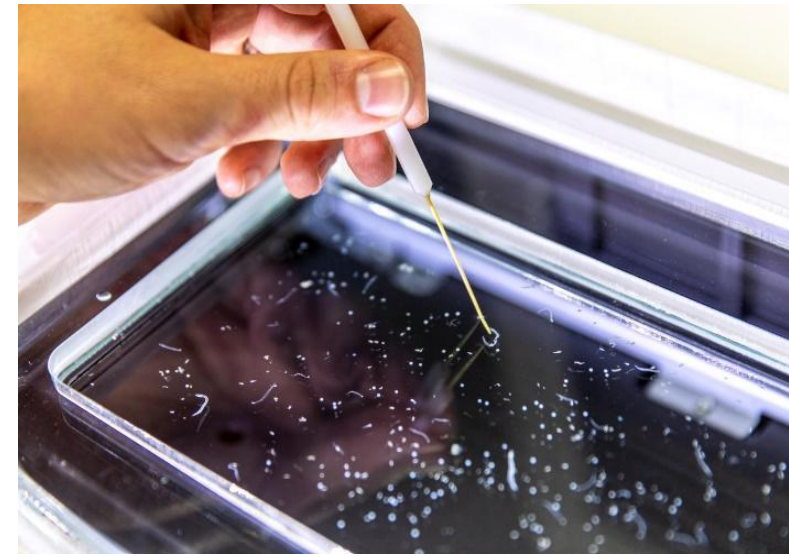
Zooscan – EcoTaxa pipeline (FR)
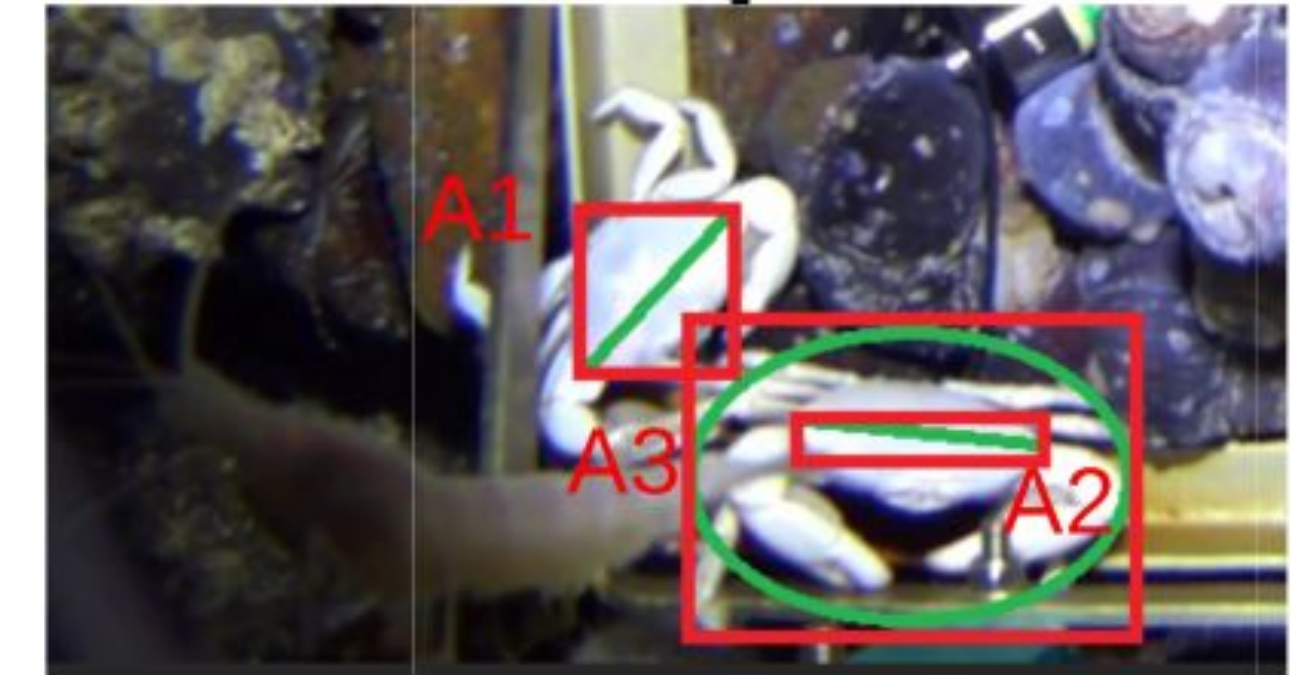
Marine Ecosystem Monitoring

EMSO ERIC:
UPC (ES), IFREMER (FR), MI (IR)
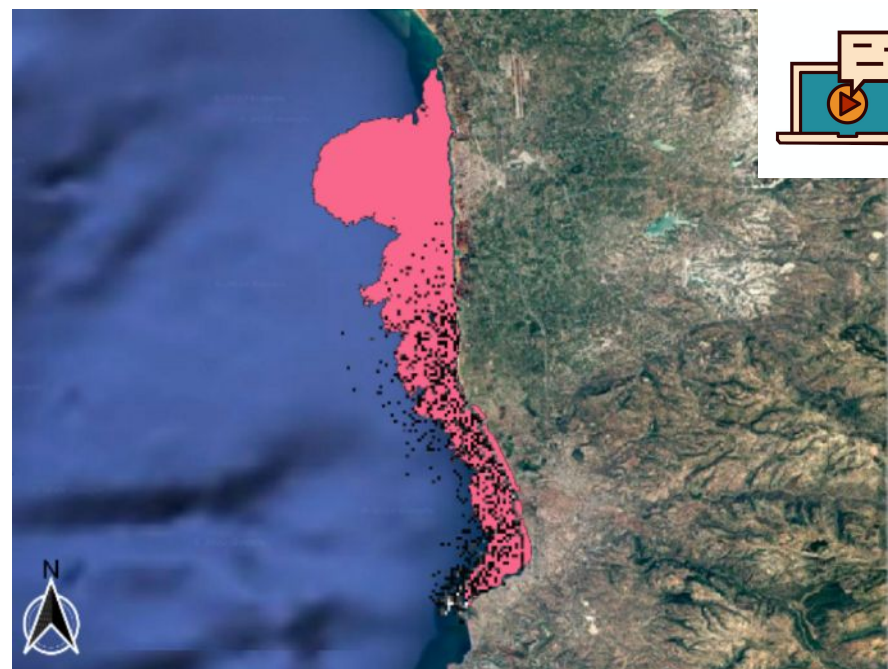
Oil Spill Detection

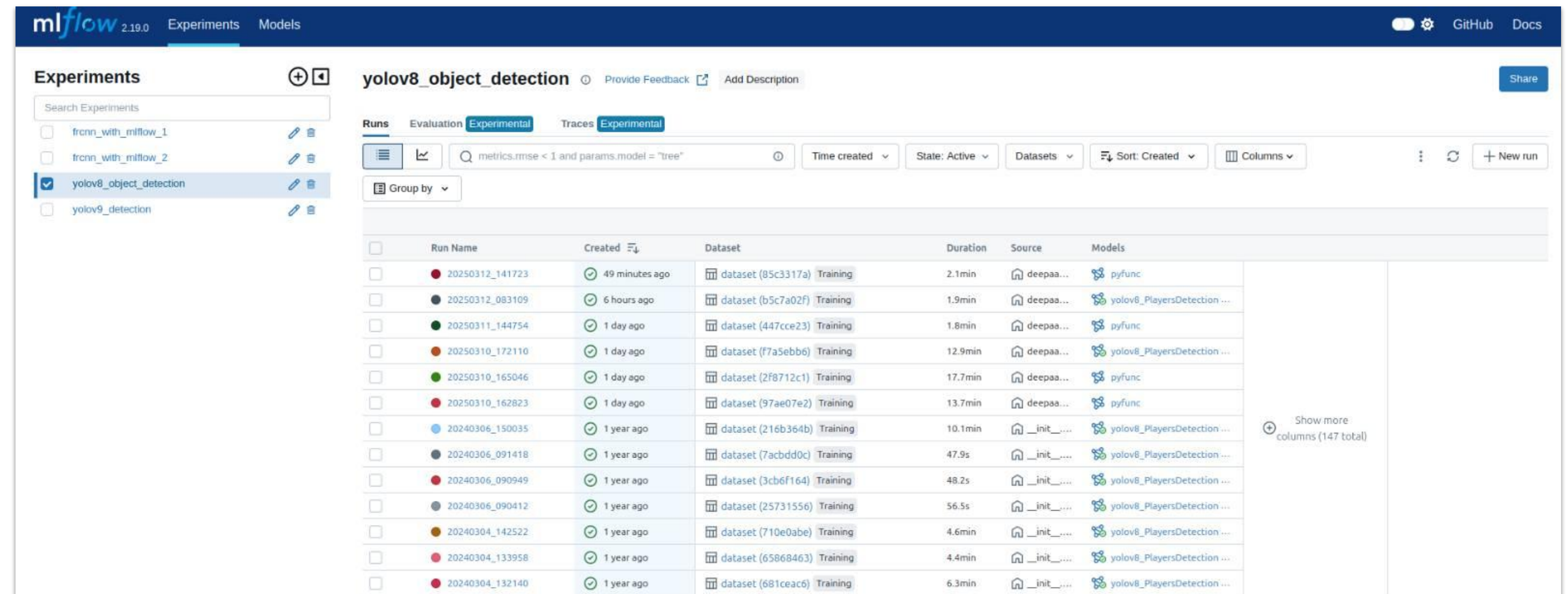CMCC, OrbitalEOS, UNITN (IT)
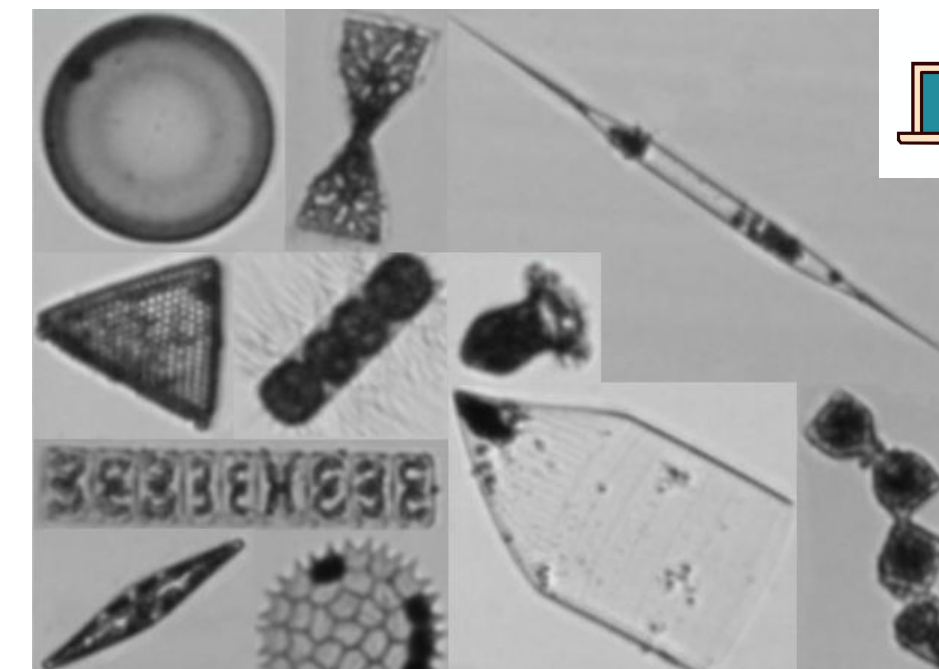
Flowcam phytoplankton identification

VLIZ (BE)

Best Practices:

Monitoring Model Performance
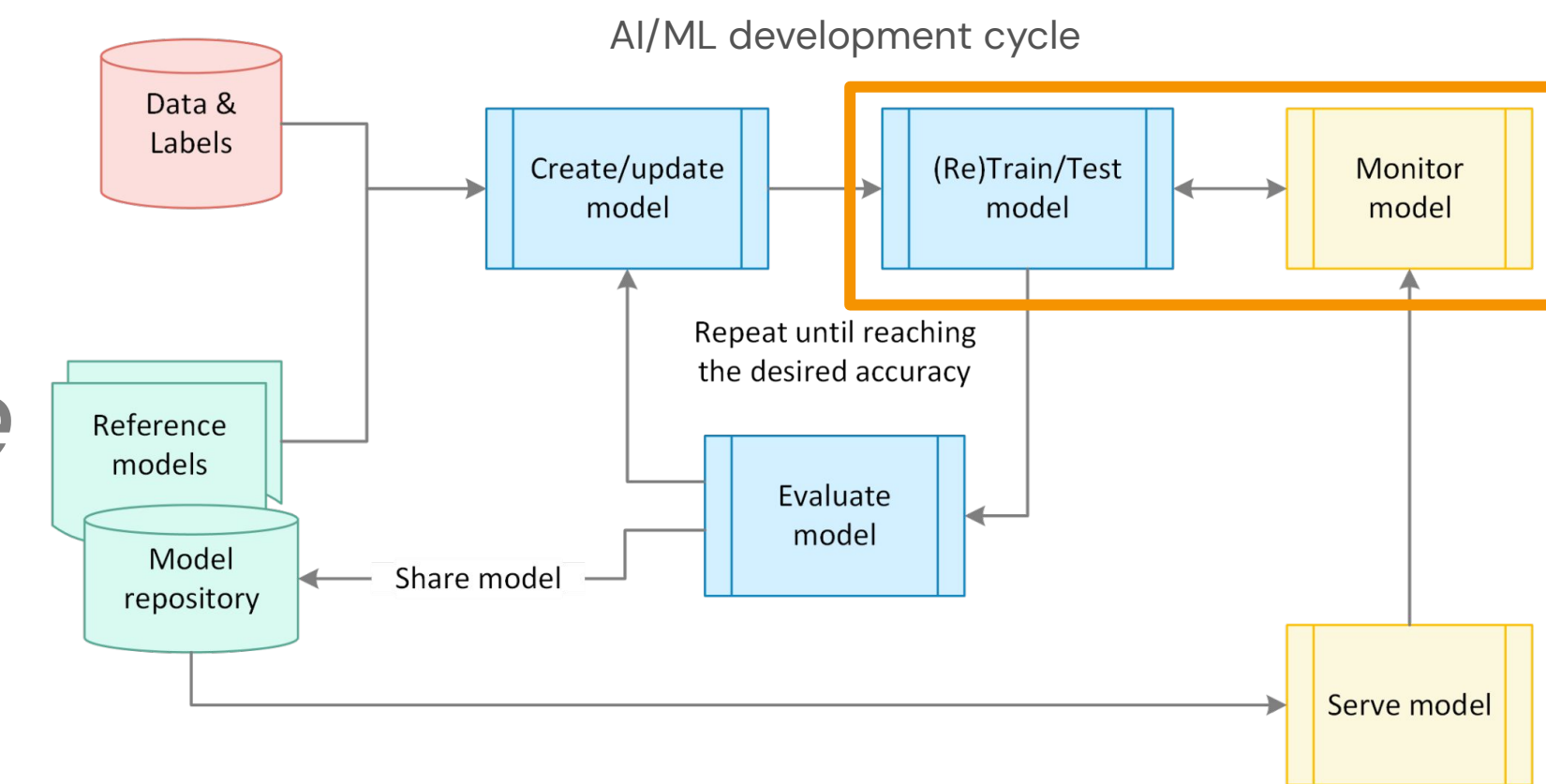
ML Experiment Tracking Tools

- **Tensorboard**: few functionality, single user and self-hosted

- **Weights and Biases**: cloud-based service

- **MLflow**: efficient management and tracking

Best Practices:
Serve Model

AI/ML development cycle
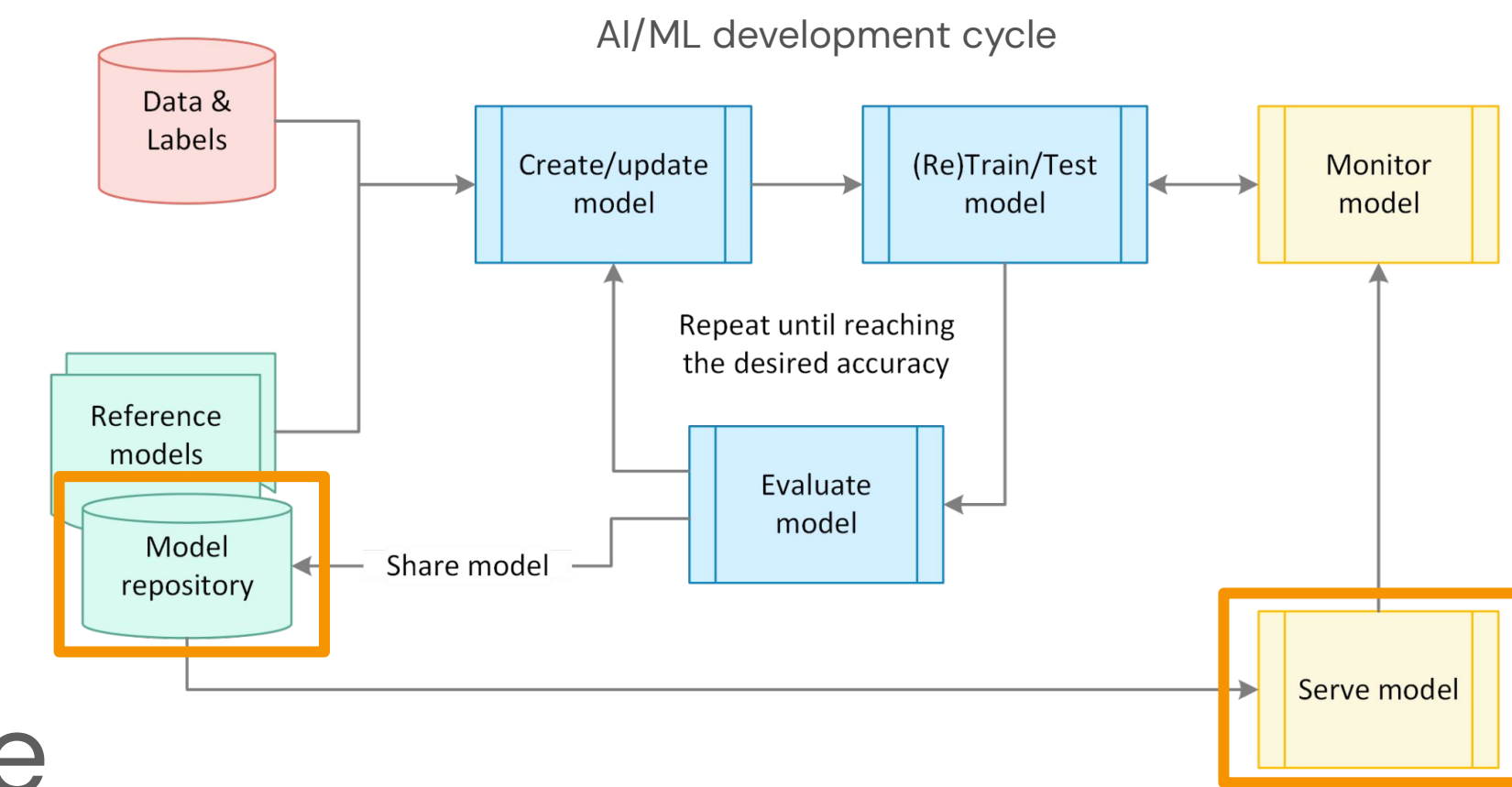
- Sharing and Serving AI Models via iMagine Marketplace
  - AI models are published as Docker images and code repositories
  - Anyone authenticated can **Try** an inference endpoint (for 10 min)

Marketplace

- Model Deployment in Production
  - OSCAR: efficient and scalable deployment for running the AI model inferences

User Interface
Create a short 10min deployment using Gradio

OSCAR   Elyra   Node-RED