The iMagine AI platform for image analysis – Opportunities for aquatic sciences

Gergely Sipos (EGI Foundation)
On behalf of the iMagine consortium
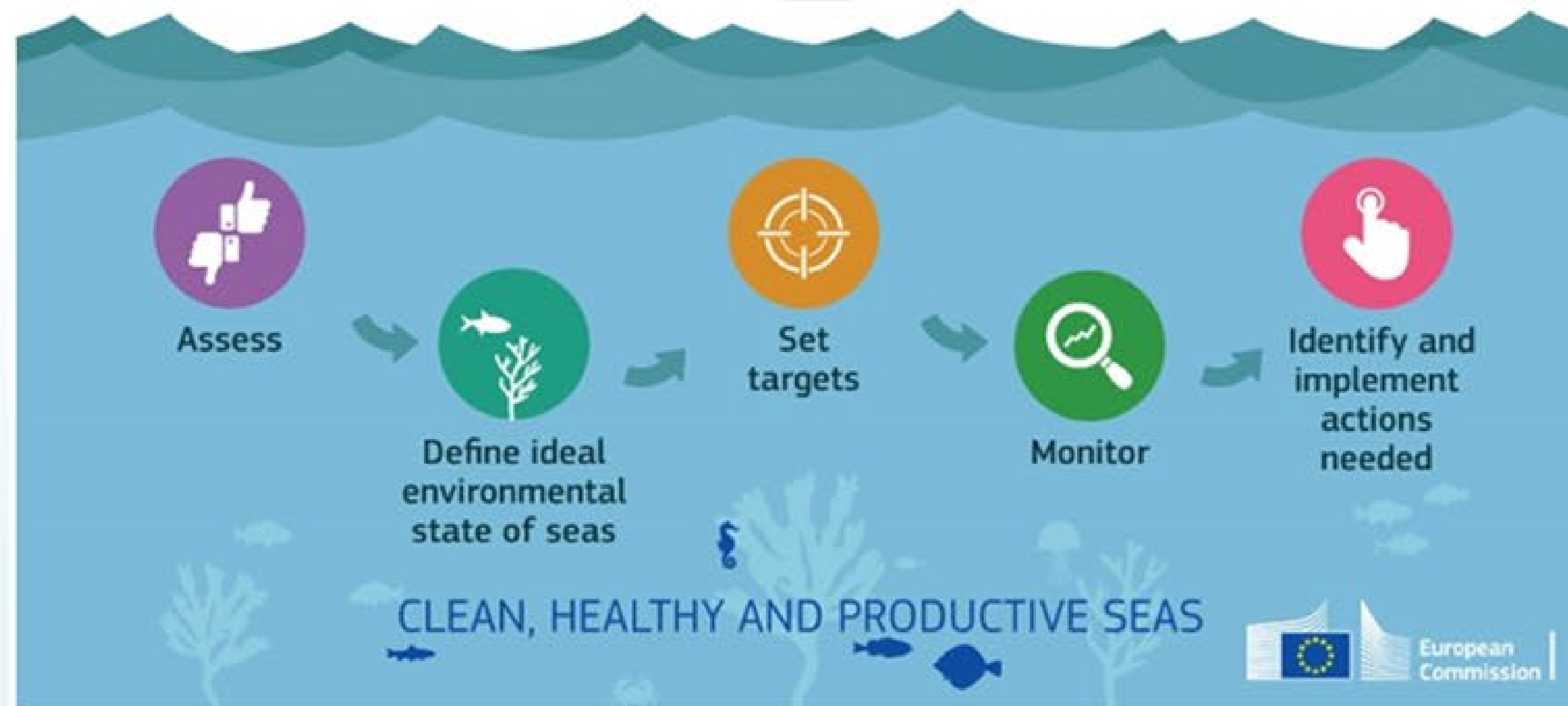
EGI 2023 Conference

21/June/23, Poznan

iMagine

Outline

- Project facts
- Technical architecture
- Imaging services and applications
- Lessons learned
- Opportunity to work with iMagine

Marine environmental management and policy making

How EU Member States develop marine strategies

Assess → Define ideal environmental state of seas → Set targets → Monitor → Identify and implement actions needed

CLEAN, HEALTHY AND PRODUCTIVE SEAS

European Commission

Source: EU MSFD website

Relevant EU Directives and initiatives for aquatic domain, such as:

- Marine Strategy Framework Directive (MSFD)
- Water Framework Directive (WFD)
- European Green Deal
- Mission Starfish 2030 "Healthy oceans, seas, coastal and inland waters"
- United Nation's 2030 Agenda for Sustainable Development
- United Nation's Decade of Ocean Science (2021–2030)

- 50% of the global oxygen production is produced by photosynthesis of marine algae.
- Oceans are home to the world's largest diversity of species and habitats.
- Annually 100 Mio tons of marine organisms are exploited as food source.

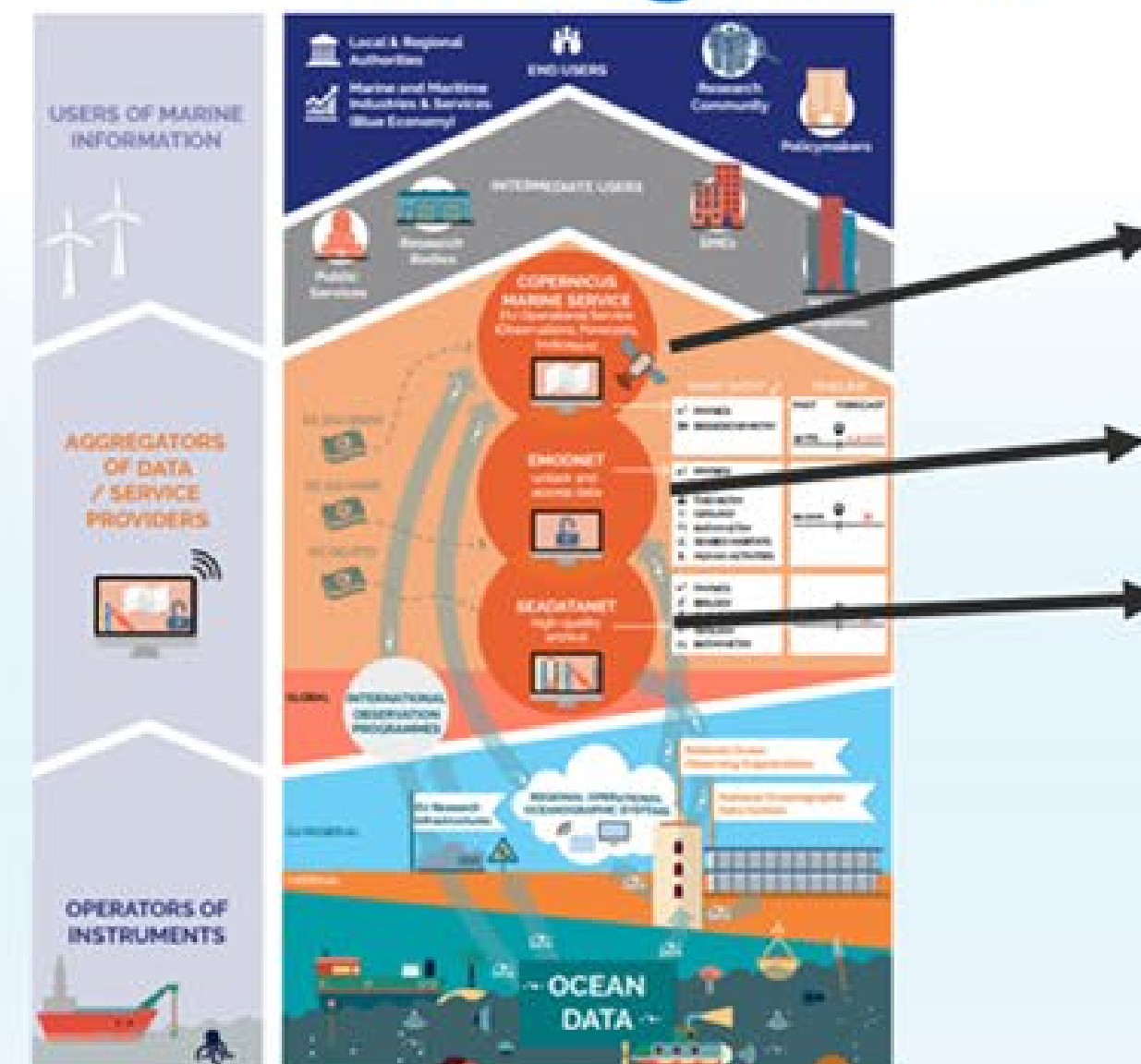# Implementation requires knowledge

The implementation requires an increase of our overall knowledge, demanding more science and improved access to observation data and analytical processing.
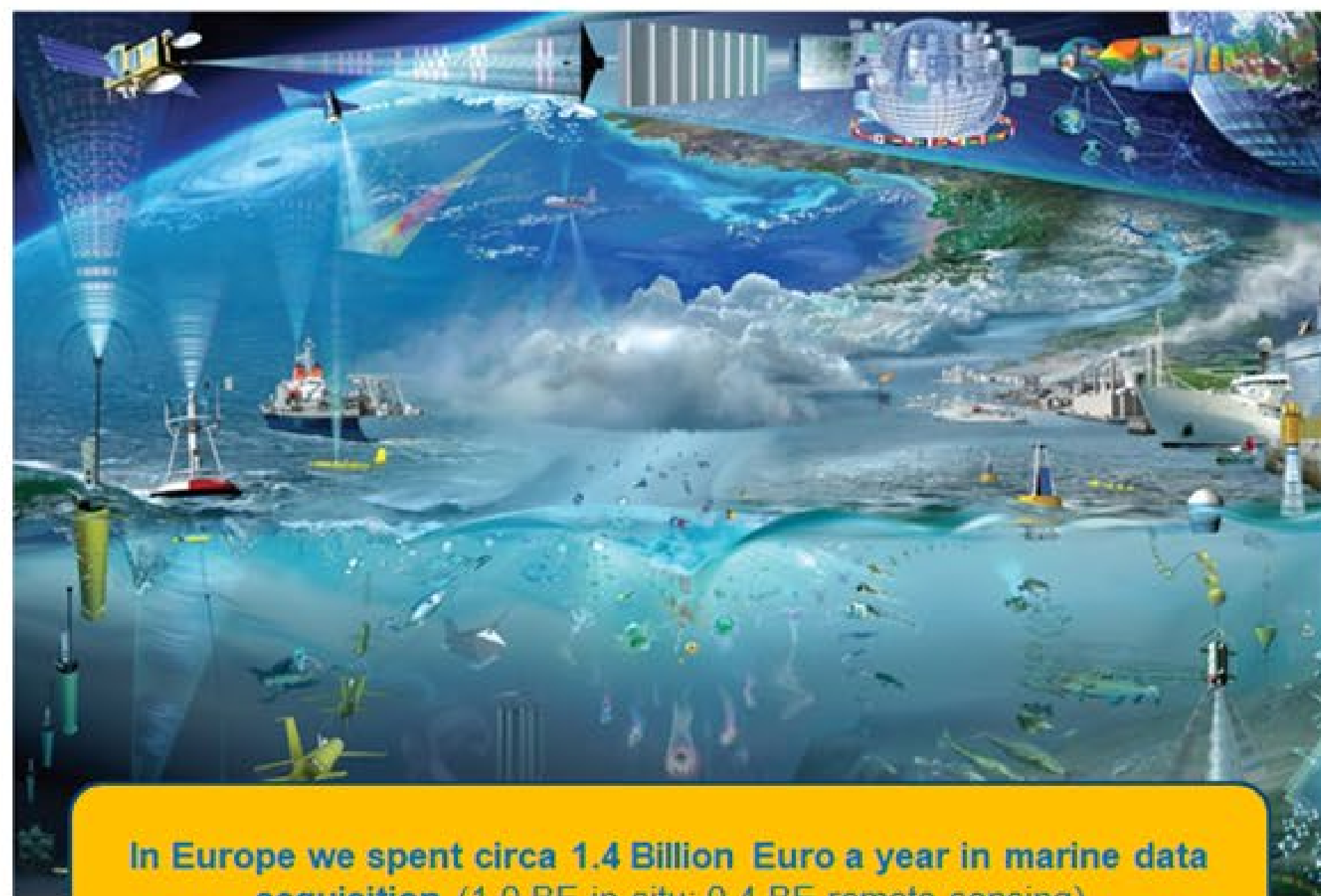
In Europe we spent circa 1.4 Billion Euro a year in marine data acquisition (1.0 BE in-situ; 0.4 BE remote sensing)

European landscape marine data management

Data aggregators and providers of data products and services

Europe already has developed an impressive capability for aquatic environmental observation, data–handling and sharing, modelling and forecasting, second to none in the world. This builds upon national environmental observation and monitoring networks and programs, complemented with EU initiatives such as the Copernicus programme (CMEMS) and EMODnet, and European Research Infrastructures (RIs).

Project factsheet

OBJECTIVE:

*To deploy, operate, validate, and promote a dedicated **iMagine AI framework** and platform, connected to EOSC and AI4EU, giving researchers in **aquatic sciences** open access to a **diverse portfolio of AI based image analysis services and image repositories** from multiple RIs, working on and of relevance to the overarching theme of 'Healthy oceans, seas, coastal and inland waters'.*

- **36** months
- From **Sept. 2022** until **Aug. 2025**
- **€4.5 million** EC funding
- **23** participants (**19** beneficiaries + **4** affiliated partners)
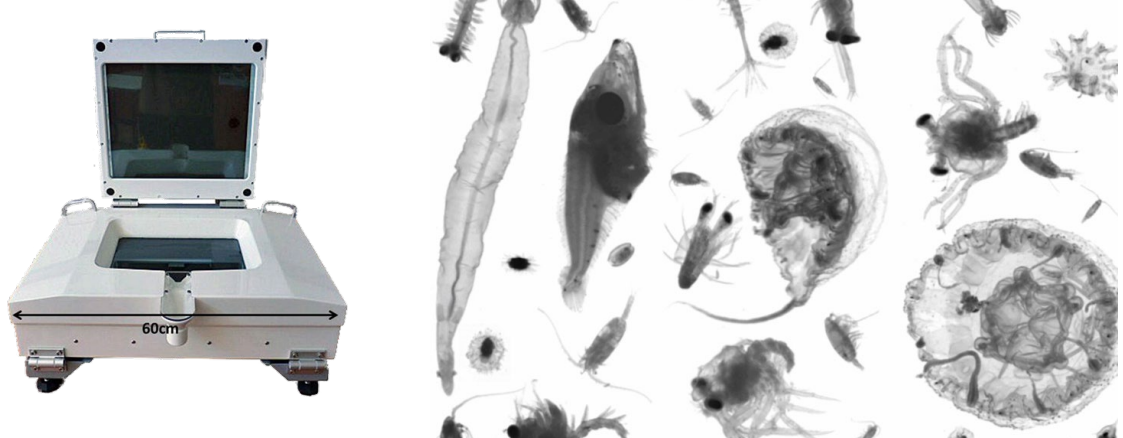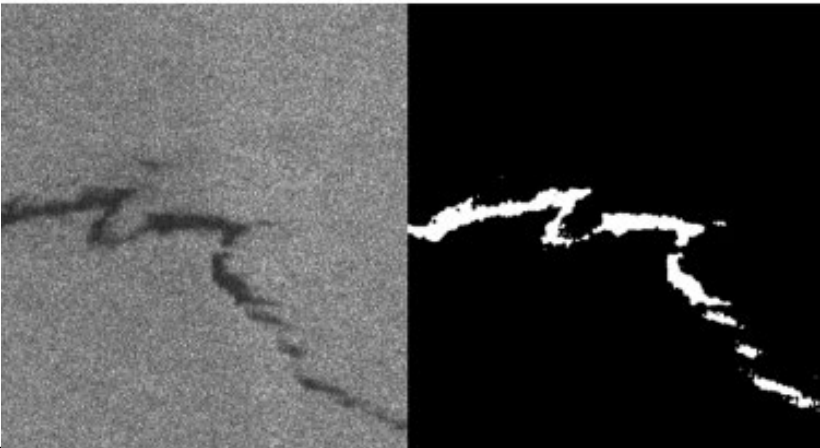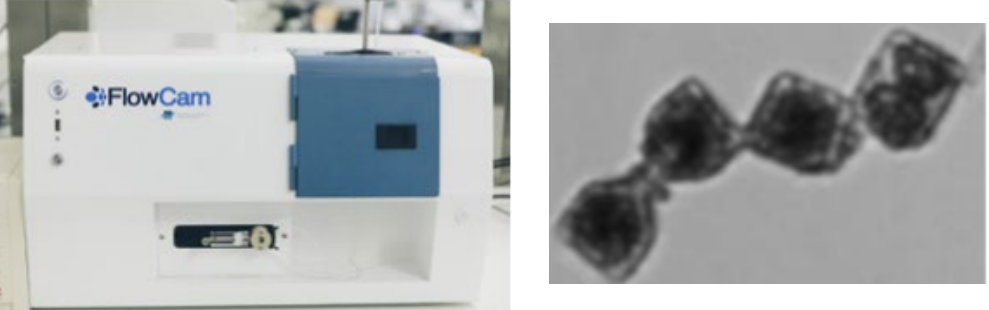- **18** service installations (Virtual Access)

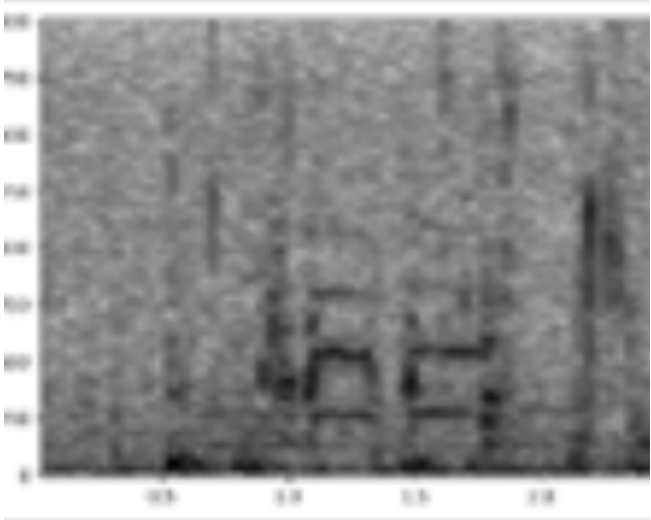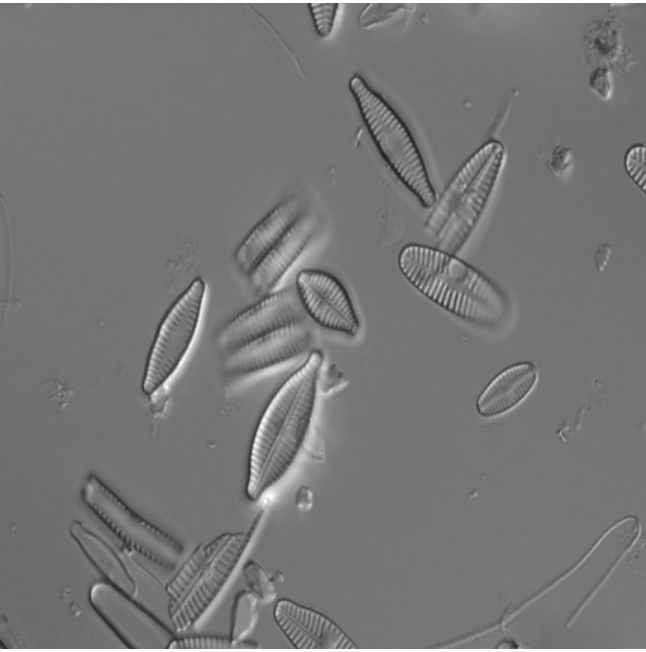# Enabling scalable AI/ML services

4 national cloud compute centres and

5 AI/ML technology development institutes supports

14 research institutes **for 12 RIs**

5 mature use cases

Aquatic Litter Drones: Aquatic Litter monitoring system using drones	**DFKI (DE)** + MARIS (NL) + OGS (IT)	Drone Survey - Level B
EcoTaxa pipeline: Taxonomic identification of zooplankton using Zooscan	**Sorbonne Université (LOV+IMEV) (FR)**	
Ecosystem monitoring at EMSO sites by video imagery	**EMSO ERIC (IT)** + UPC (ES) + Ifremer (FR) + MI (IE)	
Oil Spill Detection: Oil spill detection from satellite images	**CMCC (IT)** + OrbitalEOS (ES) + Uni. of Trento (IT)	
Taxonomic identification of phytoplankton using Flowcam images	**VLIZ (BE)**	60% oxygen on Earth!

3 prototype use cases

Underwater Noise Identification: Underwater noise identification from acoustic recordings using spectrograms	**VLIZ (BE)**	Spectrograms
Beach Monitoring: Posidonia oceanica berms and rip-currents detection from beach monitoring systems	**SOCIB (ES)**	Palma
Identification of diatoms in freshwater (unicellular microorganism)	**UL–LIEC (FR) + CNRS–IRL2958 (FR) + SU–LOCEAN (FR)**	

Specific Objectives and indicators

O1 **Objective 1**. Deliver a scalable, shared IT platform for image analysis in marine and freshwater research

Operational iMagine platform with common AI development framework

O2 **Objective 2**. Advance existing image analytical services to increase research performance in aquatic sciences

Launch of 5 aquatic AI image analytics services, running operationally on the iMagine platform

O3 **Objective 3**. Develop & prototype new image analytical services and datasets that can accelerate progress towards healthy oceans, seas, coastal and inland waters

3 AI-based imaging processing application pilots, 8 scientific image repositories

O4 **Objective 4**. Capture and disseminate development and operational best practices to imaging data and image analysis service providers

Best Practices documentation, interaction with EOSC and AI4EU platforms. + Training programme

O5 **Objective 5**. Deliver a portfolio of scientific image and image analytics services targeting researchers in marine and aquatic sciences

Portfolio: operational services, image repositories, Best Practices, iMagine framework and platform

Our technical architecture
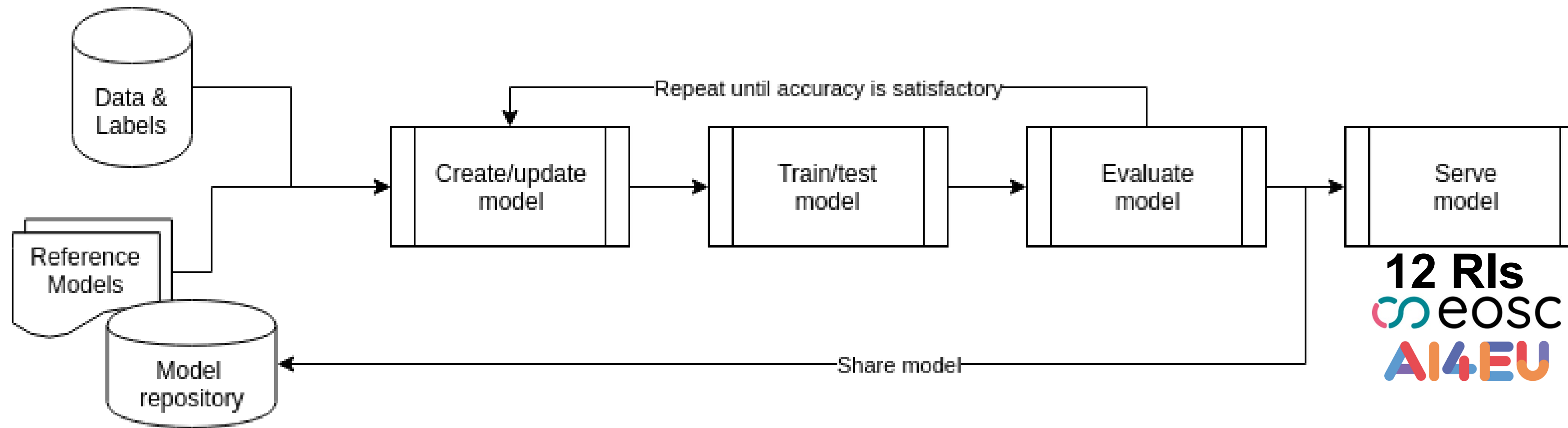
- 5 production AI services
- 3 AI application prototypes
…for aquatic sciences

12 RIs
eosc
AI4EU

Generic, scalable platform for AI/ML applications

EGI federated cloud infrastructure
OpenStack GPUs, CPUs, Storage in Spain, Portugal, Turkey, Ireland

1500 TB–months
132,000 GPU–hours
6,000,000 CPU–hours

iMagine AI Platform

Common implementation approach
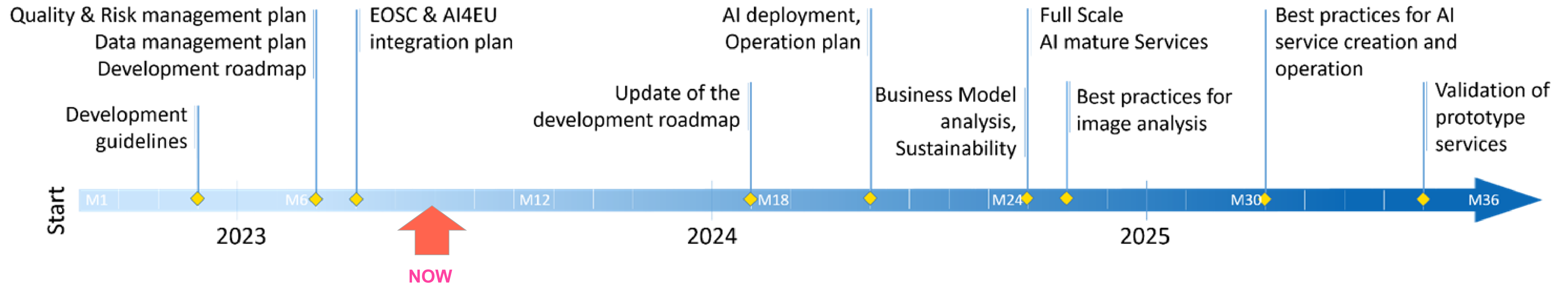
- Analysis of use cases using Persona–EPIC–User story approach
- Gap and Bottleneck analysis
- Initial Development roadmaps
- Initial collection of requirements for the platform

- Update of the roadmaps
- Building up operation plans

Summing up lessons learnt for Best practices for
- image analysis in aquatic sciences
- AI service creation and operation

Development roadmap: https://zenodo.org/record/7760413

Lessons learnt... so far

- Labelling/annotating data: various tools are being assessed and used

- Building datasets for training and testing: the real world is noisy!

  → Data quality check is of relevance and interest

- Working on metadata to improve FAIRness of data

- Selecting AI models for object detection, segmentation, further classification/taxonomy

  → Monitoring of experimentation is helpful

- Engineering of end-to-end pipelines

- Regular knowledge exchange is important

  → Bi-weekly calls with all the use cases and AI platform experts
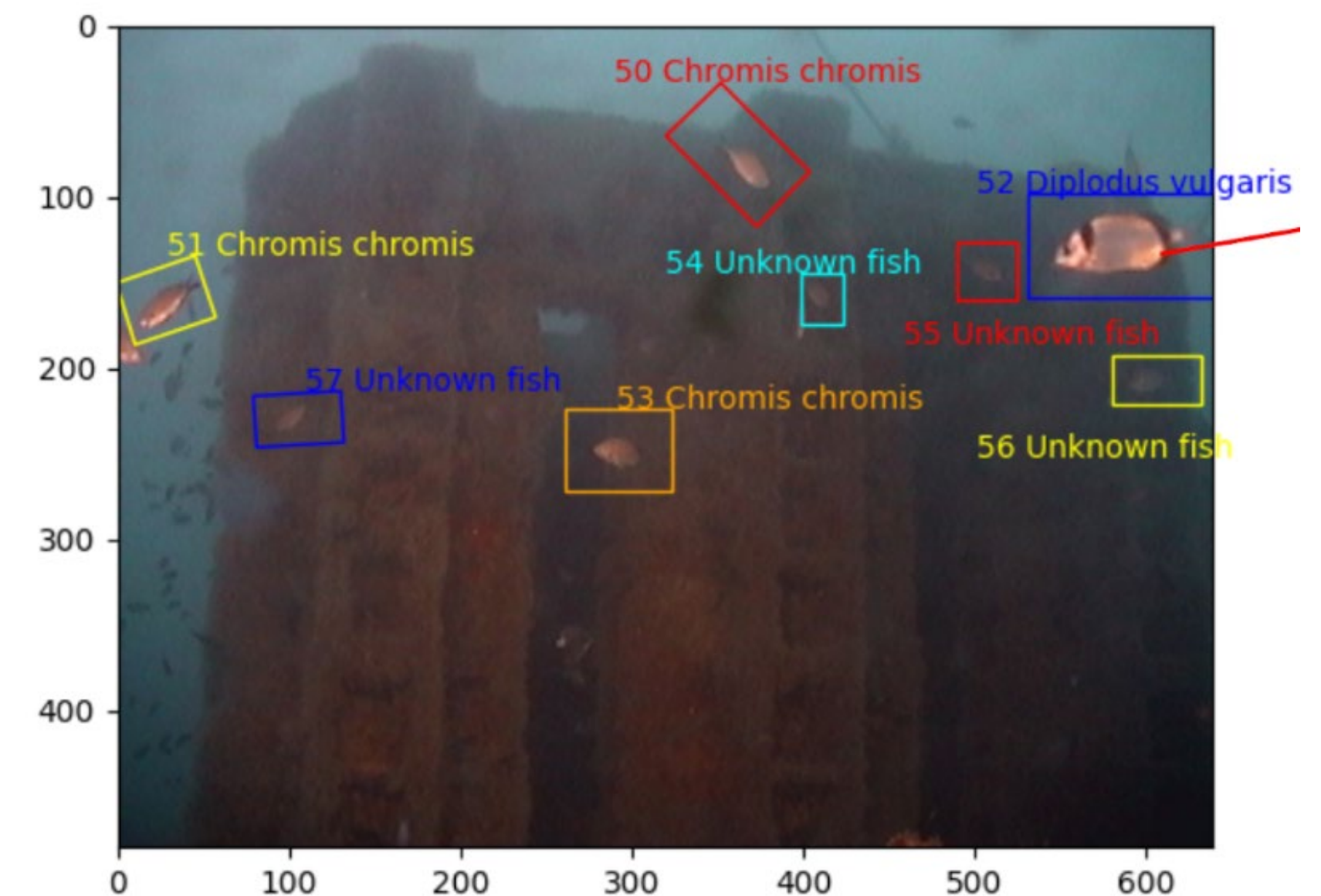
- Filling manpower gaps is hard...

Labelstud.io   Labelbox.com   Biigle.de
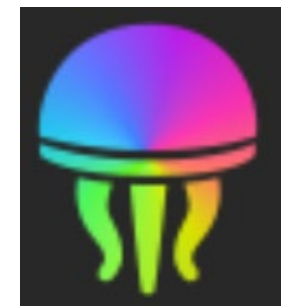
iMagine

POWERED BY
egi
POWERED BY

The iMagine Open Call for AI-powered image analysis in aquatic sciences is NOW OPEN!

We offer:
- 10-month collaborative projects

Support for
- AI model training
- Large-scale image analysis
- Use of the iMagine AI platform to develop and train AI models
- Access to cloud resources (GPUs, CPUs, storage) to store images and to scale up analysis workflows

Apply by 31 July