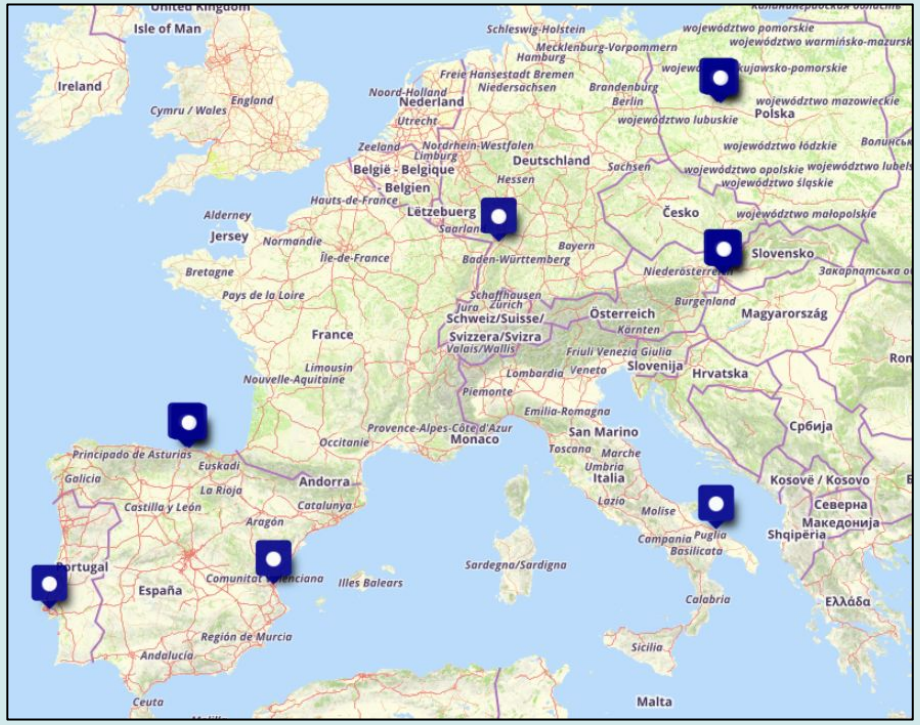
AI4 | EOSC

Empowering #EOSC scientists with AI services and tools

Álvaro López García (aloga@ifca.unican.es)

Co-funded by
the European Union



AI4 | eosc

Co-funded by the European Union

AI4EOSC

Artificial Intelligence for the #EOSC

- Evolution of the DEEP Hybrid DataCloud platform
- HORIZON-INFRA-2021-EOSC-01-04 call
- Runs September 1st 2022 – August 2025 (36 months)
- 7 academic + 2 SME + 1 non-profit organization

Advanced features for distributed, federated, composite learning, metadata provenance, MLOps, event-driven data processing, and provision of AI/ML/DL services

Objectives

Objective 1
Feature rich services and platform to build and deploy custom AI applications in the EOSC

Objective 2
Support for building AI systems on distributed datasets, with a particular focus on federated learning

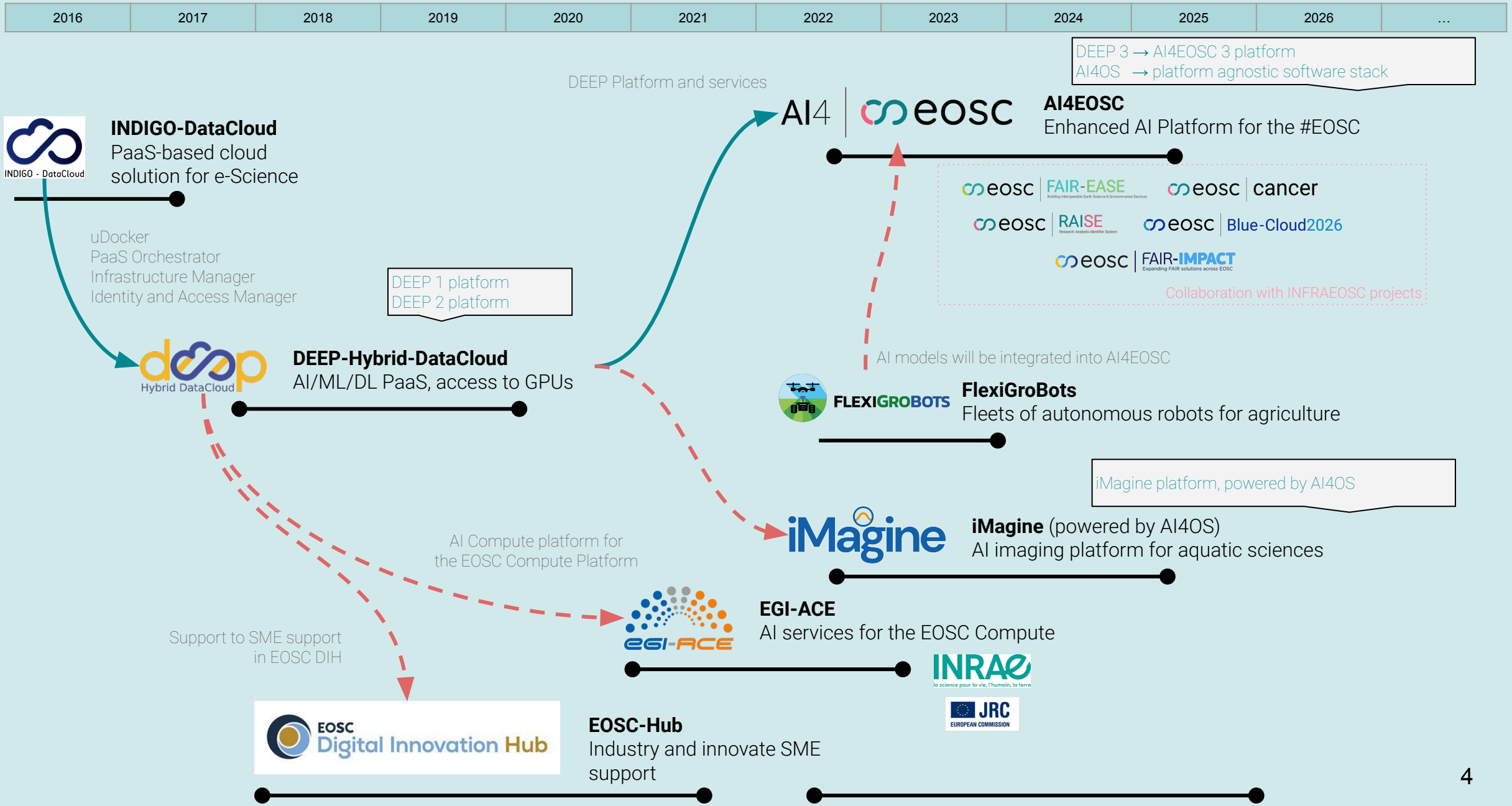
Objective 3
Services to compose AI tools, enabling the development of complex data-driven applications

Objective 4
AI exchange/hub in the context of the EOSC, enhancing and increasing the application offer currently available

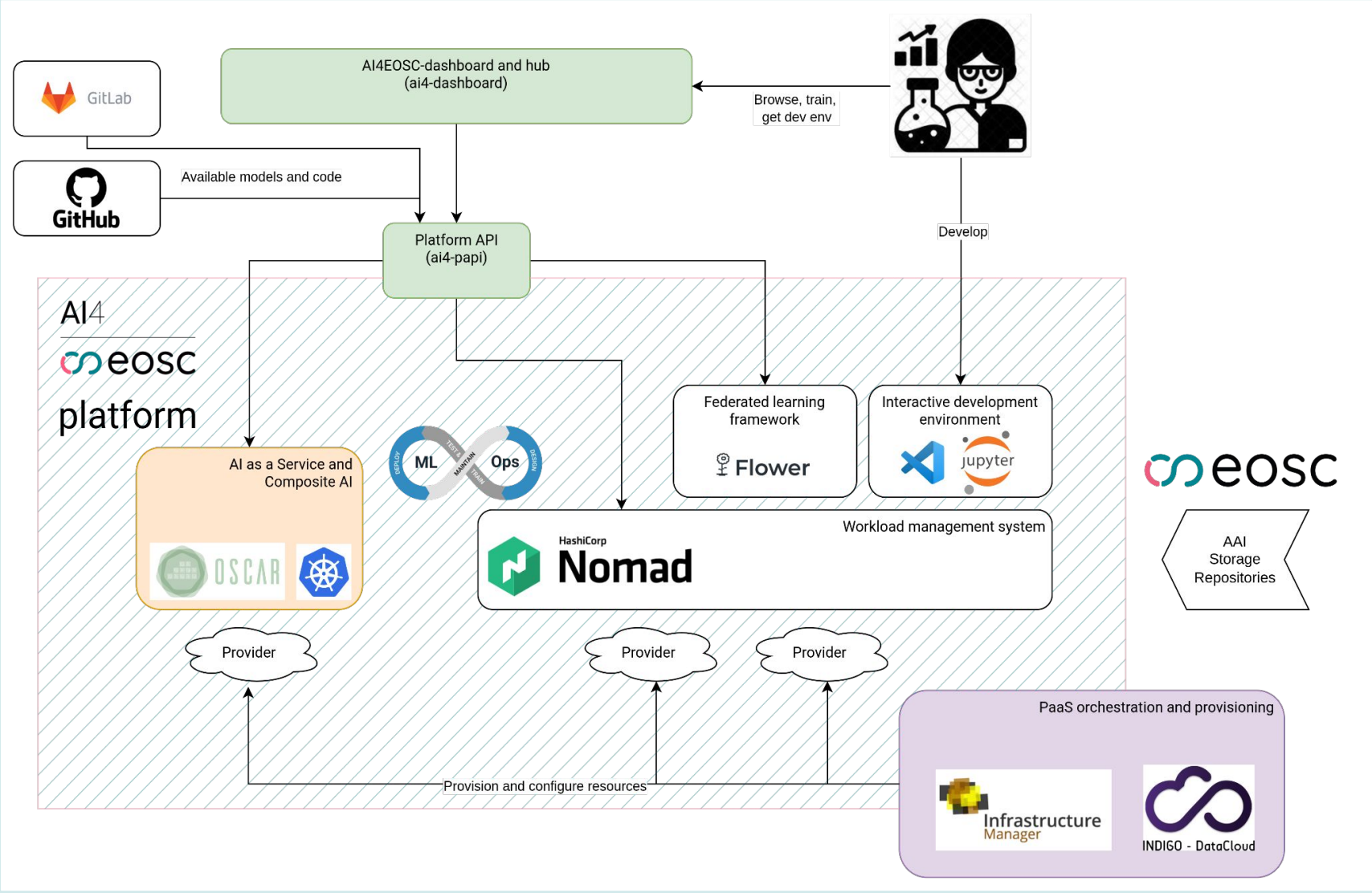
Objective 5
Extend the service offer and the capabilities being offered through the EOSC portal, with focus on AI

Goal
Foster an AI exchange in the EOSC context transforming the development of AI applications in the EOSC

Background, ecosystem, collaborations



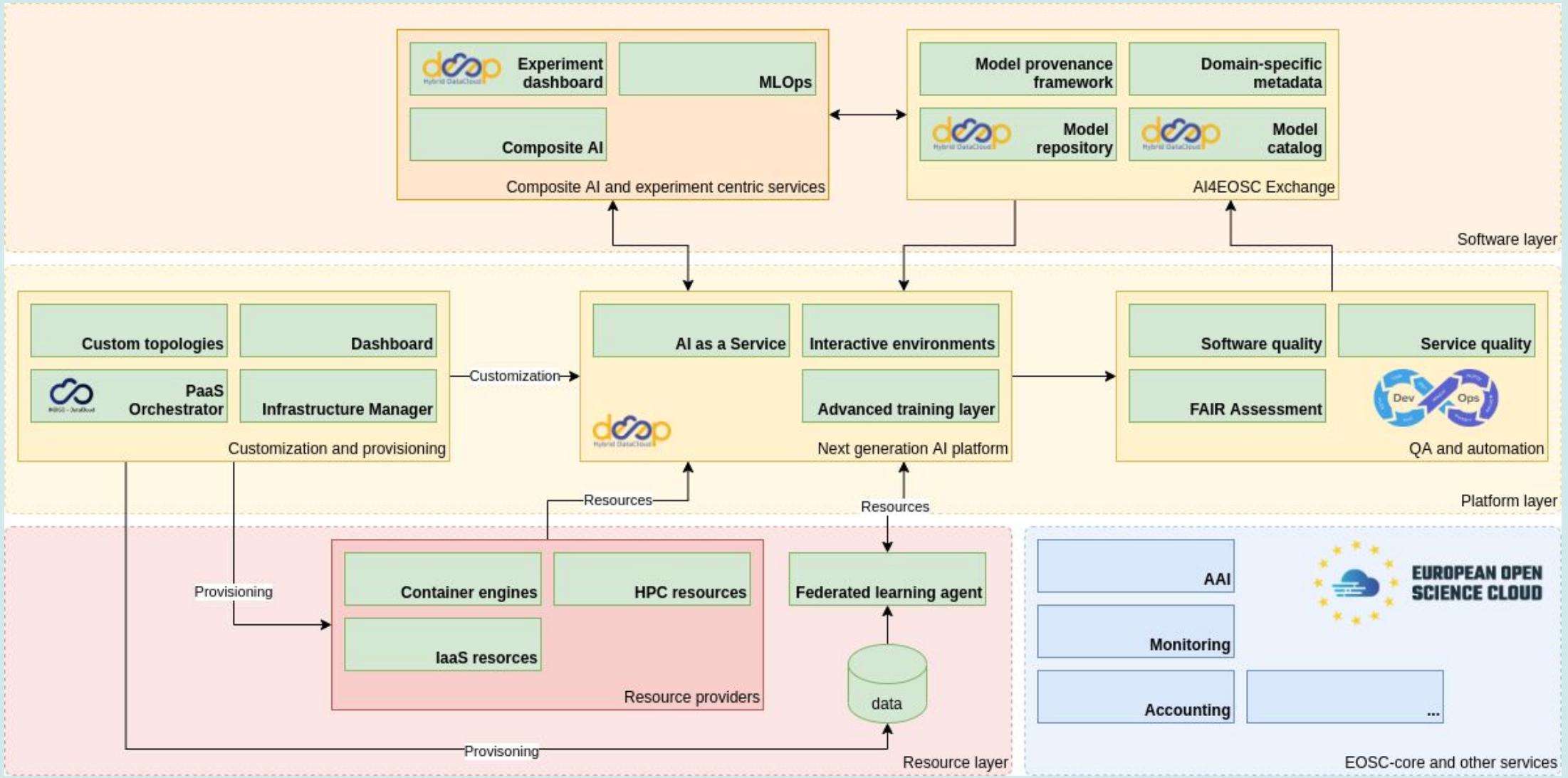
AI4EOSC high level architecture



Detailed C4 architecture can be found here:

- Workspace https://structurizr.com/share/73873/2f769b91-f208-41b0-b79f-5e196435bdb1
- Diagrams: https://structurizr.com/share/73873/2f769b91-f208-41b0-b79f-5e196435bdb1/images

AI4EOSC conceptual diagram



DEEP evolves into… AI4EOSC

Training on single site, centralized dataset expected	Federated learning, split learning, gossip learning, making possible training on decentralized datasets
Single AI application, self deployed or on serverless computing	Composite AI for complex AI tools and applications through function composition and serverless computing
Central management of onboarded sites, complex on-premises deployment	Enhanced onboarding of resources, easier deployment on-premises

(some) New features

Integration with privacy tools (differential privacy, anonymity checks)

Community standards for models API (Kserve) following OpenAPI specifications

ML pipeline composition and workflows

Enhanced web user interface for applications

MLOps tools to monitor deployed models (drift detection, concept drift, accuracy and performance

Improved development environment (VS Code, JupyterLab)

DEEP-HDC, AI4EOSC, AI4OS…

- **DEEP-1**, **DEEP-2**: Platform releases
- Platform and software tightly coupled and interlinked, difficult to self-deploy and customize

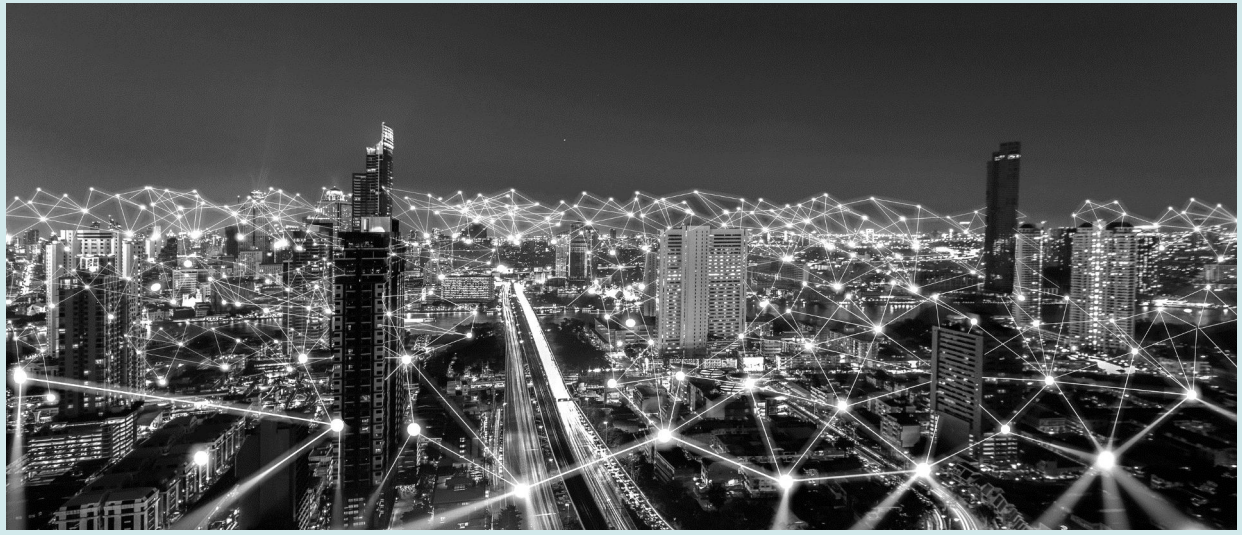
AI4 | eosc

- **AI4EOSC platform** → Platform “powered by AIOS”
  - DEEP-3 → AI4EOSC-3
- **AI4OS** → software distribution
  - Possible to build custom platforms, partially integrated with AI4EOSC platform (i.e. reusing services) or not

AI4 | eosc

Co-funded by the European Union

AI4EOSC Use Cases



Agrometeorology

Integrated plant protection

Automated thermography

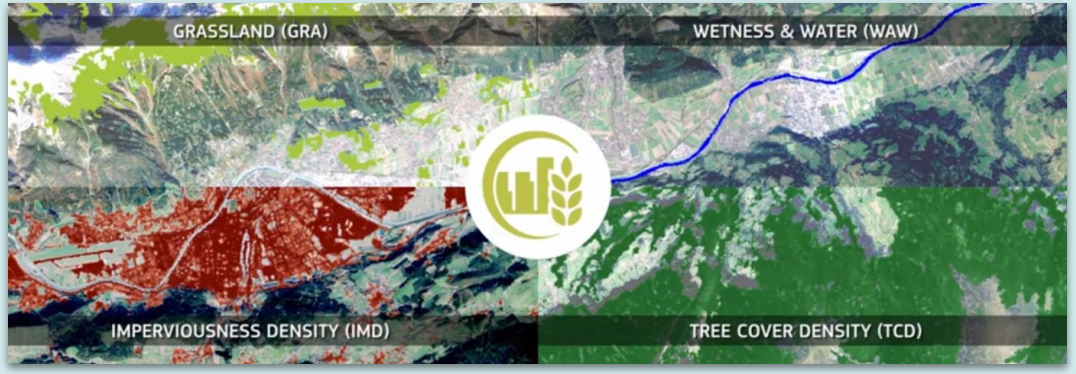
Agrometeorology

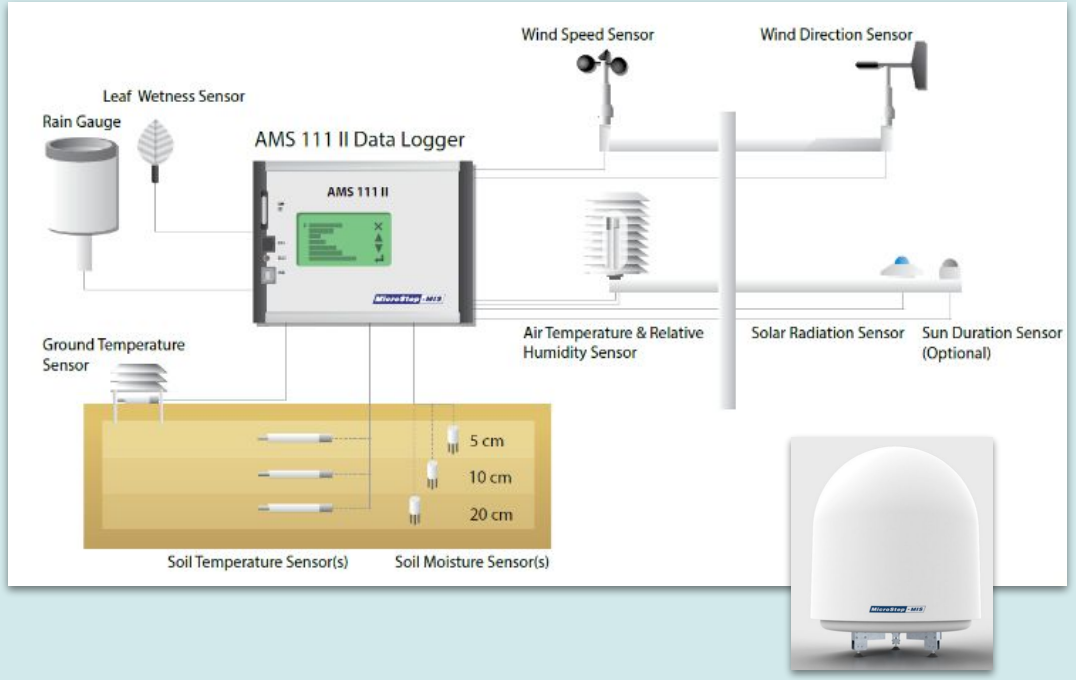
Aim: Usage of satellite imagery, in-site measurements, and weather forecasts to generate added-value products for improving farmers activity: e.g. prediction of phenological or pest development stages.

Currently: Measurement system - TRL9, prediction system - TRL3

Within AI4EOSC: Enhancement of the prediction subsystem following a Composite AI approach to combine the different machine learning models used for the different data sources

Partners: Microstep, IISAS, Predictia





Integrated plant protection

Aim: To determine the risk of disease and pests in agricultural crops and determine the phases of plant growth and the condition of crops. The developed AI models are going to be integrated into existing national advisory platforms, operated by WODR and PSNC.

Currently: WODR and PSNC operate a national advisory platform for farmers (eDWIN), which includes a network of meteorological ground stations, the Farm Management System, and ground observations of the occurrence of diseases and pests. The current solutions are based on predictive mathematical models.

Within AI4EOSC: The plan is to add to the current mathematical prediction models the ML/DL-based models used for recognition of the plant diseases and add new sources of the data. Initial focus on wheat and sugar beats and detection of the fungal diseases.

Partners: WODR, PSNC

Automated Thermography

Aim: To identify heat losses and thermal bridges in buildings and infrastructures using drone-based images and ML/DL approach in order to provide a corresponding automated AI-based service.

Currently: The group owns a dataset of drone-based images on urban districts and drone-based thermal images on a campus district (ca. 0.8TB). The identification of thermal bridges on roofs is already possible using DL (TRL 4). The identification of leakages in district heating networks is possible too (TRL 5/6).

Within AI4EOSC: Targets enlargement of the training dataset, AI model improvement, optimisation of the workflows, and creation of a cloud-based automated service

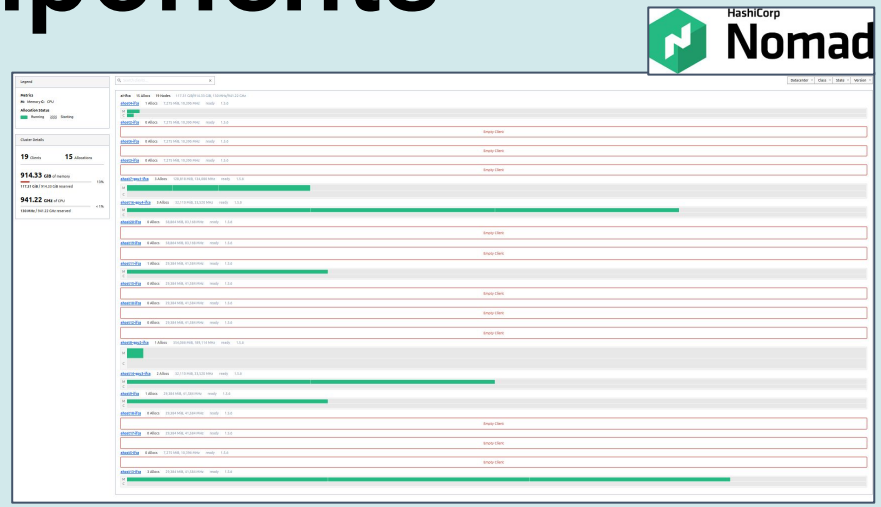
Partners: KIT (IIP, SCC)

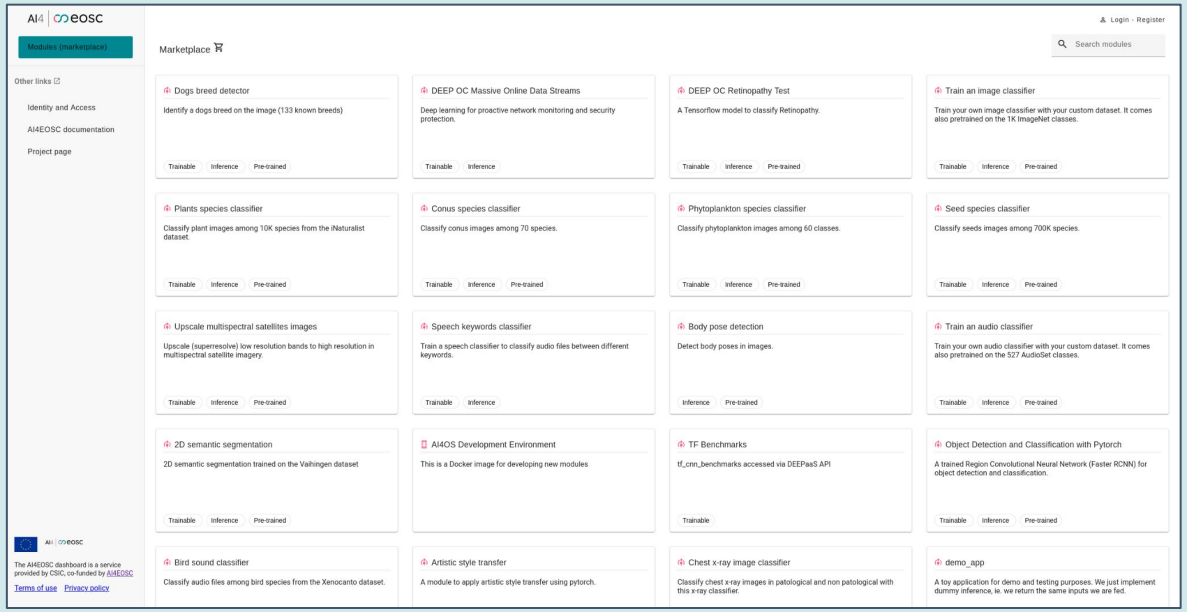
Feature-wise timeline

2022	2023	2024	2025
Project start	Federated learning	Other distributed (gossip, split) learning paradigms	Integration with EOSC storage systems and repositories
Migration from legacy components and adaptation of user tools	Enhanced interactive development environments	Experiment-centric dashboard	Production ready customizable platform and software stack
	Updated data science templates	Automated provisioning of resources	
	Serverless AI as a Service for sync/async inferences	Provenance model and metadata for ML systems	
	Federated multi-site deployments	Consolidation of composite AI tools	
	Drift detection services	MLOps pipelines available	
	Initial version of composite AI tools		

Migration from legacy components

- PaaS layer
  - Moving from Apache Mesos to Hashicorp Nomad + Consul + Traefik
  - Transparent, federated and multi-site platform deployments
  - Update of deprecated internal components at PaaS orchestrator level
- Improved accounting, security and monitoring
- Development of AI4-PAPI, new dashboard and CLI tools
  - Easier to migrate from one COE to another one (just in case)
  - Specific routes for AI and ML (e.g. Federated learning)
  - Removal of TOSCA dependency for user apps
  - Inclusion of additional sidecar tasks (i.e. storage, accounting, monitoring) at API level (i.e. not user-managed)
- Integration with EGI-Check-IN and other OpenID Connect providers

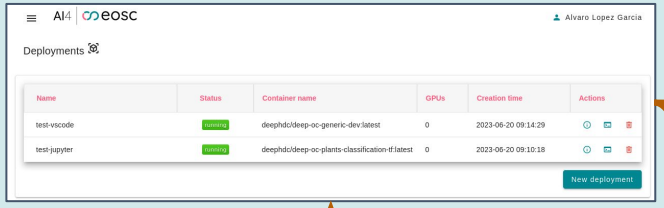


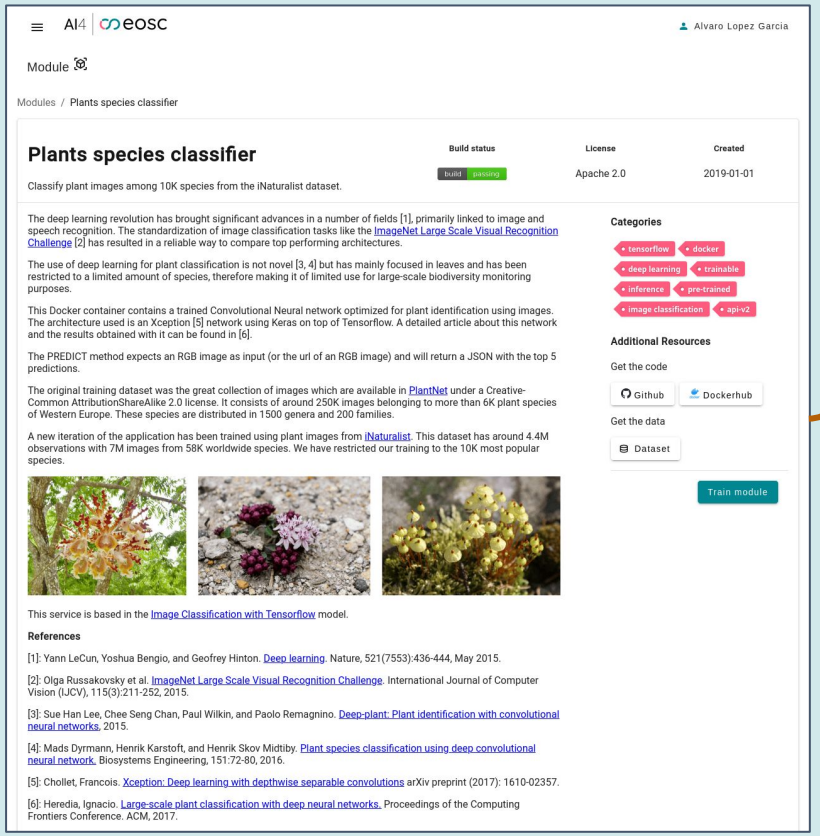


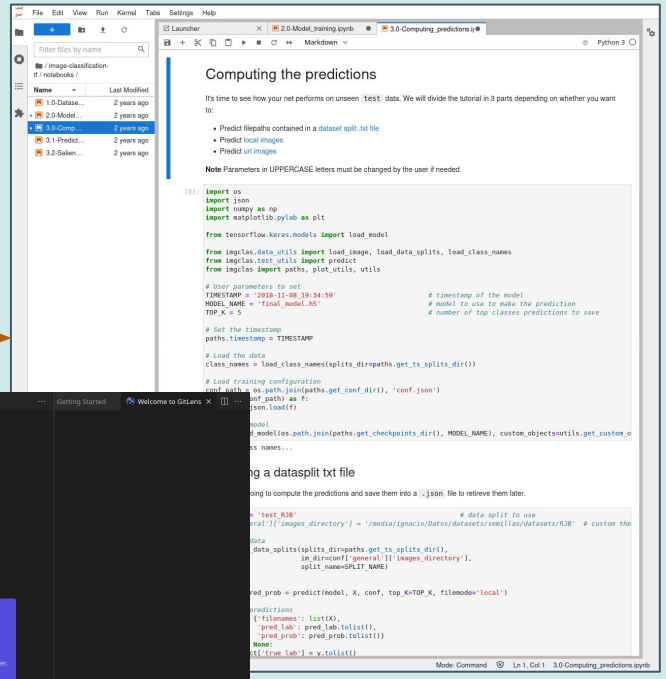
AI4 | eosc

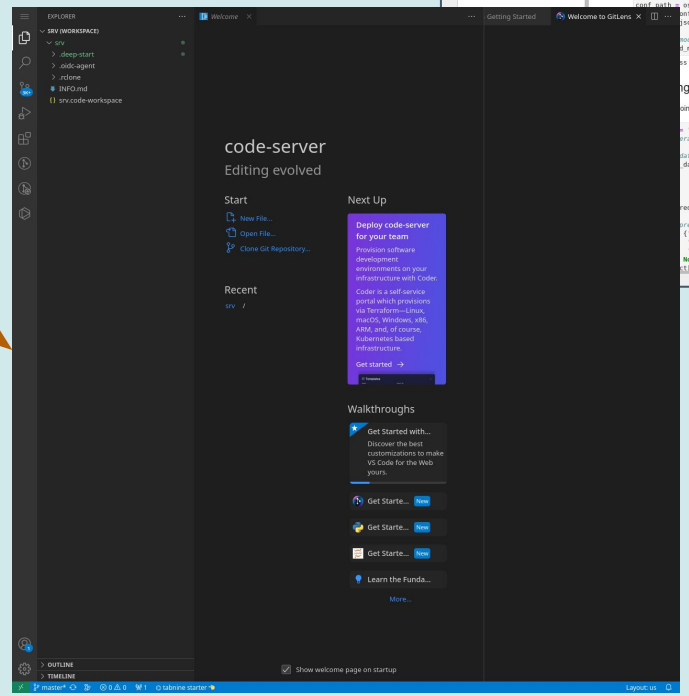
Co-funded by the European Union

Enhanced interactive environments



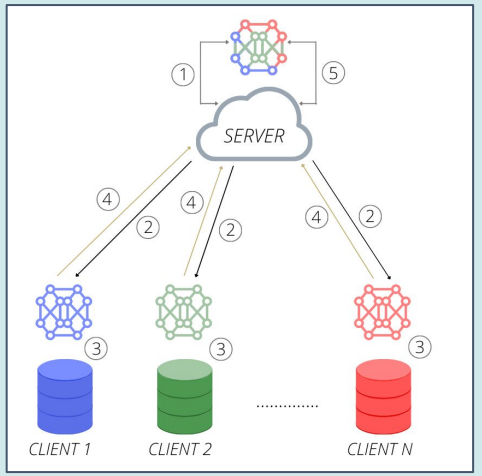


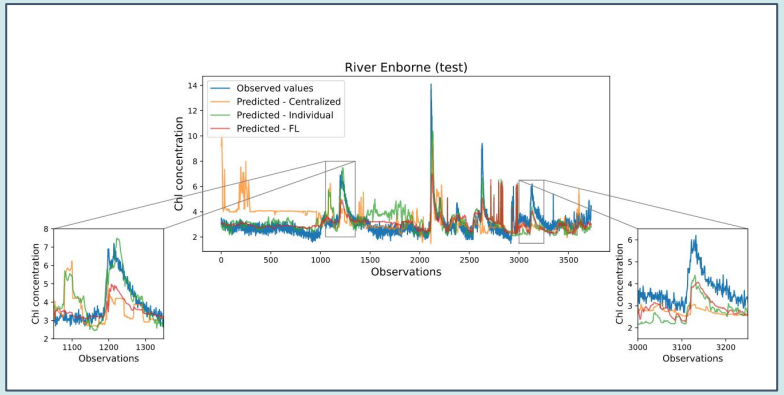


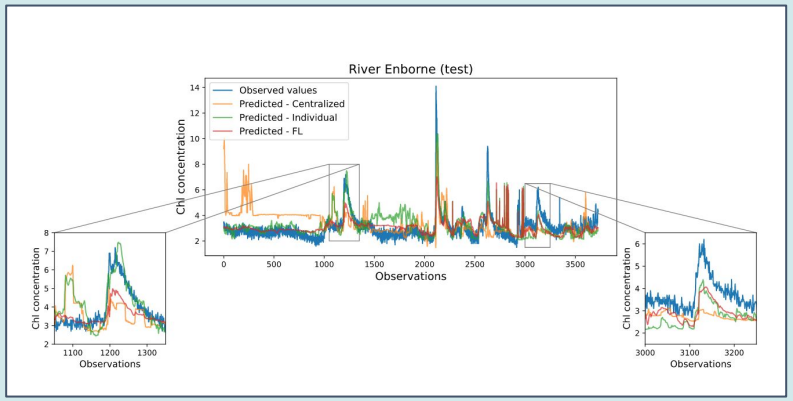


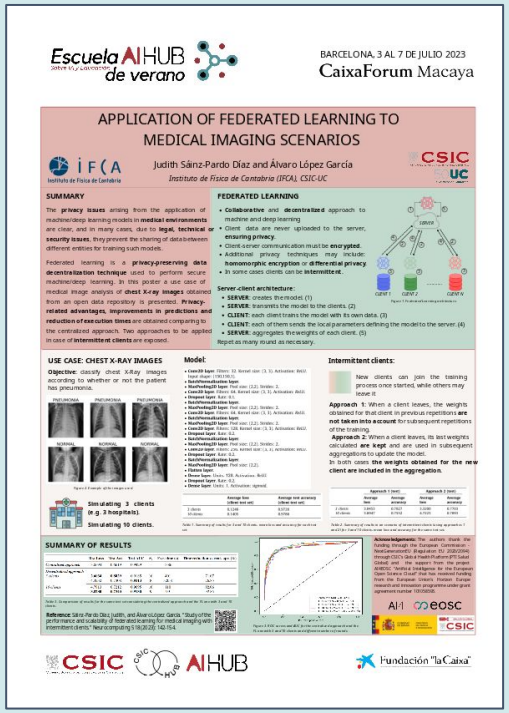
Federated learning

- Collaborative and decentralized approach to build ML models
  - No need to centralize a dataset (i.e. technical or privacy restrictions)
- Management of experiments through platform dashboard
- Participating clients both within AI4EOSC platform or external (with authentication)



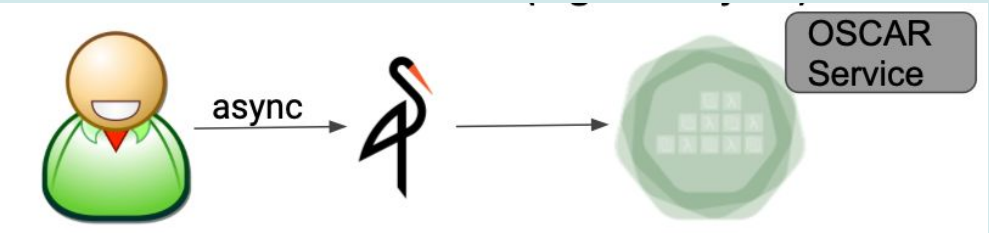




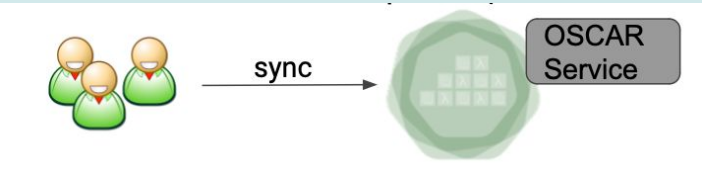


Serverless AI as a service

- OSCAR (https://oscar.grycap.net) will be used to run the AI models for inference (AI as a Service)
  - Serverless event-driven execution
    - Asynchronous Mode: Files uploaded to the object-store trigger the invocation of a data-processing script that is run inside a container (out of user-defined Docker image) within a scalable Kubernetes cluster (e.g. batch jobs)



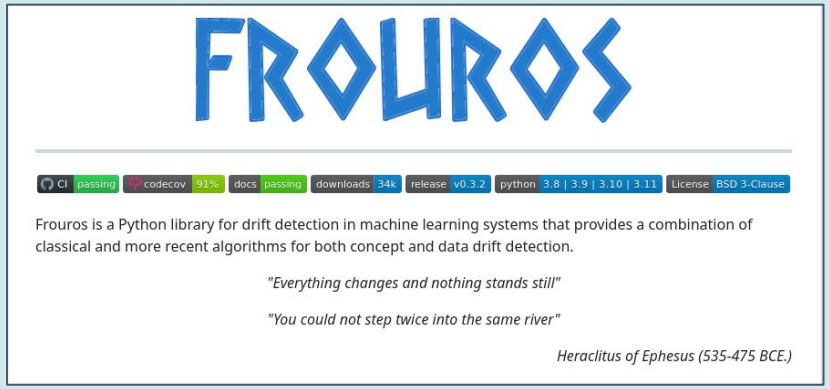
    - Synchronous mode: Scalable HTTP-based endpoints (based on KNative)

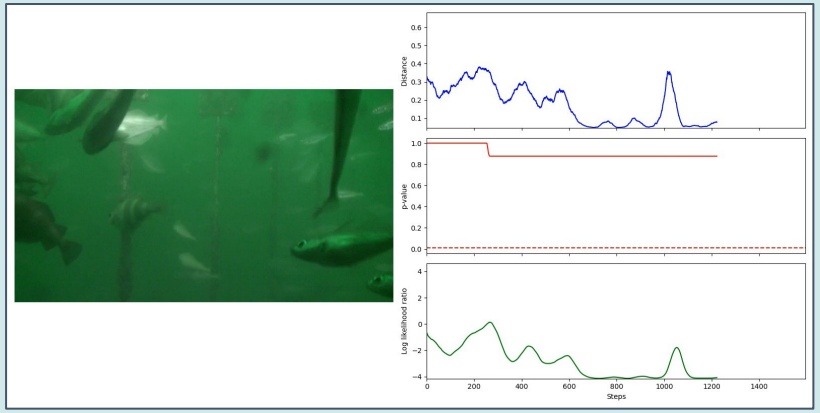


- https://inference.cloud.ai4eosc.eu/ui/#/login

Drift detection libraries and tools

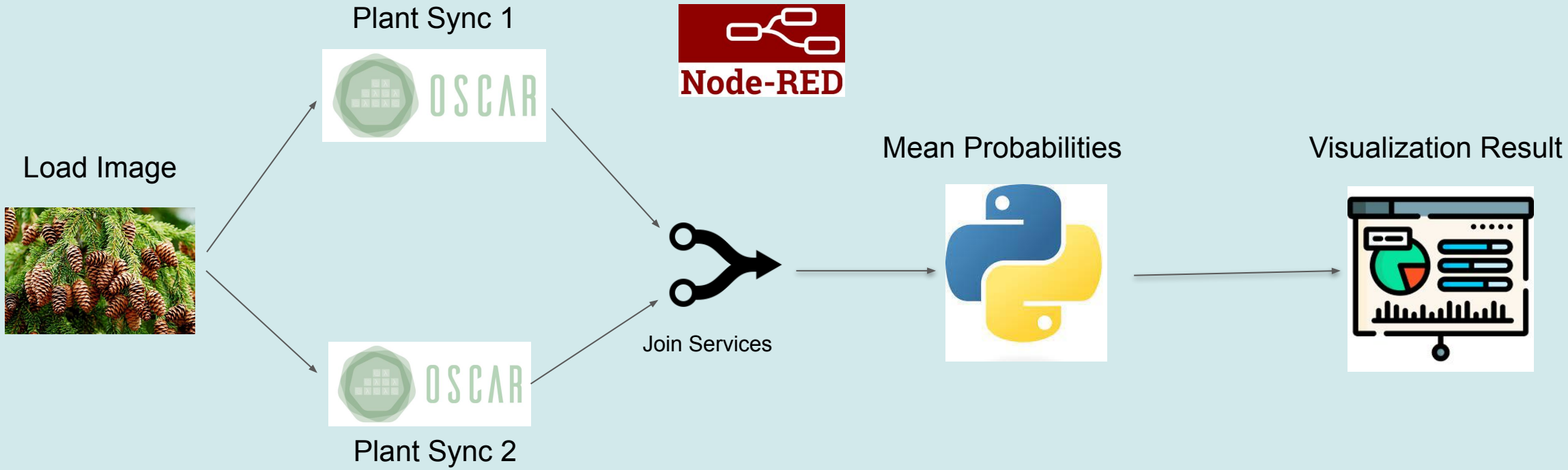
- Monitoring of models in production is not enough
  - Model learns from data, data is not stationary
  - Concept learnt by them model may change over time
- Data and concept drift detection→ essential to build more robust models
- Frouros: state-of-the-art library for drift detection in ML problems
  - https://github.com/IFCA/frouros
- Ongoing work towards online services for drift detection





Example: data drift detection in underwater video

Composite AI models, initial version



- Use case: multiple AI models can be triggered for inference and later aggregate the results for enhanced accuracy
- Reuse functions (subflow)
- Visual support (drag & drop + customization)
- Minimize orchestration costs

AI4EOSC challenges

Integration of disparate resources from different providers across EU e-Infras

Data access and privacy-preserving model training on sensitive data

Correct handling of metadata and quality aspects of AI/ML/DL assets

Exploitation of automation, infrastructure as code and automated orchestration of resources

Federated learning, gossip and split learning, together with national security measures (i.e. homomorphic encryption)

Application of RDA recommendations and participation in RDA FAIR for ML WG

AI4EOSC challenges

Data access, remote access, provider disparity	Sidecar containers providing tailored and transparent access for storage services
Access to accelerators and resources	Users do not need to deal with low-level details, just interact with preferred IDE
AI-related developments are exploding (AI bandwagon)	Focus on scientists and the EOSC ecosystem

Key impacts and outcomes

- Increased number of services in the EOSC Exchange
  - AI/ML generic (AI4EOSC platform) or community specific (e.g. iMagine, see Gergely's talk https://whova.com/portal/webapp/egi_202305/Agenda/3088789)
- 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
  - AI platform for the EOSC (AI4EOSC platform)
    - Integrated popular IDE environments
    - 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

Collaboration: INFRAEOSC and beyod

- Starting and ongoing collaborations with EOSC (i.e. INFRAEOSC) projects to provide AI-based tools and services
  - FAIR-EASE, RAISE, EOSC4Cancer, BlueCloud
  - Exploring collaborations on FAIR-ification of AI assets (RDA, FAIR-IMPACT)
- Outside the EOSC realm
  - iMagine → AI based imaging data and services for aquatic science
  - AioD → Conncectors for AI4EOSC assets
- Open for additional collaborations
  - Reach out at deep-po@listas.csic.es
  - Collaboration with industry through EOSC-DIH (https://eosc-dih.eu/)

EGI Conference 2023

AI4 | eosc

Co-funded by the European Union

AI4EOSC

ai4eosc-po@listas.csic.es

ai4eosc.eu

Reach us!

Thank you for your attention

Project Coordinator: Álvaro López García - aloga@ifca.unican.es