EOSC DevOps framework and virtual infrastructure for ENVRI-FAIR common FAIR data services

Zhiming Zhao, University of Amsterdam, ENVRI-FAIR WP7 leader

Alex Vermeulen, ICOS/Lund University, ENVRI-FAIR WP5 leader

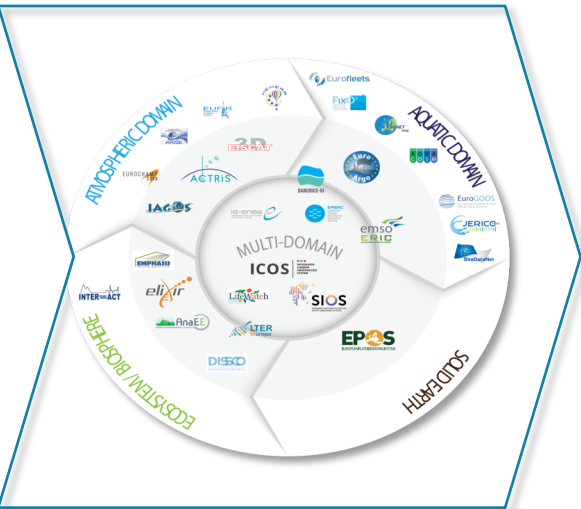**Andreas Petzold,** IAGOS, ENVRI-FAIR coordinator

ENVRI-FAIR development activities
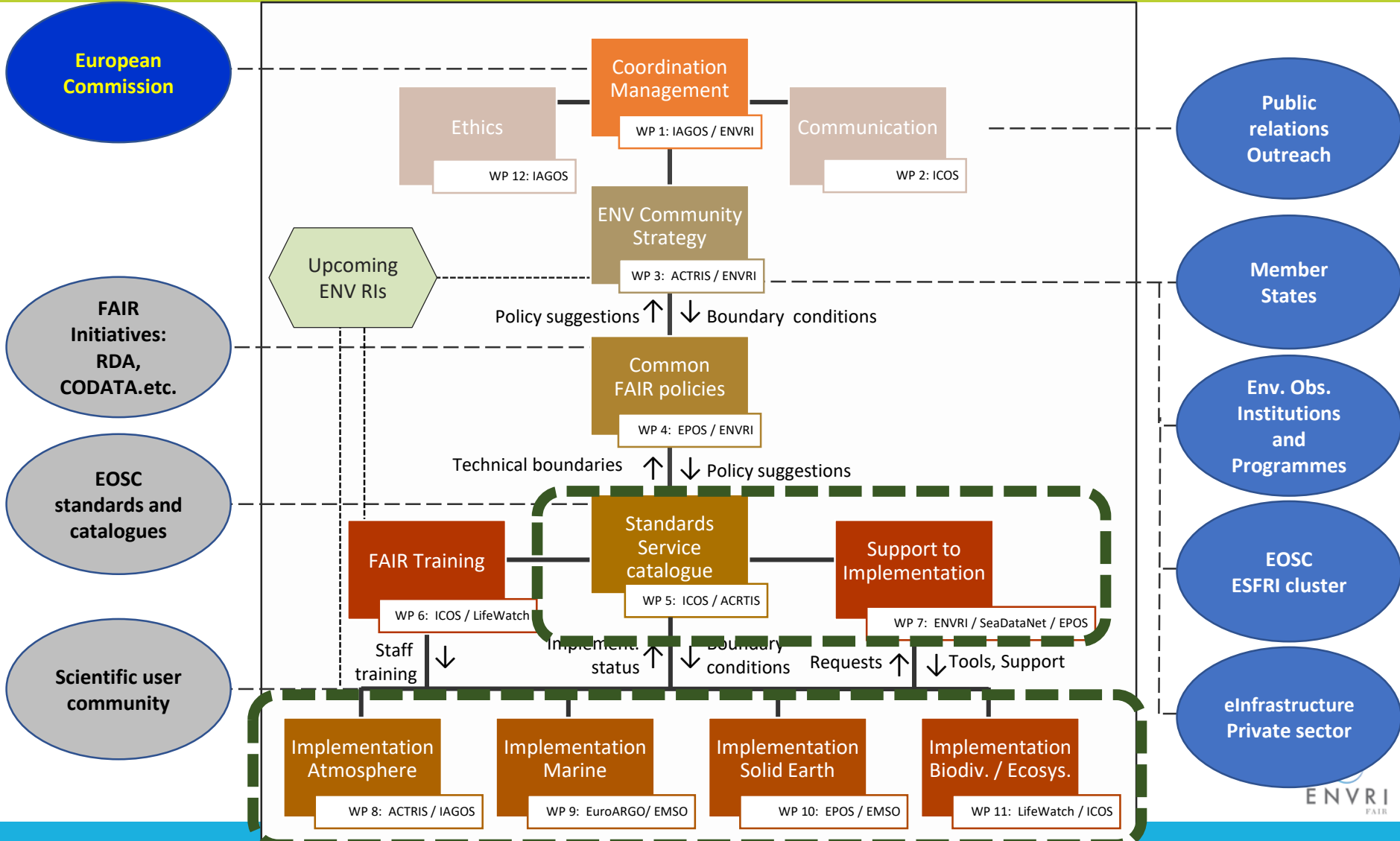
FAIR@cluster

FAIR@sub-domain

FAIR@RI

ENVRI-FAIR development activities

Why EOSC earlier adopter program?

- **A Common test bed** for testing and integrating data management services developed by ENVRI RIs/sub-domains;

- **Gain practices** for using the current EOSC services, for infrastructure (IaaS), for integrating/deploying services (Jelastic), and for scientific experiments (e.g., Jupyter notebook);

- **Preparing for the *integration*** between ENVRI results and emerging EOSC ecosystem.

ENVRI
FAIR

What are the current ENVRI-FAIR cases for EOSC eap

- Case 1: **Automated Cloud execution for data workflow**
- Case 2: **Continuously testing and integration for ENVRI services**
- Case 3: **Notebook based environment for FAIR data access and processing**

Case 1: Automated Cloud execution for data workflow

- **Scenario**: **A VRE** needs to **automate** virtual infrastructure (networked VM) provisioning, software deployment (Dockers or RESTful services), workflow execution, runtime monitoring, and provenance for **workflows** based on FAIR data and services.

- **Objective/expected output**: **demonstrate it in the VREs or in ENVRI RIs** (e.g., LifeWatch or others). It will help the ENVRI community to learn the EOSC services, and build practices for the other similar use cases.

- **Resources:**
  - 4VMs, each preferably with 4 cores and 8G memory and 100GB storage + 1 VM 12 cores, 16GB RAM and 1.5 TB storage. They will be used to configure different testing environments needed by development teams. If possible, resources should be extensible based on user demands.

- **Activities**:
  - test EOSC IaaS API
  - develop the test cases including discovering data/service from catalogues, describing workflow (using workflow system Taverna or directly describing the logic using CWL), and runtime management,
  - benchmark the performances.

Case 2: Continuously testing and integration for ENVRI services

- **Scenario**: ENVRI data management services, each update will trigger an automated testing/integration activities via the DevOps pipeline

- **Objective**: **get familiar with the DevOps/Agile methodologies** for software development, testing and operation.

- **Resources:**
  - Jelastic deployment on EGI Cloud resources
  - 2 VMs (1 for infrastructure services, 1 for user services)
    - 8 vCPU cores, 24 GB RAM, 1TB storage
    - 12 vCPU cores, 24GB RAM, 1.5 TB storage

- **Activities**:
  - set up the basic environment,
  - compare it with different DevOps options, including open source solutions (Travis, Jenkins), EOSC solutions, and AWS/Azure,
  - build best practices for the ENVRI service development

ENVRI
FAIR

# Case 3: Notebook based environment for FAIR data access and processing

- **Scenario:** ENVRI users (users of RIs/VRE) will perform their scientific research using data, software service and models. They often perform such kinds of activities using scientific workflow (like use case 1) or Jupyter notebook. This use case will focus on the second option.

- **Objective**: 1) provide the Jupyter service to users, with examples to access data sets and models, 2) users can perform customised experiments using the notebook services, access data, store the data, publish and share the results with the others.

- **Resources:**
  - Additional resources for EGI-Notebooks instance for the community to be deployed on EGI Cloud federations VMs
  - Storage on a DataHub OneProvider available from the notebooks

- **Activities**:
  - Use JupyterHub deployed by egi (notebooks.egi.eu), with extended resources
  - configure OneData service,
  - develop examples, e.g., from ENVRI RIs like ICOS

ENVRI
FAIR

How will we do?

Requests from ENVRI-FAIR communities
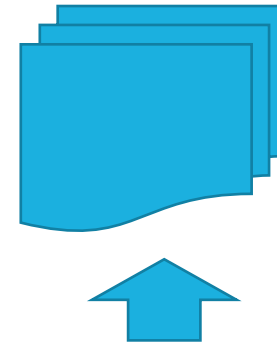
Results: software, demonstrator, reports etc.

Backlog + scheduling + support (ENVRI-FAIR, WP5/7)

Experiments in case 1

Experiments in case 2

Experiments in case 3

EOSC test bed

Time line

Phase 1	Phase 2	Phase 3	Phase 4
Describe the key use case: scenario, scope, KPI, steps etc. **Get familiar with the EOSC services**, following training and practices from the other projects etc. Get the requested resource provisioned Setup OLA with resource providers and agreement with Jelastic.	**DevOps pipeline configured**, including Git, automated testing, integration and **deployment demonstrate** in at least via two service development	**Demonstrate the initial version of the workflow from ENVRI-FAIR**, with automated workflow execution in Cloud; **Demonstrate the other common data services** identified in the ENVRI communities (optional)	**Exploit** the results to the development activities in ENVRI-FAIR sub domain **Sustain** the development by finding new opportunities, e.g., new EOSC projects etc.

ENVRI
FAIR