FAIRiCUBE Hub Architecture is finished

Milestone M3

Due date of Deliverable:	30.11.2022
Actual Submission Date:	30.11.2022
Responsible partner:	EOX
Report Author(s):	Stephan Meißl
Reviewed by:	Stefan Jetschny

Call:	HORIZON-CL6-2021-GOVERNANCE-01
Topic:	HORIZON-CL6-2021-GOVERNANCE-01-17
Project Type:	Research & Innovation Action (RIA)
Project Number:	101059238
Name of Lead Beneficiary:	NILU, Norway
Project Start Date:	1 July 2022
Project Duration:	36-Months

Document History

Version	Date	Authors/ who took action	Comment	Modifications made by
1.0	30-11-2022	Stephan Meißl	First version	

Abstract

This report provides a brief overview of the M3 milestone “FAIRiCUBE Hub Architecture is finished”. It lists the deliverables documenting and proving the milestone and describes further activities contributing to the successful M3 passing.

Table of Contents

1. Introduction

Documenting that the FAIRiCUBE Hub architecture is finished without an associated deployment or document deliverable is arguably not trivial, but documenting the steps taken so far and making forward references to the deliverables D4.1 "FAIRiCUBE Hub Architecture" as well as D4.4 "Operational FAIRiCUBE Hub" due at months 6 and 12 respectively will hopefully make the case.

2. Formal deliverables contributing to M3

There is no formal measure to validate the achievement of the M3 milestone as there are no formal deliverables are associated with this milestone. However, there are some future deliverables that are related to this milestone as summarized in Table 1.

Description	Lead Beneficiary	Type	Dissemination Level	Due Date
D4.1 FAIRiCUBE Hub Architecture	EOX	R	PU	2022-12-31
D4.4 Operational FAIRiCUBE Hub	EOX	S	PU	2023-06-30

Table 1 : Formal future deliverables related to M3

The **report deliverable D4.1** “FAIRiCUBE Hub Architecture” will provide a brief description of the FAIRiCUBE Hub architecture detailing its components, its deployment and operations strategy based on control and worker plane as well as on-boarding requirements and processes for service and app providers.

The **service deliverable D4.4** “Operational FAIRiCUBE Hub” will provide the publicly accessible operational FAIRiCUBE Hub deployment as such. Users will be able to access the publicly accessible parts of the FAIRiCUBE Hub like data and service catalogs directly and will be able to register and request access to processing or data resources.

3. FAIRiCUBE Hub Architecture Summary

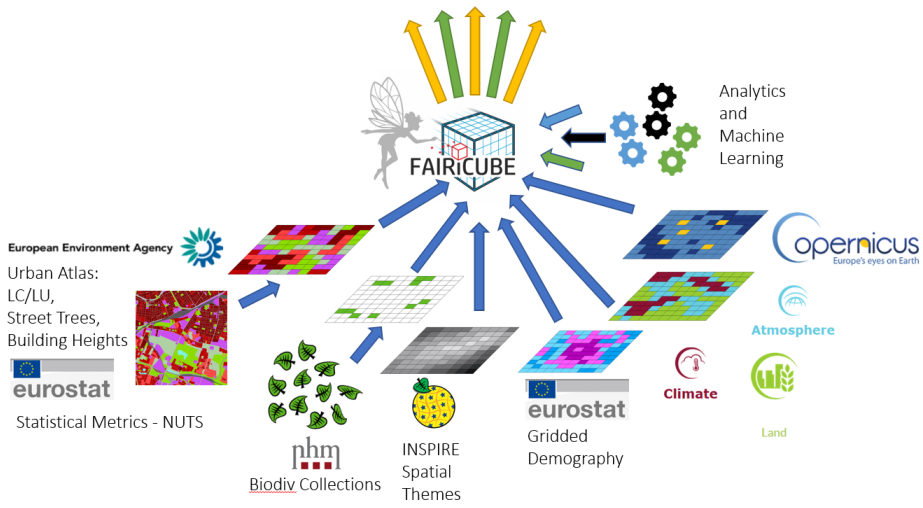
The FAIRiCUBE Hub Architecture will be detailed in the future D4.1. Nevertheless, a summary is provided here.

The core objective of FAIRiCUBE is to enable players from beyond classic Earth Observation (EO) domains to provide, access, process, and share gridded data and algorithms in a FAIR and TRUSTable manner.

To reach this objective, the FAIRiCUBE Hub, a crosscutting platform and framework for data ingestion, provision, analysis, processing, and dissemination, to unleash the potential of environmental, biodiversity and climate data through dedicated European data spaces is created. Within the FAIRiCUBE project, TRL 7 will be attained, together with the necessary governance

aspects to assure continued maintenance of the FAIRiCUBE Hub beyond the project lifespan (Figure 1). It will enable a broader range of stakeholders to focus on what they are supposed to do best: overcome technical barriers to make data-driven decisions and leverage state-of-the-art processing technologies, including machine learning (ML).

Figure 1: Schematic Project Overview



The FAIRiCUBE Hub is the central technology point of FAIRiCUBE, providing access to diverse data sources as well as processing and visualization tools. Users wishing to perform an analysis are able to bring their own data to FAIRiCUBE, whereby non-aligned sources (point, vector, but also gridded data not aligned to the European Grid) will be transformed as required before ingestion. Diverse analysis and ML tools are made available for users to tailor to their specific requirements and data sources. Finally, the FAIRiCUBE Dissemination Dashboard will allow for interactive presentation of results.

Once all required data has been put in place, analysis and machine learning can commence. In order to enable easy access to existing processing modules together with available data, we will utilize EOxHub (https://hub.eox.at/), a platform and workflow management runtime for Earth Observation services and apps. EOxHub will be branded to provide the FAIRiCUBE Hub and deployed on a kubernetes cluster.

The FAIRiCUBE Hub will be centred on the usage of Jupyter Notebooks, allowing to execute them close to the data for simple exchange and sharing of processing modules. The notebooks can either be executed interactively using JupyterLab, for example to develop an algorithm, or in a headless way using a REST API. The data can be accessed in these Jupyter notebooks via different means depending on what is best suited for the use case at hand and the skill level of the user. The available options span from direct object storage access via the xcube or xarray Python libraries to the Process API of Sentinel Hub and OGC interfaces. Jupyter notebooks can further utilize libraries like dask to parallelize and scale processing jobs.

The Figure 2 shows the schematic FAIRiCUBE Hub Architecture based on EOxHub. The App or Service Developer pushes the App or Service software to the code management repository where automatic CI/CD tests, builds, packages, and publishes the App or Service as Docker image after which it is registered at the Marketplace. The App or Service Consumer discovers the App or Service and requests or triggers the deployment of the App or Service to their workspace to be run on the Cloud Infrastructure. The App or Service is now available to be used by the Consumer within the resources provided in their workspace subscription.

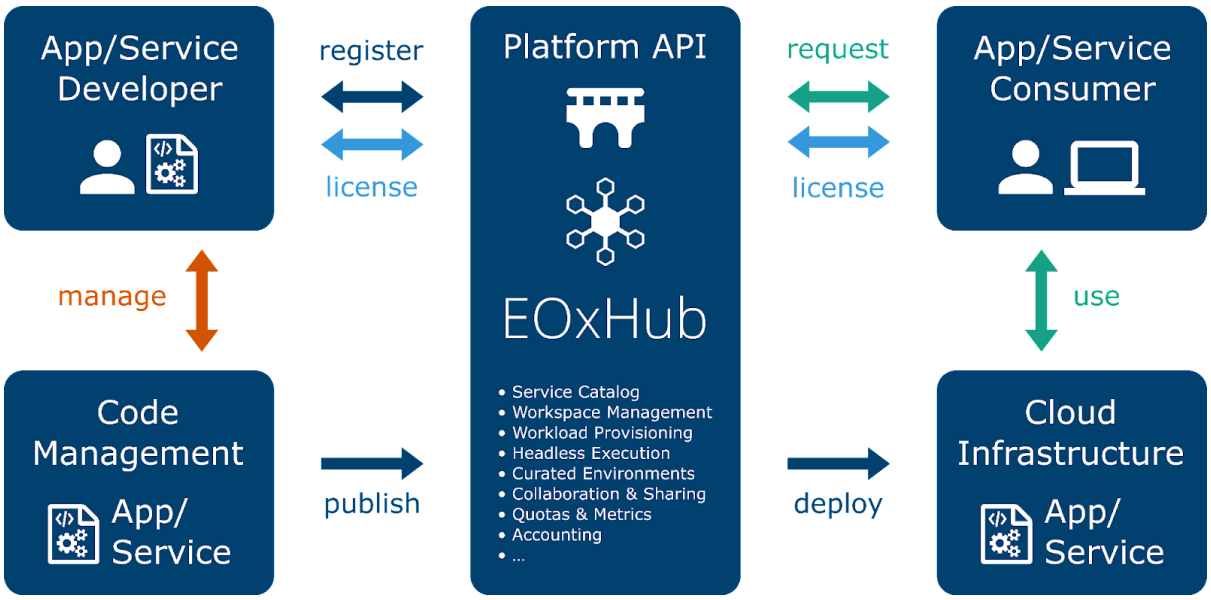


Figure 2: Schematic FAIRiCUBE Hub Architecture

Figure 3 shows the FAIRiCUBE Hub Architecture as Virtual Private Cloud (VPC) in a kubernetes cluster split into control (bottom) and worker (top) plane. The tooling deployed in the worker plane like JupyterLab, pygeoapi, or MLflow is accessed by end users whereas the control plane is accessed by API or cluster management only.

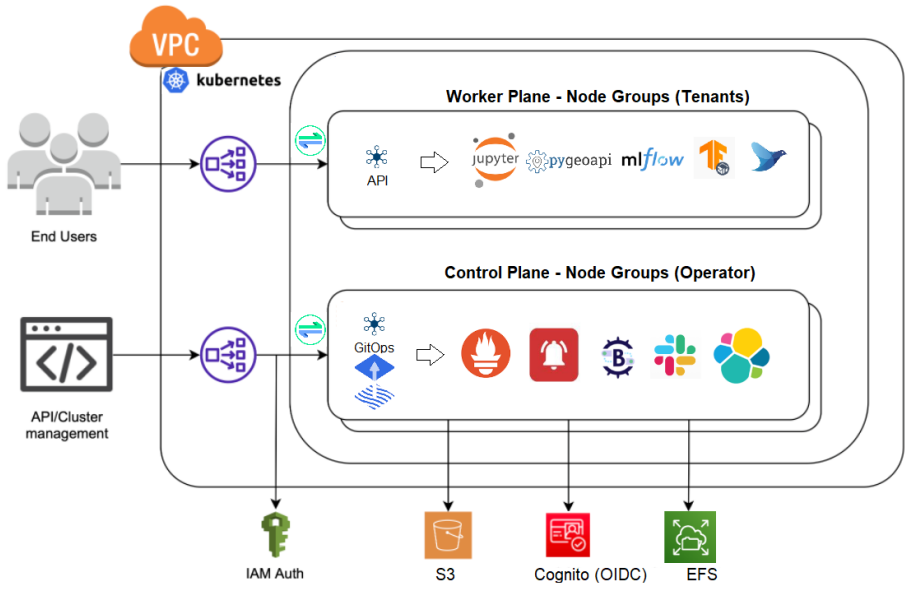


Figure 3: FAIRiCUBE Hub Architecture

Operations of an Exploitation Platform offering Big Data and Machine Learning usually involves substantial cost. It is a challenge to determine beforehand the expected consumption by AI-based Apps of cloud resources for processing and storage. The Marketplace concept therefore must include tools for parameterized cost estimation and quotations for the Consumer. When processing is underway informative and real-time cost accounting results are needed for avoiding cost overruns.

It has been decided among the project partners to use the Frankfurt region of the Amazon Web Services (AWS) as the cloud infrastructure provider.

4. Planning

The next step is to finish the report deliverable D4.1 "FAIRiCUBE Hub Architecture" by the end of the year 2022.

Work has already started to deploy the service deliverable D4.4 "Operational FAIRiCUBE Hub" which is due at the end of June 2023. However, particularly to project partners the FAIRiCUBE Hub shall be available earlier in order to work on the use cases. The current planning is to provide online workspaces to first use cases as early as January 2023.