FAIRICUBE –
F.A.I.R. INFORMATION CUBES

Work Package 4: Share
Milestone 7: FAIRiCUBE HUB in operation

Deliverable Lead: EOX
Deliverable due date: 30/06/2023

Version: 1.0
2023-11-23

Document Control Page

Document Control Page	
Title	Use cases exploratory data analysis released
Creator	Christian Schiller
Description	M7: FAIRiCUBE HUB in operation
Publisher	"FAIRICUBE – F.A.I.R. information cubes" Consortium
Contributors	Christian Schiller
Date of delivery	30/06/2023
Type	Text
Language	EN-GB
Rights	Copyright "FAIRICUBE – F.A.I.R. information cubes"
Audience	☐ Public ☒ Confidential ☐ Classified
Status	☐ In Progress ☐ For Review ☒ For Approval ☐ Approved

Revision History			
Version	Date	Modified by	Comments
0.1	04/09/2023	Christian Schiller	Initial release
0.2	21/11/2023	Christian Schiller	Misc. Updates
0.3	22/11/2023	Christian Schiller	Misc. Updates
1.0	23/11/2023	Stefan Jetschny	review

Disclaimer

This document is issued within the frame and for the purpose of the FAIRICUBE project. This project has received funding from the Horizon Europe research and innovation programme under grant agreement No. 101059238. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the European Commission.

This document and its content are the property of the FAIRICUBE Consortium. All rights relevant to this document are determined by the applicable laws. Access to this document does not grant any right or license on the document or its contents. This document or its contents are not to be used or treated in any manner inconsistent with the rights or interests of the FAIRICUBE Consortium or the Partners detriment and are not to be disclosed externally without prior written consent from the FAIRICUBE Partners. Each FAIRICUBE Partner may use this document in conformity with the FAIRICUBE Consortium Grant Agreement provisions.

Table of Contents

List of Figures

List of Tables

1   Introduction

This deliverable summarizes the implementation and the operational status of the FAIRiCUBE HUB. The core mission 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. A further goal is to leverage power of Machine Learning (ML) operating on multi-thematic datacubes for a broader range of governance and research institutions from diverse fields, who are at present cannot easily access and utilize these potent resources.

Therefore, when providing data, processing functionality and data products to relevant stakeholders all these aspects must be considered. Following the **FAIR principles**, while data is becoming increasingly **findable**, **accessible**, and **interoperable**, true **reusability** depends on the availability and functionality of suitable processing mechanisms. Especially the interoperability and reusability part of data management still causes much trouble and has frequently been found to be highly complex and very time consuming. FAIRiCUBE aims to advance the FAIRness of both data and data analysis and subsequent products by enhancing the reusability of existing data, as well as show, and hopefully overcome, short comings in this area.

The FAIRiCUBE Hub, which bundles all cloud-based services to be executed during project duration, is one of the key infrastructure elements and represents one of the two main objectives of the FAIRiCUBE project. As the use cases (UCs) are very diverse, operate on different scale length & time ranges and cover various scientific fields, we cover a wide range of potential data science tasks. Providing all online services such as accessing, ingesting, storing, processing, and sharing data, in a FAIR, efficient, streamlined, and user-friendly way is a key project component. The various UCs provide the input and feedback for the required functionalities and their correct functioning after their implementation. For a first understanding of the various components of the FAIRiCUBE Hub the following Figure provides an Architecture Overview of FAIRiCUBE. A detailed of all the components description is provided in D4_1_Deliverable_FAIRiCUBE-Hub-Architecture.
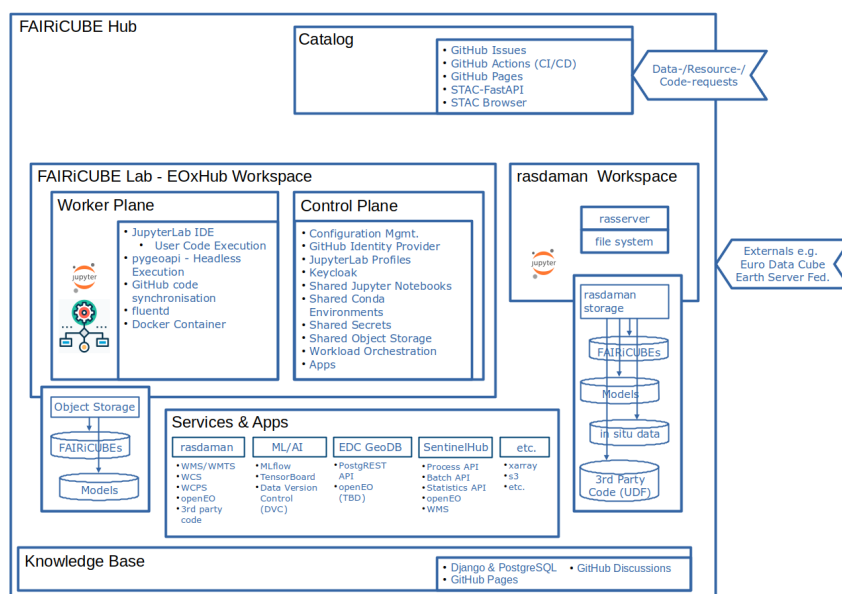

Figure 1 : FAIRiCUBE Architecture Overview

2 Deliverables contributing to M7

There are several formal deliverables contributing to M7 Milestone as listed in Table 1. Further, the milestone represents a demonstration of the already implemented workflows and applications as well as the overall functioning of the FAIRiCUBE Hub and the FAIRiCUBE Lab.

Table 1: Deliverables related to M7

Description	Lead Beneficiary	Type	Dissemination level	Due dates
D4_1_FAIRiCUBE-Hub-Architecture	EOX	DEM	Public	30.12.2022
D4_2 Public Listing (Catalog) of FAIRiCUBE data resources	EOX	R	Public	28.02.2023 31.12.2023 31.12.2024
D4_3 Public Listing (Catalogue) of FAIRiCUBE processing-analysis resources	EPS	R	Public	28.02.2023 31.12.2023 31.12.2024
D4_4_Operational_FAIRiCUBE_HUB	EOX	DEM	Public	30.06.2023
D5_2_FAIRiCUBE ingestion pipelines	JUB	R	Public	30.06.2023

The deliverable "D4_1_Deliverable_FAIRiCUBE-Hub-Architecture" provides a detailed description of the components comprising the FAIRiCUBE Hub, 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 deliverable "D4_2 Public Listing (Catalog) of FAIRiCUBE data resources" compiles the requirements for the metadata, describing the data sets to be used in FAIRiCUBE. To collect all desired field names, discuss, harmonize, streamline and explain them was a very time-consuming task resulting in quite some delay. This deliverable further provides a mapping between the collected metadata fields and the requirements for the STAC[1] standard used for the FAIRiCUBE catalogue.

The deliverable "D4_3 Public Listing (Catalogue) of FAIRiCUBE processing-analysis resources" outlines the metadata requirements necessary for the identification and use of processing/analysis resources created within FAIRiCUBE. In addition, it describes how these user requirements can be supported using STAC (Spatio Temporall Asset Catalogue) catalogue, mapping each concept from the identified metadata requirements to concepts from the STAC standard.

The deliverable "D4_4_Operational_FAIRiCUBE_HUB" describes the operational FAIRiCUBE HUB as an Exploitation platform which was deployed early in the project and gets continuously extended with concrete services, apps and data offerings as required by the use cases.

The deliverable "D5_2_Deliverable_FAIRiCUBE-Ingestion-Pipelines" describes how data sets are ingested into the catalogue and into the two data management systems, comprising the FAIRiCUBE Hub.

---

1   https://stacspec.org/en

3   Progress summary

The overall FAIRiCUBE Hub Architecture has been setup including the configuration and roll-out of the **FAIRiCUBE Lab** infrastructure, its user management, and use case relevant configurations of the respective Jupyter Notebooks according to each of the UC's requirements. For each UC a distinct setup providing the required processing resources (CPU/GPU, RAM, etc) and initial software tools have been provided. As a second important pillar for processing and data management a **rasdaman** data base and virtual machine has been setup to support users needs. This provides users with the ability for detailed queries and processing steps on ingested datacubes.

Since an initial version of the FAIRiCUBE Hub went operational in early summer, the use cases are now starting to explore the FAIRiCUBE Hub and all its possibilities. These explorations result in new requirements, which could not be foreseen a priori, which further condensed in initially unplanned updates of the Architecture document. An additional delay, which we wanted documented in the Architecture document, relates to the ingestion of the data and metadata, which is described in more detail below. However, both these facts lead to the delay of this document.

To make such a complex construct as FAIRiCUBE, for such a diverse user community, working, information about the availability and accessibility of data is a must. To support this, need a **catalogue** has been set up. As an initial version the static FAIRiCUBE catalogue, based on STAC, (https://catalog.fairicube.eu/) has been made available and has been initially filled with information, in-kind derived from Euro Data Cube (EDC). Concurrently, a dynamic STAC-API based catalogue version is adapted for FAIRiCUBE (https://catalog.eoxhub.fairicube.eu/), and scripts to fill this catalogue with the respective information, provided by the use cases, are currently under development. As initial input the same data, as avalable in the static catalogue, is also accessible via the dynamic catalogue, which provides additional features e.g. search functionality. Once the STAC-API based catalogue development reaches a higher level of maturity it will replace the current, static, version and the catalogue access will be made at https://catalog.fairicube.eu/ as the sole version.

Together with the use cases a series of required metadata input fields are currently collected and harmonized between use cases and data integrators. Once all fields are agreed upon, the implementation of the metadata and **data ingestion** scripts will be finalized and shall ultimately allow for semi-automated processing and ingestion via CI/CD pipelines of the data- & metadata-resource requests into the FAIRiCUBE catalogue. Unfortunately, this process of metadata harmonization is currently still ongoing since the highly diverse nature of the use cases in FAIRiCUBE resulted in a very complex metadata collection which have to be synced and adapted with the requirements of EO data processing.

In addition, the initially foreseen ingestion process, based on usage of GitHub Issues for metadata preparation and ingestion, showed some unexpected limitations in the number of possible fields to support. These limited number of field was far below the required number desired by the FAIRiCUBE users. We therefore had to invent and implement another solution, based on a **WebGUI** for data collection and editing and GitHub as storage and discussion platform. This system is now implemented and provides, in addition, better cross-checking and formatting options (e.g for harmonized units of measure or other predefined sysntax).

A further important component of FAIRiCUBE is the **Knowledge Base** (KB) (https://fairicube-kb.dev.epsilon-italia.it/). Its core task is to provide the community with a set of tools, documents, algorithms, code, tips and tricks, mistakes to avoid, and examples of use. A self-training library contains a set of links to web pages and project resources, appropriately selected and organised into categories, with the aim of provide the user with a basic background on FAIRiCUBE Knowledge Base topics. An interactive query tool provides the user with an extensive search possibility inside the KB. The architecture of the KB is composed by a web-application, a database and multiple data sources.

Preparation for the utilization/integration of FAIRiCUBE associated resources sponsored by the European Space Agency (ESA) via their Network of Resources (NoR) programme have been conducted and the possibility to execute processes in an asynchronous way (headless Jupyter notebook execution) have been prepared and demonstrated. The use cases are now taking advantage to fully utilize the available funded resources.