



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

Work Package 3 : Process

Milestone 5 : Use cases exploratory data analysis
released

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

Version: 1.0
2024-11-21

Document Control Page

Document Control Page	
Title	Use cases exploratory data analysis released
Creator	Stefan Jetschny
Description	M5 Use cases exploratory data analysis released
Publisher	"FAIRICUBE – F.A.I.R. information cubes" Consortium
Contributors	Mohamed-Bachir Belaid
Date of delivery	30/06/2023
Type	Text
Language	EN-GB
Rights	Copyright "FAIRICUBE – F.A.I.R. information cubes"
Audience	<input checked="" type="checkbox"/> Public <input type="checkbox"/> Confidential <input type="checkbox"/> Classified
Status	<input checked="" type="checkbox"/> In Progress <input type="checkbox"/> For Review <input type="checkbox"/> For Approval <input type="checkbox"/> Approved

Revision History			
Version	Date	Modified by	Comments
0.1	27/06/2023	Stefan Jetschny	Initial draft with headlines
0.2	21/08/2023	Stefan Jetschny, Mohamed-Bachir Belaid	Internal review and minor updates, draft for review
1.0	30/08/2023	Stefan Jetschny	Minor updates, draft for external review
1.0	13/10/2023	Jaume Targa	Final review



Disclaimer

This document is issued within the frame and for the purpose of the FAIRiCUBE project. This project has received funding from the European Union's Horizon 2020 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

- Document Control Page2
- Disclaimer.....3
- Table of Contents.....4
- List of Tables5
- 1 Introduction.....6
- 2 Deliverables contributing to M57
- 3 Summary.....8



List of Tables

Table 1: Formal deliverables contributing to M5 _____ 7



1 Introduction

The use cases executed under the FAIRiCUBE Hub and during the project duration are not formally a key delivery from the project but nevertheless serve a significant role: to test, develop, improve, and co-define the FAIRiCUBE hub, which bundles all cloud-based services to be executed while solving typical data science research questions. As the use cases (UCs) are truly diverse, operate on different scale length and time ranges, and cover various scientific fields, we thereby 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 and only by executing our UCs, we can determine which components of the FAIRiCUBE Hub are already sufficiently working and where there is room for improvement.

As part of the data science work in WP3 and in close collaboration with the WP2 domain experts, we studied the nature and characteristics of the data identified by the UC owners as being relevant to solving the UC research question(s). During the exploratory data analysis conducted for each UC, the numerical formats, statistical properties, and the general state of the data (origin, completeness, accuracy, etc.) were thoroughly investigated, which is an essential input for all further data science work. This M5 milestone is therefore the foundation for *M6: "Machine learning strategy specific for each use case released"*. Additionally, the completion of each UC exploratory data analysis has wide implications for data processing, machine learning algorithms, and the numerical and financial costs of executing them. During the data analysis, the relations between different sources are tested and will iterate back to the formulation of each research questions.

Milestones M5 and M6 function as indicators of UC progress and simultaneously as demonstrators of the data science and machine learning capabilities enabled by the FAIRiCUBE Hub.

2 Deliverables contributing to M5

There is one formal deliverable contributing to M5 Milestone and that is deliverable D3.1 as listed in Table 1. Note that the UC exploratory analysis describes the current state. Each UC, with updates foreseen as the UC execution unfolds.

Description	Lead Beneficiary	Type	Dissemination level	Due date
D3.1 UC exploratory data analysis	NIL	R	Public	30.06.2023

Table 1: Formal deliverables contributing to M5

In practice, several deliverables from WP2 “*use*” and WP5 “*ingest*” have enabled the milestone M5 by formulating and updating the analysis plan (D2.2 *Report on UC Analysis Plans*), i.e., the UC research questions, and provisioning and synchronizing the input data (D2.3 *UC Ingest_Process Synergy Report*, D5.1 *List of datacube resources*, D5.2 *Ingestion Pipelines*, D5.3 *Validation of ingestion*). Finally, all data science work is supposed to be carried out on the FAIRiCUBE Hub which is described by the deliverable D4.1 *FAIRiCUBE Hub Architecture*.



3 Summary

Each use case (UC) has defined domain specific research questions, documented as use case analysis plans, and they are being addressed as data science tasks. As part of the data science work, a thorough analysis of the suitability and characteristics of the input data was carried out (exploratory data analysis), which is the foundation for the development of a machine learning strategy. The result of this analysis is a snapshot of the progress of each UC and will be updated during the project duration if necessary. Key elements of the milestone will be made available as part of the UC scientific publishing and on our [project website](#).