

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

WP3 Processing

D3.4 Processing knowledge base services

Deliverable Lead: Stefania Morrone (EPS)

Deliverable due date: 29/02/2024

Version: 1.0

2024-03-18



Document Control Page

Document Control Page	
Title	D3.4 Processing knowledge base services
Creator	EPS
Description	D3.4 Processing knowledge base services
Publisher	"FAIRICUBE – F.A.I.R. information cubes" Consortium
Contributors	EPS
Date of delivery	29/02/2024
Type	R — Document, report
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 type="checkbox"/> In Progress <input type="checkbox"/> For Review <input checked="" type="checkbox"/> For Approval <input type="checkbox"/> Approved

Revision History			
Version	Date	Modified by	Comments
0.1	08/02/2024	Stefania Morrone, Antonio Cozzolino, Liliana Martirano, Giacomo Martirano	First Draft
0.2	23/02/2024	Mirko Gregor	Internal review
1.0	23/02/2024	Stefania Morrone	Final version



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 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 Page	2
Disclaimer	3
Table of Contents	4
List of Figures.....	5
1 Introduction	6
2 KB design	7
2.1 Content organization	8
2.1.1 KB sections.....	8
2.1.2 Relation with the community collaboration platform	9
2.2 Content Management.....	9
2.3 Components and structure.....	11
3 The Query Tool.....	12
3.1 The Query Tool interface	12
3.2 The Query tool infrastructure.....	15
4 KB implementation and next steps	16



List of Figures

Figure 1 : KB Homepage	9
Figure 2 : Tips & Tricks section	10
Figure 3 : KB components and structure	11
Figure 4 : Query Tool landing page	12
Figure 5 : Query result example	13
Figure 6 : Simplified Queries	14
Figure 7 : Custom Queries	15

List of Tables

Table 1 : List of Knowledge Base requirements	7
Table 2 : Knowledge Base sections	8



1 Introduction

This document describes the FAIRICUBE Processing Knowledge Base services (hereafter the “KB”). Driven by the needs and built on the experience and understanding of the project use cases, the KB aims to share knowledge enabling the extraction of real value and insights from large and complex data collections, with a focus on data-driven machine learning (ML) model generation.

The development of the KB is part of the activities related to Task 3.5 and should be considered in the general context of WP3, i.e. to provide guidance, recommendations, technical and implementation expertise to the FAIRICUBE use cases in terms of data analysis and processing. At the time of writing, a demo application of the [FAIRICUBE Processing Knowledge Base](#) is made available online.

While the initial users of the KB are the use case managers, the potential end users are diverse and have varying levels of technical expertise, ranging from scientists wishing to apply the models developed by the use cases to different areas, to urban policy makers interested in extracting insights from large and heterogeneous data collections to assess their city's resilience to climate change.

To meet the needs of diverse types of users, the core of the KB is a query service, namely the Query Tool, which allows the project's processing resources to be interactively searched in a variety of ways, depending on the user's level of expertise. As detailed in Section 3, the Query Tool builds on the information provided by the use cases in the metadata of their processing resources and allows keyword searches, predefined queries and, for advanced users, the ability to create complex queries by specifying parameter values. This flexibility is also valuable for users who want to take a more granular approach to their research. Complementing the Query Tool, the other parts of the KB act as a gateway to information contained in internal project resources (such as the documentation of successful or failed use case processes available in the project's GitHub repositories) as well as external resources freely available on the web, providing users with additional insight and information. The availability of an interactive tool for querying the processing resources of the use cases, coupled with online documentation that provides context and insight into the data analysis and processing procedures and tools, ensures that users can extract value from the KB with ease and efficiency.

In sharing the knowledge and experience gained in the project, the KB complements the Community Collaboration Platform (described in Deliverable 4.4). Specifically, while the Community Collaboration Platform focuses on the exchange of information between experts and guides on how to use the FAIRICUBE Hub, the KB facilitates access to and contextualisation of the resources available in the Metadata Catalog, in the GitHub repositories and in the Community Platform itself. More details are provided in Section 2.1.2



2 KB design

The design of the KB has considered the heterogeneity of potential users and the fact that some of them might not have specific or even generic technical skills while others might be specialists looking for very specific solutions. A guided path, including a searchable interface, was envisioned.

To better define the guided paths, three basic operational scenarios were considered:

- A biologist, with technical skills but no AI training, looking for a working example to analyse complex data collections and understand what factors make species grow happily.
- A researcher with an AI background, looking for details of applications and algorithms applicable in similar contexts/scenarios as FAIRiCUBE.
- The head of an environmental department, a data policy expert who is not interested in technical details but wants to understand whether the department's technical staff could reuse FAIRiCUBE results to solve similar problems.

Based on these potential scenarios, the following functional and non-functional requirements for the KB were formalised as listed in Table 1.

Table 1 : List of Knowledge Base requirements

Req. ID	Req. description
R1	GitHub will be used as a platform to store technical artefacts and documentation (code, models, etc) as well as to have discussions and document issues. An interface allows non-technical users to avoid interacting directly with GitHub.
R2	The KB landing page is available from the website home page, as main menu item or as subitem.
R3	A user interface allows "queries" to the KB. Basic (e.g., search by keywords) as well as complex customized queries must be enabled.
R4	The KB landing page contains: <ul style="list-style-type: none"> • description of the KB scope, structure and functionalities • links to different KB sections
R5	The section landing pages contain: <ul style="list-style-type: none"> • descriptive text • links to related GitHub repositories • links to specific artefacts
R6	The use cases sections must contain descriptive text as well as links to the metadata of related datasets and a/p resources. It must complement without overlapping the website uc description page as well as uc-specific repositories descriptions.
R7	The KB content should interact with and complement to the community collaboration platform



2.1 Content organization

The KB is largely based on existing content in the project's [GitHub repositories](#) and on the documentation of processing and analysis resources (a/p) in the [metadata catalog](#), but it also includes input from selected external web resources and links to the Community Collaboration platform.

While [the use case repositories](#) provide knowledge based on the experience of the relevant use cases, the [common-code](#) repository hosts documents and code that can be reused across use cases and beneficial to different types of users. Scripts to create cloud optimized GeoTIFF, step-by-step guide on how to use QGIS to create binary layers from raster data are examples of artefacts that could be found in the common-code repository. To make GitHub information more easily classifiable as relevant to the KB, appropriate categorisation of the available content has been carried out using labels.

To support the understanding and reuse of the project's results and resources, freely available external resources from the web have been selected and grouped by topic, giving users an overview of macro-categories related to data analysis and processing (pre-processing, machine learning, deep learning...)

2.1.1 KB sections

At the time of writing, the KB content is grouped into the below six sections, that can be accessed from the menu bar displayed at the top of each page.

Table 2 : Knowledge Base sections

KB section	Brief description
Use Cases	This section summarizes main aims and results (success and failures) of the use cases and provides access to more in-depth descriptions and specific resources e.g. shared via GitHub. In this section, for each use case, there is a set of "how to use" and "try me" pairs to provide an access point to the most relevant or ready-to-use resources.
Query Tool	This is the engine of the KB. It allows to query KB processing resources metadata based on predefined or customized queries, e.g. based on the type of algorithm used.
Self-training Library	This section links to several training resources freely available from the web and/or from the project repositories to support understanding and reuse of the project outcomes and resources.
Tips & Tricks	This section is aimed to share solutions adopted / workarounds to overcome various challenges faced in the UC lifetime
Metadata Catalog	This section gives access to the Metadata Catalog.
GitHub Project	This section gives access to the FAIRiCUBE GitHub repositories.

The KB landing page landing page is shown in Figure 1 below and describes the scope and structure of the FAIRiCUBE Knowledge Base.

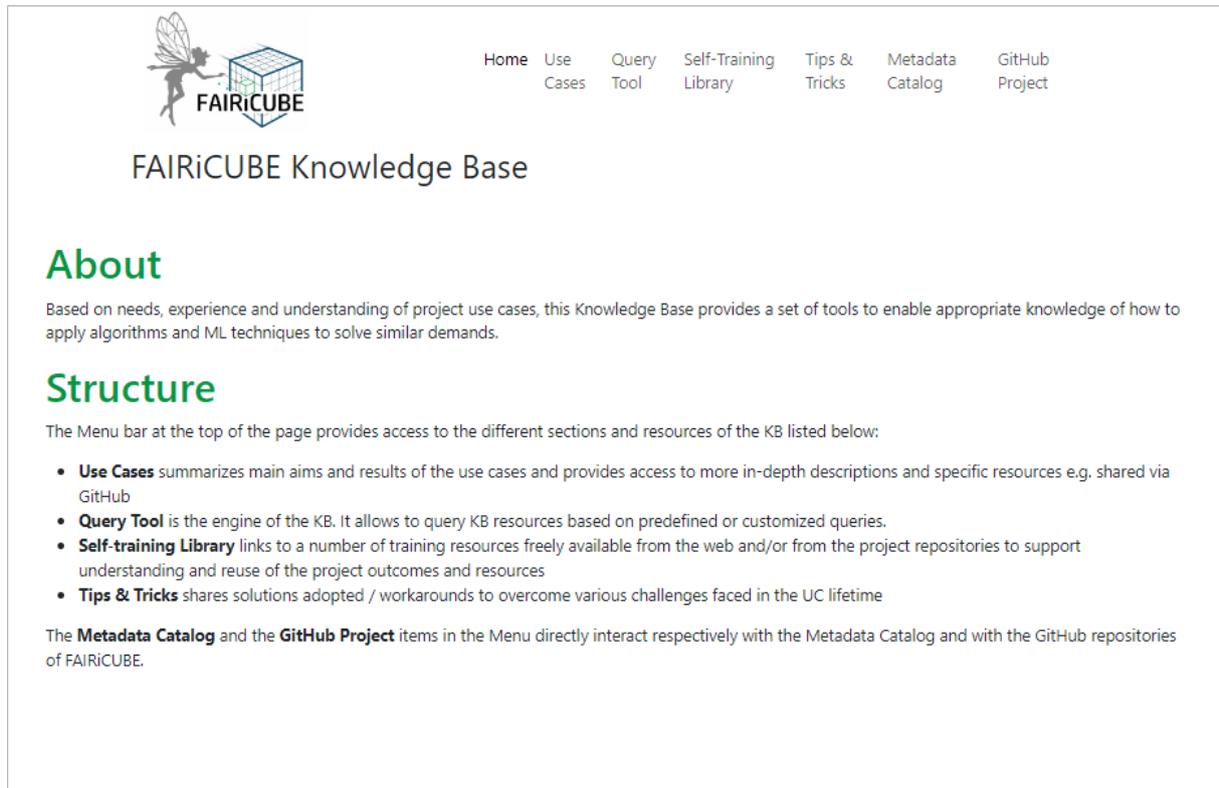


Figure 1 : KB Homepage

2.1.2 Relation with the community collaboration platform

Both the Knowledge Base and the Community Collaboration platform aim to provide guidance, recommendations, technical and implementation expertise. The [Community Collaboration platform](#) is specifically deputed to the exchange of information between experts using or willing to use data cubes for environmental studies. Additionally, it provides guidance on how to use FAIRiCUBE Hub.

The primary purpose of the KB is to help users with diverse levels of expertise understand if and how it is possible to reuse the experience of and the resources developed by the use cases for their specific needs. It's being discussed whether it would be better to make all training documents available on the same platform to further improve user experience. In particular, it is being considered whether the content of the KB Self-Training section should be moved to the Community Collaboration platform.

At the time of writing, no final decision has been taken, partly due to the still ongoing implementation of both platforms. The content of the Self-Training section is currently being reviewed and reorganised to avoid overlap, including cross-referencing with content on the Community Collaboration Platform.

2.2 Content Management

Partners can contribute to the KB content through:

- Documentation of the a/p resources (algorithms, pipeline, preprocessing, etc.) developed by their use cases. This happens via the creation of related a/p metadata. Metadata creation (detailed in Deliverable D4.3) is performed using a Python webform where, after authentication, the users input all required information. At the end of the metadata ingestion process, the related information is immediately available to the KB users using the Query Tool.



- Provision of additional information related to the processing resources. For example, it is possible for the use cases to document their way to success, as well as their failures or specific workaround adopted. This can be done by opening issues in the issue tracker of the [FAIRiCUBE/Knowledge-Base](#) and by labelling these issues as “relevant for KB”. The “Tips & Tricks” section of the KB (illustrated in Figure 2) will collect this information and make it available to the user in the FAQ form.

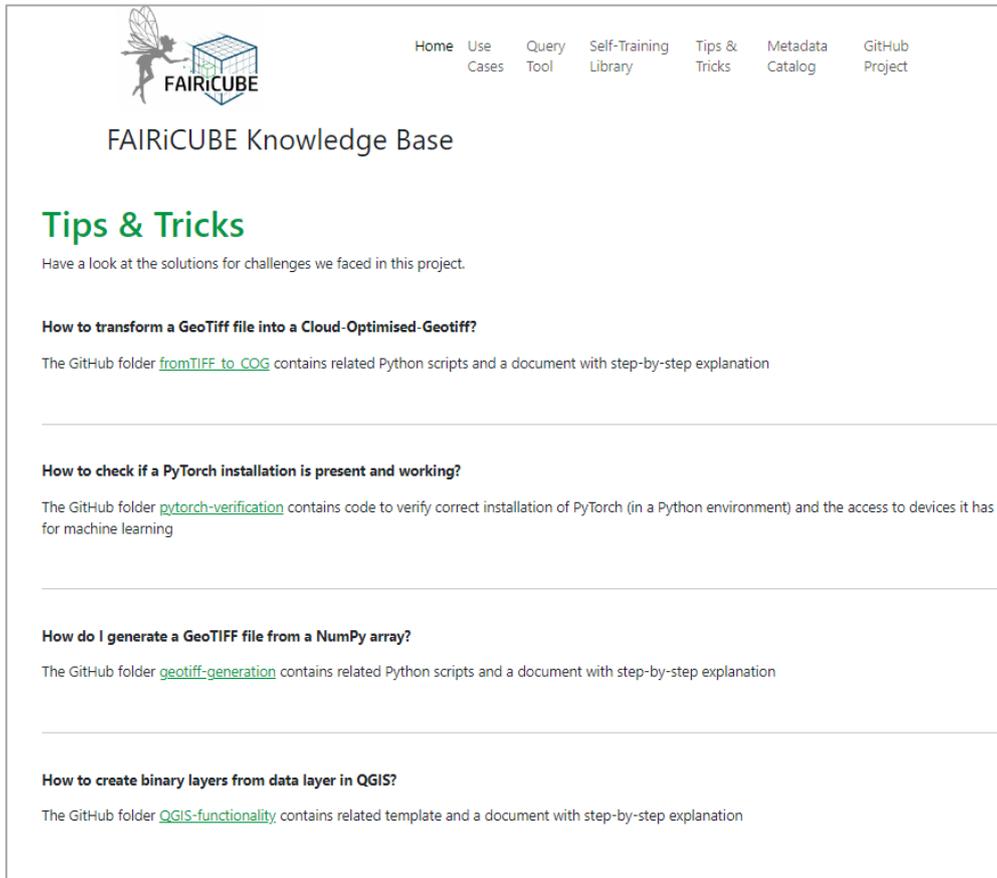


Figure 2 : Tips & Tricks section

2.3 Components and structure

The components and structure of the KB are illustrated in Figure 3 below.

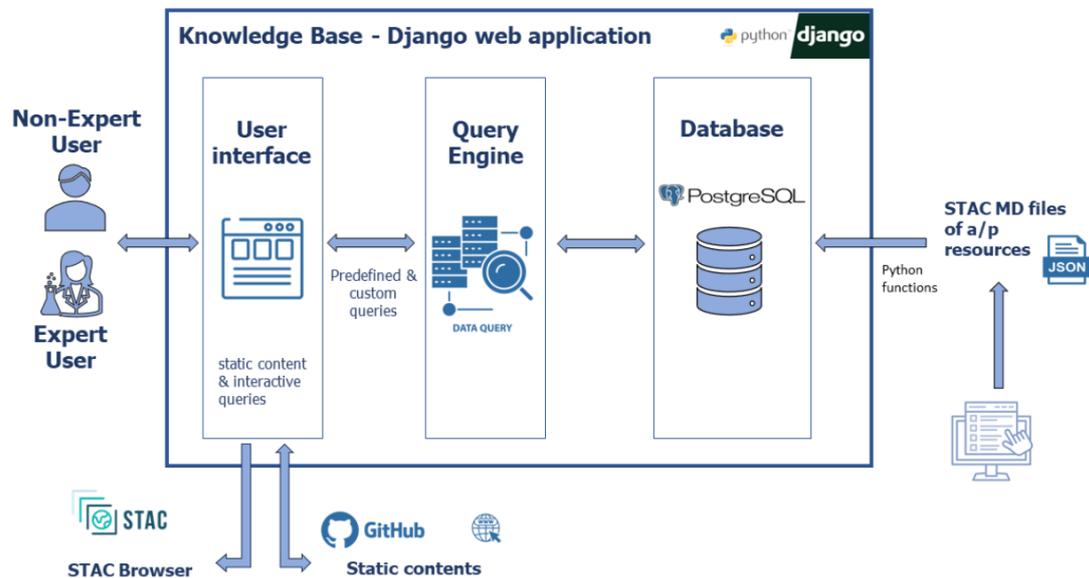


Figure 3 : KB components and structure

The KB relies on a web-application, providing:

- a user interface, through which the users can both access static content web pages and interact dynamically with information on the processing resources of the project (via Query Tool).
- a query engine
- links to multiple external sources: first and foremost, the project’s metadata catalog and GitHub repositories. As needed, however, links to thematic or technical web sites are provided in the static web pages.

The KB web-application is coded in Python and using the Django web-framework.



3 The Query Tool

The Query Tool has been specifically designed to facilitate the understanding, access and reuse of FAIRiCUBE processing resources. It is a service that allows users to interactively search for processing resources in different ways, according to their level of expertise and the desired level of granularity.

3.1 The Query Tool interface

The Query Tool interface, whose landing page is illustrated in Figure 4, enables filtering a/p resources based on:

- simplified queries, predefined or via keywords, for non-expert users
- custom queries created by expert users.

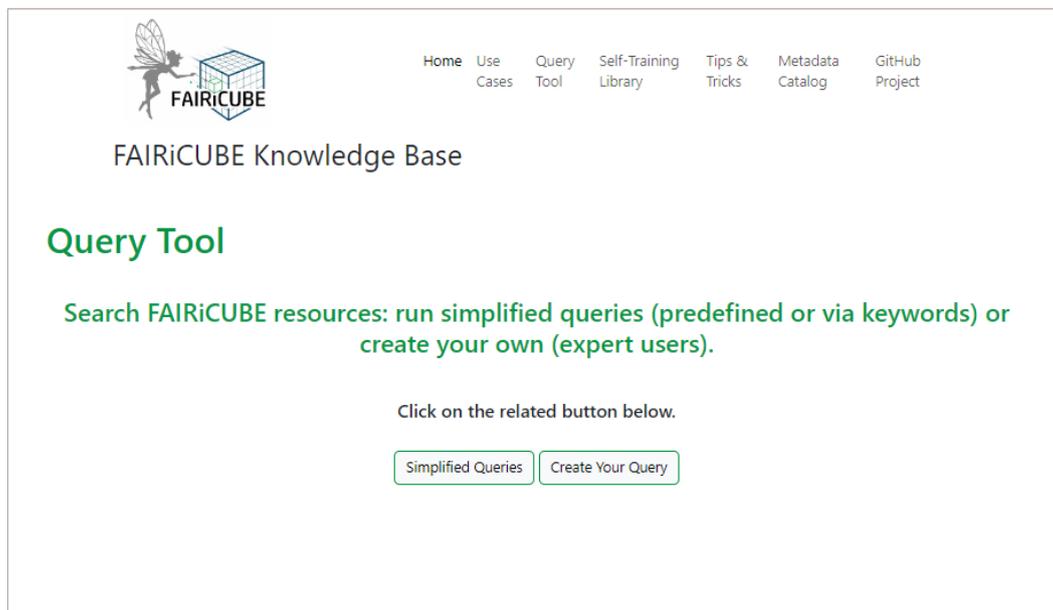
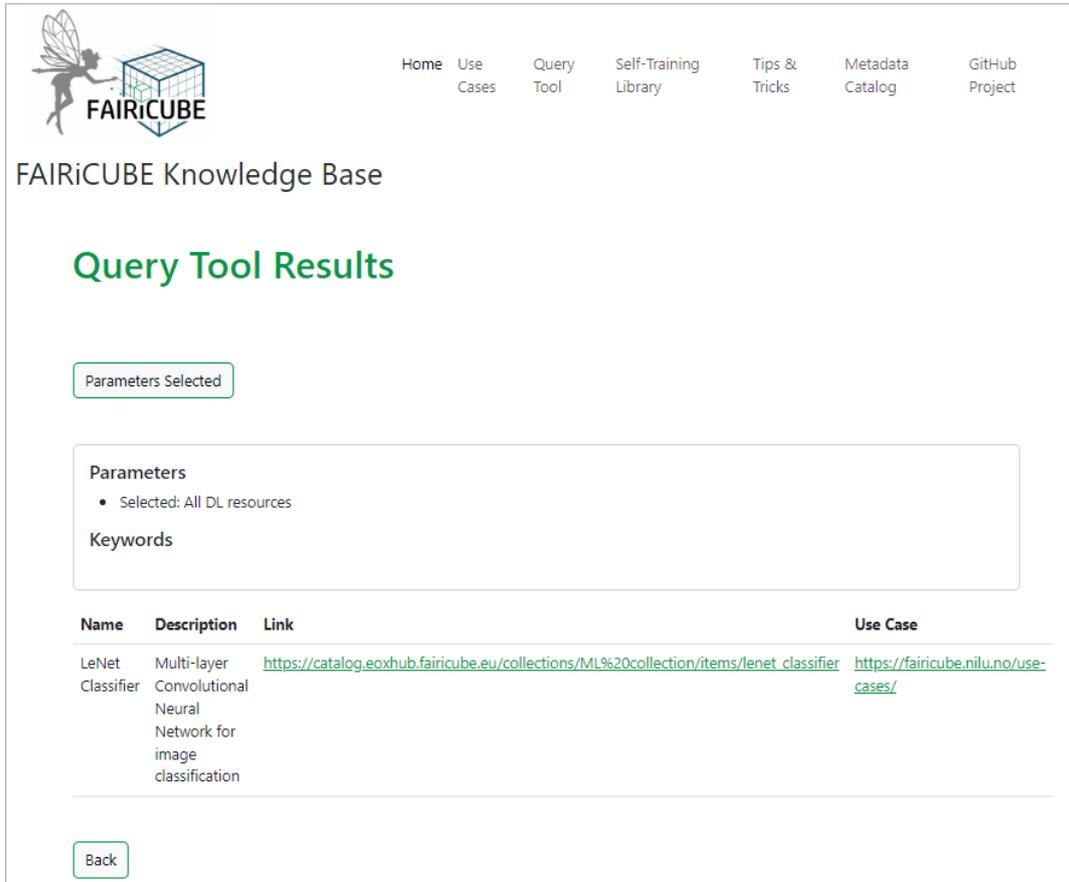


Figure 4 : Query Tool landing page

Query results

Query results (see Figure 5) consist of a list of a/p resources matching all the query criteria. For each result item, the name, a brief description, the link to the related metadata in the metadata catalog and the link to the related use case(s) page(s) of the Community Collaboration Platform will be displayed.



FAIRiCUBE Knowledge Base

Query Tool Results

Parameters Selected

Parameters

- Selected: All DL resources

Keywords

Name	Description	Link	Use Case
LeNet Classifier	Multi-layer Convolutional Neural Network for image classification	https://catalog.eoxhub.fairicube.eu/collections/ML%20collection/items/lenet_classifier	https://fairicube.nilu.no/use-cases/

Back

Figure 5 : Query result example

Simplified Queries

Through predefined queries made available as drop-down list, the Simplified Queries interface (shown in Figure 6) allows users to search by most meaningful parameters and parameters values. The users can also search for one or more keywords.



Home Use Cases Query Tool Self-Training Library Tips & Tricks Metadata Catalog GitHub Project

FAIRiCUBE Knowledge Base

Simplified Queries

Use this parameter to look for resources associated to a specific keyword.

Keyword

OR

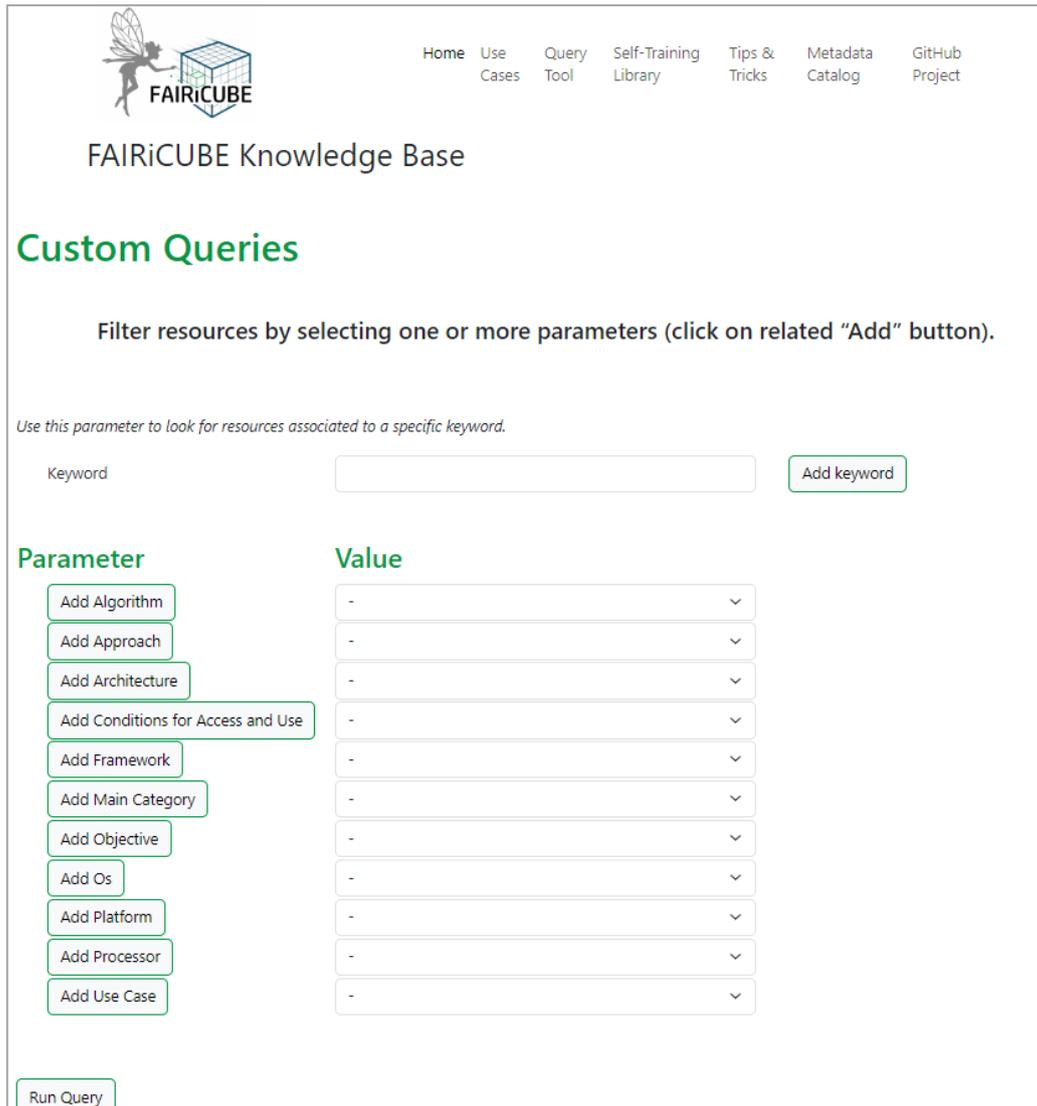
Select from the below list of all common UC resources queries to quickly get resources based on most utilized search criteria.

Query list:

Figure 6 : Simplified Queries

Custom Queries

Custom Queries (see Figure 7) allow the users to specify values for one or more query parameters selecting from drop-down lists. For each parameter, the drop-down list shows the set of values for it in the resource database.



The screenshot shows the FAIRiCUBE Knowledge Base interface. At the top, there is a navigation menu with links for Home, Use Cases, Query Tool, Self-Training Library, Tips & Tricks, Metadata Catalog, and GitHub Project. The main heading is "FAIRiCUBE Knowledge Base" followed by "Custom Queries" in green. Below this, a instruction reads: "Filter resources by selecting one or more parameters (click on related 'Add' button).". A sub-instruction says: "Use this parameter to look for resources associated to a specific keyword." Below this is a text input field labeled "Keyword" and a green "Add keyword" button. The main area is divided into two columns: "Parameter" and "Value". The "Parameter" column contains a list of green buttons: "Add Algorithm", "Add Approach", "Add Architecture", "Add Conditions for Access and Use", "Add Framework", "Add Main Category", "Add Objective", "Add Os", "Add Platform", "Add Processor", and "Add Use Case". The "Value" column contains a list of dropdown menus, each with a "-" sign and a downward arrow. At the bottom left, there is a green "Run Query" button.

Figure 7 : Custom Queries

3.2 The Query tool infrastructure

The Query Tool is powered by a PostgreSQL database. All SQL statements issued by the Query Tool are executed against this database, which contains information derived from the metadata of the processing resources. Each time a processing resource metadata is ingested into the metadata catalog (for details see D4.3), a web application automatically populates the database tables with the related information. Specifically, the database tables do not store all the information contained in the metadata records, but only those pieces of information required to filter resources according to the user's request.

The possibility of querying metadata resources directly via the STAC FASTAPI of the metadata catalog is still under evaluation. Depending on the results, it may be decided to discontinue this database-based approach and move to dynamic queries via APIs.



4 KB implementation and next steps

At the time of writing, a demo application of the [FAIRICUBE Knowledge Base](#) is available. The [FAIRICUBE Knowledge Base repository](#) has been created to discuss among partners and get feedback on the proposed architecture, structure and functionalities of the KB as well as related look & feel. Discussions are underway on possible further improvements in the organization of the KB content so that it streamlines:

- the UC descriptions on the website, as well as in the UC-specific repositories.
- the content available on the FAIRICUBE Community Collaboration Platform.

As use cases begin to use the KB more systematically, updates may be made to the structure and organisation of the KB contents described in this document. If so, these changes will be reflected in updated version(s) of this deliverable.