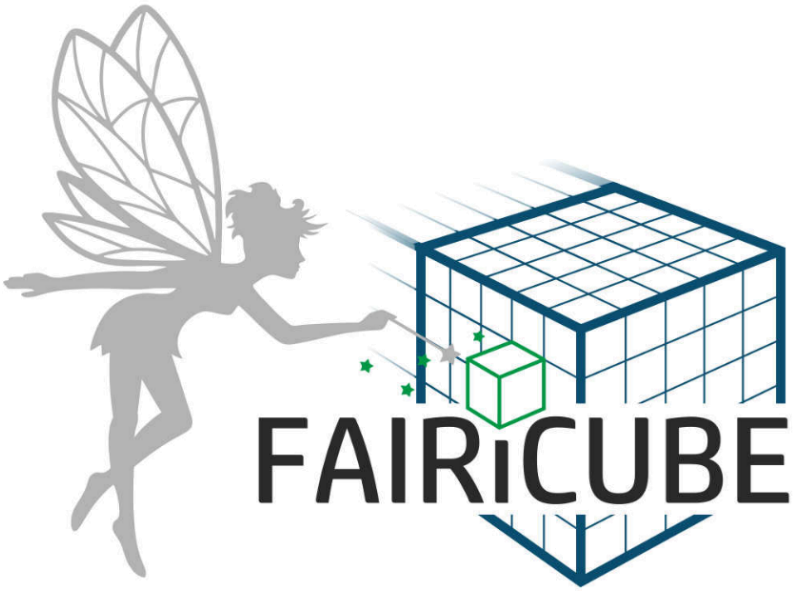

FAIRICUBE – F.A.I.R. INFORMATION CUBES

Work Package 2: Use

Deliverable 2.1: Report on UC data source synergies

Deliverable Lead: 4sfera
Deliverable due date: 30/06/2024

Version: 2.1
2024-07-01

Document Control Page

Document Control Page	
Title	Report on UC data source synergies
Creator	Cristina Carnerero and Jaume Targa
Description	Deliverable D2.1 report on UC data source synergies
Publisher	"FAIRICUBE – F.A.I.R. information cubes" Consortium
Contributors	
Date of delivery	30/06/2024
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
0.0	20/11/2022	Cristina Carnerero and Jaume Targa	Internal draft
0.1	13/12/2022	Cristina Carnerero and Jaume Targa	First draft
0.2	19/12/2022	Mirko Gregor	Review
1.0	19/12/2022	Cristina Carnerero and Jaume Targa	Final document
1.1	14/06/2024	Jaume Targa and Maria Colina	Updated draft
1.2	14/06/2024	Kathi Schleidt and Stefan Jetschny	Internal Review
2.0	14/06/2024	Jaume Targa	Minor edits and format check
2.1	21/06/2024	Stefan Jetschny	Final internal 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 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.

Table of Contents

List of Figures

List of Tables

1 Contextualisation of this deliverable

1.1 Overall objective of WP2

The overall objective of Work Package 2: Use (WP2) is to ensure efficient execution of the use cases, assuring those potential synergies pertaining to both data and processing are identified and leveraged.

1.2 Description of WP2 work

WP2 focuses on the execution and cross-coordination of the different use cases (UCs) on the FAIRiCUBE Hub. Acting with an "outsider" role, the supervision crosscuts through all UC activities ensuring harmonisation with both upstream (data sources, ingestion and processes) and downstream (results, promotion and distribution of outputs) activities.

1.3 Description of Task 2.2

Task 2.2 (Synchronization of data sources) focuses on the management of common data sources across UC and the collaboration with other WPs for all UCs. This is to ensure efficiency on data ingestion and harmonisation of data. This task identifies and utilizes synergies on data source across UCs.

2 Data source requirements

4SF coordinated the elaboration of an initial inventory of required data sources with all 4 Use Cases (now 5). A spreadsheet with the inventory as of 13.12.2022 was uploaded to FAIRiCUBE website (see also **Error! Reference source not found.**). The aim of the inventory was to gather an initial list of all t he different data sources that all UCs envisage to be using, thus identifying overlaps between the different use cases, enabling synergies in gaining access to these jointly used datasets. This initial deliverable was submitted at the end of M6. Since then, the procedure for data request has been updated (see later in this section).

The identification of data source synergies is essential for efficiently executing the use cases, not only to minimise data access but to cut down on processing/ingestion time. The data source inventory compiled includes the following information:

- Name
- Source
- Origin
- Responsible maintenance
- Coverage
- Spatial resolution
- Time resolution
- Time coverage
- Documentation
- Comments
- Priority for use cases
  - Climate Change urban adaptation
  - Biodiversity & agriculture
  - Biodiversity occurrence cubes
  - Neighbourhood building stock

No.	Name	Source	Origin	Responsible maintenance	Coverage	Spatial Resolution	Resolution	Time Coverage	Documentation	Comments	Use cases (1=high priority, UC1 (S4E)	UC2 (WEB)	UC3 (OBIS)	UC4 (DELI)
1	Open Street Map (OSM)	www.openstreetmap.org		OSM	World	vectorized data				OSM provides 2D-level data for buildings	2	2	1	1
2	INSPIRE Buildings geoportal	https://inspire-geoportal.ec.europa.eu/overview.html?view=themeOverview&theme=bu		EEA	EEA-38	vectorized data				INSPIRE can be used as an additional source. For instance, some EU nations provide limited covered 3D-level data for built environment	2			1
3	EPISCOPE (material intensity)	https://episcope.eu/fileadmin/tabula/public/calc/TABULA-xlsm_Download.zip		EPISCOPE	21 EU nations	txt				EPISCOPE material intensity contain masses and types of building materials for each building archtype				1
4	EPISCOPE (energy calculation)	https://episcope.eu/fileadmin/tabula/public/calc/tabula-calculator.xlsx		EPISCOPE	21 EU nations	.xlsx files			https://episcope.eu/fileadmin/tabula/public/doc/docs/report/TABULA_CommonCalculationMethod.pdf	EPISCOPE energy calculation presents simplified energy calculation for certain building archtypes				1
5	Corine Land Cover	https://land.copernicus.eu/pan-european/corine-land-cover	EO-based, produced by countries	EEA	EEA-38+UK	gridded 100*100m		1990 2000 2006 2012 2018	https://land.copernicus.eu/pan-european/corine-land-cover	The present 100m raster dataset is the 2018 CLC status layer modified for the purpose of consistent statistical analysis in the land cover change accounting system at EEA. CORINE Land Cover (CLC) data are produced from 1986 for European (EEA member or cooperating) countries. Altogether five mapping inventories were implemented in this period, producing five status layers (CLC1990, CLC2000, CLC2006, CLC2012, CLC2018) and four CLC-Change (CLCC) layers for the corresponding periods (1990-2000, 2000-2006, 2006-2012, 2012-2018). Pan-European CLC and CLCC data are available as vector and raster products.	1	2	1	
6	High Resolution Layer - Imperviousness	https://land.copernicus.eu/pan-european/high-resolution-layers/imperviousness/status-maps	EO-based	EEA	EEA-38+UK	10m 20m		2006 2009 2012 2015 2018	https://land.copernicus.eu/pan-european/high-resolution-layers/imperviousness/status-maps	The high resolution imperviousness products capture the percentage and change of soil sealing. Built-up areas are characterized by the substitution of the original (semi-) natural land cover or water surface with an artificial, often impervious cover. These artificial surfaces are usually maintained over long periods of time. A series of high resolution imperviousness datasets (for the 2006, 2009, 2012, 2015 and 2018 reference years) with all artificially sealed areas was produced using automatic derivation based on calibrated Normalized Difference Vegetation Index (NDVI). The series of imperviousness layers constitutes the main status layers. They are per-pixel estimates of impermeable cover of soil (soil sealing) and are mapped as the degree of imperviousness (0-100%). Imperviousness change layers were produced as a difference between the reference years (2006-2009, 2009-2012, 2012-2015, 2015-2018 and additionally 2006-2012, to fully match the CORINE Land Cover production cycle) and are presented 1) as degree of imperviousness change (-100% – +100%), in 20m and 100m pixel size, and 2) a classified (categorical)	1	2	2	

Figure 1: Snapshot of a sample of the inventory list as of 13/12/2022

As the Excel inventory sheet was just a intermediate and improvised solution, a similar overview of the open ingestion items can now be gained by an online listing of the GitHub pull requests[1] using the WebGUI data request procedure as described in the following.

[1] https://github.com/FAIRiCUBE/data-requests/tree/main/stac_dist

2.1 Data request (via WebGUI)

The data request process is uniform for both the rasdaman and EOX pillars. When a Use Case requests an additional dataset, a new record is created on the FAIRiCUBE data requests GitHub repository through a new issue. This process has been streamlined with a Web GUI available at https://data-request.fairicube.eu , where requesters can provide necessary information via a form. This form generates a machine-readable request that initiates the ingestion process.

The Web GUI functions as a frontend to a GitHub repository. When the form is submitted, it creates a GitHub Pull Request as a new branch, with the user automatically following the branch for updates based on their GitHub notification settings. Progress, issues, and discussions are documented in a GitHub issue associated with the Pull Request, allowing all interested parties to monitor and contribute feedback.

For organising and searching within the FAIRiCUBE HUB, a clear set of required metadata concepts has been established (for details, see D4.2), which includes details such as mapping campaign, reference year, inventory year, and version number. These concepts will be added to the expert search functionality of the catalog. For each new data set, a corresponding entry in the catalog at https://catalog.fairicube.eu is created with all necessary descriptive metadata. In this manner, all datasets are exposed and visible.

The FAIRiCUBE catalog includes data requested and ingested in the different platforms. The different data sets can be accessed here:

- https://catalog.eoxhub.fairicube.eu/ (all available collections)
- https://fairicube.rasdaman.com/rasdaman/ows#/services

3 Data source synergies

The initial aim of compiling an inventory of data sources from all UCs was to identify initial data source synergies across these UCs. All 5 UCs initially identified a total of 54 data sources. Once the initial draft of data sources inventory was compiled, the team worked with all UCs to identify those data sources which could be of use, despite not being initially listed as a required data source. The identification of further data source synergies is a clear example of knowledge sharing across UCs.

The exercise detected 25 data sources which were identified as potentially useful sources with high or low priority further use by at least 2 UCs. A summary of these sources is presented in Table 1. The synergetic data sources are related to geographical parameters, land cover, data on buildings, biological parameters and climate variables. It is worth noting that all of the 25 synergetic sources have been identified as high priority by at least one UC. Additionally, 3 data sources (Open Street map, Temperature data and Digital Elevation Model) have been identified as further use by all 5 UCs.

Data sources are classified as high priority (1), low priority (2) and a blank meaning that the UC does not plan to use the source. This has assisted in identifying the common priorities across UCs.

3.1 High priority data sources

For UC1, the datasets marked as high priority are predominantly derived from EO-based sources and maintained by organisations such as the EEA or Copernicus. Some notable datasets in this category include Corine Land Cover, Imperviousness, various forest-related datasets, the Urban Atlas, Temperature, Climate extremes indices, and the Copernicus DEM. These datasets are characterised by detailed spatial and temporal resolution. These provide global or regional coverage, making them essential for comprehensive environmental monitoring and urban analysis.

In UC2, datasets often reflect broader environmental and urban indicators. Important datasets for this use case include INSPIRE Buildings, Corine Land Cover, Imperviousness, Urban Atlas, Temperature, and Copernicus DEM. These datasets offer vector or grid data at varying resolutions. In addition, UC2 utilises the Occurrence Cube Paradigm.

For UC3, the use case shows limited usage, due to issues encountered in accessing and utilizing data on the rasdaman platform. Now that this UC has been shifted to EOX infrastructure, significant progress has been made, working on datasets including INSPIRE Buildings, Corine Land Cover, Imperviousness, and various forest-related datasets. These datasets encompass both EO-based and in-situ observations with specific spatial resolutions.

UC4 exhibits a mix of high, low, and non-usage across various datasets. Key datasets for this use case include Open Street Map, INSPIRE Buildings, Corine Land Cover, Urban Atlas, and Climate extremes indices. The datasets feature vector data with varying temporal coverage.

In UC5, priorities are assigned to datasets focusing on environmental and biodiversity aspects. Crucial datasets for this use case include Corine Land Cover, EUNIS Habitats, Imperviousness, various forest-related datasets, and the Copernicus DEM. These datasets cover diverse spatial resolutions and temporal extents. In addition, UC5 utilizes the Occurrence Cube Paradigm.

Corine Land Cover, Imperviousness, and forest-related datasets are high priority across multiple use cases (UC1, UC3, UC5). This indicates their broad applicability in environmental and urban analysis. Additionally, most datasets provide European or global coverage, which is crucial for comprehensive regional analysis. A large number of these datasets are derived from Earth Observation (EO) sources. The datasets also offer a variety of spatial resolutions (e.g., 10m, 20m, vector), providing flexibility for different analytical needs across the use cases. Moreover, datasets with extensive temporal coverage (e.g., spanning multiple years) are prioritized for trend analysis and monitoring changes over time.

Both UC2 and UC5 are working on the Occurrence Cube concept, a new paradigm arising from the biodiversity community. This approach is also being utilized within two of the FAIRiCUBE Sibling projects B-Cubed and AD4GD. Collaboration on this exciting new approach to biodiversity data across Sibling projects is underway.

In summary, high priority in UC1 is consistently assigned to datasets that are essential for broad environmental monitoring and urban analysis. However, UC2 to UC5 show varied priorities, reflecting specific application needs such as climate impact analysis (UC4) or biodiversity studies (UC5). While major entities like EEA and Copernicus maintain many of the datasets, entities such as Eurostat and the GBIF Secretariat also contribute crucial datasets.

Table 1: Summary of synergetic data sources found. Columns UC1, UC2, UC3, UC4 and UC5 represent the priority of use of each data source, with 1 meaning high priority, 2 meaning low priority and a blank meaning that the UC does not plan to use the source.

Name	Source	Origin	Resp. maint.	Coverage	Spatial res.	Time cov.	UC1	UC2	UC3	UC4	UC5
Open Street Map	www.openstreetmap.org		OSM	World	vector		2	2	1	1	2
INSPIRE Buildings	https://inspire-geoportal.ec.europa.eu/overview.html?view=themeOverview&theme=bu		EEA	EEA-38	vector		2			1	
Corine Land Cover	https://land.copernicus.eu/pan-european/corine-land-cover	EO-based, produced by countries	EEA	EEA-38+UK	100m	1990 2000 2006 2012 2018	1	2	1		2
Imperviousness	https://land.copernicus.eu/pan-european/high-resolution-layers/imperviousness/status-maps	EO-based	EEA	EEA-38+UK	10m 20m	2006 2009 2012 2015 2018	1	2	2		2
Forest type	https://land.copernicus.eu/pan-european/high-resolution-layers/forests/forest-type-1/status-maps	EO-based	EEA	EEA-38+UK	10m 20m	2012 2015 2018	1	2	2		2
Forest - tree cover	https://land.copernicus.eu/pan-european/high-resolution-layers/forests/tree-cover-density/status-maps	EO-based	EEA	EEA-38+UK	10m 20m	2012 2015 2018	1	2	2		2
Forest - Dominant Leaf Type	https://land.copernicus.eu/pan-european/high-resolution-layers/forests/dominant-leaf-type/status-maps	EO-based	EEA	EEA-38+UK	10m 20m	2012 2015 2018	1	2	2		2
Grassland	https://land.copernicus.eu/pan-european/high-resolution-layers/grassland	EO-based	EEA	EEA-38+UK	10m 20m	2015 2018	1	2	2		2
Water & Wetness	https://land.copernicus.eu/pan-european/high-resolution-layers/water-wetness	EO-based	EEA	EEA-38+UK	10m 20m	2015 2018	1	2	2		2
Small Woody Features	https://land.copernicus.eu/pan-european/high-resolution-layers/small-woody-features	EO-based	EEA	EEA-38+UK	10m 20m	2015 (2018)	1	2	2		2

Name	Source	Origin	Resp. maint.	Coverage	Spatial res.	Time cov.	UC1	UC2	UC3	UC4	UC5
European Settlement Map	https://land.copernicus.eu/pan-european/GHSL/european-settlement-map	EO-based	JRC	EEA-38+UK	2,5m 10m 100m	2012 2015	1		1		
CLMS Urban Atlas	https://land.copernicus.eu/local/urban-atlas	EO-based	EEA	EEA-38+UK	vector	2006 2012 2018	1	2		2	2
CLMS Urban Atlas Street Tree Layer	https://land.copernicus.eu/local/urban-atlas	EO-based	EEA	EEA-38+UK	vector	2012 2018	1	2			2
Population by Urban Atlas polygon	https://land.copernicus.eu/local/urban-atlas/population-estimates-by-urban-atlas-polygon	EO-based	EEA	EEA-38+UK			1			2	
Natura 2000 landcover/land use	https://land.copernicus.eu/local/natura	EO-based	EEA	EEA-38+UK	vector	2006 2012 2018	1	2			2
NUTS regions	https://ec.europa.eu/eurostat/web/nuts/background	national reporting	Eurostat	EEA-38+UK, -BA, XK	vector		1	2	2		2
Urban Audit city delineations (FUA, city, commuting zone)	https://ec.europa.eu/eurostat/web/cities/background		Eurostat	EU-27+EFTA	vector		1			2	
Temperature	https://cds.climate.copernicus.eu/cdsapp#!/dataset/derived-near-surface-meteorological-variables	modelled	C3S	global	grid		1	2	1	2	2
Physiologically Equivalent Temperature	https://climate.copernicus.eu/thermal-assessment-tool		C3S	global	grid		1			2	
Climate extremes indices and heat stress indicators	https://cds.climate.copernicus.eu/cdsapp#!/dataset/sis-extreme-indices-cmip6?tab=overview		C3S		grid		1	2	1		2

Name	Source	Origin	Resp. maint.	Coverage	Spatial res.	Time cov.	UC1	UC2	UC3	UC4	UC5
Copernicus DEM	https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1	EO-based	Copernicus	global	grid 10m 30m 90m		2	2	1	2	2
Global Biodiversity Information Facility	https://www.gbif.org/	in-situ observations	GBIF Secretariat	World	various	2014-2022		1	1		1
HR VPP (NDVI, PPI, FAPAR, LAI)	https://land.copernicus.eu/pan-european/biophysical-parameters/high-resolution-vegetation-phenology-and-productivity	Sentinel 1, 2	CLMS	Europe	10 m	2014-2022	2	2			2
Soil grids	https://soilgrids.org/		ISRIC	World	250m	2014-2022	2	1			1
3D building model (LOD2)	https://geoe3platform.eu/geoe3	GeoE3	GeoE3	NO, ES, NL, FI, EE	vector		2			1	

4 Data source management and ingestion

FAIRiCUBE catalogue includes around 100 data sources with diverse temporal and spatial resolution providing data related to meteorology, oceans and other water bodies, land use and land cover, geography, demographics, vegetation and air quality. Some of these datasets have been identified by UCs to be useful for their future analysis. https://catalog.eoxhub.fairicube.eu/?.language=en

In case new data must be added, a Data Request WebGUI has been created. More detailed information can be found in deliveries D4.1 and D5.2.

4.1 New data sources

Due to the nature of the UC implementation, the team is aware that the list of data sources will increase as each UC works on their ML implementation under WP3.

In addition, it has to be considered that the UCs have very specific characteristics that require the identification and incorporation of additional data sources over time. For example, the two UCs focusing on urban areas will collocate their analysis in a set of specific cities (Oslo, Vienna and Barcelona) . Once the specific locations se have been determined, the UCs might get additional local data that will need to be incorporated to the repository of existing data sources. Following the initial data inventory, new data requests need to follow the procedure specified here: https://catalog-editor.eoxhub.fairicube.eu/

This section will be compiled as part of the final delivery in month 24. This aims to identify new data sources that each team identifies during the project development.

4.2 Other data sources

Rasdaman and EOxHub services already provide access to several data sources. The data source inventory (see spreadsheet uploaded in FAIRiCUBE website) includes a column specifying whether a data source has already been included or not in the FAIRiCUBE catalog. The data already ingested in each platform and readily available. Particular catalogs of already available data can be found in:

- https://catalog.eoxhub.fairicube.eu (dynamic, using fastAPI)
- https://catalog.fairicube.eu (static)
- https://collections.eurodatacube.com/
- https://opensciencedata.esa.int/
- https://stacindex.org/catalogs

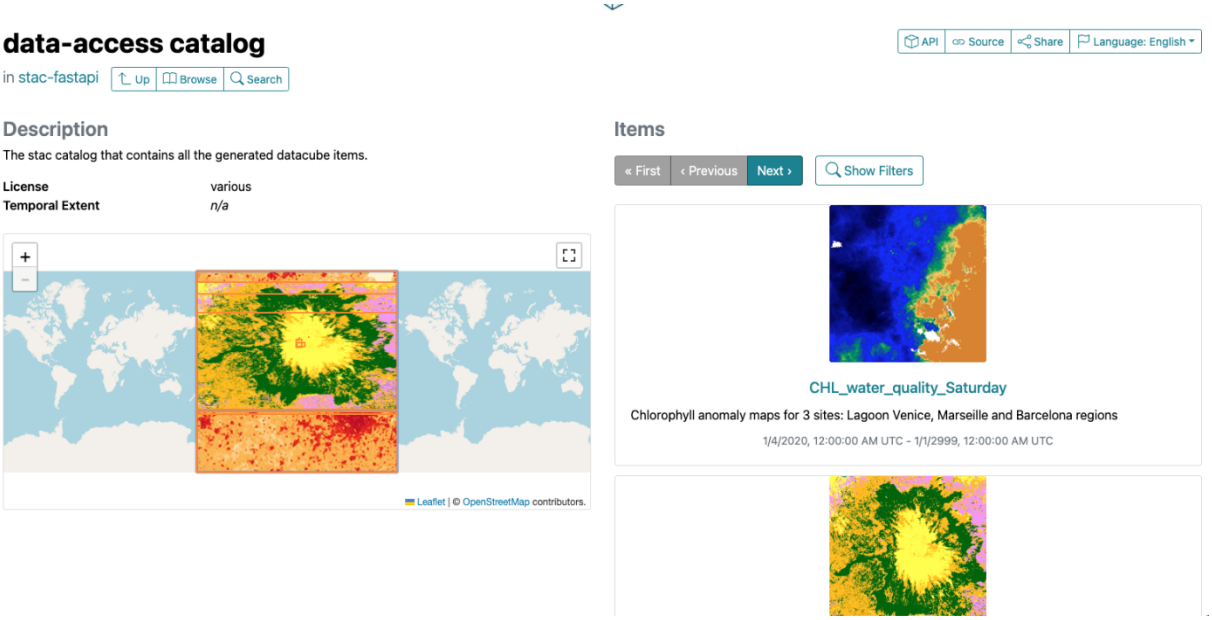

Figure 2: FAIRiCUBE catalogue - https://catalog.eoxhub.faricube.eu/

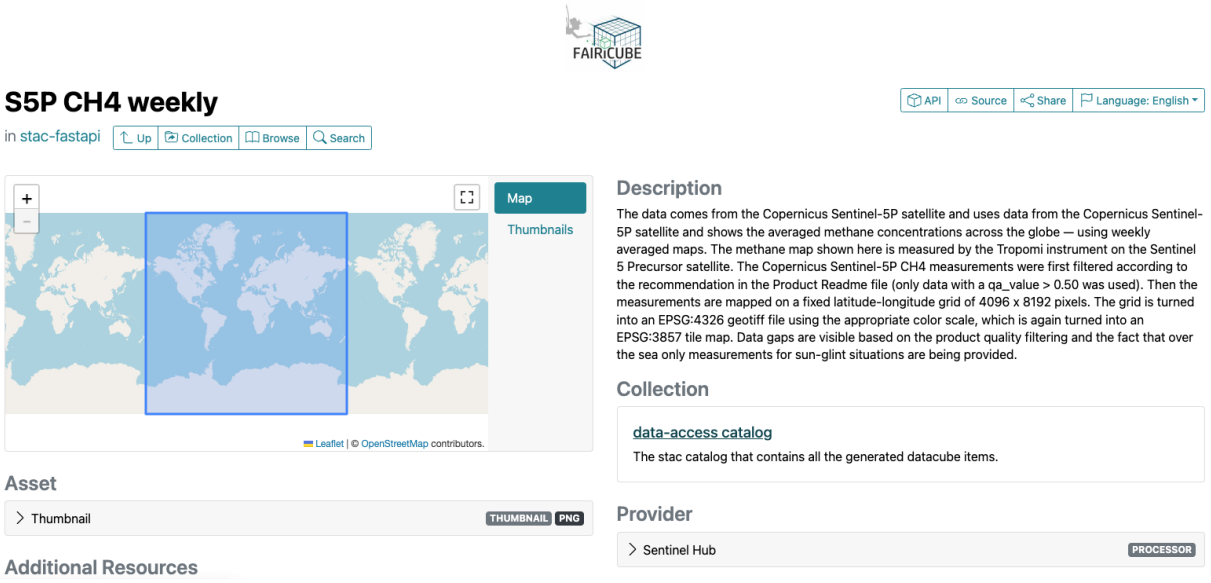

Figure 3: Example of dataset in FAIRiCUBE catalogue