Ref. Ares(2022)2610717 - 06/04/2022

D1.9 Agricultural Research Data Index Tool (ARDIT)

agricore

Deliverable Number	D1.9
Lead Beneficiary	AAT
Authors	AAT, IDE, UNIPR
Work package	WP1
Delivery Date	M31
Dissemination Level	Public

www.agricore-project.eu

The Agricore project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No. 816078

Document Information

Project title	Agent-based support tool for the development of agriculture policies
Project acronym	AGRICORE
Project call	H2020-RUR-04-2018-2019
Grant number	816078
Project duration	1.09.2019-31.8.2023 (48 months)
Deliverable Authors	Juan Carlos Castillo (AAT), Lisa Baldi (UNIPR), Matteo Rasponi (UNIPR), Pablo Báez (IDE)
Deliverable Reviewers	Pablo Gallegos (IDE), Carlos Leiva (IDE)

Version History

Version	Description	Organisation	Date
0.1	ToC definition	AAT	03 Feb 2022
0.2	Development of sections	AAT	14 Feb 2022
0.3	Review of the established requirements	AAT, IDE	16 Feb 2022
0.4	First development review	AAT, IDE, UNIPR	16 Mar 2022
0.5	Second development review	AAT, IDE	22 Mar 2022
0.6	Final review	IDE	27 Mar 2022
0.7	Changes implemented	AAT, IDE	04 Apr 2022
1.0	Deliverable exported, formatted and completed	IDE	05 Apr 2022

Disclaimer

All the contributors to this deliverable declare that they:

- Are aware that plagiarism and/or literal utilisation (copy) of materials and texts from other Projects, works and deliverables must be avoided and may be subject to disciplinary actions against the related partners and/or the Project consortium by the EU.
- Confirm that all their individual contributions to this deliverable are genuine and their own work or the work of their teams working in the Project, except where is explicitly indicated otherwise.
- Have followed the required conventions in referencing the thoughts, ideas and texts made outside the Project.

Executive Summary

The Agricultural Research Data Index Tool (ARDIT) is a platform created in the framework of the AGRICORE Project to index characterisations of datasets that can be used for the analysis and research of the agri-food chain.

These characterisations of datasets (or dataset catalogues) can be incorporated by registered users through web-based forms built on the basis of the AGRICORE-DCAT 2.0 ontology, which allows characterisation down to the level of the variables contained in each dataset.

The aim of this document is to explain what ARDIT is, how it has been constructed, how it works and its importance within the AGRICORE project. For this purpose, a detailed description of its architecture is provided together with annexes that include user manuals for operating the platform and characterising agricultural datasets.

This document is organised as follows:

The introduction section presents how ARDIT fits into the AGRICORE suite and its data flow in combination with the data warehouse. The second section outlines the requirements (functional and non-functional) defined for the ARDIT tool. The third section presents the architecture of the tool, distinguishing the front-end from the back-end. It also introduces the format of the different requests including the built-in security configuration. The fourth section shows the development process of the tool, including its links with the development of the underlying AGRICORE-DCAT 2.0 ontology, as well as decisions related to the graphical interface and workflow. The fifth section presents the plan for the future governance of the tool, comprising both the IT maintenance part and the editing and enhancement of existing characterisations and those that may be added in the future. The sixth section summarises the conclusions of the document.

The annexes include:

ANNEX 1: The ARDIT Tool User Manual.

ANNEX 2: Brief overview of the libraries, frameworks and programming languages involved in the development of ARDIT.

ANNEX 3: Guide to the ARDIT data model.

Abbreviations

Abbreviation	Full name
ABM	Agent Based Model
ABM-e	Agent-based Model simulation Engine
API	Application Programming Interface
ARDIT	Agricultural Research Data Index Tool
CSS	Cascading Style Sheets
CSV	Comma-Separated Values
DB	Database
DEM	Big Data Extraction Module
DFM	Big Data Fusion Module
DWH	Data Warehouse
ETL	Extract-Load-Transform
EU	European Union
GUI	Graphical User Interface
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
IT	Information technologies
JSON	JavaScript Object Notation
JWT	JSON Web Token
LDAP	Lightweight Directory Access Protocol
PDF	Portable Document Format
REST	Representational State Transfer
RSS	Really Simple Syndication
SPG	Synthetic Population Generator
SQL	Structured Query Language
URL	Uniform Resource Locator
XML	Extensible Markup Language

List of Figures

List of Tables

Table of Contents

1 Introduction

Nowadays, the use of technology and the digital transformation that the world is experiencing allows institutions, authorities, governments and companies to create spaces that provide universal access to the large amount of data that is generated, as a means of transparency, redistribution of information, accessibility and availability and, in short, to provide society with resources to facilitate research, innovation and analysis through decision-making. These spaces are known as data portals and can be defined as:

"*Data portals are web-based interfaces designed to make it easier to find reusable information. Like library catalogues, they contain metadata records of datasets published for reuse, mostly relating to information in the form of raw, numerical data rather than textual documents. In combination with specific search functionalities, they facilitate finding datasets of interest.*"[1]

AGRICORE is a Research and Innovation Action funded by the European Commission under the call RUR-04-2018 whose objective is to design and build an Agent-based-modelling tool to carry out impact assessments of different measures and instruments alternatively implementable within the Common Agricultural Policy (CAP). These impact assessments fall into three main categories (socio-economic, environmental and delivery of ecosystem services) and will serve both to assess a posteriori the effects of measures already implemented (ex-post analysis) and to predict a priori the impact of measures in the design phase (ex-ante analysis). Depending on the policy instrument whose impact is to be assessed, it is necessary to detect and access data that enable the assignation of values to the set of attributes of interest that make up each of the agents. This information must be statistically representative of the set of real farm holdings to be synthesised, precisely those that are affected by the policy measure under study in terms of their typology (technical-economic orientation), size (economic dimension) and location (geographical and administrative scope). For this purpose, the AGRICORE project provides the Agricultural Research Data Index Tool (ARDIT), previously referred to European data sources index module, as a web application with two well-defined objectives for the project. The first one is to deliver an open data portal where researchers can publicly access an index of available agricultural data sources. The platform serves the metadata of the datasets and the links to their origin or download URLs; this means that ARDIT does not store the indexed datasets. The second objective is to use this tool within the project to provide data necessary to fill the attributes of the farm agents that compose the synthetic populations used in the ABM.

AGRICORE deliverable D1.1, Standardised Methodology and Set of Ontologies for the Characterisation of Data Sources describes the methodology carried out to identify the properties that allow the universal identification of all agricultural datasets. This process is known as characterisation and not only helps in the task of identifying and adding datasets to the indexer, but also, supports the supply of data to the project's simulation tools. This deliverable defines the following:

"*Developing the AGRICORE ABM requires an in-depth knowledge of the datasets providing the information necessary to model the primary and the many domains the farmer has to consider in his decision-making behaviour...To gather this required knowledge and to store it correctly, the AGRICORE project planned for the design and implementation of a characterisation methodology. This approach would allow analysing each data set systematically and providing comparable results across them... Moreover, to facilitate the generation of this characterisation and the dissemination and reuse of results (making the defined characterisations explicitly available to other researchers), the AGRICORE project foresaw the development of an index tool. This tool initially named in the project as the "EU Index Tool", is currently being designed as an Agricultural Research Data Index Tool (ARDIT) and will provide easy access to interested researchers and policymakers to its database. The platform will assist the user in the task of identifying proper data sources that could be used to perform different types of analyses in the domain of agriculture (and related ones).*

The characterisation ... covers characteristics such as spatial scope and resolution, aggregation level, update frequency, last update available, privacy level of the data, and accessibility, among others. Additionally, it will also inform users on how to retrieve the data, including the potential need for finalising an agreement with the corresponding data owner(s) or provider(s). Finally, technical details for retrieving this data from the original data sources will be also included, potentially enabling the automatic extraction of the relevant datasets and/or variables for preparing the actual dataset employed for quantitative analysis through the inclusion of pre-defined Extraction Transformation and Loading scripts (ETLs)"

The main purpose of this document is to present the design and development of the ARDIT, as well as its functions and objectives within the AGRICORE project architecture. Such as its connection with the DWH or the implementation of the characterisation methodology defined for agricultural datasets. Firstly, the document sets out the list of requirements defined for the tool, followed by some chapters where the architecture of the platform and the different phases that the development has undergone are reviewed. Then, the document includes the results and conclusions obtained, the next steps to be taken and finally, the user manuals of the platform and the descriptions of each tool, technology and library used in ARDIT.

Figure 1 shows the different modules that compose the AGRICORE architecture, where ARDIT can be observed as a component necessary to prepare the data required to run the simulations (D.1). For a detailed description of the remaining modules, the interested reader is referred to deliverable 'D6.1 - AGRICORE architecture and interfaces'.

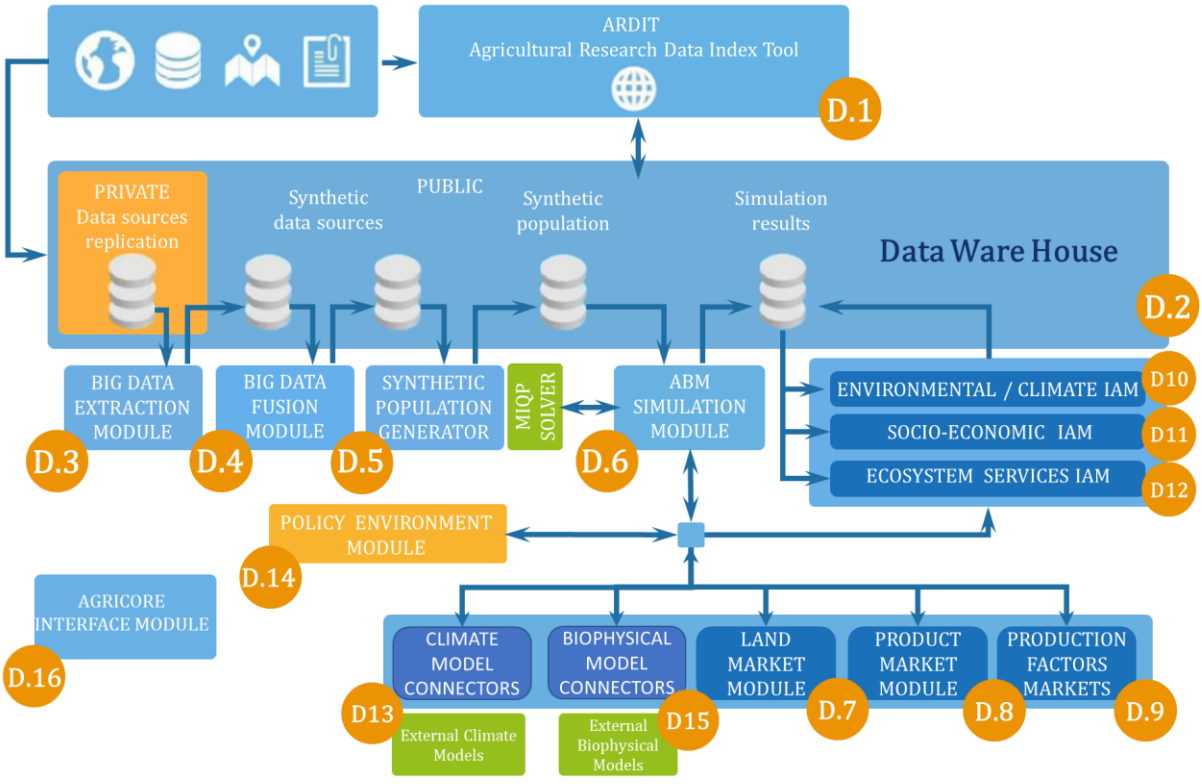


Figure 1 AGRICORE Modular Architecture

The use of the ARDIT tool in the AGRICORE project is especially linked to the Data Warehouse (D2), which is the tool responsible for the ingestion, storage, management and processing of data within the AGRICORE project. The combination of both architectures defines the complete process of data preparation and gathering. Figure 2 shows the joint architecture between them, where two different versions of ARDIT can be identified. The **Global Indexer** addresses the need

to create a fully public and accessible portal of datasets related to agriculture, where users will be able to perform searches and filter datasets based on multiple attributes and properties of them. In addition, to be able to link to their origins in order to download them in different formats.

On the other hand, given the open source nature of the AGRICORE project and the fact that the complete suite of tools can be downloaded and installed by interested users and academic or research centres, the **Local Indexer** provides the tools necessary for storing new datasets in the DWH. The catalogue of datasets available in the Local Indexer will be managed completely by the user, who will be able to update it, synchronising its database with the Global Indexer and obtaining any new registered or added dataset in the Global version. More detailed descriptions and functions of each component or tool that composes the joint architecture between ARDIT and the DWH can be found in the deliverable D2.1 Data Warehouse (DWH) for agriculture policy impact assessment data management.

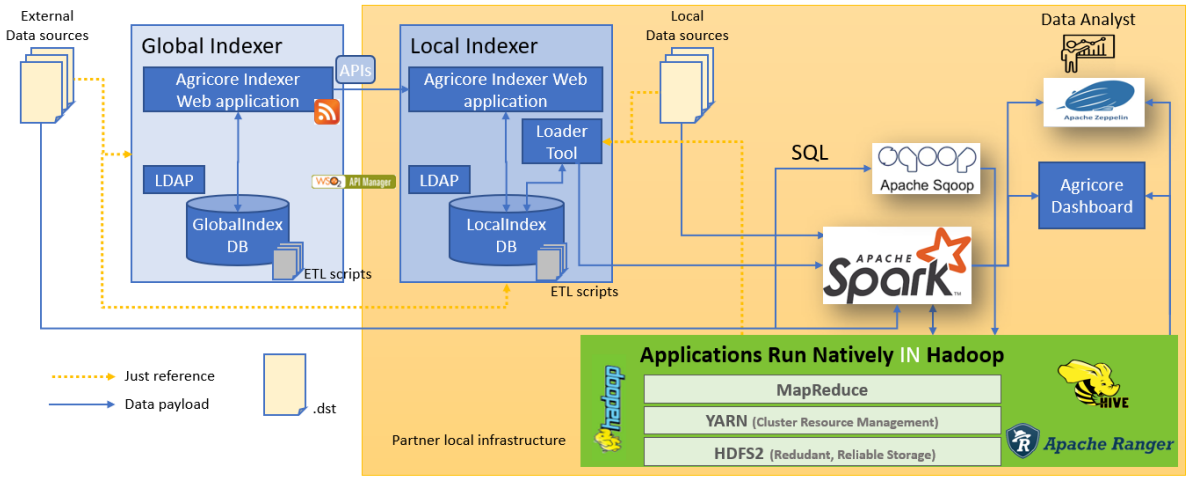


Figure 2 Complete architecture of ARDIT and the DWH

As described above, the purpose of storing the datasets is required to feed the simulations with useful data. Such behaviour is done by the Synthetic Population Generator (SPG). The component responsible for creating the population of agents for the ABM, representing these agents, the crop farms selected for the simulation. Each individual agent possesses a number of attributes of interest that need to be populated with values extracted from the datasets. This extraction and assignment process involves ARDIT, the DWH, the SPG, the **Big Data Extraction** (DEM) and **Big Data Fusion** (DFM) modules. When setting a synthetic population in the AGRICORE interface module, in order to run a specific simulation, the user will have to select a dataset for each of the attributes that compose the agents. The process of finding the most appropriate datasets for each attribute, storing them in the DWH and extracting the necessary data, is performed by the DEM. Then, the DFM search for correlations between the attributes of the agents and the variables, the data extracted from the datasets, in order to obtain probability distributions which are later used to generate random values to populate the attributes of the agents. The following Figure 3 represents the data flow between the different modules described above during the synthetic population generation process:

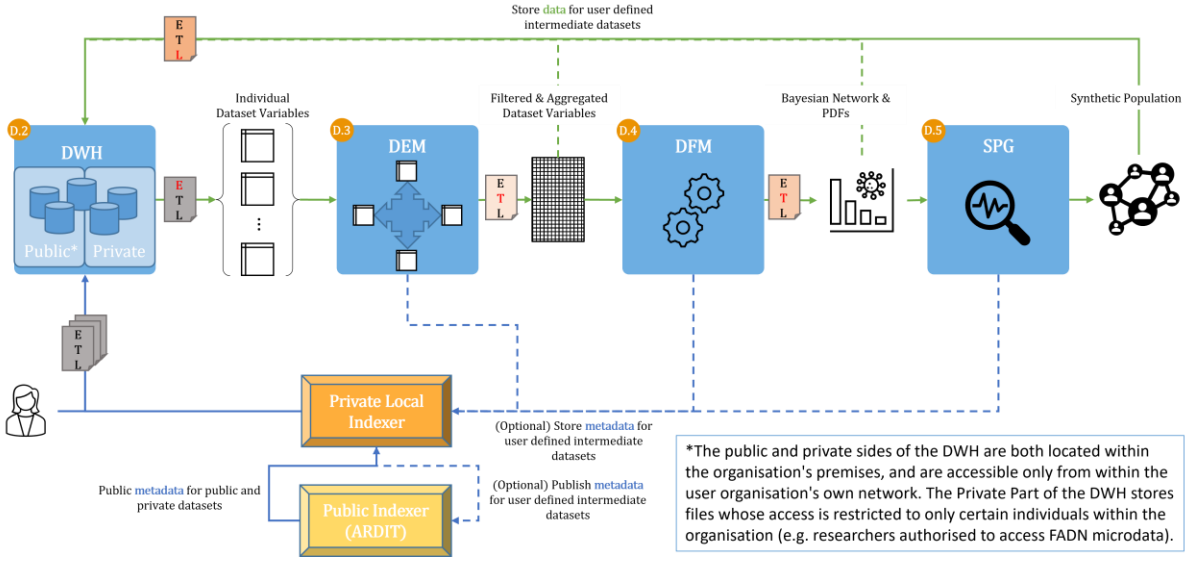


Figure 3 Data flow among ARDIT-DWH-DEM-DFM-SPG during Synthetic Population Generation procedure

2 Requirements of the ARDIT tool

2.1 Functional requirements

Functional requirements (FR) describe features that the software/system must offer or comply to meet the objectives of a project, defining its internal behaviour [2]. The functional requirements for the ARDIT tool are the following:

AG.D1.FR.001. Provide a publicly accessible index of agriculture data sources: ARDIT must provide a publicly accessible index of data sources available for agriculture policy assessment.

- **AG.D1.FR.001-1. User registration service**: A user can register into the ARDIT tool.
  - **AG.D1.FR.001-1-1. Register a new user**: A new user can be registered himself. The user should have a user role where some operations are restricted.
  - **AG.D1.FR.001-1-2. User verification**: A new user must be verified by email.
- **AG.D1.FR.001-2. Login service**: A user can login into the system using its credentials.
  - **AG.D1.FR.001-2-1. Login**: A user can login into the system using its credentials.
  - **AG.D1.FR.001-2-2. Account recovering**: A user can recover its account due to forgotten credentials. An email will be sent to the user to modify its password.
- **AG.D1.FR.001-3. Administration service**: An administrator has a menu to perform administrative operations.
  - **AG.D1.FR.001-3-1. Administrator creates a user**: An administrator can create a new user.
  - **AG.D1.FR.001-3-2. Administrator modifies a user**: An administrator can modify a user.
    - **AG.D1.FR.001-3-2-1. Modify the basic user information**: An administrator can modify the user's basic information such as username and email.
    - **AG.D1.FR.001-3-2-2. Assign a predefined role to a user**: An administrator can assign a predefined role to a user (User, Editor, Maintainer or Admin).
    - **AG.D1.FR.001-3-3. Delete a user**: An administrator can delete a user from the system.

AG.D1.FR.002. Available for all stakeholders: The ARDIT platform must be publicly available for all stakeholders (from data analysts to policy makers and researchers).

- **AG.D1.FR.002-1. Anyone can list data sources**: Any ARDIT user and guest (anonymous user) can list all the data sources.
- **AG.D1.FR.002-2. Anyone can download ETL**: Any ARDIT user and guest user (anonymous) can download an ETL associated with a data source.
- **AG.D1.FR.002-3. Anyone can access to global legal notice section**: Any user can access to the Global legal notice section:
  - Disclaimer.
  - Copyright notice.
  - Privacy statement (European Union Open Data Portal).

- Cookie policy.

- **AG.D1.FR.002-4. Anyone can access to a contact form**: A user could contact the organization with a contact form. The form is sent as an email to the ARDIT organization.
- **AG.D1.FR.002-5. Anyone can access to other websites section**: Any user can access to other related websites such as:
  - EU publications.
  - Web pages related to AGRICORE.
- **AG.D1.FR.002-6. Anyone can access to a help section:** Any user can access to a help section that inform about how to use the application. This section could have other services such as:
  - How it works.
  - Suggestions.
  - Error reports. They must be sent to an external ticketing service.

AG.D1.FR.003. Store relevant information of the data sources: ARDIT must allow the registration and modification of data sources according to the ontology (AGRICORE DCAT-AP 2.0 extension), as well as to be easily updated to adapt to the evolution of the ontology. Relevant data from the data sources such as fields, spatial scope, resolution, aggregation level, update frequency, last update available, privacy level of the data or accessibility.

- **AG.D1.FR.003-1. Register a new data source**: Any user with an account, can register a new data source in the application. `[Admin | Maintainer| Editor]` can fully register and publish data sources on the platform. `[User]` may propose new characterisations (drafts), which must be reviewed by editors or maintainers before full publication.
  - **AG.D1.FR.003-1-1. Add information in a data source**: A user `[Admin | Maintainer | Editor | User]` can add all the required information to register a new data source, based on the ontologies defined in AGRICORE DCAT-AP 2.0.
  - **AG.D1.FR.003-1-2. Add and remove a maintainer of a data source**: A user `[Admin | Maintainer]` can add and remove others maintainers and editors from a data source.
  - **AG.D1.FR.003-1-3. Add an ETL in a data source**: A user `[Admin | Maintainer | Editor | User]` can add an ETL script in a data source.
- **AG.D1.FR.003-2. Edit a data source**:
  - A user `[Admin | Maintainer]` can edit any data source.
  - A user `[Editor]` can edit data sources in which he/she has been designated as editor.
  - A user `[User]` can only edit data sources of which he/she is the owner.
- **AG.D1.FR.003-3. Remove a data source**:
  - A user `[Admin | Maintainer]` can remove any data source.
  - A user `[Editor]` can remove data sources created by him/her.

- **AG.D1.FR.003-4. Set ETL as correct or incorrect**: A user `[Admin | Maintainer | Editor | User]` could mark an ETL as correct or incorrect to prevent an error when the ETL is going to be executed in the local indexer.
- **AG.D1.FR.003-7. Display a data source**: Any user could display the details of a specific data source, including all the properties defined in the ontology AGRICORE DCAT-AP 2.0 extension.
    - **AG.D1.FR.003-7-1. Navigate to the data source link**: Any user can navigate to the data source link.
    - **AG.D1.FR.003-7-2. Display the number of views of a data source**: Any user can see the number of views of a data source.
- **AG.D1.FR.003-8. Update notifications**: A user could be subscribed to an update notification service if a new data source is registered, modified or removed.
    - RSS Feed.
    - Email.

AG.D1.FR.004. Researchers will be able to extend its scope with additional data sources: ARDIT must have a mechanism for extending the scope of the tool with the addition of further data sources to the tool by researchers. Researchers may add or suggest new data sources depending on the level of authorisation they have.

- **AG.D1.FR.004-1. Anyone can suggest a new data source:** Any registered user could suggest a new data source in a dedicated suggestion form.

AG.D1.FR.005. Semantic search will be allowed: ARDIT allows semantic searches using a dedicated API developed in the WP4 to identify datasets and other similar ones. However, ARDIT does not allow finding data within a dataset. For example, a search could be done using natural language such as: 'PH data from Italy between 2018 and 2019'.

- **AG.D1.FR.005-1. Search a data source using natural language**: Any user can search for a data source by typing any text using an input. Then, the result is displayed based on the text typed. E.g.
    - Query: PH data in Europe for the period 2000 – 2020.
    - Result: Dataset, where PH is a variable, in the period requested and the type of match (Exact, Related Variable, Derived variable).

AG.D1.FR.006. Advanced search will be allowed: ARDIT allows advanced search for experienced researchers, using the GUI (Graphical User Interface) to select specific attributes and values to retrieve the datasets indexed in the tool.

- **AG.D1.FR.006-1. Search the results by its title and any of its properties:** Any user can search a data source by any of its properties, including its title.
- **AG.D1.FR.006-2. Filter the results by date:** Any user can sort the results of a search operation by date.
- **AG.D1.FR.006-3. Filter the results by any property:** Any user can filter the results of a search operation by any property.

AG.D1.FR.007. Local deployment capability: ARDIT could provide a way to deploy the tool in a private environment, allowing the possibility to be synchronised with the ARDIT global tool to retrieve the new public datasets indexed.

- **AG.D1.FR.007-1. DWH connection**: An administrator can configure the DWH connection parameters in ARDIT.
- **AG.D1.FR.007-2. ETL execution in the DWH**: A user can execute an ETL in their local DWH.
  - **AG.D1.FR.007-2-1. Add ETL to the job queue**: A user can add an ETL to the job queue.
  - **AG.D1.FR.007-2-2. Remove ETL from the job queue**: A user can remove the ETL from the job queue.
  - **AG.D1.FR.007-2-3. Display ETL job queue**: A user can display the ETL job queue.
  - **AG.D1.FR.007-2-4. Launch an ETL in the job queue launcher**: A user can execute an ETL added to a job queue for the DWH.
  - **AG.D1.FR.007-2-5. ETL execution feedback**: A user could see the feedback on the ETL execution process.
  - **AG.D1.FR.007-2-6. ETL launched feedback**: A user can check if an ETL has been executed and stored in the DWH.
  - **AG.D1.FR.007-2-7. ETL execution only if it is correct**: A user can launch the ETL only if the ETL has been marked as correct. If the ETL is incorrect, the user is not allowed to execute it.
  - **AG.D1.FR.007-2-8. ETL isolated tracked environment**: The application must provide an isolated environment for its execution. An ETL should have tracked its links as well as its names, its output table name to avoid collisions between its data, folder names and common files.
- **AG.D1.FR.007-3. Local indexer database synchronization**: A user can synchronize the local indexer database from the global indexer. Synchronizing the data sources and ETLs by Timestamp.

AG.D1.FR.008. Comments Management: The characterisation of each dataset will have a comment section to suggest and discuss corrections or changes, thus allowing peer review of the ARDIT tool.

- **AG.D1.FR.008-1. Create a new comment:** Registered users will be able to post new comments in the platform.
- **AG.D1.FR.008-2. Reply to a comment:** Registered users will be able to reply to any existing comment in the platform.
- **AG.D1.FR.008-3. Update a comment:** The user who has posted a comment will be able to modify its content at any time.
- **AG.D1.FR.008-4. Delete a comment:** The user who has posted a comment will be able to delete it. If the comment has no replies, it will be completely removed from the system. If the comment has replies, it will be marked as deleted. However, a user `[Admin | Maintainer]` will be able to completely remove any comment from the system directly.

2.2 Non-functional requirements

Non-functional requirements (NFR) define how the system should be or how it should perform through constraints or quality objectives that it must satisfy [3]. The non-functional requirements for the ARDIT tool are the following:

AG.D1.NFR.001. Predefined roles: The user roles (and authorizations) are predefined to:

- Admin: they have access to all the resources available, including all IT aspects.
- Maintainer: users who are able to manage the entire collection of datasets of the platform.
- Editor: users in charge of reviewing dataset characterisations made by users with more restricted roles.
- User: most basic role, which allows creating Draft characterisation proposals and interacting through comments to existing characterisations.

AG.D1.NFR.002. Authentication information: In the global indexer, no authentication information of 3rd sites access. In the local indexer, authentication information to 3rd site access is included.

AG.D1.NFR.003. ETL architecture version: An ETL must have stored the architecture version where it has been executed to track which ETLs are supported for each architecture.

AG.D1.NFR.004. Pop-up alerts: On any user interaction that involves creating, updating or deleting any resource controlled by the platform, a pop-up alert will be displayed to give feedback to the user on the success or failure of the operation.

AG.D1.NFR.005. Help popovers: The interface will provide users with help buttons, which, when clicked, display information about sections, functionalities or actions to be performed.

AG.D1.NFR.006. Open-sourcing: Following the open-source paradigm of the AGRICORE Project, the platform and its code will be publicly available for download to any interested user.

3 Architecture

The ARDIT tool has been designed using a REST (Representational State Transfer) architecture [4]. This type of architecture is characterised by the complete separation between web client (frontend) and web server (backend), unlike monolithic web applications, where the frontend and the backend are fully embedded, combining the user interface, business logic and data access layer in the same project or code base [5]. In REST, the server exposes an API (Application Programming Interface) which is consumed by one or more clients of different types. In this way, different applications, users or developers, whether through other websites, mobile applications or any other software of any kind, can access or modify the same data regardless of the platform they use. Today, there are hundreds of large companies using this type of architecture, where users have both web and mobile applications to access the same type of data. In addition, many companies make their APIs public in order to monetise their data, to access new markets or simply to establish connections with their business partners.

The importance of providing ARDIT with a REST API lies in the possibility of giving users and researchers a way to access and redistribute the data stored on the platform without the need to use a user interface or enabling the possibility to build their own interfaces that consume the API and process the characterisations of the datasets available on the platform freely. Promoting research, innovation, reutilisation and the creation of a community of users specialised in agricultural datasets.

Figure 4 illustrates the different uses that the ARDIT API provides for the users and the project, through the characterisations of the agricultural datasets, defining its three possible applications:

1. **Connection to the ARDIT GUI:** to provide not only the benefits of the API described above, but also, a more user-friendly and accessible mechanism, as it does not require IT experts to use it. Facilitating the inclusion of new characterisations and the search and filtering of them.
2. **Direct use of the API**: the fact of providing an API allows users, researchers and developers to access and manage the platform's resources with tools such as Postman or Swagger, without the need to use the ARDIT GUI. This factor also allows other GUIs to consume the API.
3. **Connection to other AGRICORE project modules**: the API allows other project modules to access the resources that ARDIT has available. An example would be the DWH or the semantic search services that will be enabled on the platform.

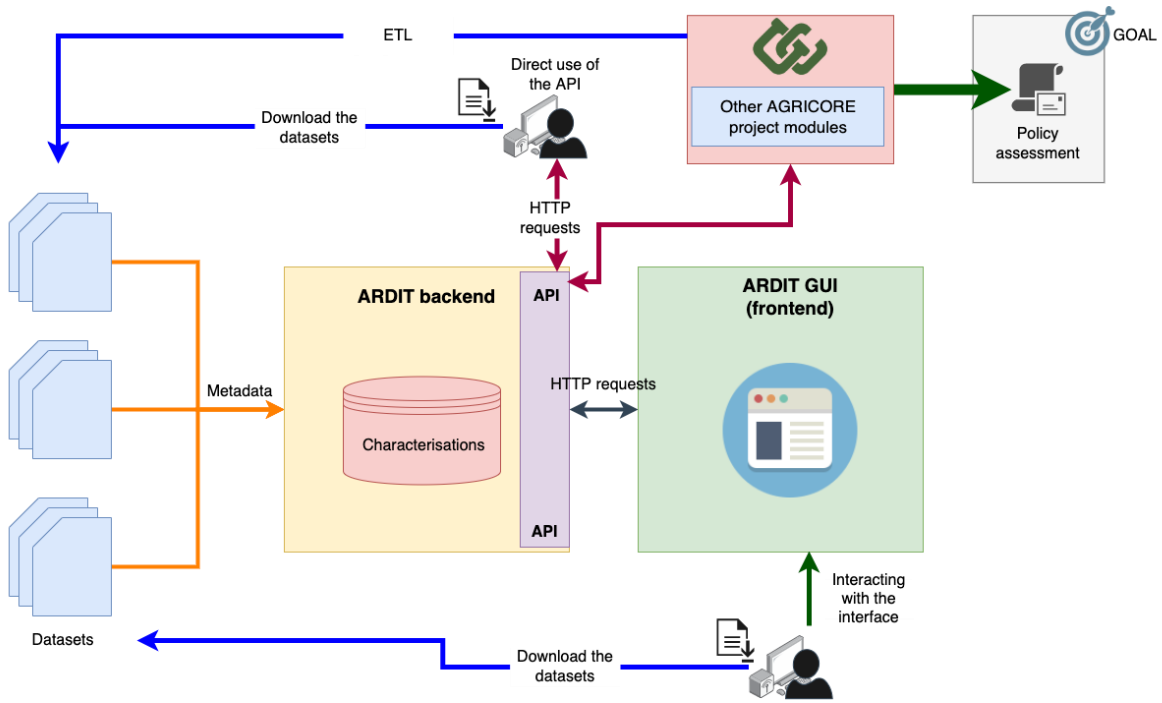


Figure 4 ARDIT API

In the REST architecture, clients send HTTP requests to the server to access or modify available resources (data), while the server responds to the requests with a result (more data) and/or the status of the request. The Hypertext Transfer Protocol (HTTP) is a communication protocol used by the World Wide Web that defines how messages are transmitted between web servers and web clients. HTTP requests consist of the following parameters [6][7]:

- **HTTP verb**: defines which operation is going to be performed. There are multiple verbs depending on the action to be performed, but the four main ones used by ARDIT are:
    - **GET**: used to retrieve data from the server.
    - **POST**: used to create new resources.
    - **PUT**: used to update or modify an existing resource.
    - **DELETE**: deletes a resource.
- **Header**: includes information about the request.
- **Path**: in a REST API is known as endpoint, it defines the URL to access the specific resource.
- **Body**: contains data from the request.

Once the operation is completed, the server sends the result together with the status. Responses also share some request parameters, such as the headers or the body. Usually, the transfer of information between client and server is done by including data in JSON format within the body of the requests and responses. The following Figure 5 shows an example of the structure of the JSON format: [8].

```
{
  "squadName": "Super hero squad",
  "homeTown": "Metro City",
  "formed": 2016,
  "secretBase": "Super tower",
  "active": true,
  "members": [
    {
      "name": "Molecule Man",
      "age": 29,
      "secretIdentity": "Dan Jukes",
      "powers": [
        "Radiation resistance",
        "Turning tiny",
        "Radiation blast"
      ]
    },
    {
```


Figure 5 JSON format example

The status of the responses is identified with numerical codes, allowing the client to determine the success or failure of the operation. Some of the most common codes are:

- **Successful operations**: from code 200 to 299. Some examples:
    - GET request with 200 response code (Ok). The resource is retrieved successfully.
    - POST request with 201 response code (Created): the resource has been created successfully.
    - DELETE request with 204 response code (No content): the resource has been deleted successfully.
- **Client error**: from code 400 to code 499. Some examples:
    - Any request with 400 response code (Bad request): the server cannot process the requests due to errors by the client when building it.
    - Any request with 401 response code (Unauthorized): if the server has authentication, this code is associated with errors when verifying user credentials or if the session has expired.
    - Any request with 403 response code (Forbidden): if the server has authorization, this code is associated with restricted permissions to certain resources depending on the user's role.
    - Any request with 404 response code (Not found): the resource to be retrieved, modified or deleted could not be found in the server.
- **Server error**: from code 500 to code 599. For example:
    - Any request with 500 response code (Internal server error): the request made has produced an error that the server has been unable to handle.

Figure 6 shows an example of communication with the ARDIT REST API using the HTTP protocol. First, a user accesses the ARDIT platform, whose interface will be described in this document, using the *http://www.ardit.agricore-project.eu/datasets* URL in his/her browser. This page

should display a list of datasets available in ARDIT. At that moment, the web client creates an HTTP GET request to the backend's API endpoint associated with dataset resources. The backend processes the request and sends the resulting datasets in a JSON file in the response body, with the status 200 (successfully executed). The ARDIT web client receives the JSON and processes it to display it in an HTML document that is rendered by the browser of the user who made the request.

In the same way, other clients or platforms, either internal or external to the project, can consume the API to access the same resources and be able to process them. The role of the ARDIT API, in this project in particular, is to supply data to other modules, applications or platforms that integrate the AGRICORE suite and not be only linked to the web client presented in this document.

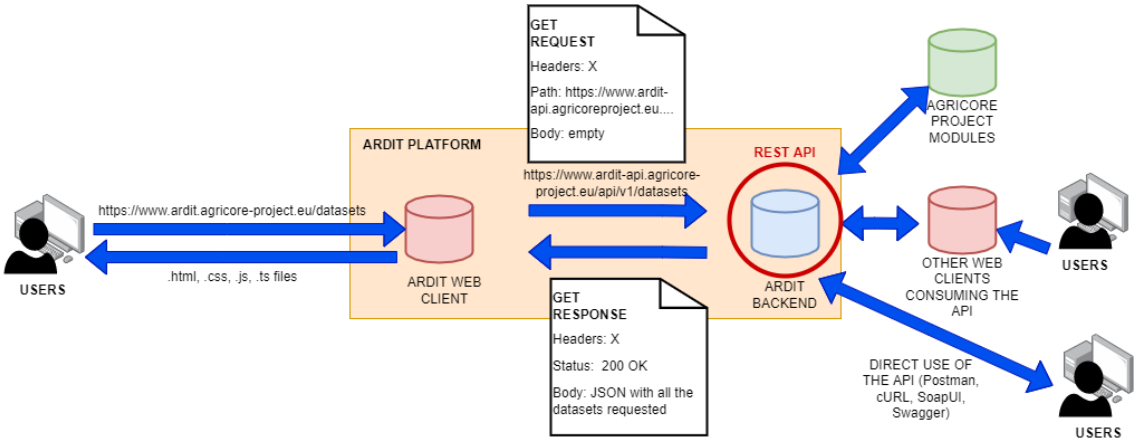


Figure 6 HTTP communication with ARDIT

This communication scheme can be translated into the complete architecture of the platform, which can be seen in Figure 7. The backend and REST API of ARDIT is built using Java, as the programming language, through the Spring framework. PostgreSQL is used as a relational database management system (RDBMS), storing all the data linked to the platform. In addition, given the need for ARDIT to require user accounts in order to record or track user actions, define who can or cannot do certain actions according to their role, or to simply enable communications with the users via email or internal notifications, the platform uses LDAP (Lightweight Directory Access Protocol) as a tool for storing the credentials and account data of the registered users on the platform, for the authentication of users and for the definition of the different roles or groups of users, which ultimately define who can perform certain actions or not.

In addition, the architecture has a web client developed using the Angular framework that consumes the ARDIT API. This client has an interface from which users can register new dataset characterisations and add them to the platform's complete catalogue. It also has management sections where advanced users can manage the platform properly.

In order to comply with the principles defined for the overall architecture of the project, in terms of distributed deployment and use of microservices and containerisation technologies, the complete deployment of the architecture and its tools is configured using Docker.

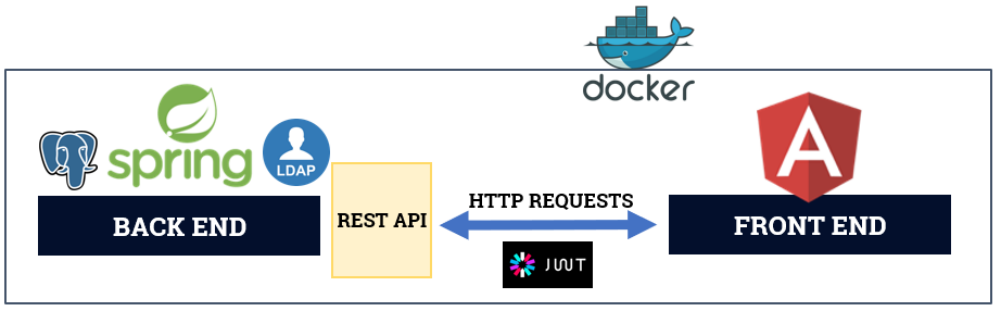


Figure 7 ARDIT architecture

The stateless nature of REST APIs means that the server does not store any information about the client. Each HTTP request or response sent is completely independent and isolated from others. If the API resources require authentication or authorization in order to be retrieved or modified, every request must contain the necessary data to carry it out. So, the client is responsible for storing and managing this information on its side. ARDIT uses the JSON Web Token (JWT)[9] standard for the creation of tokens or keys that allow the authentication and authorization of the user (identity and privileges held). When a user submits a login request in ARDIT along with their login credentials (username/email and password), the backend, first, verifies that the user exists by querying LDAP. Once verified, a token or key specific to the user is generated, which is simply a string of numbers and letters representing encrypted information signed by the HMAC-SHA256 algorithm. Figure 8 shows an example of an encrypted JWT token:

```
eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.ey
JzdWIiOiIxMjM0NTY3ODkwIiwibmFtZSI6Ikpva
G4gRG9lIiwiaWF0IjoxNTE2MjM5MDIyfQ.SflKx
wRJSMeKKF2QT4fwpMeJf36POk6yJV_adQssw5c
```

Figure 8 JWT token example

ARDIT tokens contain the following information:

- **Header** (red in Figure 8): contains information about the token, such as the algorithm that signs it (HMAC-SHA256).
- **Payload** (purple in Figure 8): information regarding the user to whom the token belongs and the expiration date. ARDIT token stores the user's nickname and the list of roles he/she owns.
- **Signature** (blue in Figure 8): the signature is generated by encrypting the header, the payload and a secret that is stored in the backend, making use of the HMAC-SHA256 algorithm.

With the user verified and the token generated, the backend sends the token in response to the login request. ARDIT does this in two different ways, the first is to include the generated token in the body of the response, and the second is to send the token inside a cookie. With the cookie option, browsers can store and send it back to the backend automatically, relieving the user of this responsibility. Instead, through the response option, the user would have to work directly with the token, extracting it from the response and adding it manually to each new request.

Now that the user's API key is on the client-side, it must be added to all the new requests to access restricted resources. With the cookie option, the token is sent automatically on every request to

the backend, but this will always happen if the user is using the web client developed for the project. If the user is not working with ARDIT through the browser, he/she may not be able to handle the cookie. For this reason, the backend also enables the possibility for the user to include the token in the **Authorization** header of the HTTP request (Figure 9).

Figure 9 Example of sending the token in the header with Postman

Once a new request with the token is received, either via cookie or in the header, the backend first validates it using the signature and the secret, checking that the token content has not been modified and that the expiration date of the token has not yet expired. Then, the backend takes the user information contained in the token and verifies that the user is valid and has the necessary permissions to access the requested resources.

The logout process in ARDIT is rather symbolic, only used by the user interface. Both token and cookie are defined, by default, with a maximum expiration period of three hours. After these three hours, neither cookie nor token are valid and if the user tries to make a request to the backend, it will return an authentication error (401), which the web client will interpret by redirecting the user to the login screen. When a user wants to log out from the GUI by clicking on the log out button, the client sends a request to the backend asking to delete the cookie. This process is achieved by simply modifying the cookie with an expired date. When the web client receives the cookie, the browser will automatically delete all information referring to that session, redirecting the user to the login screen.

Extrapolating the technologies and tools that compose the ARDIT to the joint architecture with the DWH, described in the introductory chapter, Global and Local Indexers would share the same ARDIT architecture. Understanding the Global Indexer as the public portal to support agricultural researchers in the identification of the most suitable datasets, Local Indexers would be a complete and independent replication of the Global Indexer architecture, which would mean that each Local Indexer would have the same tools and technologies but, its own user and platform management, database and, therefore, a different collection of datasets than the one provided by the Global Indexer.

Communication between the two indexers is done with RSS (Really Simple Syndication) and the REST API of the Global Indexer. RSS allows Local indexers to subscribe to the Global one to receive notifications and content when new dataset characterisations are added. Allowing users to visualise the characterisations from their Local Indexers and providing the possibility to transfer them to the Local Indexer database using the REST API of the global one. Once in the Local Indexer, users may decide to store such dataset in the DWH.

Figure 10 depicts the replication of ARDIT architecture in both Global and Local Indexers:

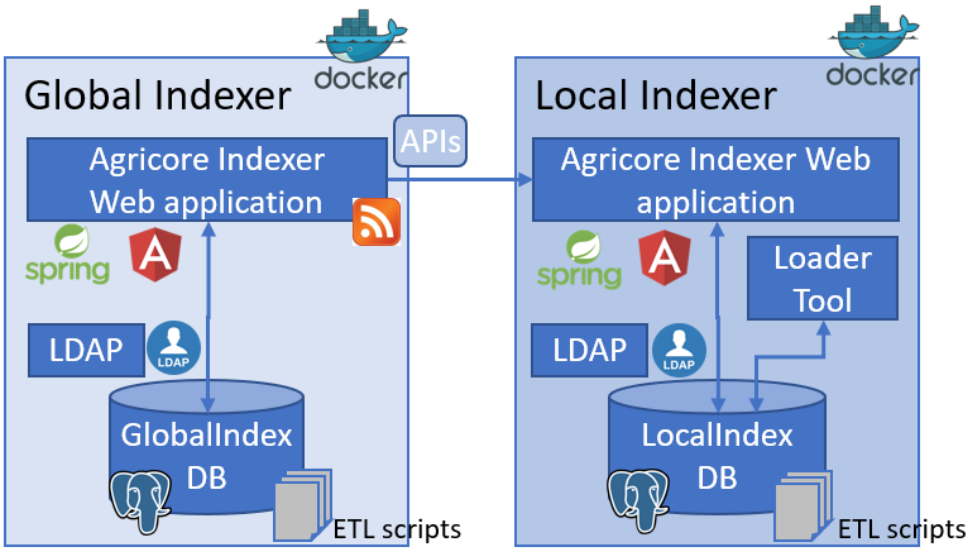


Figure 10 Global and local indexer joint architectures

More information about the technologies and tools that compose the architecture can be found in the annexes of the document.

4 ARDIT development process

This chapter reviews all the stages that the tool has undergone through its development. From its design, GUI construction, first characterisations, to the release of the platform into a production environment. The process of building ARDIT has been a process of continuous evaluation, analysis and monitoring carried out by the partners working on AGRICORE Work Package 1: Data sources and participatory research (WP1). The continuous testing and analysis conducted during the work of the partners and the periodic monitoring meetings that have been held throughout the development phase of the platform have contributed to detect possible improvements, errors or any necessary changes to ensure that the platform meets its defined objectives.

4.1 Development stages

The stages through which the project has passed could be summarised as follows:

4.1.1 Requirements and preliminary architecture definition (mid-2020s)

The first stage of the tool is in which its requirements and architecture were defined.

4.1.2 Generation of the code projects and Docker files (August 2020)

Phase in which the Spring and Angular code projects were initialised and the Docker files were defined with all the tools and technologies that compose the architecture.

4.1.3 User access control (September and October 2020)

The connection between Spring and LDAP with the creation of a REST service for the registration and authentication of users on the platform, generation of JWT tokens and CRUD (Create-Read-Update-Delete) operations on users.

4.1.4 ARDIT GUI mockups and interface design (October 2020)

In this phase, a number of mockups were created with MarvelApp[10] as a very preliminary version of the platform design. Also, the navigation flow that the user would have to perform on the platform was defined. Figure 11 and Figure 12 illustrate some of the mockups defined. The first one represents a screen with the details of a dataset characterisation, and the second one, with the screen where the datasets available in the platform are listed, together with a section with search filters.

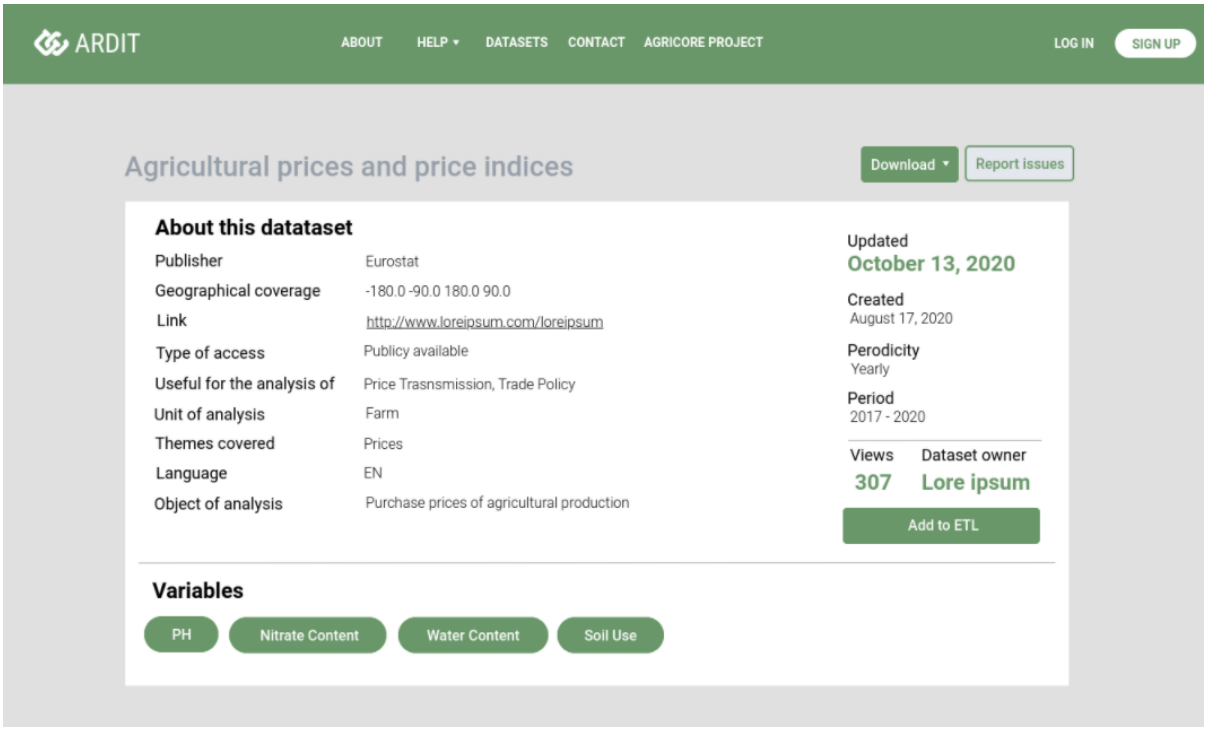


Figure 11 ARDIT mockup with a dataset characterisation details

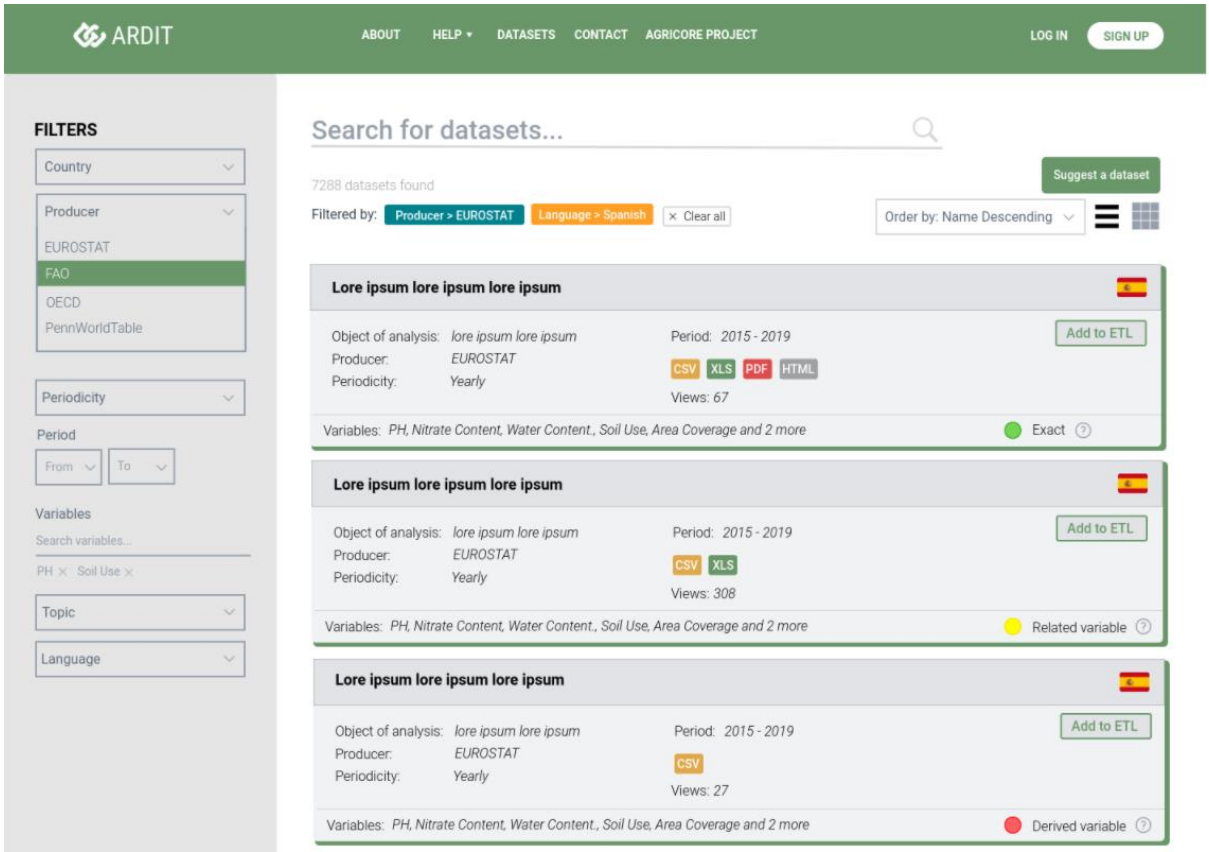


Figure 12 ARDIT mockup with a list of dataset characterisations

4.1.5 AGRICORE DCAT-AP ontology addition (From November 2020 to November 2021*)

Translation of the developed ontology into the ARDIT database schema. Although this process was mainly carried out during the first months of this phase, due to the continuous evaluations and tests that were conducted during the year 2021, some changes were introduced to both the ontology and the database schema to adapt them to the needs that emerged with the first characterisations performed on the platform. For this reason, it was not until November 2021 that a stable and definitive version of the ontology and schema was achieved.

4.1.6 GUI development and characterisation forms (From December 2020 to March 2022*)

In this phase, the development of the ARDIT interface with Angular and the characterisation forms was started, using the properties defined by the ontology. As in the previous phase, due to the improvements and changes introduced throughout the development, this phase was subject to modifications during the whole process.

4.1.7 Deployment on production environment (February 2021)

In February 2021, the ARDIT platform was deployed for the first time on the servers of the project partner and coordinator IDENER. With access restricted to a limited set of user accounts, this process allowed the rest of the partners to start using the tool.

4.1.8 First attempts of characterisations (From March 2021 to September 2021)

With the platform accessible through the web, partners started to characterise datasets and test the functionalities directly on the tool. This allowed to generate discussions and proposals for changes and possible improvements on the ontology and the methodology to characterise datasets.

4.1.9 AGRICORE characterisation deliverables (From November 2021 to January 2022)

During this phase, the partners used the platform to generate the characterisations of about three hundred datasets of different types (statistics, geo-referenced, national and regional datasets), which allowed the drafting of four deliverables on how to characterise each type of dataset.

4.1.10 Release (March 2022)

Migration of the platform to Ayesa AT servers and completion of some pending functionalities, such as a comment system associated with each dataset.

In addition, the entire development process followed the software quality guidelines defined in deliverable D6.6 Software Quality Assurance measures for AGRICORE, with the application of the Git workflow for version control and the use of Continuous Integration (CI) with the automatic execution of tests in the repositories on each deployment in the production environment.

4.2 DCAT-AP data model

The DCAT Application profile for data portals in Europe (DCAT-AP) is a specification based on W3C's Data Catalogue vocabulary for describing public sector datasets in Europe. Its main objective is to define a set of properties in the metadata of the datasets allowing to homogenise the way they are identified and enabling a cross-data portal search. The methodology implemented in AGRICORE D1.1 and the ontology developed based on the DCAT-AP, serves as a basis for developing the data scheme for ARDIT. From the ontology developed, the following properties can be extracted as the most representative or important when characterising a dataset:

- **Dataset**: is the main entity of the model, representing the datasets to be characterised together with all their associated properties.
- **Catalogue**: entity representing a collection of dataset metadata or data services that are related to each other, e.g. all datasets included are generated by the same entity or the data they record are related to each other
- **Unit of analysis**: entity usually referred to something spatially defined, such as the holdings, a region or a country, but it can also be something different such as a commodity.
- **Variable**: entity representing any physical or socio-economic characteristic of the Units of Analysis contained in the dataset, that can be measured or counted.
- **Distribution**: entity representing the set of different formats/shapes in which the data can be retrieved. (e.g. PDF, CSV), as well as the procedures to access the data.
- **Data service**: represents a collection of operations that provides access to or more datasets.

Some of these properties and the attributes that compose them are directly extracted from the DCAT-AP specification, others have been generated specifically for the project to meet the necessary objectives and requirements. More detailed information on how the AGRICORE version of DCAT-AP has been generated and the new properties created for it can be found in deliverable D1.1. Figure 13 shows the UML schema of classes, attributes and relationships that compose the AGRICORE DCAT-AP model:

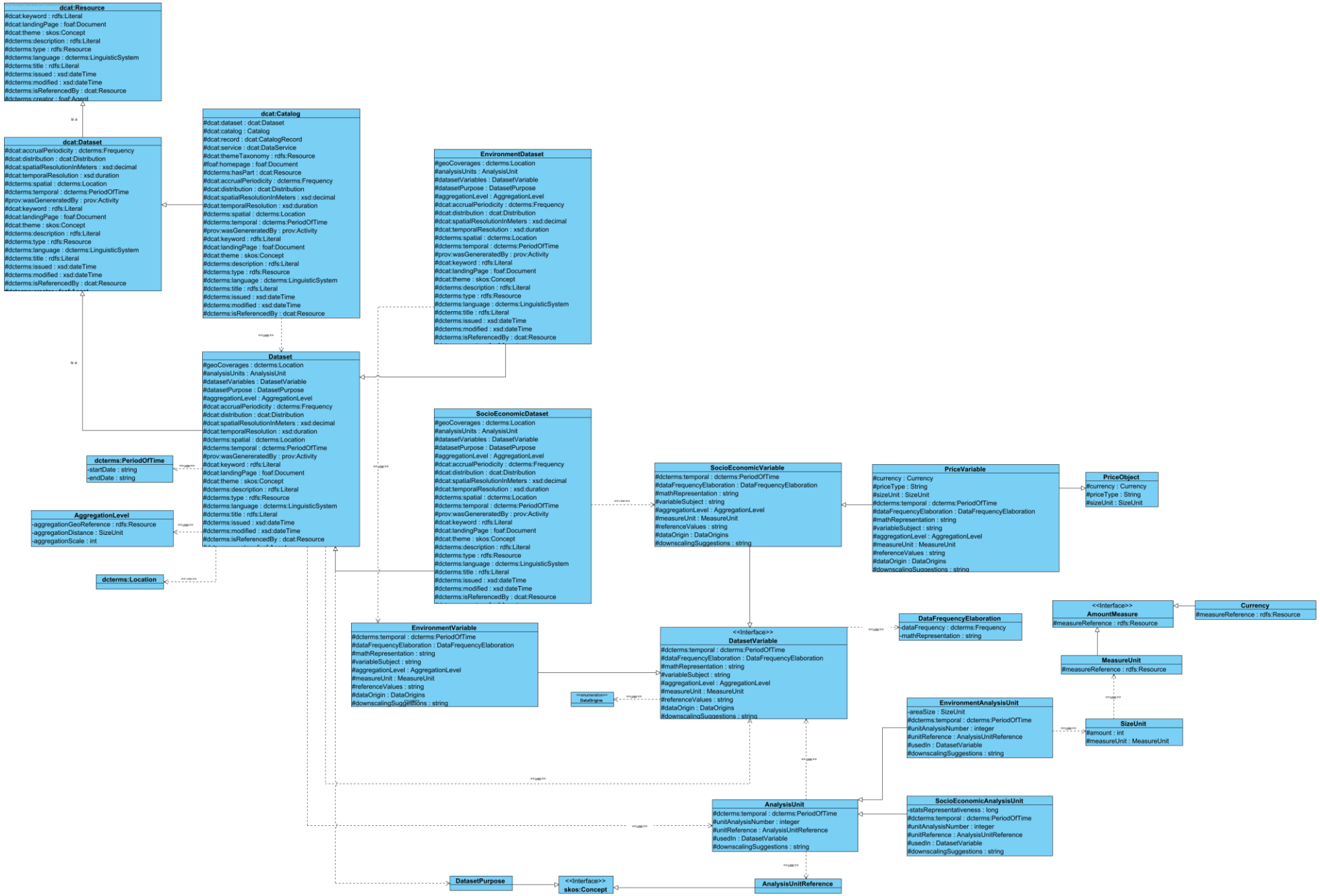


Figure 13 AGRICORE DCAT-AP Data Model Representation

One of the first steps of the development of ARDIT platform consisted in translating the ontology into the database schema, i.e. the transformation of each class and property defined in the DCAT-AP into programming classes and objects interpretable by the ARDIT backend, defining the information models which the platform operates. During the translation process, a new property that is not part of the AGRICORE DCAT-AP was defined, the vocabulary. There are certain properties on the developed ontology that need to take values from existing public vocabularies, which the EU website defines as follows:

"*In order to harmonise and standardise the codes and their associated tables used in the Publications Office and in the inter-institutional exchange of data (i.e. between the institutions involved in the legal decision-making process gathered in the Interinstitutional Metadata Maintenance Committee [IMMC]), a series of lists of National Authority Lists (NALs) have been defined.*" [11]

Vocabularies ensure that all institutions and users can access and make use of the same set of values, including multilingual vocabularies with their composing terms in different languages. Figure 14 depicts an example of a public vocabulary that contains the different types of currencies existing in the world [12]:

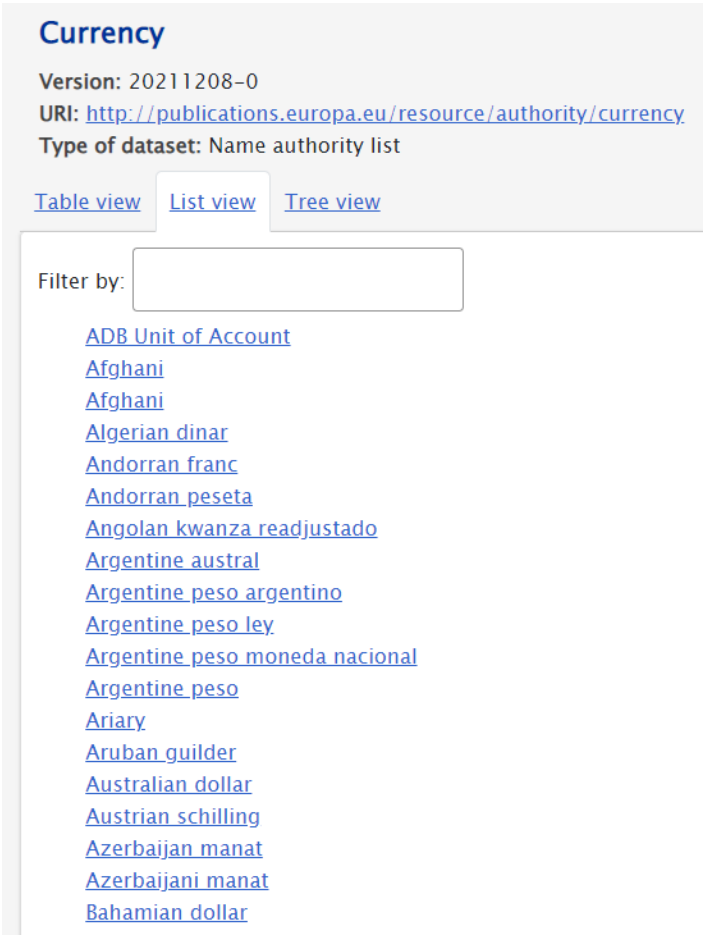


Figure 14 EU Authority tables example for currency vocabulary

The vocabularies in ARDIT have two main uses. The first is to provide users who perform the characterisations with some lists of values associated with each property that requires it, so that they can select the necessary values to assign. Secondly, vocabularies allow the user to populate search filters with values, in order to perform queries to the ARDIT backend to search for

characterisations based on one or multiple parameters. Figure 15 depicts the use of the vocabularies in ARDIT platform to support the characterisation processes:

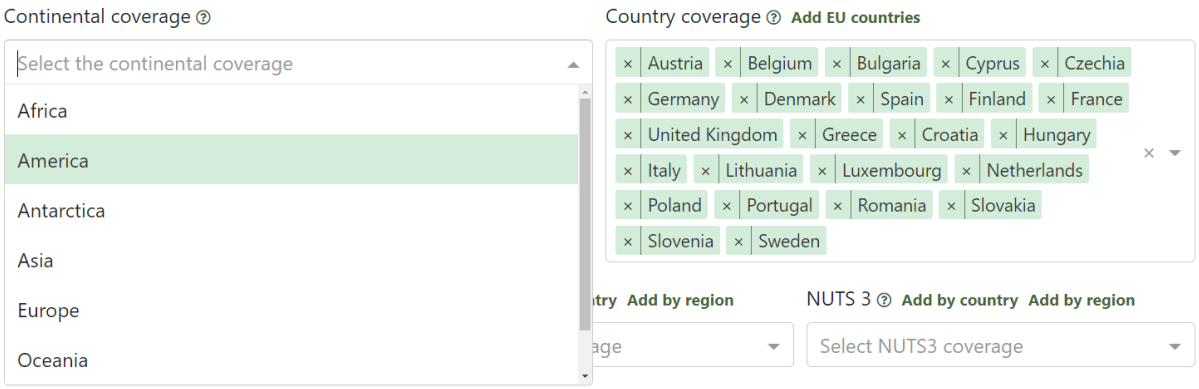


Figure 15 Example of the use of vocabularies in ARDIT platform

To provide ARDIT with the necessary vocabularies to populate the drop-down menus, it was necessary to extract them from their origins. To this end, a small tool was developed in Python[13], given the facilities provided by this language and its libraries to process files in different formats. The process of extracting and adding the vocabularies to ARDIT through this Python utility is summarised in the following steps:

1. **Extract the vocabularies from their origins**: many of the pages on which the vocabularies are hosted allow them to be downloaded directly as JSON or XML files. Others, however, do not. In this case, the HTML code of the pages was downloaded and processed with Python.
2. **Processing of vocabularies with Python**: depending on the format in which the vocabulary was extracted, the processing is different. JSON files are easily parseable. For HTML and XML files, Python's Beautiful Soup library [14] was used, which provides the necessary tools and utilities to parse this type of files and extract information from the tags.
3. **Generation of output files:** once the necessary information has been extracted, the output files are generated, which subsequently allow the vocabularies to be loaded into the database. The tool generates files in three different formats: JSON, XLSX and SQL. Being the last one used for loading them. The structure of the files is defined by the properties created for the vocabularies and their associated values. In ARDIT these are:
    a. **Vocabulary**: entity representing the vocabulary. It is composed of the following attributes:
        i. **Id**: unique identifier for the database (primary key).
        ii. **Name**: name that identifies the vocabulary over others.
        iii. **Description**: a brief description of the vocabulary.
        iv. **Topic**: dataset property to which it is related to.
        v. **URL**: link to the vocabulary on the internet.
    b. **Vocabulary value**: entity representing each value that composes the vocabulary. It is composed of the following attributes:
        i. **Id**: unique identifier for the database (primary key).
        ii. **Code**: A short identifier of the value. It can be a combination of numbers and letters of 3 or 4 characters. Such as "v01" or "vl31".

iii. **Label**: The text to be displayed in the dropdown menus.

iv. **URL**: link to the vocabulary value on the internet.

v. **Extra data**: Additional information that differentiates this value from the rest of the values in the vocabulary.

vi. **Vocabulary id**: unique identifier of the vocabulary to which the value belongs to.

4. **Uploading these files to the ARDIT database:** once the files are generated, they are added to a folder in the backend, where Liquibase[15], a tool for track, version and deploy database changes, load them.

The following Figure 16, shows an example of a SQL file generated by the Python utility:

```
INSERT INTO public.vocabulary (id,description,name,topic,url) VALUES (9, 'Dublin Core Frequency vocabulary', 'Dublin Core Frequency',

INSERT INTO public.vocabulary_value (id,code,extra_data,label,url,vocabulary_id) VALUES
(2327, "triennial", NULL, "Triennial", "http://purl.org/cld/freq/triennial", 9)
,(2328, "biennial", NULL, "Biennial", "http://purl.org/cld/freq/biennial", 9)
,(2329, "annual", NULL, "Annual", "http://purl.org/cld/freq/annual", 9)
,(2330, "semiannual", NULL, "Semiannual", "http://purl.org/cld/freq/semiannual", 9)
,(2331, "threeTimesAYear", NULL, "Three times a year", "http://purl.org/cld/freq/threeTimesAYear", 9)
,(2332, "quarterly", NULL, "Quarterly", "http://purl.org/cld/freq/quarterly", 9)
,(2333, "bimonthly", NULL, "Bimonthly", "http://purl.org/cld/freq/bimonthly", 9)
,(2334, "monthly", NULL, "Monthly", "http://purl.org/cld/freq/monthly", 9)
,(2335, "semimonthly", NULL, "Semimonthly", "http://purl.org/cld/freq/semimonthly", 9)
,(2336, "biweekly", NULL, "Biweekly", "http://purl.org/cld/freq/biweekly", 9)
,(2337, "threeTimesAMonth", NULL, "Three times a month", "http://purl.org/cld/freq/threeTimesAMonth", 9)
,(2338, "weekly", NULL, "Weekly", "http://purl.org/cld/freq/weekly", 9)
,(2339, "semiweekly", NULL, "Semiweekly", "http://purl.org/cld/freq/semiweekly", 9)
,(2340, "threeTimesAWeek", NULL, "Three times a week", "http://purl.org/cld/freq/threeTimesAWeek", 9)
,(2341, "daily", NULL, "Daily", "http://purl.org/cld/freq/daily", 9)
,(2342, "continuous", NULL, "Continuous", "http://purl.org/cld/freq/continuous", 9)
,(2343, "irregular", NULL, "Irregular", "http://purl.org/cld/freq/irregular", 9);
```

Figure 16 Example of a vocabulary SQL output file

Regarding the data schema used by ARDIT, in general, each DCAT-AP main entity has a specific class/table. However, the main difference with respect to the DCAT-AP data model is that the latter divides the different types of datasets, units of analysis and variables (socio-economic and geo-referenced) into different classes for each type, e.g. the SocioeconomicDataset class and the EnvironmentalDataset (geo-referenced) class. In ARDIT, this division does not exist, and all types are compacted in the same class, and therefore in the same database table, where there is an attribute known as datasetType, variableType or analysisUnitType that stores the type of entity to differentiate it from the rest.

The following Figure 17 shows the ARDIT database scheme after the translation of the AGRICORE DCAT-AP model.

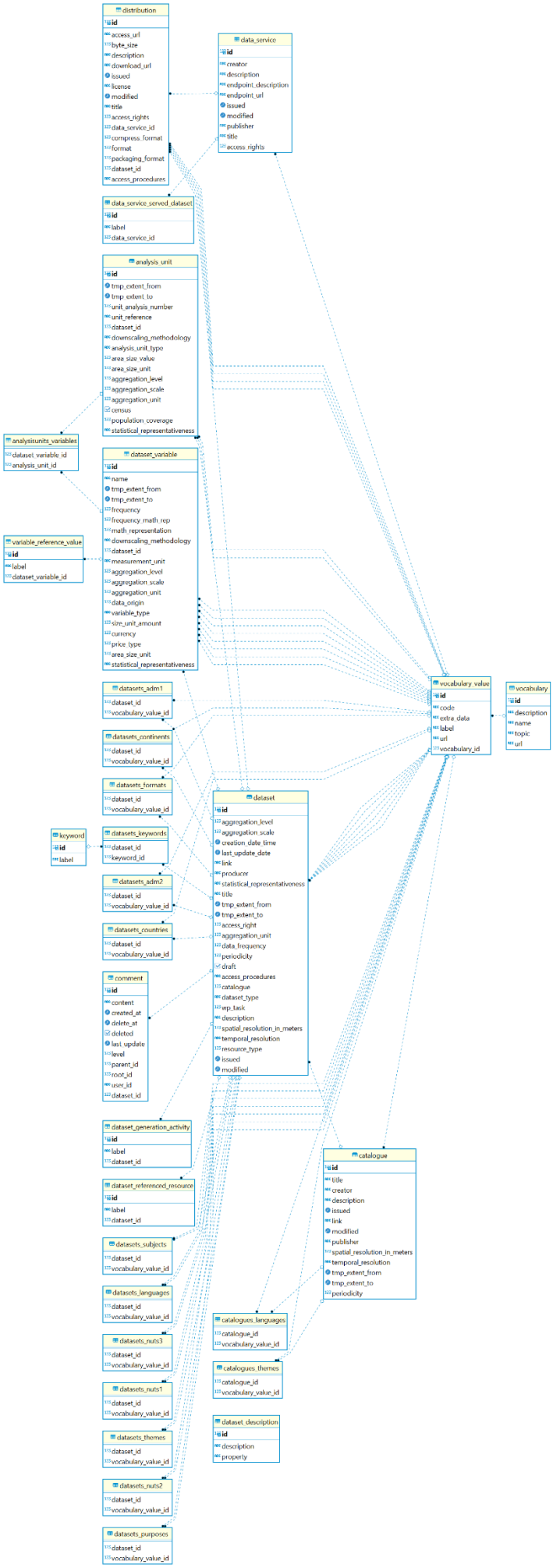


Figure 17 ARDIT database scheme

The classification of the tables is the following:

- **Vocabulary**: table representing the vocabularies (*See "Vocabularies" section in ARDIT user manual*).
- **Vocabulary_value**: table representing the values that compose the vocabulary (*See "Vocabularies" section in ARDIT user manual*).

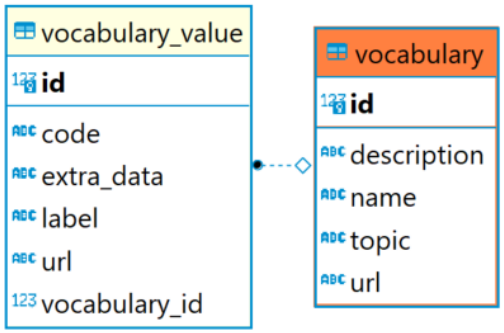


Figure 18 Database tables representing the vocabularies

- **Dataset**: table representing the agricultural datasets (See "Description of ARDIT Sections" in ARDIT user manual and "Dataset" in the guide to ARDIT data model).
- **Distribution**: table representing dataset distributions (See "Distributions" in ARDIT user manual and "Distribution" in the guide to ARDIT data model).
    - **Data_service**: table representing data services associated with a distribution (See "Distributions" in ARDIT user manual and "Data service" in the guide to ARDIT data model).
- **Analysis_unit**: table to store unit of analysis related to the dataset (See "Unit of analysis" in ARDIT user manual and "economic unit of analysis" and "Geo-referenced unit of analysis" in the guide to ARDIT data model).
- **Dataset_variable**: table to store the variables related with a dataset (See "Variables included" in ARDIT user manual and "economic and geo-referenced variables" and "Price variables" in the guide to ARDIT data model).
- **Catalogue**: table to store the catalogues of agricultural datasets (See "Catalogues" in ARDIT user manual and "Catalogue" in the guide to ARDIT data model).
- **Dataset_description**: table to store the help popover descriptions (See "Property description" in ARDIT user manual).
- **Keyword**: table to store tags or keywords related to the existing datasets in the platform (See "Keywords" in ARDIT user manual).

4.3 Graphical User Interface

As a preliminary step to the design of the ARDIT interface itself, other data portals were explored to determine the structure of the platform to be built. Some of the platforms analysed were the following:

- **The European Data Portal [16]**: is the main access point to data published by the institutions, agencies and other bodies of the European Union. With more than one million registered datasets, its aim is to make the EU institutions and other bodies more transparent and accountable.

- **The Spanish Government's Open Data Initiative portal [17]**: backed by the Economy and Business and the Territorial Policy ministries, its main objective is to promote the opening of public information and development of advanced services based on data. This portal has several tens of thousands of datasets from different categories and publishers.
- **The U.S. Government's open data portal [18]**: official US government data portal, containing more than three hundred thousand datasets.
- **The NASA open data portal [19]**: official NASA data portal contains tens of thousand datasets. Similarly to ARDIT, the majority of datasets only hold the metadata. The actual data of them are held on other NASA sites.

As a result of the analysis of these platforms, a number of conclusions were drawn, defining the common structure of these types of portals, which is:

- **Main page**: landing page or main page of the platform, containing information about the platform and menus and links to the other sections, especially those dedicated to datasets.
- **Dataset list page**: page containing a list of datasets with menus and search filters so that the user can perform queries based on multiple dataset properties and attributes.
- **Dataset details page**: the metadata and properties of each dataset resulting from the list on the previous page can be consulted on a separate page, where users can also find download links or access to the dataset data itself.
- **Others**:
    - **Contact**: page containing a contact form to comment or report any information, suggestions, problems or errors.
    - **Help/About**: page containing guides or help about the platform or how to make use of it
    - **API documentation**: page specially designed for developers or users with technical knowledge with documentation about the use of the platform API to retrieve data.

4.3.1 Colour palette

The ARDIT interface uses a colour range traditionally associated with the rural and agricultural world, which evoke the colour of forests and cultivated fields, combined with various greyscale colours, which evoke the political-institutional character of the AGRICORE tools and also complement and highlight the content appropriately. Figure 19 shows the complete colour palette for ARDIT GUI:

Figure 19 ARDIT colour palette

4.3.2 User workflow and navigation

The organisation and navigation process of the ARDIT interface is based on the web portals explored as a previous step to its design and depends on two main elements:

1. **Public/private sections**: there are sections and features of the website that are completely open to the public and others that require a user account and login to access or use them.
2. **User role**: once logged in, there are also certain sections or functionalities that are blocked to users who do not have the necessary permissions. Such as platform management sections that can only be accessed by administrators or users with maintenance permissions.

Figure 20 shows the operational flow that the user should follow to interact with the platform. The green squares represent independent views or screens within the platform, while the yellow squares represent sections or parts of a screen such as buttons, navigation bars, overlapping windows or any other element that does not assume a page change. The squares with padlocks represent pages or elements that either require a login to access them or, if logged in, require a specific role.

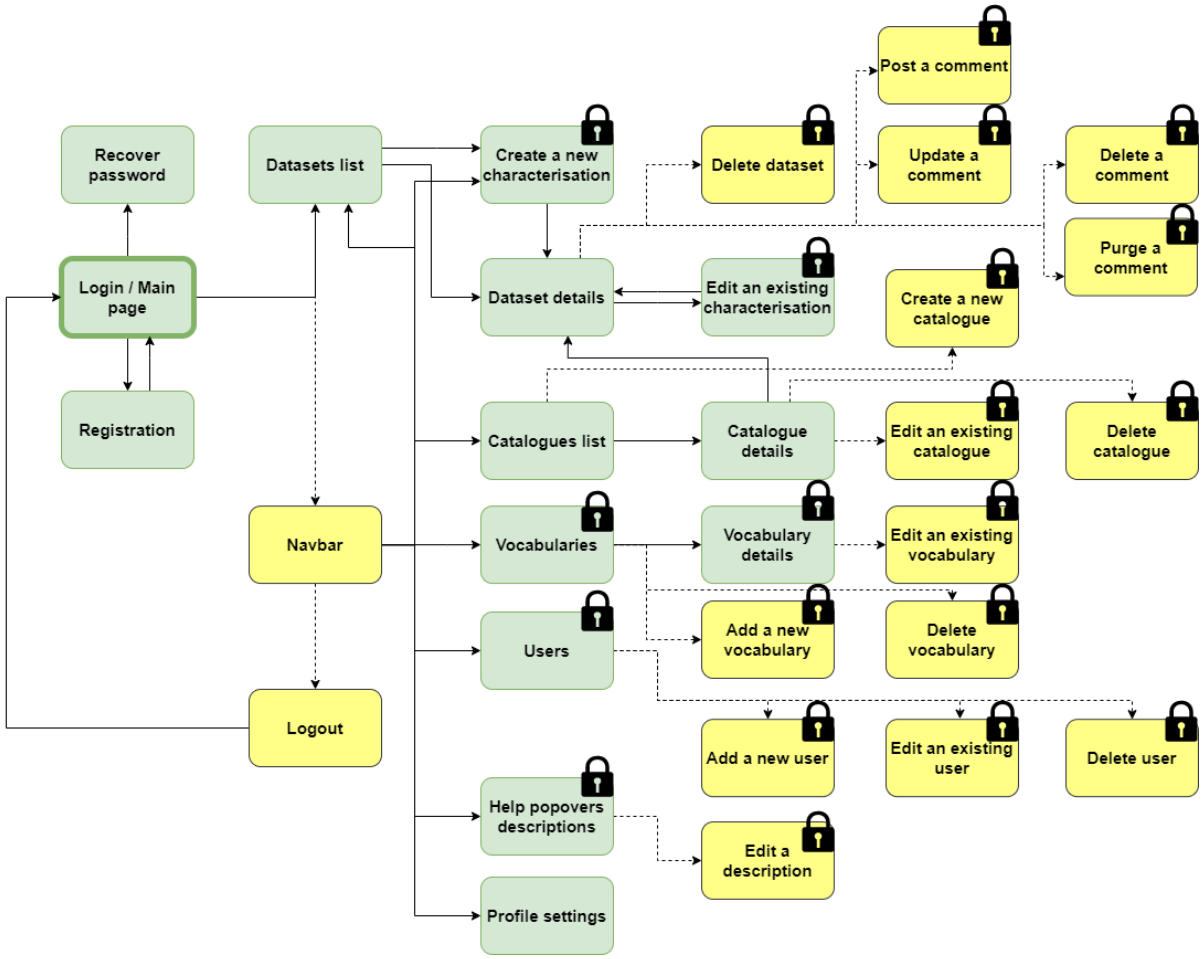


Figure 20 ARDIT user workflow

In general, the platform is divided into the following sections:

- **User access**: includes login, user registration or forgotten password retrieval screens.

- **Dataset**: comprises all platform pages used to visualise or manage the dataset collection available on the platform.
    - **Dataset list**: this fully public screen, accessible without login, allows the user to list the characterisations registered in the platform and to filter them by performing searches based on multiple parameters.
    - **Dataset details**: by selecting any of the datasets listed on the previous page, the user navigates to a detailed view showing all the metadata of the dataset, which is defined by the AGRICORE ontology based on the DCAT-AP.
        - **Comments**: each characterisation registered on the platform has a comment system where users can discuss any doubts, suggestions, errors or possible improvements about a particular characterisation. Any user with an account on the platform can add comments to every one of the datasets available on the platform. The comment system is managed internally by users with more advanced permissions, like admins and maintainers (*see the section on governance of the platform resources*). These users can delete comments that other users have made. Meanwhile, authors of comments can edit and delete their own comments.
    - **Create a new dataset characterisation**: in this view, only available after logging in, users can create new dataset characterisations using the forms provided for this purpose.
    - **Edit an existing dataset characterisation**: in this view, also available after logging in, users can use the characterisation forms to update or modify the existing dataset characterisations.
- **Catalogues**: catalogues are collections of datasets, which have common characteristics, such as the type, the scope, the organisation compiler or if all data contained in them are related to each other in some way. Through this section, users can access the list of the catalogues registered in the platform and visualise the details of each one, including the list of datasets that compose it. In addition, specific users can manage catalogues by editing their properties.
- **Management**: in the management sections, a registered user with maintenance permissions will be able to manage properties such as:
    - **Vocabularies**: in this section, the vocabularies that control certain properties can be managed, being able to create new vocabularies or edit the existing ones.
    - **Users**: in this section, registered user accounts can be managed, being able to create new users accounts, edit or delete the existing ones.
    - **Help popovers descriptions**: from this page, certain users can manage the content of the help popovers that appear on the platform to guide and clarify other users about the properties and attributes of the dataset characterisations.
- **Profile settings**: on this page, logged-in users will be able to access their user account settings. Here, they can edit their data, change their password or delete their account on the platform.
- **Help**: section containing user guides of the platform (**Q&A**) on how to use ARDIT, how to characterise datasets, how to use the API, etc.
- **Contact:** section with a form to contact the site administrator about possible suggestions, improvements, questions or bug reports.

When landing on the platform, users would be presented with the main screen of the platform. On this first screen, information about the project and the ARDIT platform is displayed, along with a login form and a guest access link.

The platform has a navbar from which you can access each and every one of the sections available in the tool. However, the content of the navbar may depend on whether the user is logged in to the platform or if he/she has acceded as a guest, in which case some sections and content of the platform will not be accessible or visible. Figure 21 depicts ARDIT navbar if the user has logged in:

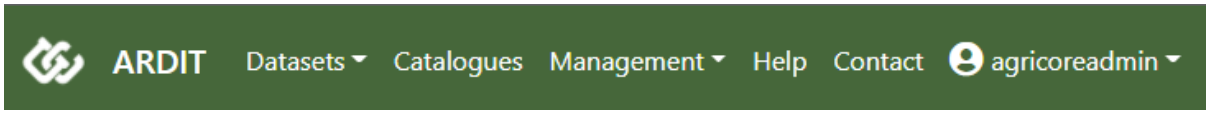


Figure 21 ARDIT navbar

Meanwhile, Figure 22 shows the appearance of the navbar when accessing:

Figure 22 ARDIT navbar when accessing as a guest

In addition to the navigation bar, the platform has a secondary navigation system via breadcrumbs (depicted in Figure 23), a set of text links to guide the user through the platform and its internal structure.

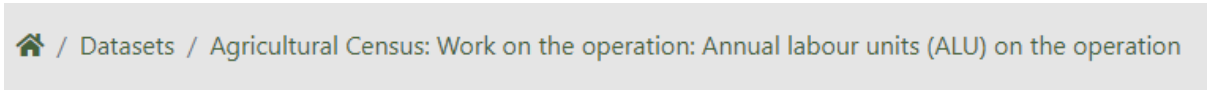


Figure 23 ARDIT breadcrumbs

As a last alternative form of navigation, users can also make use of the website URL in the browser. Taking the base URL of the platform as a basis (ardit.agricore-project.eu), the paths to the available resources can be added to it. For example:

- **Base URL + "/login"**: redirects the user to login page.
- **Base URL + "/datasets"**: redirects the user to a page where all the datasets available on the platform are listed.
- **Base URL + "/datasets/" + "id"**: redirects the user to the details page of a specific dataset.
- **Base URL + "/catalogues"**: redirects the user to a page where all the catalogues available on the platform are listed.

Figure 24 depicts the platform's URL to access the details page of a specific dataset:

Figure 24 ARDIT URL navigation example

4.4 Resources governance

Access to the different resources available on the ARDIRT platform is limited or restricted depending on the role of the user who wishes to access or modify them. In order to control the authorisation of access to resources, five roles have been defined:

- **ADMIN**: superusers, they have access to all the resources available, as well as being able to manage all IT aspects.
- **MAINTAINER**: role created for those users who are able to manage the entire collection of datasets of the platform, among other things.
- **EDITOR**: users in charge of reviewing dataset characterisations made by users with more restricted roles, such as basic users. When users with the basic role (USER) create new characterisations, these are added as drafts to the platform and can only be visualised by editors or maintainers, who have specific permissions to publish them on the platform, after a review process that verifies that the characterisation is valid and complete. In addition, editors can also modify any existing characterisation in the platform.
- **USER**: most basic role. They can create and edit their own characterisations (drafts characterisations), but these must be supervised by an editor or another user with a higher role before they can be fully published on the platform. In addition, they can add comments in any other existing characterisation.

In addition, to these roles managed and maintained from LDAP through user groups, unlogged users (guests) should be added. Although the fact of not having a user account or not having entered credentials is related to authentication, not having logged in also affects access to certain platform resources.

The following Table 1 presents the permissions that the different roles have for each available resource and functionality of the platform:

Table 1 List of permissions according to user role

	ADMIN	MAINTAINER	EDITOR	USER	NON-REGISTERED USER
Login, registration and password recovery screens	Yes	Yes	Yes	Yes	Yes
Dataset list	Yes	Yes	Yes	Yes	Yes
Dataset details	Yes	Yes	Yes	Yes	Yes
Create a new dataset characterisation (fully public)	Yes	Yes	Yes	No	No
Create a new dataset characterisation proposal (draft datasets)	Yes	Yes	Yes	Yes	No
Edit an existing dataset characterisation	Yes	Yes	Yes	Yes*	No

				(only if he/she is the creator)	
Delete an existing dataset characterisation	Yes	Yes	Yes* (only if he/she is the creator)	No	No
Search for draft dataset characterisations	Yes	Yes	Yes	No	No
Publish dataset characterisation proposals	Yes	Yes	Yes	No	No
Reject dataset characterisation proposals	Yes	Yes	Yes	No	No
List the comments of a dataset characterisation	Yes	Yes	Yes	Yes	Yes
Create new comments	Yes	Yes	Yes	Yes	No
Update comments	Yes* (only if he/she is the author)	Yes* (only if he/she is the author)	Yes* (only if he/she is the author)	Yes* (only if he/she is the author)	No
Delete comments	Yes	Yes	Yes* (only if he/she is the author)	Yes* (only if he/she is the author)	No
Purge comments	Yes	Yes	No	No	No
Catalogue list	Yes	Yes	Yes	Yes	Yes
Catalogue details	Yes	Yes	Yes	Yes	Yes
Create a new catalogue	Yes	Yes	Yes	No	No
Edit a catalogue	Yes	Yes	Yes	No	No
Delete a catalogue	Yes	Yes	Yes	No	No
Vocabulary list	Yes	Yes	Yes	No	No
Vocabulary details	Yes	Yes	Yes	No	No
Create a new vocabulary	Yes	Yes	Yes	No	No
Edit a vocabulary	Yes	Yes	Yes	No	No
Delete a vocabulary	Yes	Yes	Yes	No	No
Registered user list	Yes	Yes	Yes	No	No
Create a new user	Yes	No	No	No	No
Edit an existing user data	Yes	No	No	No	No
Edit an existing user roles	Yes	Yes* (except adding the ADMIN role to a user)	No	No	No
Edit an existing user password	Yes	No	No	No	No
Delete an existing user	Yes	Yes	No	No	No
List help popover descriptions	Yes	Yes	No	No	No
Edit a help popover description	Yes	Yes	No	No	No

Delete a help popover description	Yes	Yes	No	No	No
Profile settings	Yes	Yes	Yes	Yes	No

Authorisation management on the platform is applied in three layers, two on the frontend side and one on the backend:

- **Frontend management**:
    a. Hiding/showing interface elements and content according to the user role.
    b. Blocking navigation to certain pages of the platform through a "*guard*", which checks the user's role before starting the navigation and blocks it if it does not comply with the necessary permissions.
- **Backend management**:
    a. Each API endpoint has an associated list of authorised roles. In each request to the backend, it is checked if the role of the user that sends it is among the authorised ones. If it is not, the backend returns a HTTP 403 error (forbidden).

4.5 Results obtained

Thanks to the effort made to align ARDIT with the dataset characterisation methodology developed in previous AGRICORE tasks and the deployment of the platform, the partners were able to directly use and test the tool for the drafting of deliverables D.13 to D1.6 of the project:

- D1.3 Characterisation of EU and other Supranational statistics datasets.
- D1.4 Characterisation of Geo-referenced datasets.
- D1.5 Characterisation of national and regional data sources.
- D1.6 Characterisation of datasets from previous research projects.

These deliverables were conceived with two objectives in mind. On the one hand, to validate and confirm that the developed methodology is suitable for characterisation processes and applicable to different types of agricultural datasets. And on the other hand, to populate the ARDIT database with all the characterisations produced. Both processes were unified and made easier for the partners thanks to the release of the tool. The results of the deliverables allowed a total of 310 datasets, classified into socio-economic and geo-referenced datasets, to be added to the platform. Being 86% of the datasets characterised socio-economic and the rest, 14%, geo-referenced. Furthermore, if the total number of characterised datasets is divided according to the deliverable to which it belongs (Figure 25), the result would be:

- 53% of datasets belongs to D1.3 Characterisation of EU and other Supranational statistics datasets.
- 32% to D1.5 Characterisation of national and regional data sources.
- 12% to D1.4 Characterisation of Geo-referenced datasets.
- The remaining 4% belongs to D1.6 Characterisation of datasets from previous research projects.

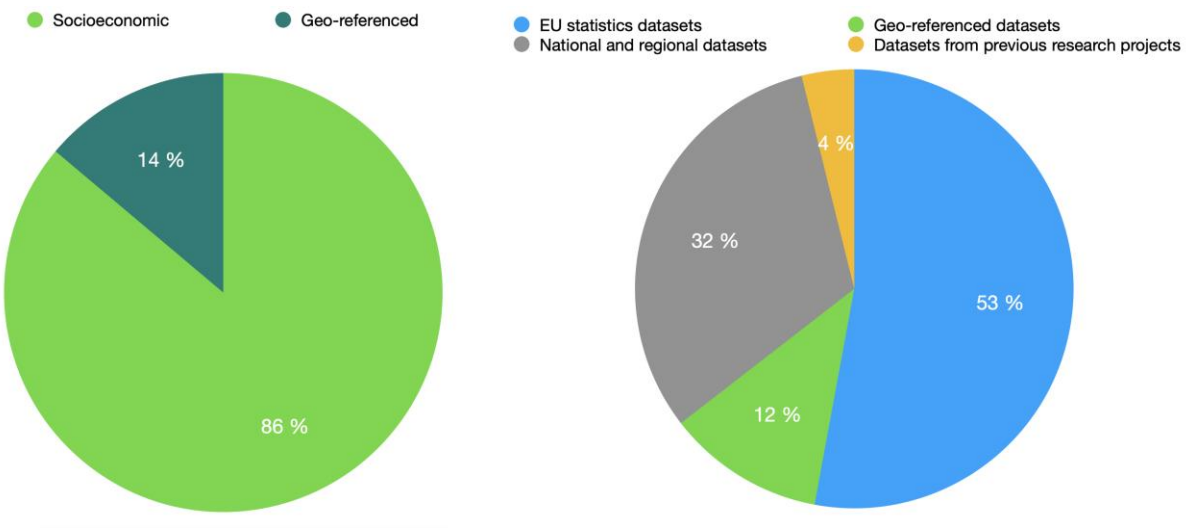


Figure 25 Types of datasets characterised on the platform

The following Figure 26, Figure 27 and Figure 28 show some screenshots of dataset characterisations and their data included in ARDIT:

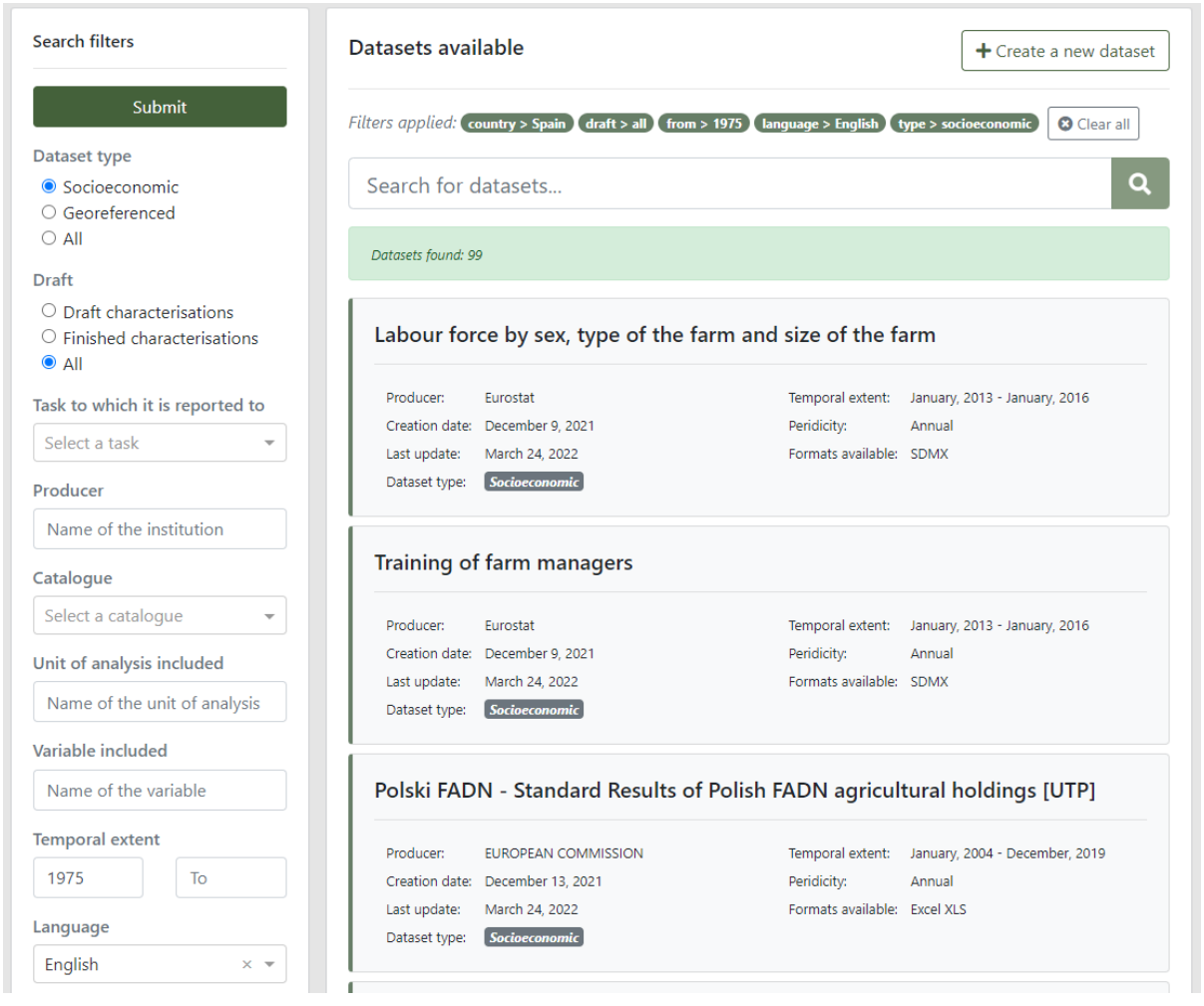


Figure 26 Example of the search and filtering screen for characterisations in ARDIT

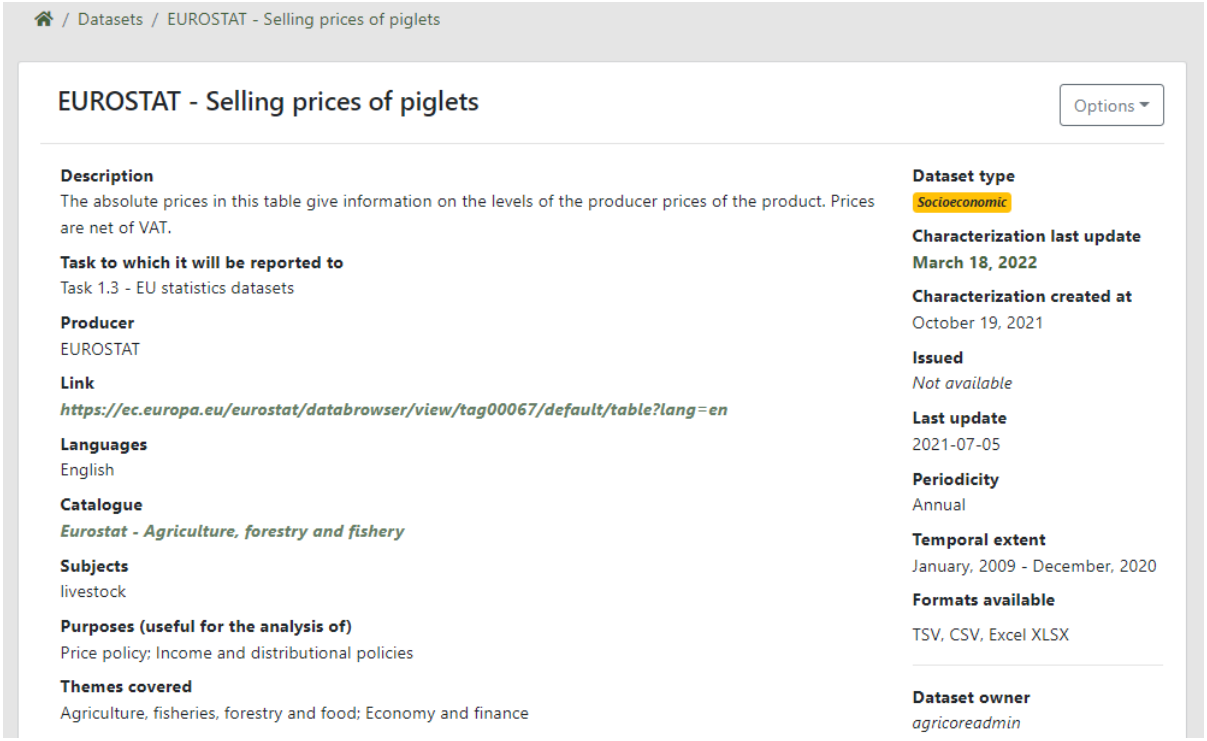


Figure 27 Example of a dataset characterisation in ARDIT platform

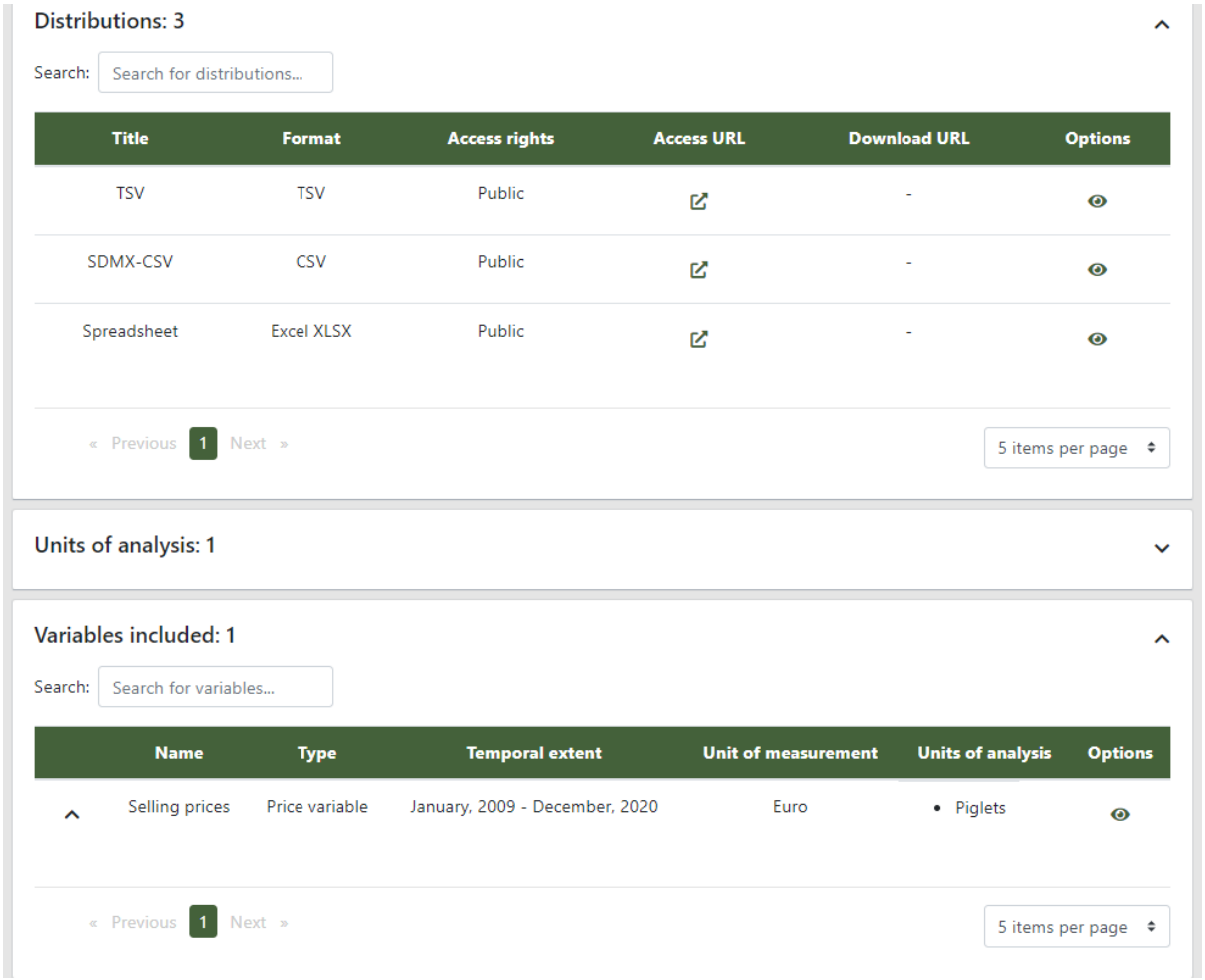


Figure 28 Content of a characterised dataset in ARDIT platform

On the other hand, Figure 29, Figure 30 and Figure 31 show some screenshots of the deliverables submitted in January 2022. The characterisations included in each one were added as annexes, as can be seen:

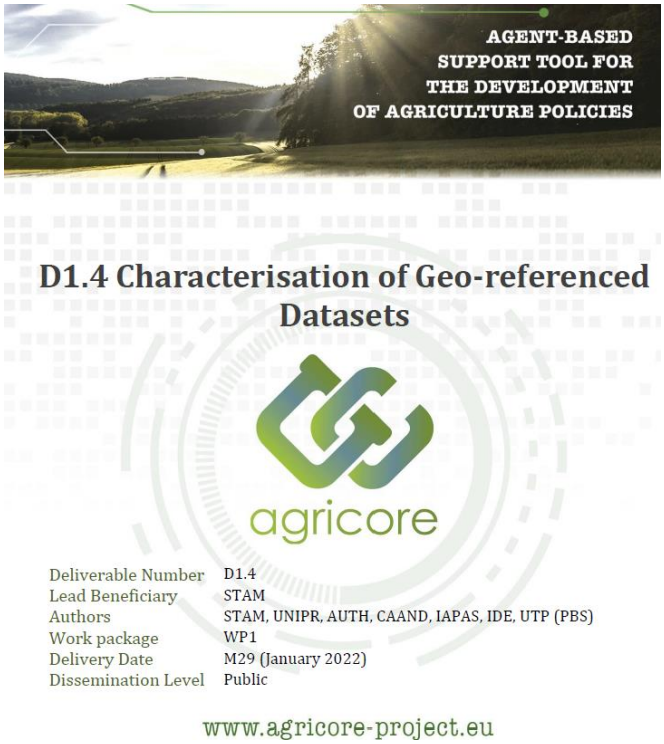


Figure 29 Cover of one of the submitted deliverables

Start of vegetation growing season 2000-2016

General information

Description: The raster files are the time series of the start of the vegetation growing season (day of the year) and the derived linear trends (in day / year). The start of the growing season time-series is based on the time series of the Plant Phenology Index (PPI) derived from the MODIS BRDF-Adjusted Reflectance product (MODIS MCD43 NBAR). The PPI index is optimized for efficient monitoring of vegetation phenology and is derived from the source MODIS data using radiative transfer solutions applied to the reflectance in visible-red and near infrared spectral domains. The start of season indicator is based on calculating the start of the vegetation growing season from the annual PPI temporal curve using the TIMESAT software for each year between and including 2000 and 2016.

Producer: European Environment Agency

Link: https://www.eea.europa.eu/data-and-maps/data/annual-start-of-vegetation-growing

Languages: English

Catalogue: BISE - Biodiversity Information System for Europe

Subjects: crop production, agricultural holdings, agricultural product, plant breeding, Agriculture, fisheries, forestry and food, Environment

Useful for the analysis of: Crop production, Farm structure, land use, production indices, Social Sustainability - Work force characteristics, Use of agricultural area, Use of water

Themes covered:

Spatial resolution (in meters): 500.0

Temporal resolution: Annual

Resource type: Dataset

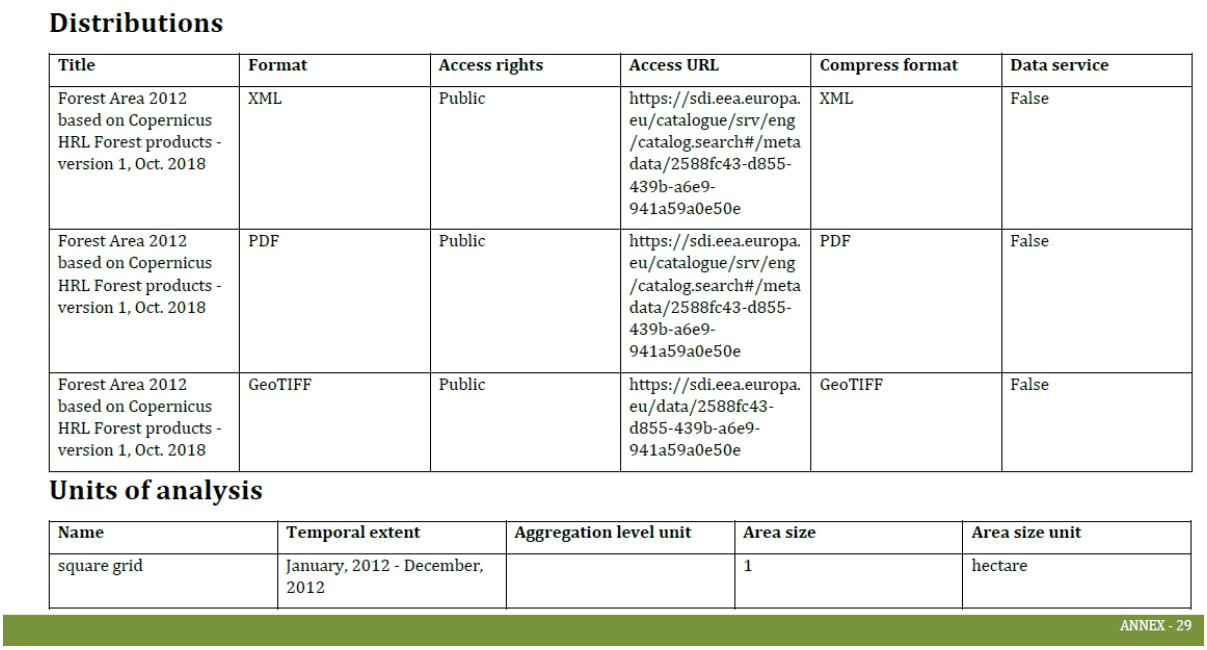
Was generated by:

Is referenced by:

Geographical coverage: Kosovo, Albania, Austria, Belgium, Bulgaria, Bosnia and Herzegovina, Switzerland, Cyprus, Czechia, Germany, Denmark, Spain, Estonia, Finland, France, United Kingdom, Greece, Croatia, Hungary, Ireland, Iceland, Italy, Liechtenstein, Lithuania, Luxembourg, Latvia, North Macedonia, Malta, Montenegro, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Sweden, Turkey

Dataset type: GEOREFERENCED

Figure 30 Example of one dataset included in the submitted deliverables



Distributions

Title	Format	Access rights	Access URL	Compress format	Data service
Forest Area 2012 based on Copernicus HRL Forest products - version 1, Oct. 2018	XML	Public	https://sdi.eea.europa.eu/catalogue/srv/eng/catalog.search#/metadata/2588fc43-d855-439b-a6e9-941a59a0e50e	XML	False
Forest Area 2012 based on Copernicus HRL Forest products - version 1, Oct. 2018	PDF	Public	https://sdi.eea.europa.eu/catalogue/srv/eng/catalog.search#/metadata/2588fc43-d855-439b-a6e9-941a59a0e50e	PDF	False
Forest Area 2012 based on Copernicus HRL Forest products - version 1, Oct. 2018	GeoTIFF	Public	https://sdi.eea.europa.eu/data/2588fc43-d855-439b-a6e9-941a59a0e50e	GeoTIFF	False

Units of analysis

Name	Temporal extent	Aggregation level unit	Area size	Area size unit
square grid	January, 2012 - December, 2012		1	hectare

ANNEX - 29

Figure 31 Content of one dataset included in the submitted deliverables

In addition, the launch of the tool and its API, opens the door to other users and researchers external to the project who wish to use the tool as a unified portal to search and link to datasets related to agriculture. As well as, the possibility of linking the platform with other modules and components of the project architecture, such as the DWH or the AGRICORE interface, which need ARDIT for its correct operation and application within the objectives defined by the project.

5 Governance implementation

ARDIT governance seeks to define the roles and responsibilities of each project partner in relation to the platform throughout the AGRICORE project lifecycle. These responsibilities provide guidance on how to keep the maintenance of the platform or implement future changes. As defined in the project's Grant Agreement document, the consortium members commit themselves to apply for the following roles during the development of the project and for a minimum period of five years after the end of the project. After which, these responsibilities will be transferred to an EU institution or body.

5.1 Technical maintenance

Being AAT the main developers of the platform, it is also responsible for the technical maintenance of the tool, which would include security measures or managing the repositories and any possible error, problem, crash, migration or upgrade related to the platform and its frameworks, libraries and tools. This responsibility will also include all the deployment processes of the platform on production software environments, that make the tool accessible to the public and the open sourcing activities that will make the code publicly available through the AGRICORE repositories.

5.2 Characterisations maintenance

The project partners who participated in the dataset characterisation deliverables (AKD, AUTH, CAAND, IAPAS, STAM, UNIPR and UTP) will initially be in charge of maintaining the characterisations and the content available in the platform. With the release of the platform to the public and the dissemination processes that will help to create a larger community on it, new institutions and entities will be gradually added to collaborate in the process of adding new characterisations, as well as managing the content of the tool.

5.3 Monitoring and support

Monitoring and support include all the tasks to supervise that the tool complies with the objectives set within the project and to provide assistance by reporting possible improvements, changes or errors that may be detected. This task is mainly conducted by IDE who, as project coordinating partners, ensure that the AGRICORE projects objectives are achieved, but also concerns the partners involved in the platform development and characterisation processes.

5.4 Retrospective

Includes all the tasks of reviewing the developed solution with the aim of analysing the possible new requirements, libraries or tools to promote a continuous improvement, if possible, of the ARDIT platform. This task will be mainly conducted by AAT, as the main developers, supported by IDE, as the project coordinators and main supervisors.

6 Conclusions and next steps

The deliverable 'D1.9 - Agricultural Research Data Index Tool' provides a guide to understanding how the ARDIT platform has been built, how it works and how the methodology defined in the project has been applied to characterise agricultural datasets. In addition, the work done on the characterisation deliverables allowed to successfully test the application of the AGRICORE DCAT-AP 2.0 ontology to define common properties in the metadata by homogenising hundreds of datasets of different types. The release of the tool opens a new stage for the growth of its community. Current and future users can now include new datasets and access a complete collection of datasets linked to the websites where they are stored to download some of them. Additionally, the ARDIT API enables the distribution and re-use of the work done by AGRICORE project partners in other European projects or in the daily work of other institutions and researchers.

Although the development of the ARDIT core is now complete, it may still receive some changes and incorporate new functionalities to allow its use connected with other modules of the AGRICORE suite that are being developed in parallel. These connections are necessary to achieve the overall objectives and goals defined for the project. The next steps involving ARDIT include:

- Implementing semantic searching and natural language processing capabilities to the platform.
- Finishing the development of the Local Indexer. This is the ARDIT version with the necessary tools to launch the ETL processes to store datasets in the DWH (so that they can be used during the synthetic population generation process) or in the generation of new datasets from others already stored.
- Organising a workshop with the aim of disseminating the platform and thereby, trying to increase its community of users.
- Following the open-source paradigm of the AGRICORE Project, the ARDIT code will be accessible in due course through a dedicated repository in GitLab.

Finally, it is important to clarify that, like any software development, the designs, tools and functionalities depicted in this document might not constitute the final version of the tool. Until the delivery of the final product of the project (AGRICORE Suite) the platform could undergo some small changes if needs require it.

References

[1] ^ E. Commission, "Open data portals." [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/open-data-portals

[2] ^ Nuclino, "A Guide to Functional Requirements." [Online]. Available: https://www.nuclino.com/articles/functional-requirements

[3] ^ T. R. V. of Wikipedia, "Non-functional requirement." [Online]. Available: https://thereaderwiki.com/en/Non-functional_requirement

[4] ^ I. C. Education, "REST APIs." [Online]. Available: https://www.ibm.com/cloud/learn/rest-apis

[5] ^ MuleSoft, "Microservices vs Monolithic Architecture." [Online]. Available: https://www.mulesoft.com/resources/api/microservices-vs-monolithic

[6] ^ M. Foundation, "An overview of HTTP." [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/HTTP/Overview

[7] ^ M. Foundation, "HTTP Messages." [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/HTTP/Messages

[8] ^ M. Foundation, Working with JSON. [Online]. Available: https://developer.mozilla.org/en-US/docs/Learn/JavaScript/Objects/JSON

[9] ^ auth0, "JSON Web token." [Online]. Available: https://jwt.io

[10] ^ M. P. LTD, "MarvelApp official site." [Online]. Available: https://marvelapp.com

[11] ^ P. O. of the European Union, "EU vocabularies: Authority tables." [Online]. Available: https://op.europa.eu/en/web/eu-vocabularies/authority-tables

[12] ^ P. O. of the European Union, "Authority tables: Currency vocabulary." [Online]. Available: https://op.europa.eu/en/web/eu-vocabularies/concept-scheme/-/resource?uri=http://publications.europa.eu/resource/authority/currency

[13] ^ P. S. Foundation, "Pyhton official site." [Online]. Available: https://www.python.org/downloads

[14] ^ L. Richardson, "Beautiful Soup official site." [Online]. Available: https://www.crummy.com/software/BeautifulSoup

[15] ^ L. Inc., "Liquibase official site." [Online]. Available: https://liquibase.org

[16] ^ P. O. of the European Union, "The official portal for European data." [Online]. Available: https://data.europa.eu/en

[17] ^ S. Economy, Business, and T. P. ministries, "Open data initiative of the Government of Spain." [Online]. Available: https://datos.gob.es/en

[18] ^ U. S. Government, "U.S. Government open data." [Online]. Available: https://data.gov

[19] ^ N. Aeronautics and S. A. (NASA), "A catalog of publicly available NASA datasets." [Online]. Available: https://data.nasa.gov

[20] ^ S. Software, "Swagger official site." [Online]. Available: https://swagger.io

[21] ^ Docker, "Docker official website." [Online]. Available: https://www.docker.com

[22] ^ Docker, "Overview on Docker Compose." [Online]. Available: https://docs.docker.com/compose

[23] ^ VMware, "Spring framework official site." [Online]. Available: https://spring.io

[24] ^ VMware, Overview of Spring Framework. [Online]. Available: https://docs.spring.io/spring-framework/docs/5.0.0.RC2/spring-framework-reference/overview.html

[25] ^ P. G. D. Group, "PostgreSQL database official site." [Online]. Available: https://www.postgresql.org

[26] ^ O. Foundation, OpenLDAP official documentation. [Online]. Available: https://www.openldap.org/doc/admin26/intro.html

[27] ^ O. Foundation, "OpenLDAP official site." [Online]. Available: https://www.openldap.org/software

[28] ^ Google, "Angular framework official site." [Online]. Available: https://angular.io

[29] ^ O. Foundation, "Node.js official site." [Online]. Available: https://nodejs.org/es

[30] ^ F. Networks, "Nginx official site." [Online]. Available: https://www.nginx.com

[31] ^ Bootstrap, "Bootstrap official site." [Online]. Available: https://getbootstrap.com

[32] ^ N. Bootstrap, "NG Bootstrap official site." [Online]. Available: https://ng-bootstrap.github.io

[33] ^ Ng-Select, "Angular ng-select - Lightweight all in one UI Select, Multiselect and Autocomplete." [Online]. Available: https://www.npmjs.com/package/@ng-select/ng-select

[34] ^ M. Bromley, "Pagination for Angular." [Online]. Available: https://www.npmjs.com/package/ngx-pagination

[35] ^ Fonticons, "Font awesome official site." [Online]. Available: https://fontawesome.com

Apart from the references mentioned along the document, the current deliverable took into account the following inputs:

- AGRICORE Proposal: project proposes a novel tool for improving the current capacity to model policies dealing with agriculture by taking advantage of the latest progresses in modelling approaches and ICT.
- AGRICORE Grant Agreement ANNEX 1 Part A and B, Research and Innovation action, Number-816078: Official Grant Agreement of the AGRICORE project, which defined the terms and conditions of the project, as well as the main requirements of the project.

Deliverable Number	Deliverable Title	Lead beneficiary	Type	Dissemination Level	Due date	Reason
D1.1	Standardised Methodology and Set of Ontologies for the Characterisation of the Data Source	UNIPR	Report	Public	M11	Understand the AGRICORE DCAT-AP data model and ontology for the characterisation of data sources
D2.1	Data Warehouse (DWH) for agriculture policy impact assessment data management	AAT	Report	Public	M24	Understand the connection between ARDIT and the DWH to store data sources potentially useful in the simulation processes
D6.1	AGRICORE Architecture and Interfaces	IDE	Report	Public	M24	Understand the connection between different modules of the project and the AGRICORE interface
D6.6	Software Quality Assurance measures for AGRICORE	AAT	Report	Public	M15	Development and testing guidelines for software quality assurance

ANNEX I: End-users manual

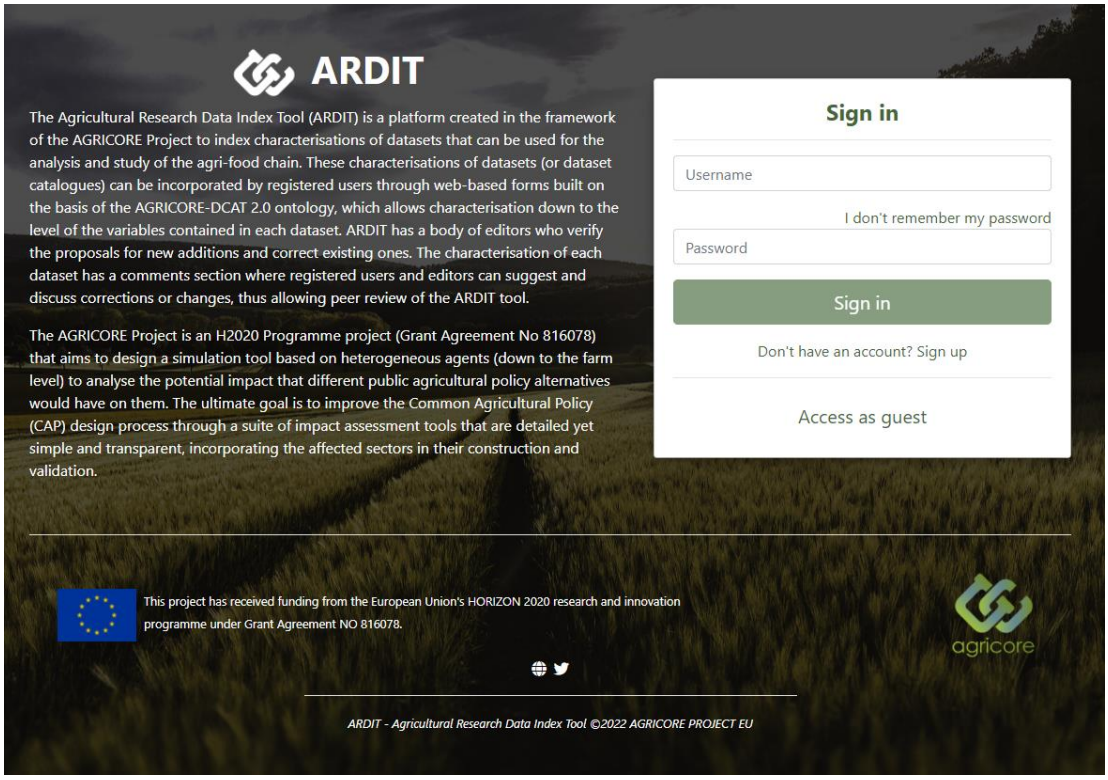
Version History

Version	Description	Organisation	Date
0.1	User Manual Template	IDE	15/11/2021
0.2	Inclusion content regarding the characterisation of socio-economic datasets	UNIPR	17/11/2021
0.3	Inclusion content regarding the characterisation of georeferenced datasets	STAM	17/11/2021
0.4	Review and updated platform screenshots	AAT	24/03/2022

A1.1 Description of ARDIT Sections

A1.1.1 Main page/login

To access the ARDIT tool click on this link: https://ardit.agricore-project.eu, and insert your credentials.



A1.1.2 Dataset

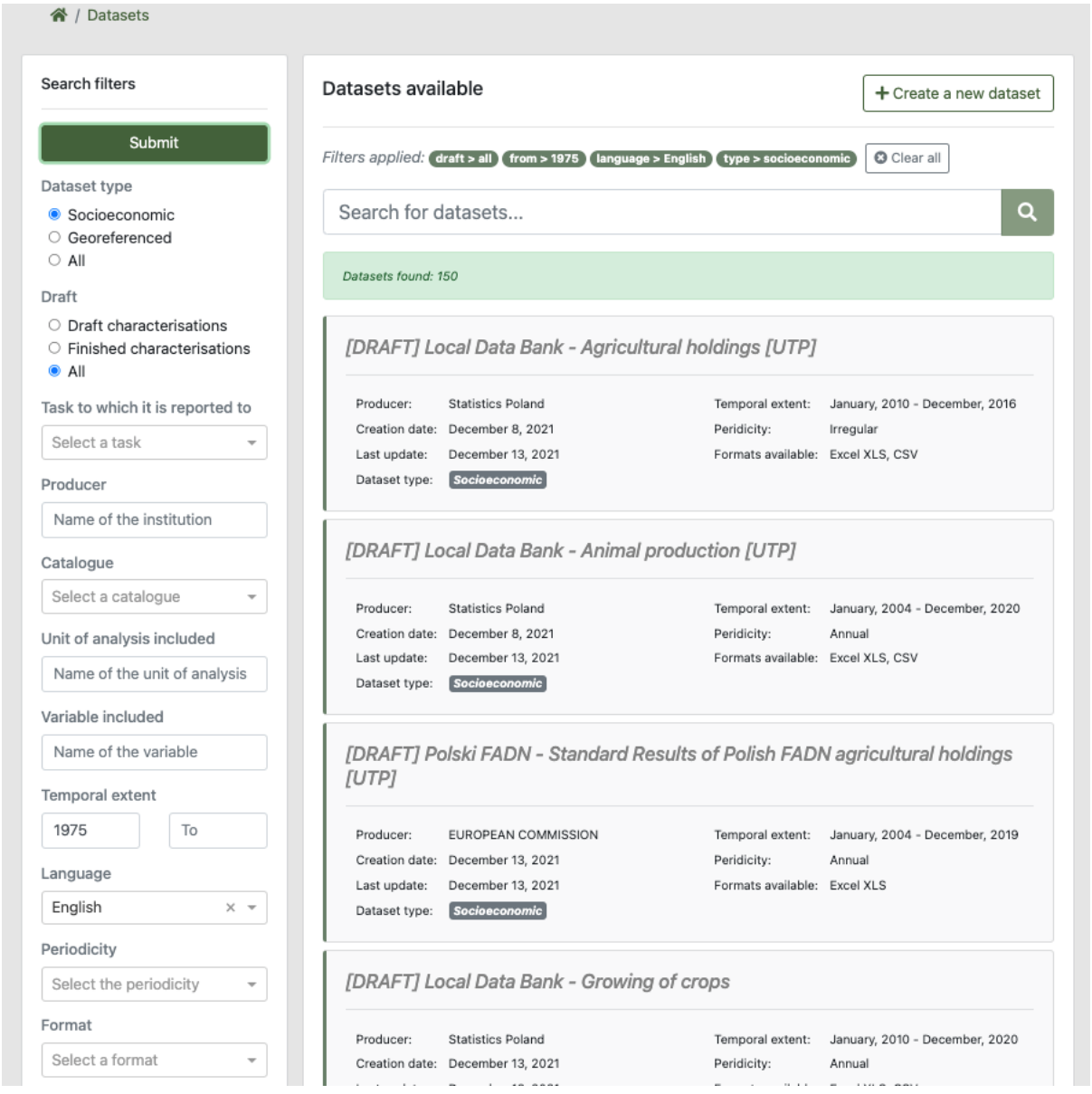
After logging in, the landing page is the Dataset page, where there are several search filters:

- **Dataset type:** Socioeconomic or Geo-referenced
- **Producer:** the organization, the statistical center who collected the data.

- **Catalogue**
- **Variables included:** When needed specify the variables covered by the dataset (e.g., in FAOSTAT, the dataset "Production" and the sub-dataset "Crops" covers the Area harvested, Yield, and Production quantity for each crop. The latter 3 could be listed here instead of creating 3 further sub-sub-sections). It could also be the type of nutrients, rainfalls, temperature etc. It may also include the units of measurement of the variables
- **Temporary extent**
- **Language**
- **Periodicity**
- **Format**

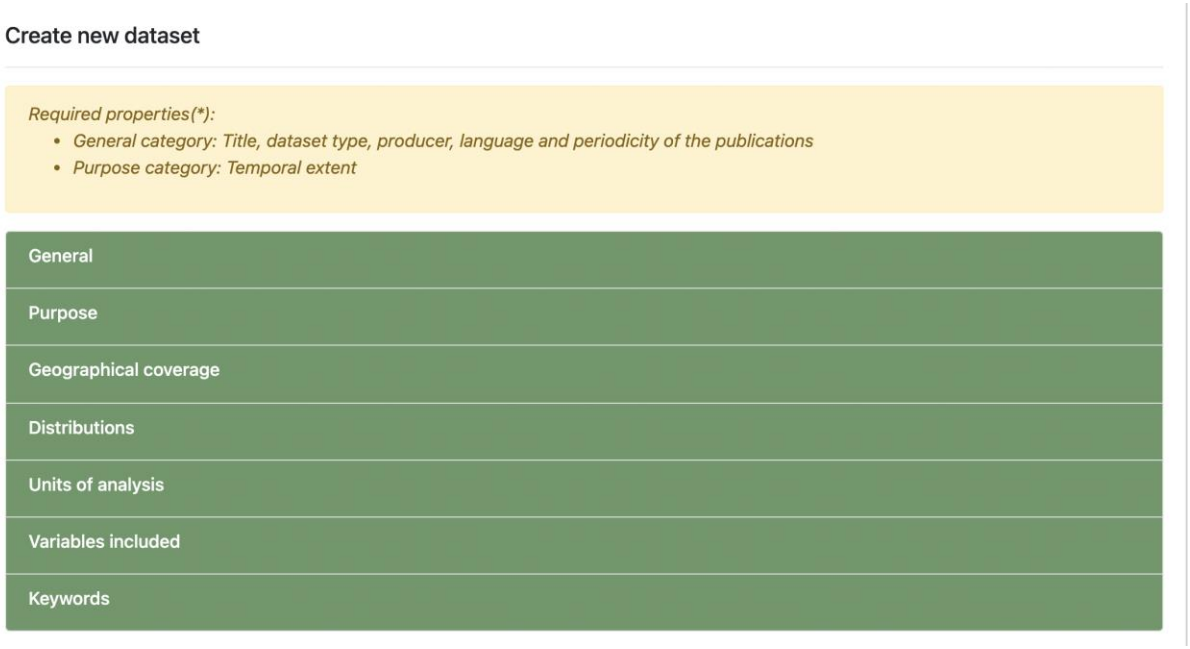
Searches can also be performed based on geographical coverage / aspects, namely: Continent, Country or region (**NUTS1, NUTS2 and NUTS3**).

In addition, datasets can be searched by name.

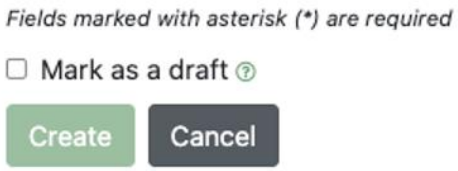


A1.1.3 Create a new dataset

To create a new dataset, the user is requested to fill in the following tabs:



While doing the data entry, the dataset can be saved as a draft.



Once all the mandatory fields have been filled in with the available data, the dataset can be saved and the draft flag removed.

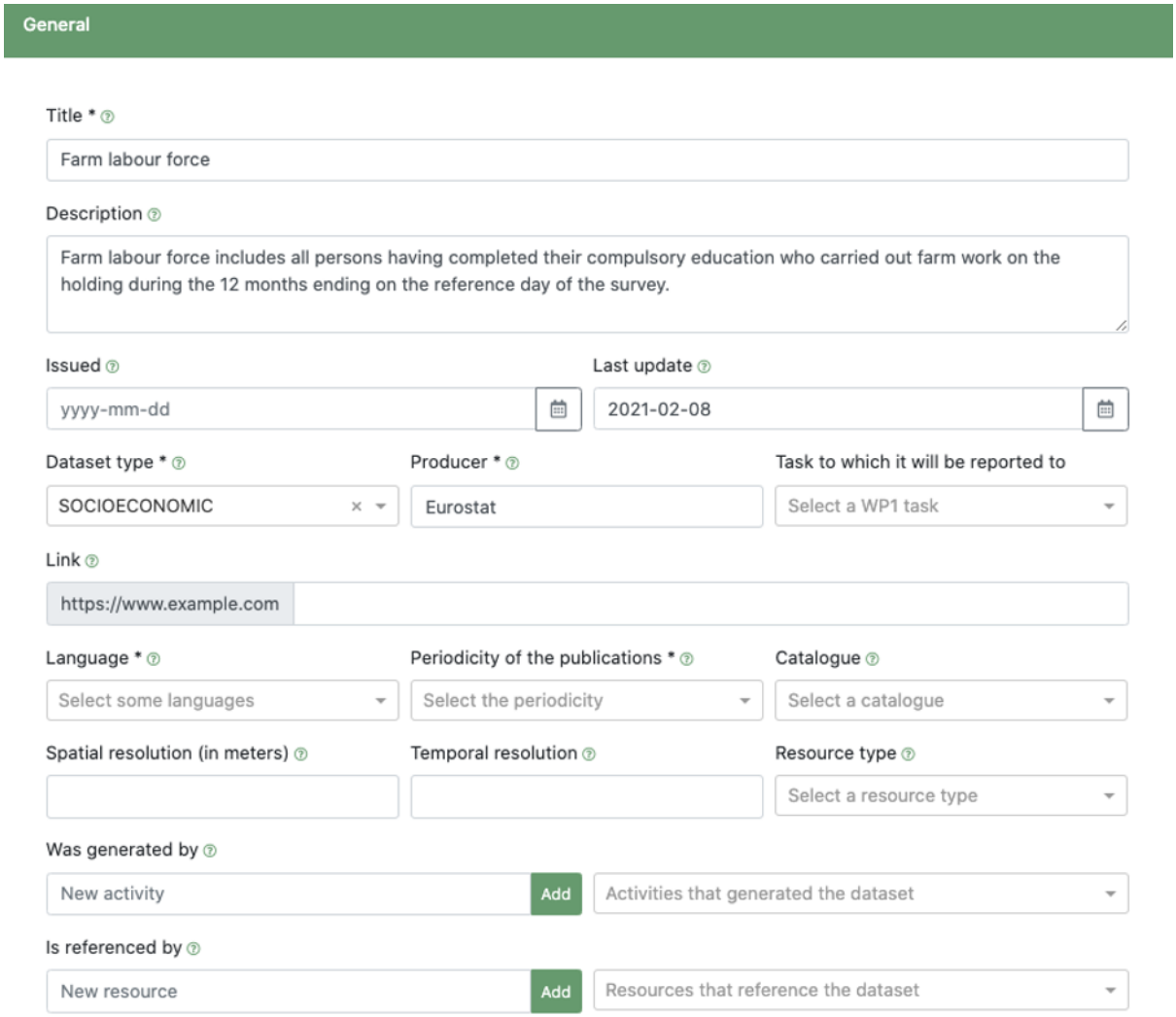
Every item with the asterisk is required.

If the information is not available and not required, it's possible to leave a field blank.

The question mark next to each field label is meant to explain the information required in that specific field.

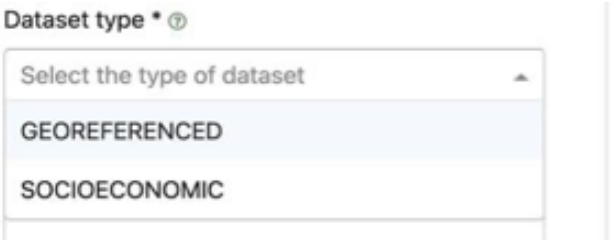
A1.1.4 General information about the dataset

This tab contains the main general information about the dataset.

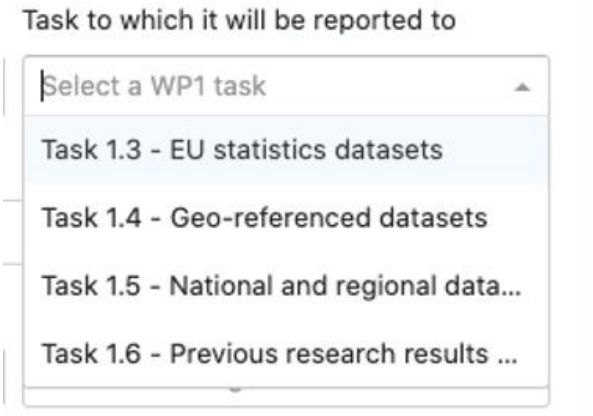


The required fields are:

- **Title**: the actual name of the table. It's a required free text property. We can usually find the date or time to which the specific dataset relates.
- **Description:** a brief description of the dataset, its purpose or the data collected; this can be found on the website of the data producer (es. Eurostat).
- **Issued**: date of formal issuance or publication of the specific dataset.
- **Last update**: most recent date on which the dataset was changed, updated or modified.
- **Dataset type**: the dataset can be of 2 types:
    - GEO-REFERENCED: to indicate a geo-referenced dataset.
    - SOCIOECONOMIC: for statistical or economic datasets.

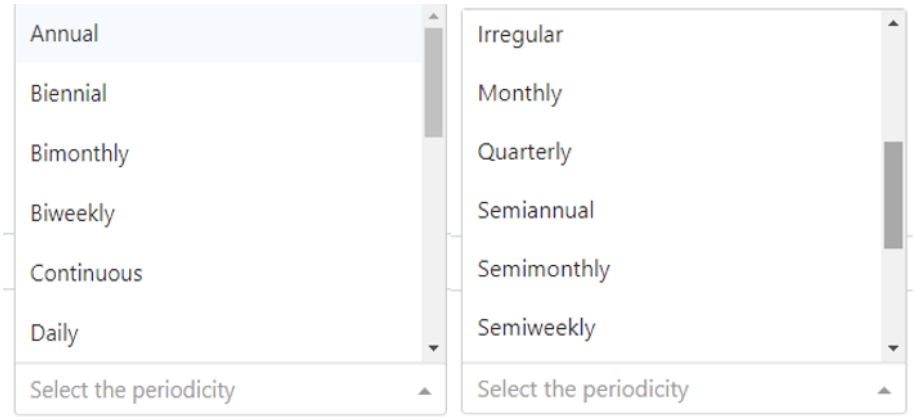


- **Producer**: the institution which publishes or maintains the dataset. It's a required free text property.
- **The Task to which it will be reported (included only for AGRICORE characterisation processes, to be eliminated in the final ARDIT version)**:
    - **Task 1.3:** EU statistics datasets: The goal of this task is to produce a characterisation of the EU-wide statistics datasets most useful for the purpose of assessing the impact of agricultural policies.
    - **Task 1.4:** Geo-referenced datasets: Description: This task covers the analysis of the geo-referenced datasets available in the EU at a global and local level that can provide useful information for the impact assessment of agricultural policies.
    - **Task 1.5**: National and regional datasets: In general, but especially in the case of micro-level policy analysis at the farm level, the information available at national and regional level can be key for conducting proper analysis, as they may provide more information (e.g., higher detailed information at lowers spatial scale) than the EU-wide data sources.
    - **Task 1.6**: Previous research results as information sources: The goal of this task is to ensure a proper exploitation of previous research efforts in the agricultural policy impact assessment field. This includes the identification and characterization of suited previous project results.

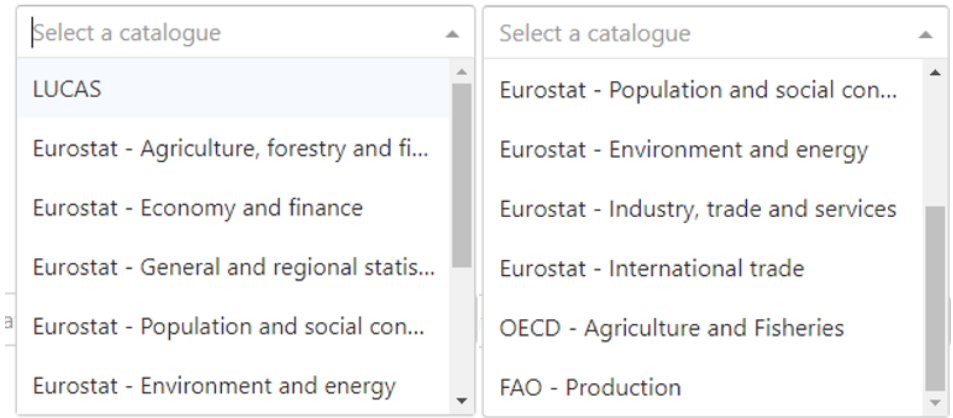


- **Link**: enter here the hyperlink of the referenced dataset.
- **Language**: the original language in which the data and metadata are available. In the case of a national dataset in the local language, the characterization will be done in English and the original name of the table will be reported between brackets, next to the original title

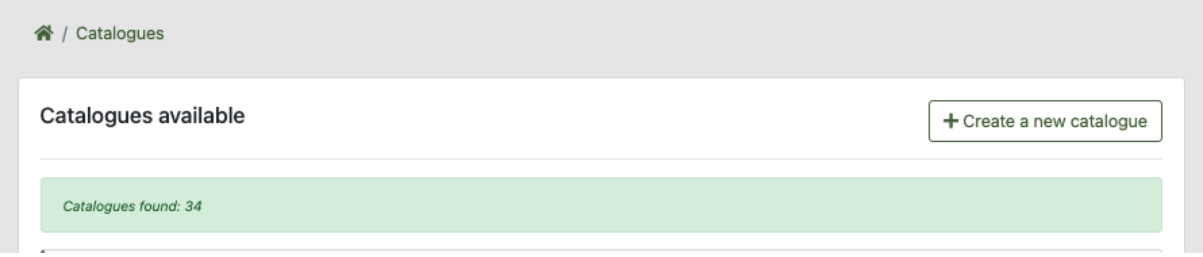
- **Periodicity of publication**: frequency with which each data is updated. We can find different periods of time, but it also includes the item "*irregular*", which is used when there is no regularity in the publication of data.



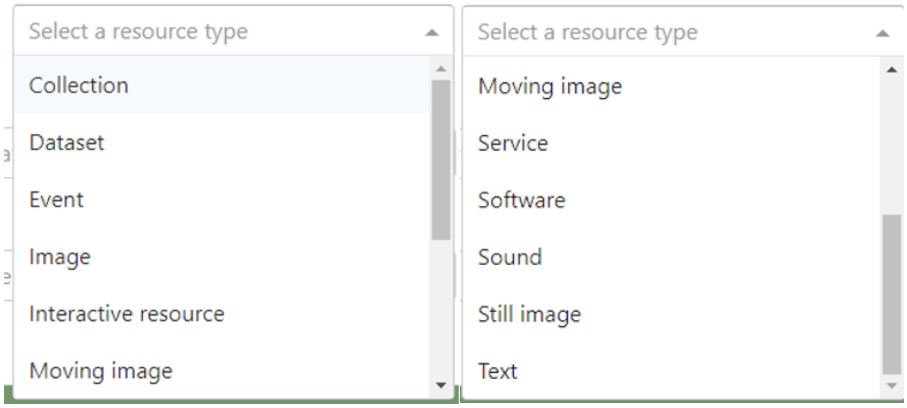
- **Catalogue**: this field depicts the catalogue to which the dataset belongs to. Examples of catalogues are:
    - **LUCAS**: survey (across all Member States) to gather information on land cover and land use. Estimates of the area occupied by different land use or land cover types are computed based on observations taken at more than 250,000 sample points throughout the EU rather than mapping the entire area under investigation.
    - **EUROSTAT**: is the statistical office of the UE, responsible for the publication of statistics and quality indicators to allow comparisons between Countries and Regions.



If the catalogue is not yet present in the drop-down list, it can be added by clicking on "**Create a new catalogue**", (**https://ardit.agricore-project.eu/catalogues**) and by filling the required fields (*See "Create a new catalogue" section*):



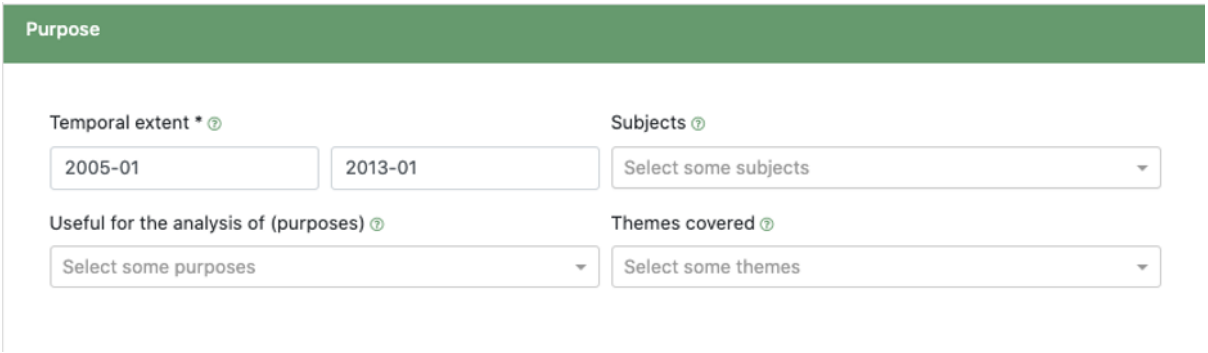
- **Spatial resolution**: only for georeferenced datasets: minimum spatial separation resolvable in a dataset, measured in meters.
- **Temporal resolution**: it is intended to provide a summary indication of the temporal resolution of the data distribution as a single value. More complex descriptions of various aspects of temporal precision, accuracy, resolution and other statistics can be provided using the Data Quality Vocabulary [VOCAB-DQV] (https://www.w3.org/TR/vocab-dcat-2/#Property:dataset_temporal_resolution).
- **Resource type**: this field depicts the nature or genre of the resource; the table below lists the possible values.



- **Was generated by**: this field is meant to define the activities that generated or provided the business context for the creation of the dataset. An activity is something that occurs over a period of time and acts upon or with entities; it may include consuming, processing, transforming, modifying, relocating, using, or generating entities. The activity associated with the generation of a dataset will typically be an initiative, project, mission, survey, ongoing activity ("business as usual") etc.
- **Is referenced by:** if the dataset is cited in the literature, or on other datasets or used by some other related resources, such as publications, etc.

A1.1.5 Purpose of the dataset

This session depicts what kind of analysis the dataset was created for.



The purpose property has been added in the AGRICORE ontology and it identifies the purpose of the dataset, i.e., for what kind of analysis that dataset has been created: there is an internal vocabulary of terms that will be enriched over time.

- **Temporal extent**: primarily refers to the data collection period. It indicates the period of time to which the specific data refers.
- **Subjects**: this property has been added in the AGRICORE ontology and it should identify the object of the dataset, i.e., the type of data that has been analysed and represented: the chosen vocabulary is that of the Digital Europa Thesaurus, which is not an "internal vocabulary", so it's not possible to add more subjects to the list. Users can only choose the ones already inserted.
- **Useful for the analysis of (purposes)**: purposes of the dataset and the data. An initial list of purposes is stored in an internal vocabulary, and it is possible to add new values whenever required. The list is constantly evolving.
- **Themes covered**: the theme derives from DCAT-AP and represents the themes covered by the dataset. The chosen vocabulary is the Data theme. Each dataset can have one or more themes. The values can be the URL of the Eurovoc controlled vocabulary or Eionet Data Dictionary.

A1.1.6 Geographical coverage

This session depicts the geographic features of a dataset. It's a spatial property. It's a list of spatial regions or named places. It can be represented using a controlled vocabulary or with geographic coordinates.

Only one level of spatial coverage is needed. If users indicate 'Europe' as continental coverage, there is no need to report each European country; the only important thing is to make sure that the dataset contains data for all the countries in that continent (as per regions of each country).

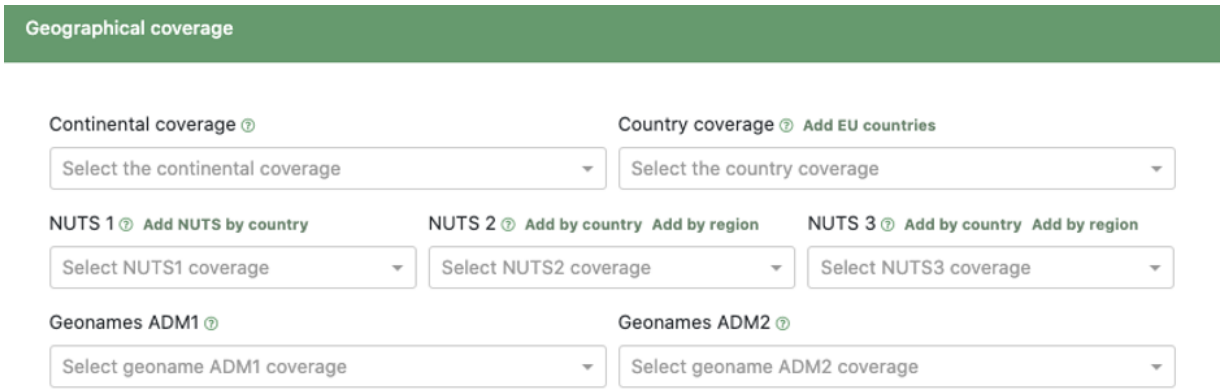
- **Continental coverage**: list of continents. Users can add all the continents to which the specific dataset refers to.
- **Country coverage**: in addition to the "*continental coverage*", users can also enter the specific countries.

In the table that contains the data divided by States, those that do not report any value should not be entered; however, those that report even a single value (of a single year) or if the reported *value is zero*, must be entered in the field "*Country coverage*".

- The NUTS classification (Nomenclature of territorial units for statistics) is a hierarchical system for dividing the economic territory of the EU and the UK. The current NUTS 2021 classification is valid from 1 January 2021 and lists 92 regions at NUTS 1, 242 regions at NUTS 2 and 1166 regions at NUTS 3 level. (Source: Eurostat). For more details see the definition below.

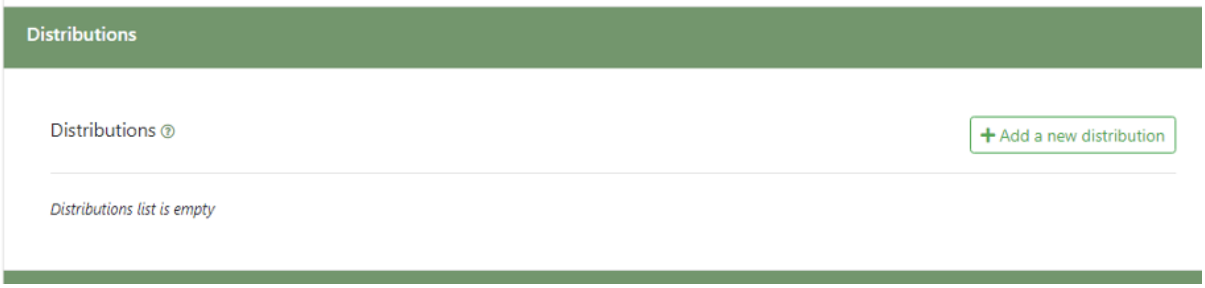
		upper limit	**minimum threshold**
NUTS1	major socio-economic regions	7 000 000	3 000 000
NUTS2	basic regions for the application of regional policies	3 000 000	800 000
NUTS3	small regions for specific diagnoses	800 000	150 000

- The GeoNames geographical database covers all countries and contains over eleven million placenames that are available for download free of charge.
    - **Geonames ADM1**: first-order administrative division; a primary administrative division of a country such as United States (ADM1 values: https://www.geonames.org/advanced-search.html?q=ADM1&country=&featureClass=&continentCode=EU).
    - **Geonames ADM2**: second-order administrative division; a subdivision of a first-order administrative division (ADM2 values: https://www.geonames.org/advanced-search.html?q=ADM2&country=&featureClass=&continentCode=EU).



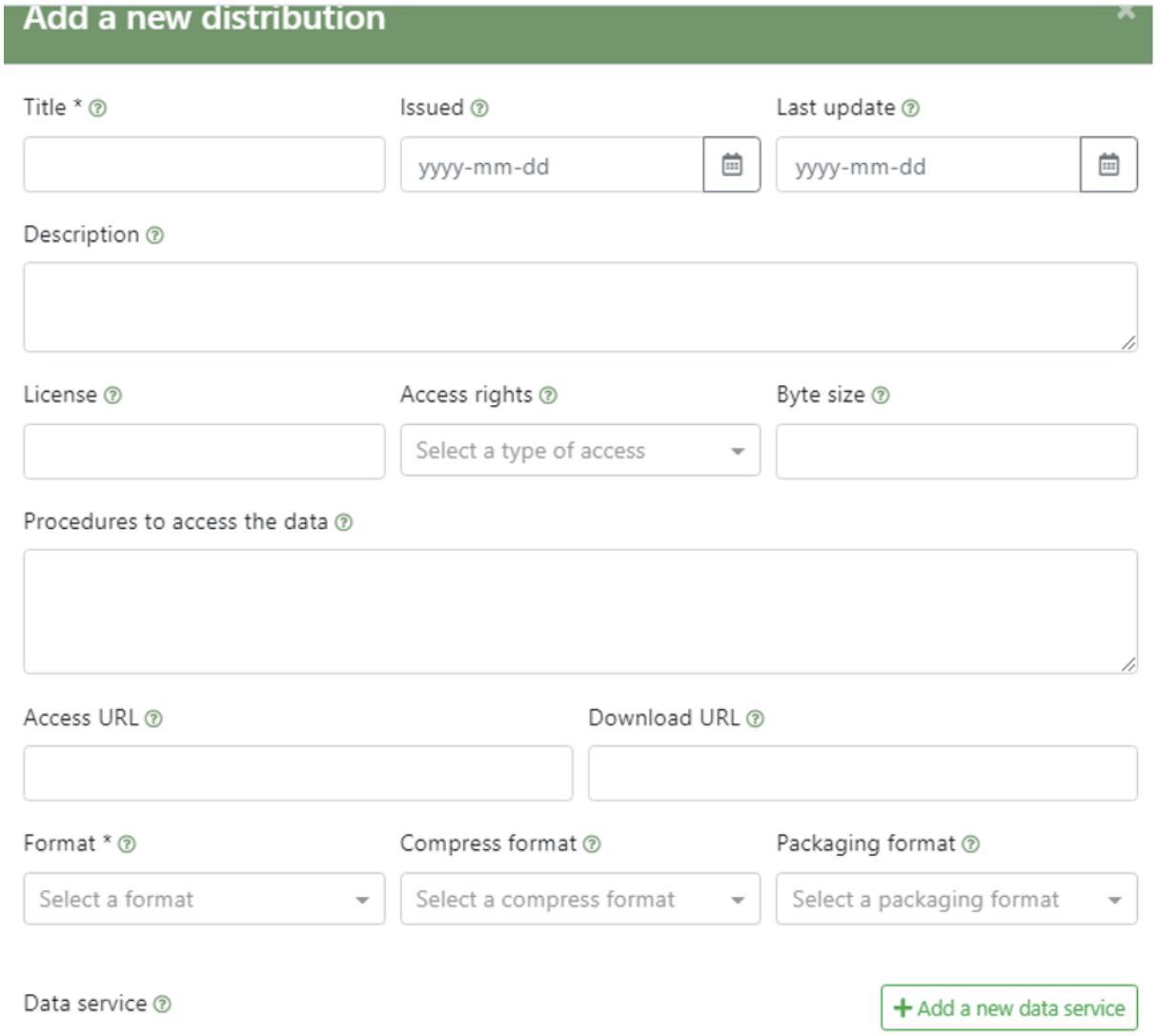
A1.1.7 Distributions

The distribution of a dataset is the format in which the data can be retrieved. A dataset might be available in multiple formats, such as spreadsheet, pdf, CSV, etc.

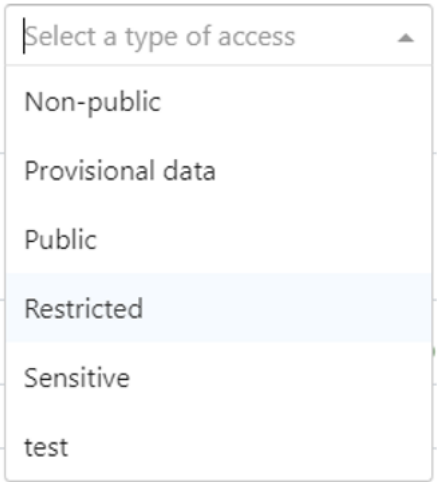


A1.1.7.1 Add a new distribution

To add a new distribution, the following fields must be filled:



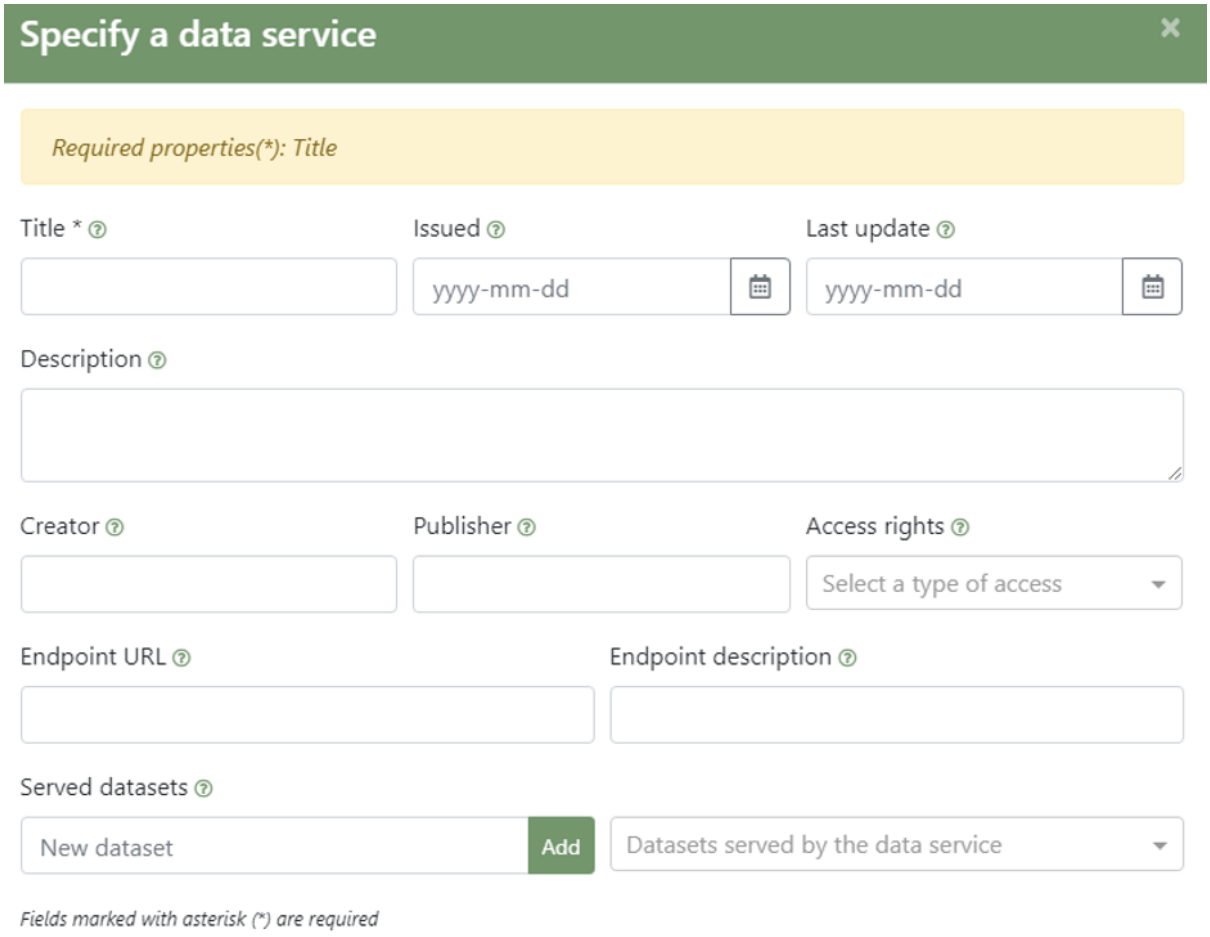
- **Title**: name given to the distributions
- **Issued**: date of formal issuance or publication of the distribution
- **Last update**: most recent date on which the distribution was changed, updated or modified
- **Description**: a free-text account of the distribution
- **License**: a legal document under which the distribution is made available
- **Access rights**: a declaration of the rights that concerns how the distribution is accessed. Publicly available / Access request required *(i.e., Publicly available, publicly available (Registration Required), Publicly available (API or download procedure possibly available))*. If an access request is required, please describe the main characteristic of the authorisation access procedure (use a link to the information page of the actual dataset/link to the provider).



- **Byte size**: the size of a distribution is generally measured in bytes. A byte represents the unit of digital information that most commonly consists of eight bits; it is a utility class that makes byte size representation in code easier.
- **Procedures to access the data**: this field depicts the guidelines for accessing private datasets. Manually created within the project (we can reuse existing documents) detailing how we gather access to the specific data source. Only applicable to non-fully public datasets. It's a free text field which may contain external documents or hyperlinks.
- **Access URL**: in this field the URL of the resource of the referenced dataset (that gives access to a distribution of the dataset) must be entered
- **Download URL**: the URL of the downloadable file in a given format
- **Format**: this field depicts the file format of the distribution. xls, csv, html, pc axis, spss, tsv, (pdf only if it is the only format available). Only one format Is selectable, hence the creation of a distribution for each format type is needed.
- **Compress format**: the compression format of the distribution in which the data is contained, if the distribution is in a compressed form
- **Packaging format**: the package format of the distribution in which one or more data files are grouped together
- **Data service**: a data service that gives access to the distribution of the dataset

A1.1.7.2 Add a new data service

To add a new data service, the following fields have to be filled in:



- **Title**: name given to the data service
- **Issued**: date of formal issuance or publication of the data service
- **Last update**: most recent date on which the data service was changed, updated or modified
- **Description**: a free-text account of the distribution
- **Creator**: the entity responsible for producing the data service
- **Publisher**: the entity responsible for making the data service available
- **Access rights**: a declaration on the rights that concerns how the data service is accessed (we can find the same different types of "access" that we saw in the distribution)
- **Endpoint URL**: a link for the root location or primary endpoint of the service. It is a web resolvable IRI
- **Endpoint description**: a link to the description of the services available via endpoints, including their operations and parameters
- **Served dataset**: a collection of datasets that this data service can distribute. It could be filled in with links to characterized dataset ARDIT

A1.1.8 Unit of analysis

The **Unit of Analysis (UA)** is the object which the variables are referred to.

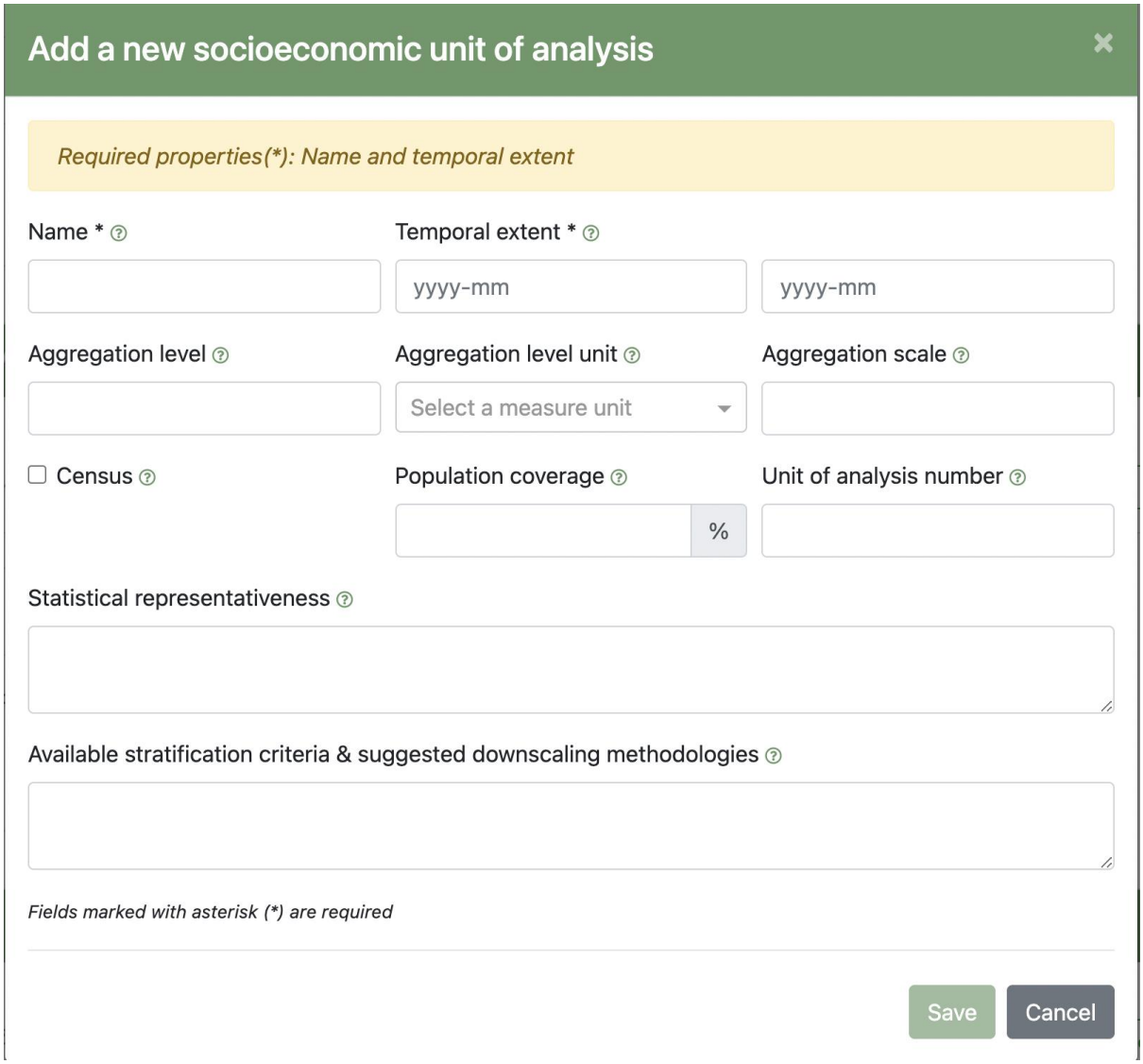
Every dataset has at least one UA, but in one dataset there could be more than one.

The UA is generally something spatially defined, such as the holdings, a region or a country, but it can also be something different such as a commodity. Here below some examples of UA:

- AQUASTAT: COUNTRY, the actual list of all the countries therein, most likely the 194 FAO Members)
- EUROSTAT - Farm Labour Force (https://ec.europa.eu/eurostat/databrowser/view/tag00020/default/table?lang=en). The Unit of Analysis of this dataset is the "Agricultural Holdings" (the farm).
- EUROSTAT - Agricultural Prices and Price Indices: Unit of analysis of a price could be the country, like in this case where the UA is the COUNTRIES (list of the EU countries). This dataset has selling and purchase prices, but for a series of different products.
- EUROSTAT - Crop production in EU standard humidity: COUNTRY LIST THEREIN (exceeding the EU Member States given it considers Switzerland, Albania, Montenegro, sometimes Turkey etc.) and COUNTRY GROUPINGS (EU, EU28, EU_2020)
- EUROSTAT - Economic accounts for agriculture - values at current prices COUNTRY LIST therein and COUNTRIES GROUPINGS (i. e., EU 15, EU25, EU27, EU28, EU27 (post-Brexit), Euro Area 11, Euro Area 12, Euro Area 16, Euro Area 19)
- EUROSTAT - Economic Accounts for Agriculture by NUTS 2 Regions: NUTS0, NUTS1, NUTS2 codes and labels
- OECD-FAO Agricultural Outlook https://stats.oecd.org/Index.aspx?DataSetCode=HIGH_AGLINK_2021: The Unit of Analysis of this dataset is the commodity (e.g., the wheat)
- OECD-FAO (AMIS dataset): Policy Measure(s) (in Agriculture and/or Biofuels)
- AgCFSR Climate Forcing Dataset for Agricultural Modelling: grid cells
- European Climate Assessment & Dataset: meteorological stations
- Global Summary of the Day: meteorological stations

A1.1.8.1 Add a new Unit of Analysis

To add a new unit of analysis, the following fields have to be filled in (in this screenshot there's a socio-economic unit of analysis, which is equal to georeferenced one; it depends on the dataset type chosen above):



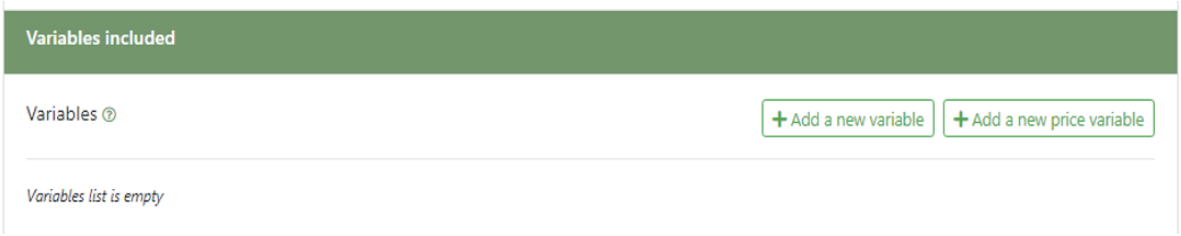
- **Name:** this field indicates the name of the unit of analysis.
- **Temporal extent**: this field depicts the time frame covered by the unit of analysis; it could have a different value from one of the datasets.
- **Aggregation level / unit / scale**: the aggregation level (Aggregation Level, Aggregation Level Unit and Aggregation Scale) is both a unit of analysis and variables property. Only if any of the variables have different values, then the analysis unit will have different values too.
- **Census**: it indicates whether the dataset is a census
- **Population coverage**: when Census is not true, add 'Population Coverage' to indicate (when known - Non-mandatory) the proportion of Units of Analysis sampled with the size of the total population of Units of Analysis. Could be either a float between 0 and 1 or a percentage between 0 and 100.
- **Statistically representative**: this is a free-text property that indicates, when known, information on how the sample was built or which "Unit of Analysis" are included. It is a value between 0 and 1 that refers to what percentage of the total real population of units of analysis is represented by the sample used to generate the dataset. A value of 1 would indicate 100%

(census); a value of 0.2 would indicate that the sample represents 20% of the total actual units of analysis. It is a non-mandatory field and it is inherited top-down.

- **Unit of analysis number:** it should be included when the exact number of analysis units included in the sample is known. It represents the totality of the sample not in percentage terms, e.g.: a sample of 150, the 2% is removed; the unit of analysis number will be the 98% of 150.
- **Downscaling methodology suggestions**: in this field the downscaling strategy, if any, should be described, for example, to move from aggregated data by country to data disaggregated by region.

A1.1.9 Variables included

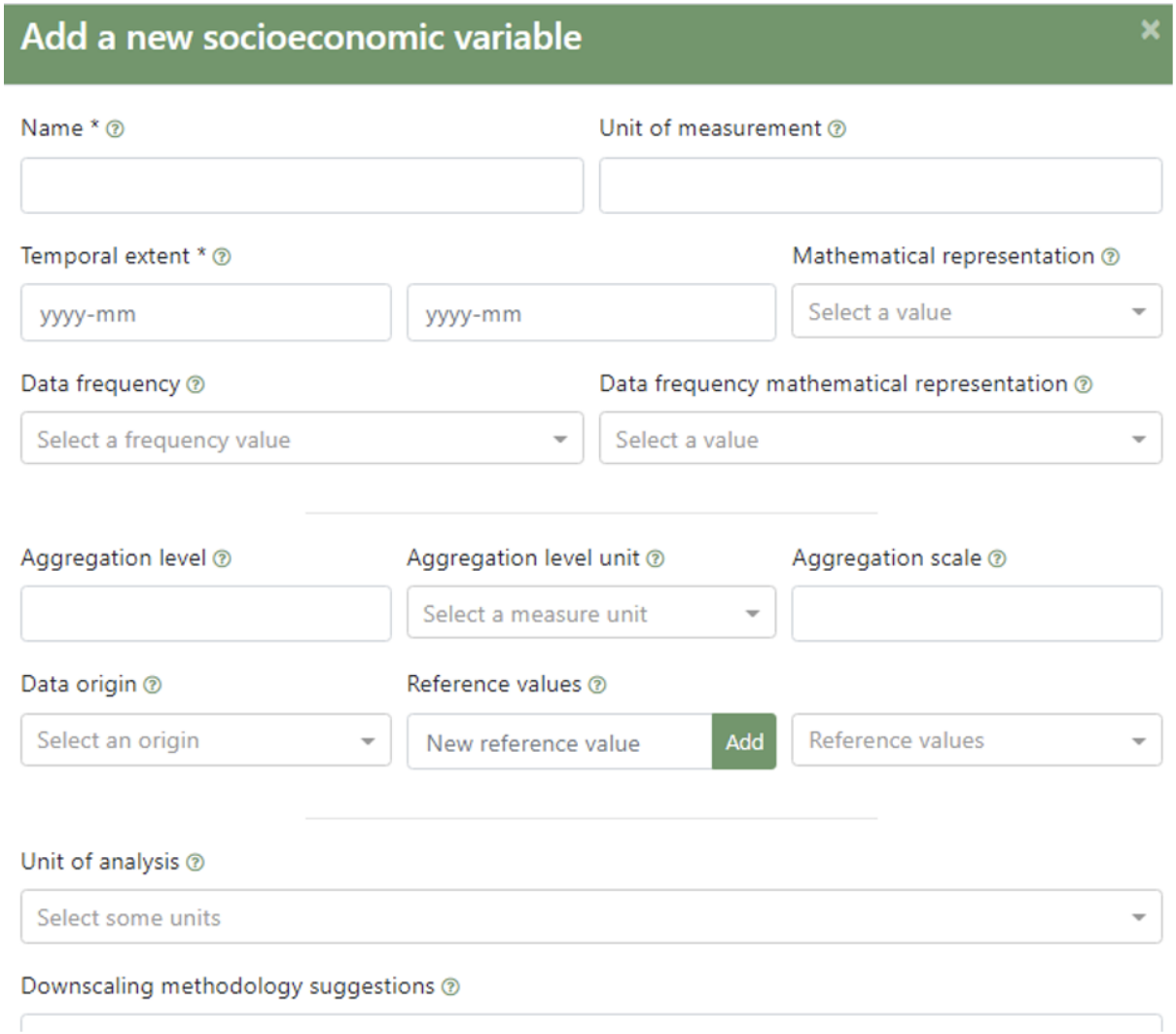
Once the Analysis Unit has been created, the variable related to it must be characterized by filling the fields below. The variables, depending on the dataset type chosen, can be socioeconomic or georeferenced. In the case of a socioeconomic dataset, variables can also be price-related.



A1.1.9.1 Socioeconomic or georeferenced variable

- **Name**: name of the variable which identifies it.
- **Unit of measurement**: it's the units in which the variable is measured (Hectare, Thousands, Person, Day, Litre). It is referred to the variable and not the Unit of Analysis
- **Temporal extent**: time frame covered by the data, e. g., From 2008-01-01T11:45:30 to 2008-12-31T09:10:00
- **Mathematical representation:** indicates the way the variable is represented: average, instant value, etc.
- **Data frequency**: this field represents the frequency with which data are collected and processed during the "temporal extent". e.g.: dataset with two variables for temperature, one is the monthly average (DF = monthly, DF_MR = average), the other is the daily maximum (DF = daily, DF_MR = Max). Annual, quarterly, monthly, weekly, daily, intraday, hourly, 3-hourly (more than one entry if frequency varies within the dataset)
- **Data frequency mathematical representation**: indicates the way data are represented.
- **Aggregation level**: this field specifies the spatial units of the data (e. g., NUTS1, NUTS2, NUTS3 or other administrative regions/units, 50000 (e.g., 1:50000 scale map), 0.25 degrees). The spatial resolution refers to the level of detail of the dataset/ analysis. It shall be expressed as a set of zero to many resolution distances (typically for gridded data and imagery-derived products) or equivalent scales (typically for maps or map-derived products).
- **Aggregation level unit**: variable aggregation level unit of measurement
- **Aggregation scale**: variable aggregation scale

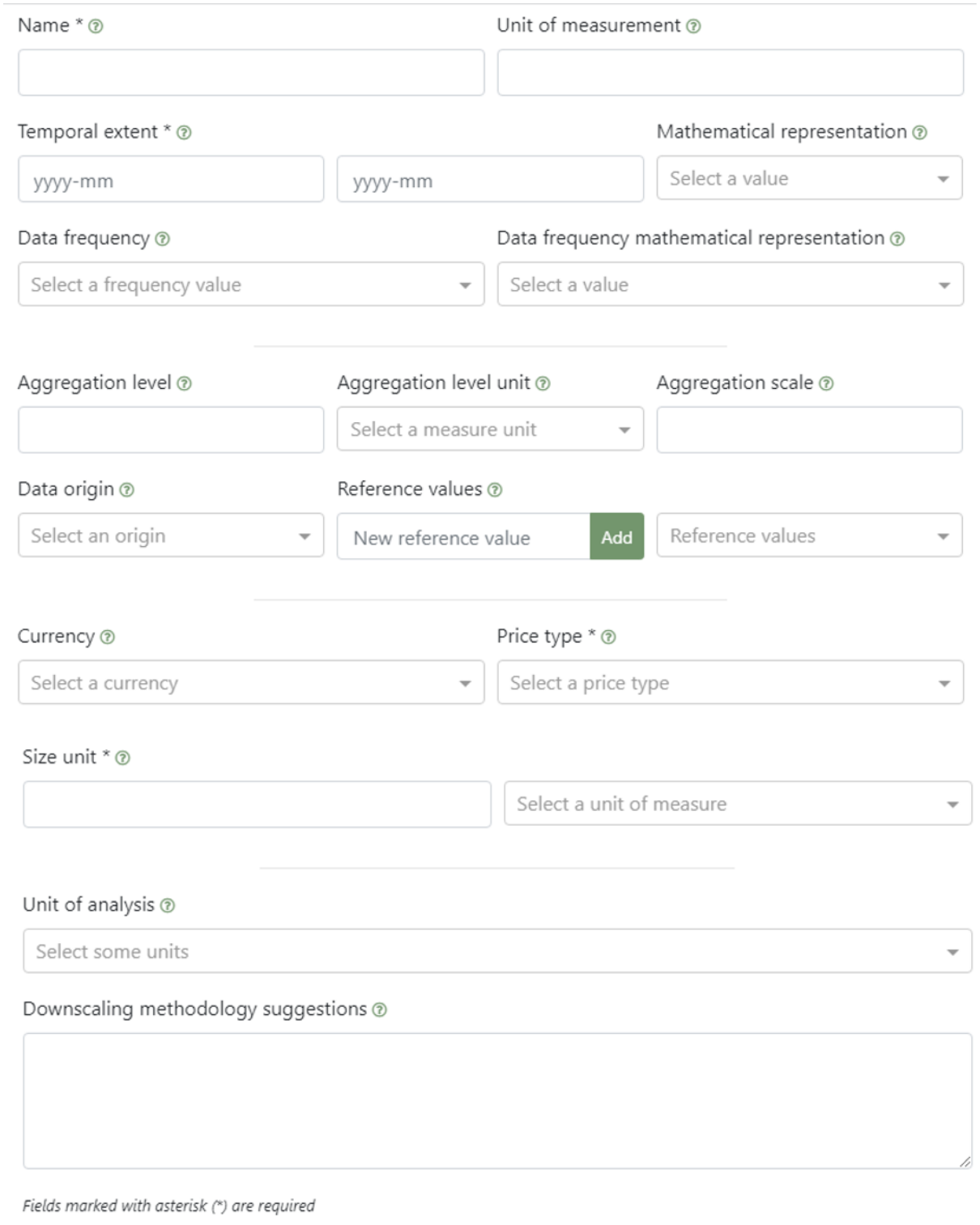
- **Data origin**
- **Reference values**
- **Unit of analysis**: this field represents the unit of analysis the variable is related to. The list of Unit of Analysis, created in the appropriate section, will be displayed in a drop-down menu, from which the user has to select one or more.



A1.1.9.2 Price variable

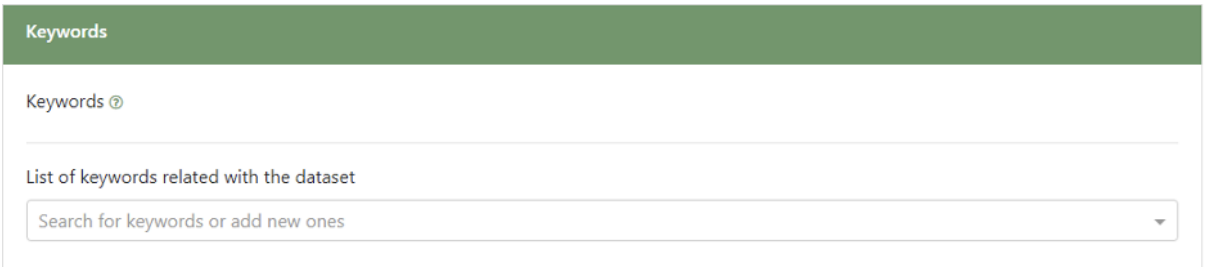
Includes three new properties:

- **Currency**: price variable currency property
- **Price type**: price variable, price type property
- **Size unit**: price variable size unit property; at what quantity the price refers to



A1.1.10 Keywords

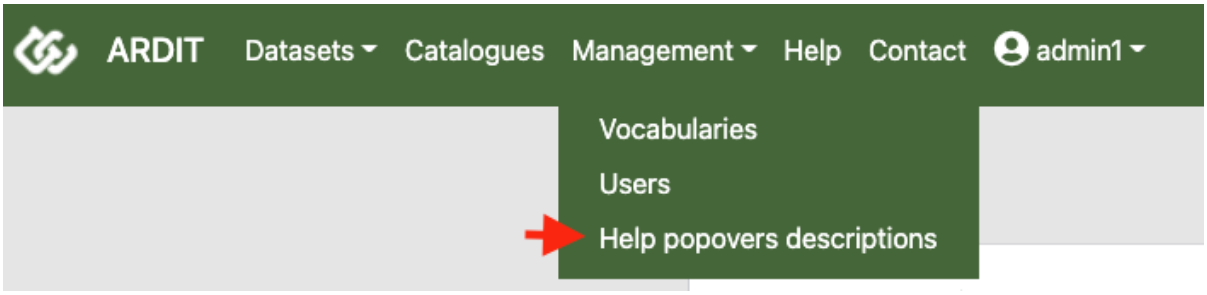
This section is to list the keywords describing or representing the dataset (it is a sort of tagging). It can be filled with existing keywords inserted in other datasets or with new ones.



A1.1.11 Property description

While in Edit mode, fields labels in ARDIT are described by mean of a pop-up:

The contents in these description fields can be updated in the "**Help popovers descriptions**" section in the management menu (only accesible by administrator or maintainer users).



A1.1.12 Catalogues

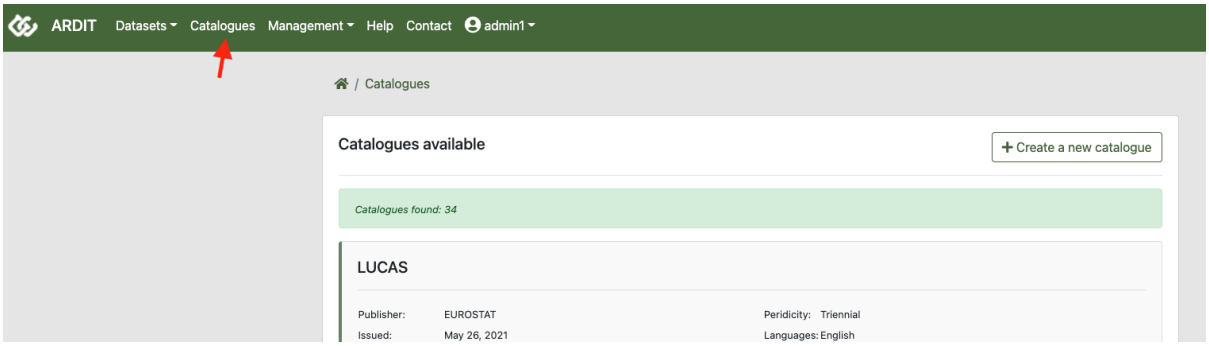
One of the fields required while characterizing a new dataset is the "Catalogue".

When clicking on the "Catalogue" label, a list of predefined catalogues will appear. If the catalogue does not yet exist, it can be created with the "Create a new catalogue" function.

A1.1.12.1 Create a new Catalogue

One of the fields required while characterizing a new dataset is the "Catalogue".

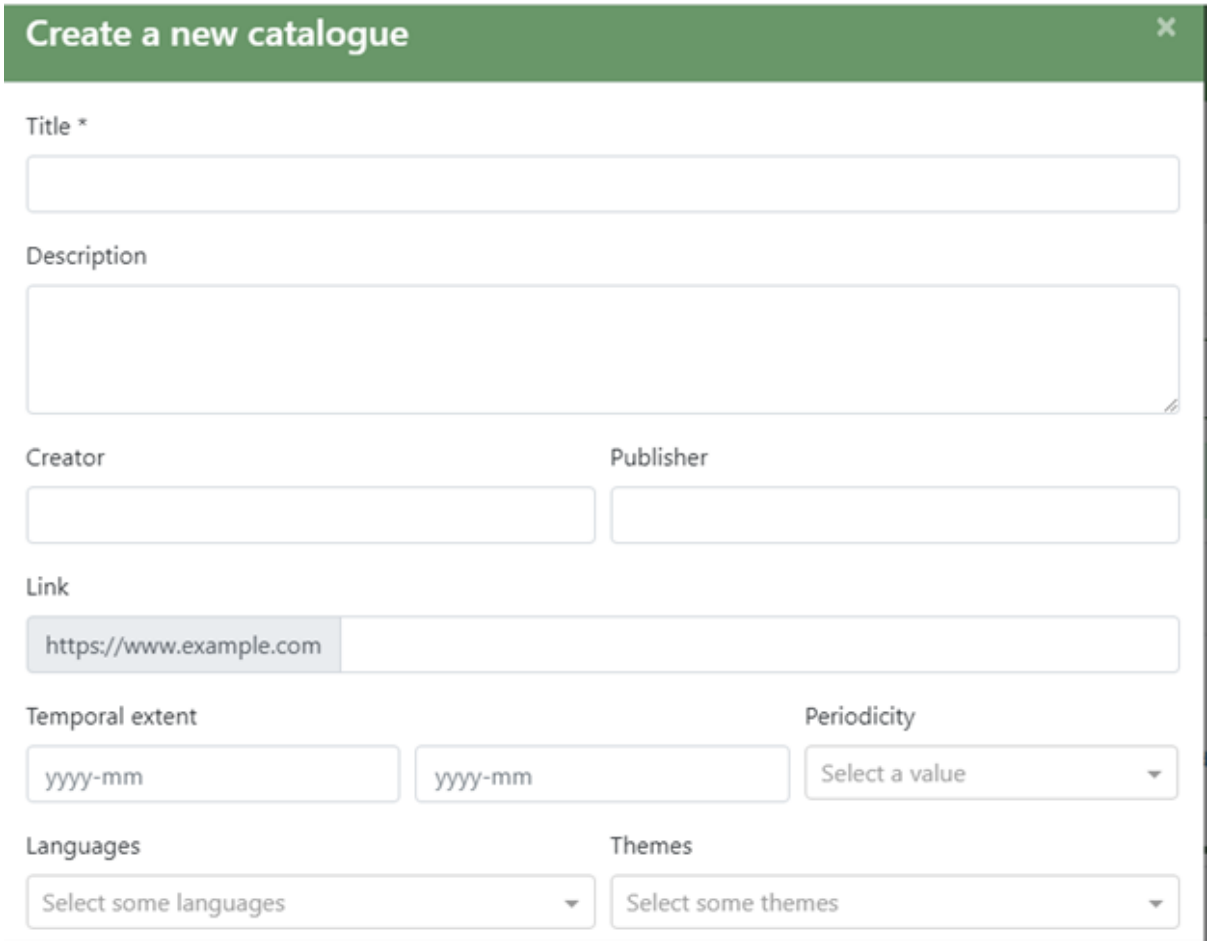
When clicking on the "Catalogue" label, a list of the predefined catalogue will appear. If the catalogue does not yet exist, it can be created with the "Create a new catalogue" function.



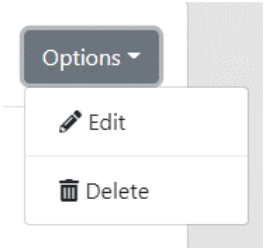
To create a new catalogue, the following fields have to be filled in:

- **Title**: the name of the catalogue must be entered (es: Eurostat, FAO) followed by the name of the catalogue of interest
- **Description**: this is a free-text field describing what is included in the catalogue

- **Creator**: creator of the catalogue and name of the database, if available
- **Publisher**: name of the publisher of the catalogue (Eurostat, OECD)
- **Link**: hyperlink of the catalogue that we can find directly on the website of the database
- **Temporal extent**: the first and latest published catalogue survey dates
- **Periodicity**: period of time in which the catalogue is carried out
- **Languages**: languages in which the catalogue is presented
- **Themes**: themes covered and included in the specific catalogue

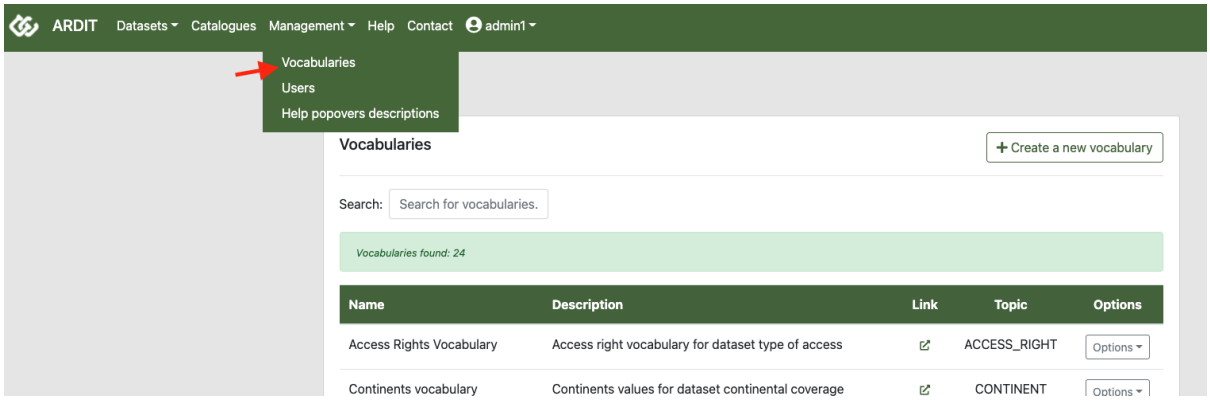


Once it is done, by using the button "*Options*" in a specific catalogue, this can be edited and changed whenever there is a necessity.



A1.1.13 Vocabularies

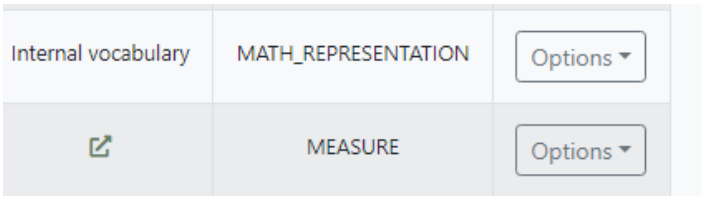
Some properties in ARDIT retrieve their values from specific vocabularies. Vocabularies can be public (and cannot be modified by ARDIT users) or internal. For example, "*Math representation*" can be edited, while "*Measure*" cannot be modified because it comes from a vocabulary / glossary taken from online external publications/organizations. Values in internal vocabulary can be updated and new internal vocabulary can be created by ARDIT users.

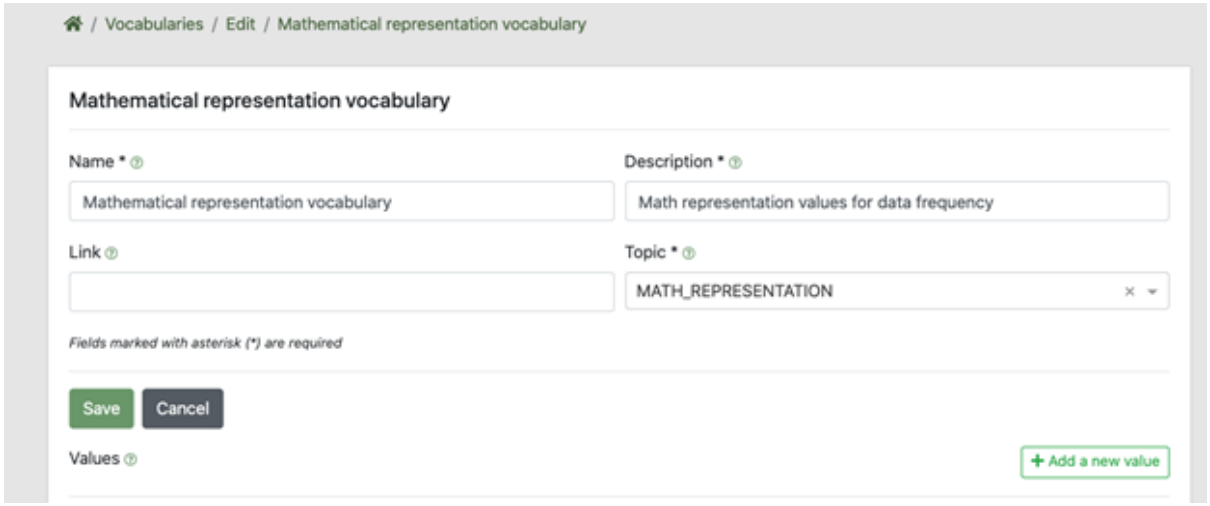


A1.1.13.1 Add a Vocabulary item

The "**Vocabularies**" tab displays the list of existing vocabularies, with respective name, description, link (public or internal) and topic.

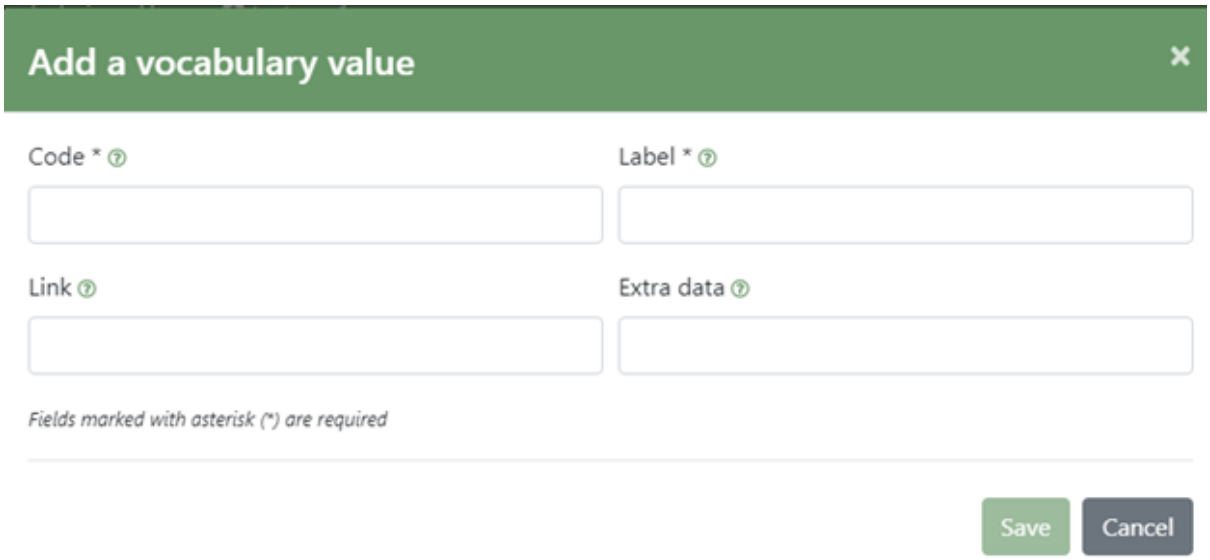
To add new values to an existing internal vocabulary, click on the "**Options**" tab of the vocabulary that has to be updated and click on the "Add a new value" tab.



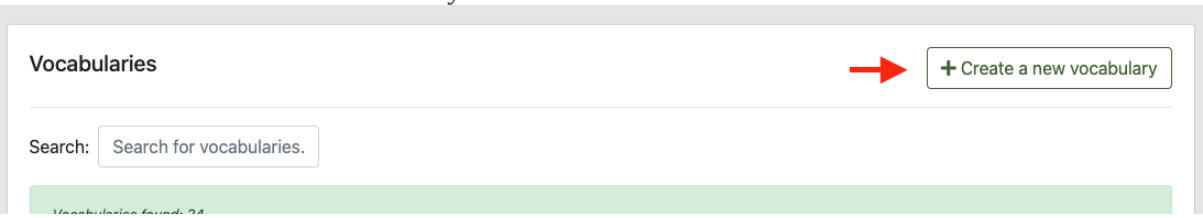


To add a new value in a specific vocabulary, the following fields have to be filled in:

- **Code**: a short identifier of the value. It can be a combination of numbers and letters of 3 or 4 characters. It is a mandatory value however there are no rules on how this code should be represented
- **Label**: in this field the text to be displayed in the vocabulary dropdown menu must be filled in
- **Link**: an external link to the value; if it's an internal vocabulary, it has to be left blank.
- **Extra data**: additional information that differentiates this value from the rest of the values in the vocabulary

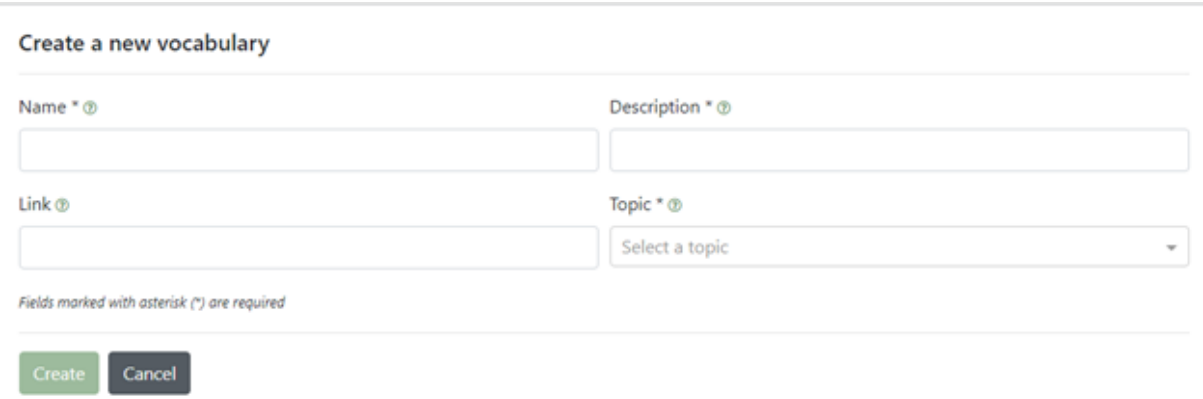


A1.1.13.2 Create a new Vocabulary



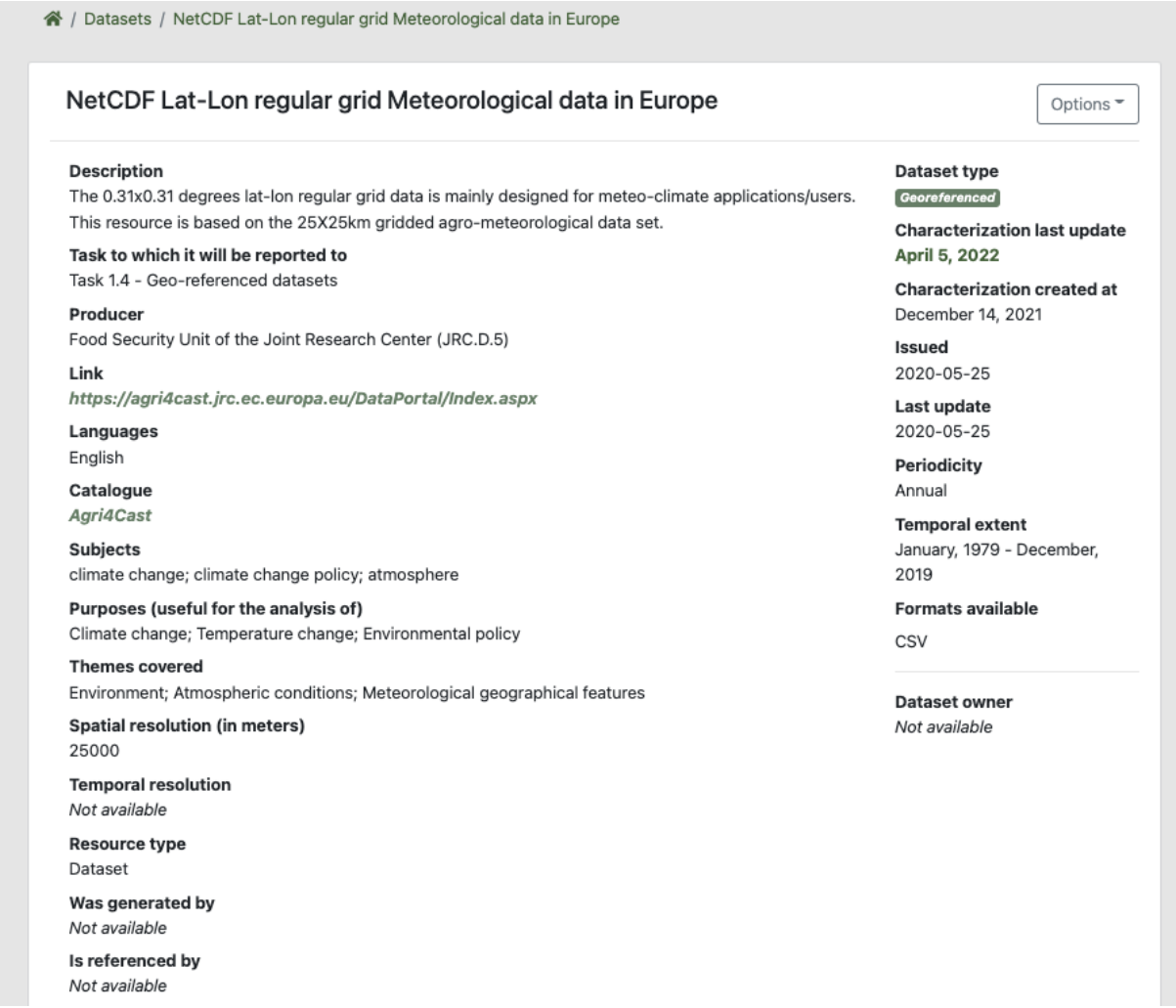
To create a new vocabulary, the following fields have to be filled in:

- **Name**: name of the vocabulary, it is a free text property
- **Description**: a brief description of its use.
- **Link**: an external link to the vocabulary. Leave it blank if it is an internal vocabulary.
- **Topic**: property of the dataset to which the vocabulary is related to

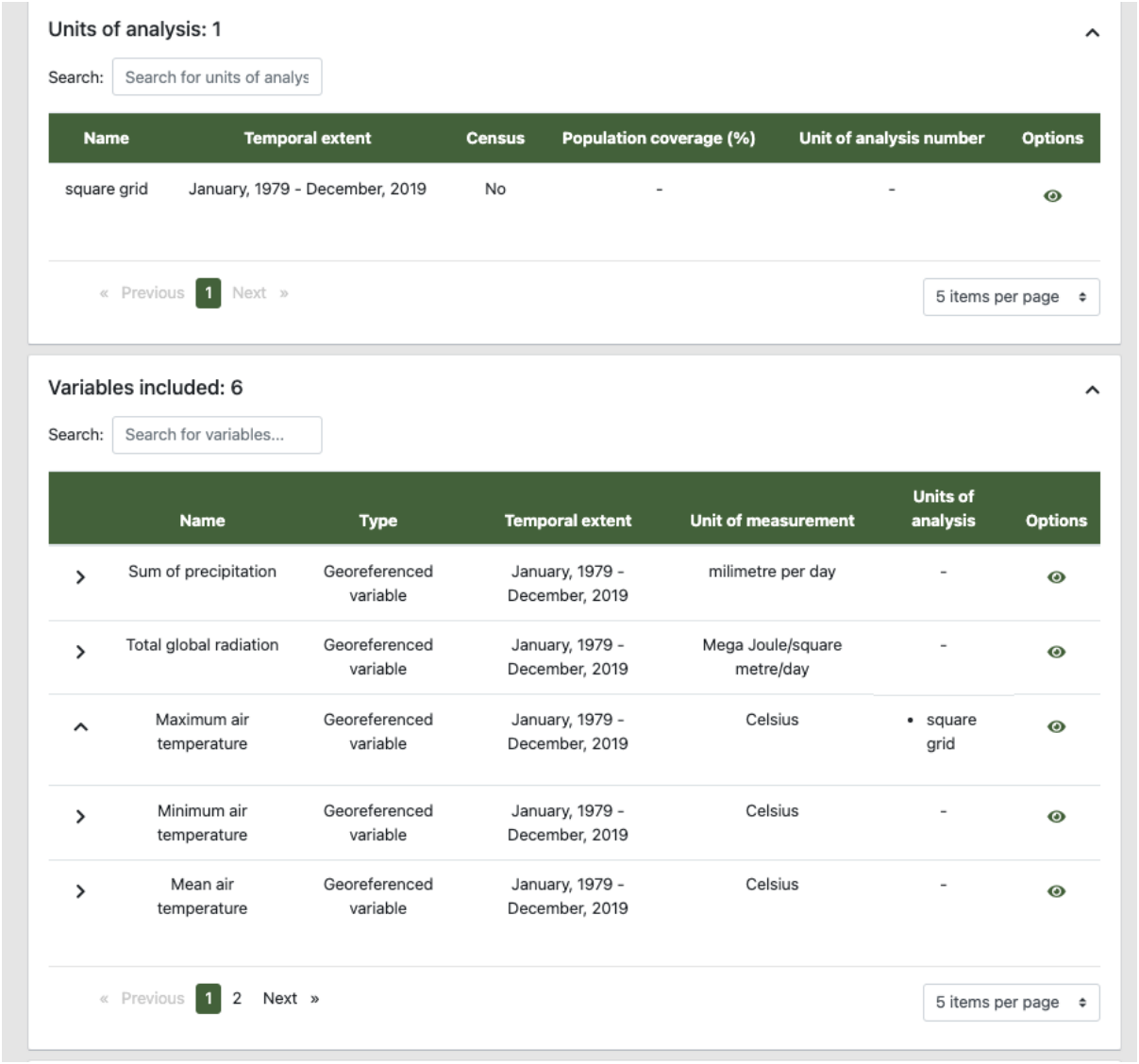


A1.1.14 Details of the dataset characterisations

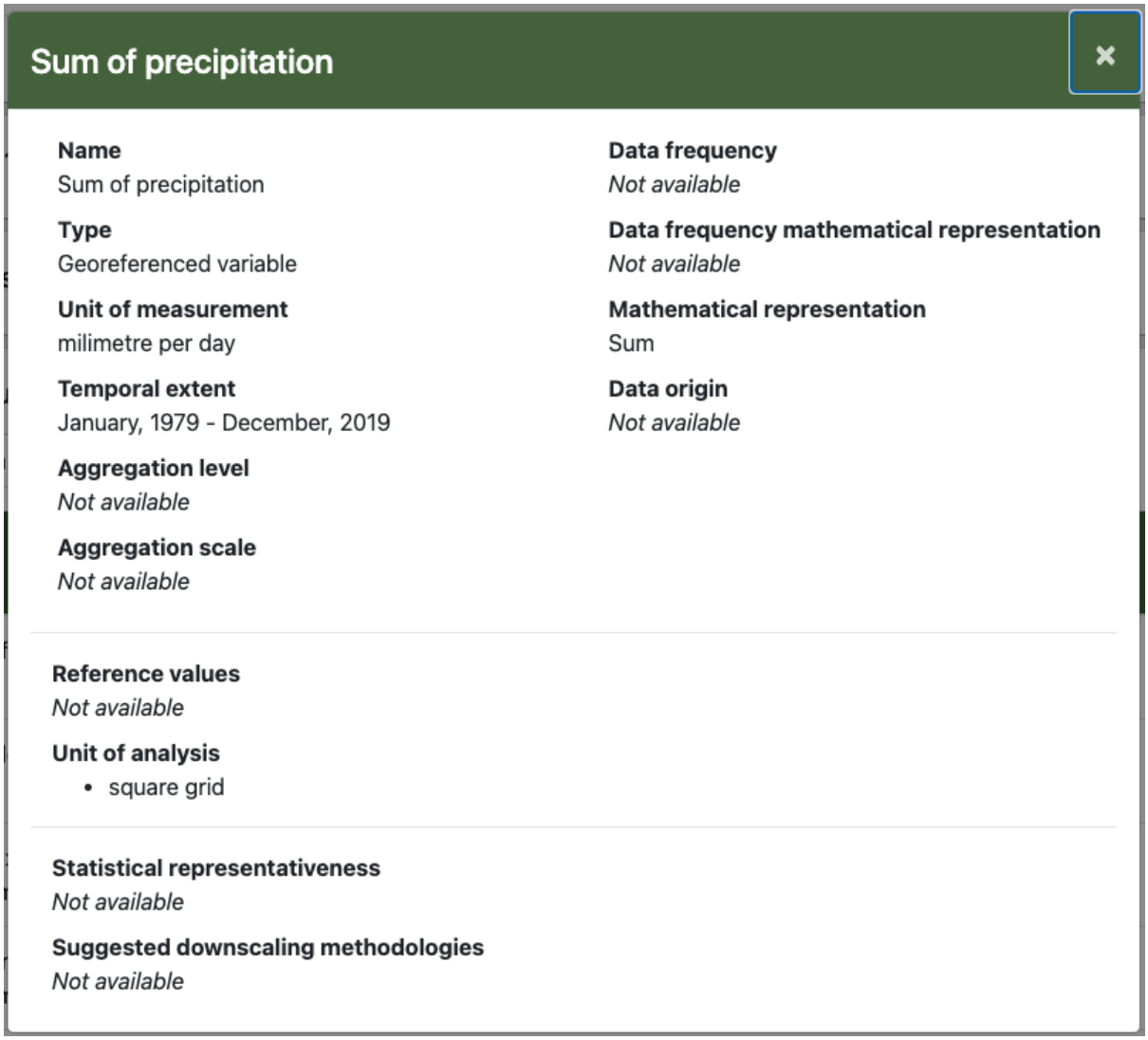
By selecting any of the existing datasets on the platform, the user navigates to a detailed view showing all its metadata defined by the AGRICORE DCAT-AP 2.0 ontology.



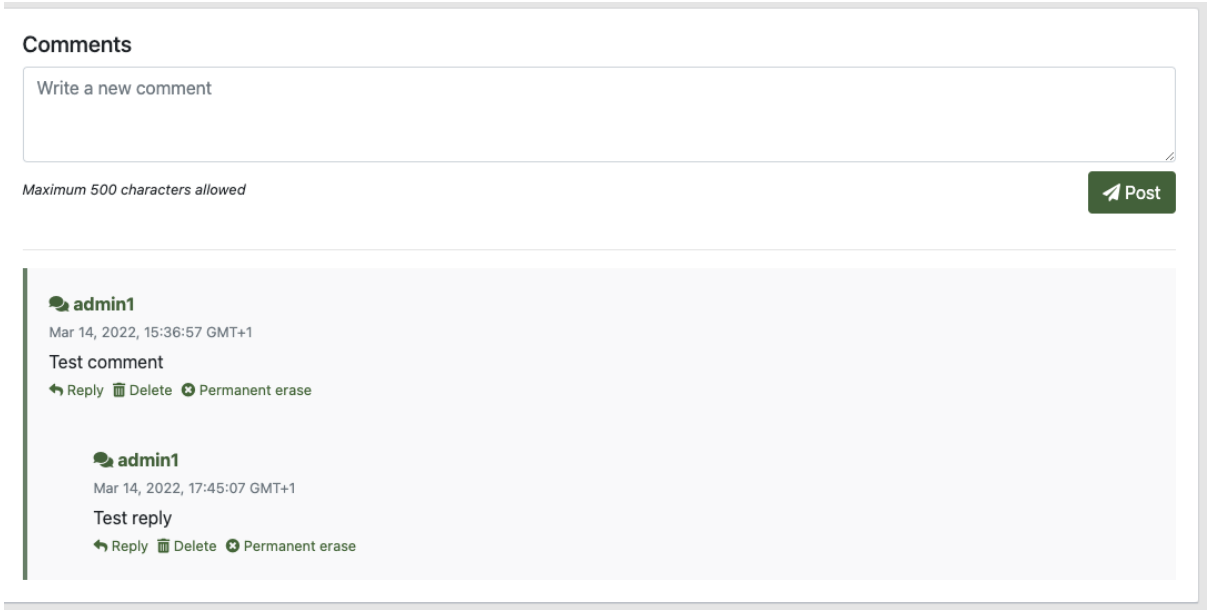
Complex properties such as units of analysis, distributions and variables are shown in tables along with some of their properties. The user can filter these properties directly on the table using a search box.



In addition, by clicking on the "**eye**" icon of each property, the user can open a menu or window where all the attributes or sub-properties of the variables, distributions or units of analysis are shown in more detail.



Finally, at the bottom of this screen, the user can find the comment section associated with the dataset. From there, registered users can post new comments or reply to existing ones.



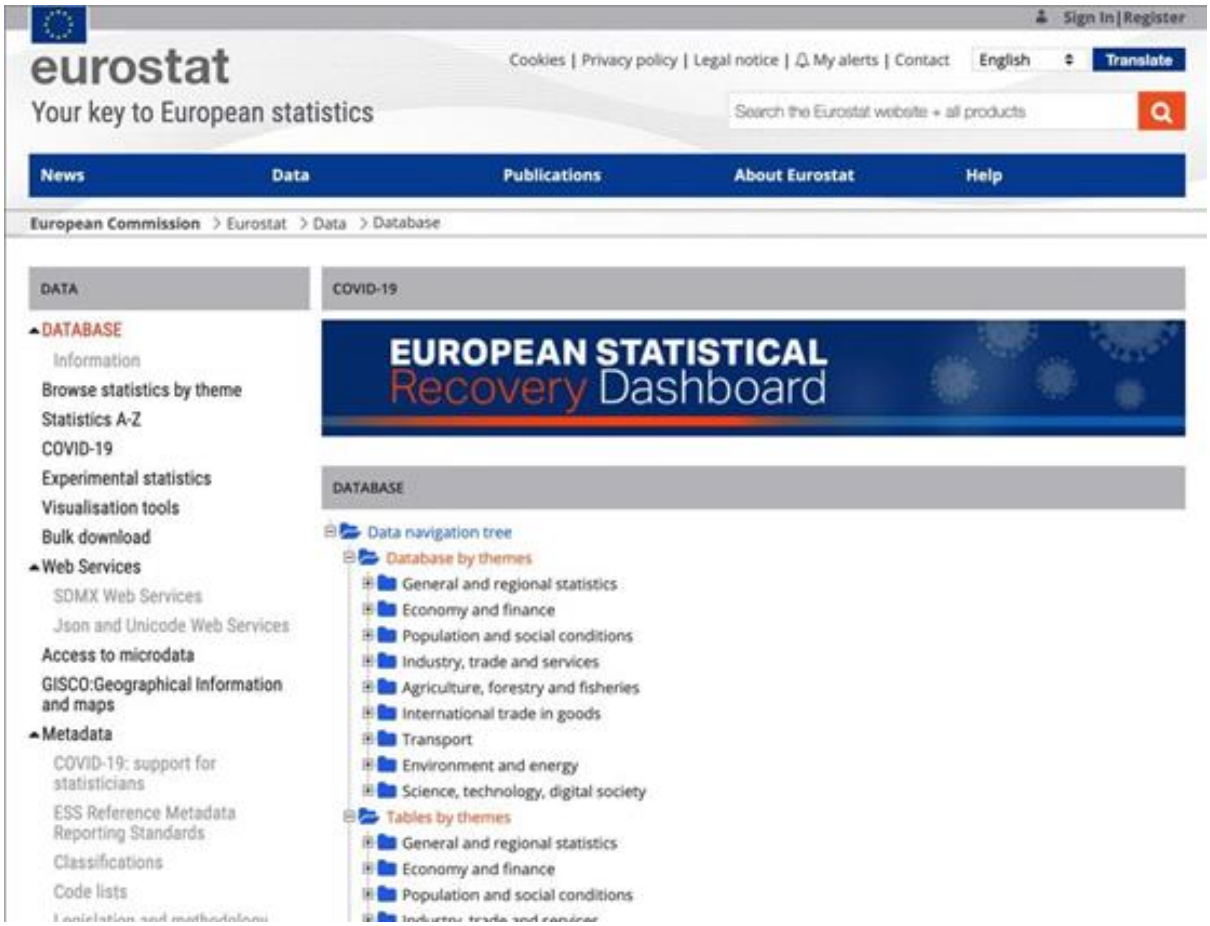
A1.2 Characterisation of 'Socio-Economic' Datasets

A1.2.1 Example 1: EUROSTAT Database

Eurostat's main role is to process and publish comparable statistical information at the European level. Eurostat does not collect data. This is done in EU countries by their statistical authorities. They verify and analyze national data and send them to Eurostat. Eurostat's role is to consolidate the data and ensure they are comparable. Eurostat is the only provider of statistics at the European level and the data issued are harmonized as far as possible.

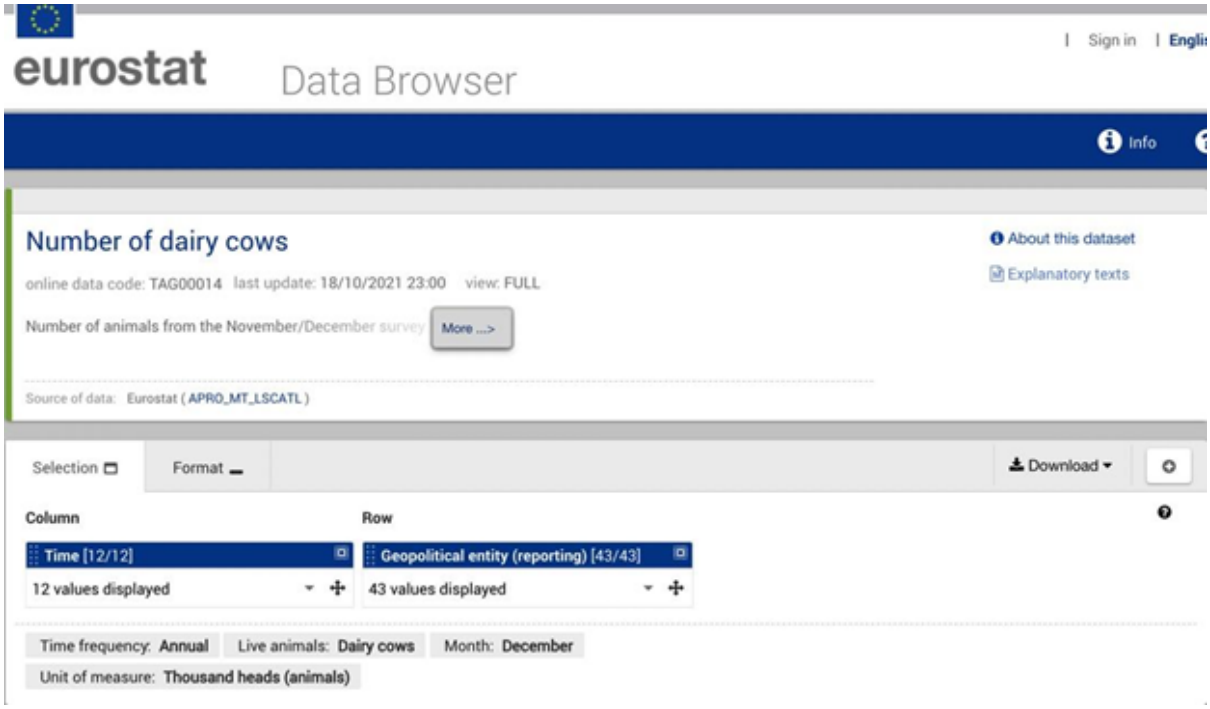
https://ec.europa.eu/eurostat/data/database

Dataset can be viewed by clicking on Data > Database.



In the context of the AGRICORE project, it has been decided to characterize Eurostat *Tables by themes*, as the *Database by themes* are mostly dynamic data views. However, for "Environmental and Energy" datasets it has been decided to characterize using *Database by themes* because only in these sections the various sectors (NACE activities) can be differentiated, and the AGRICORE scope focuses on agricultural activities.

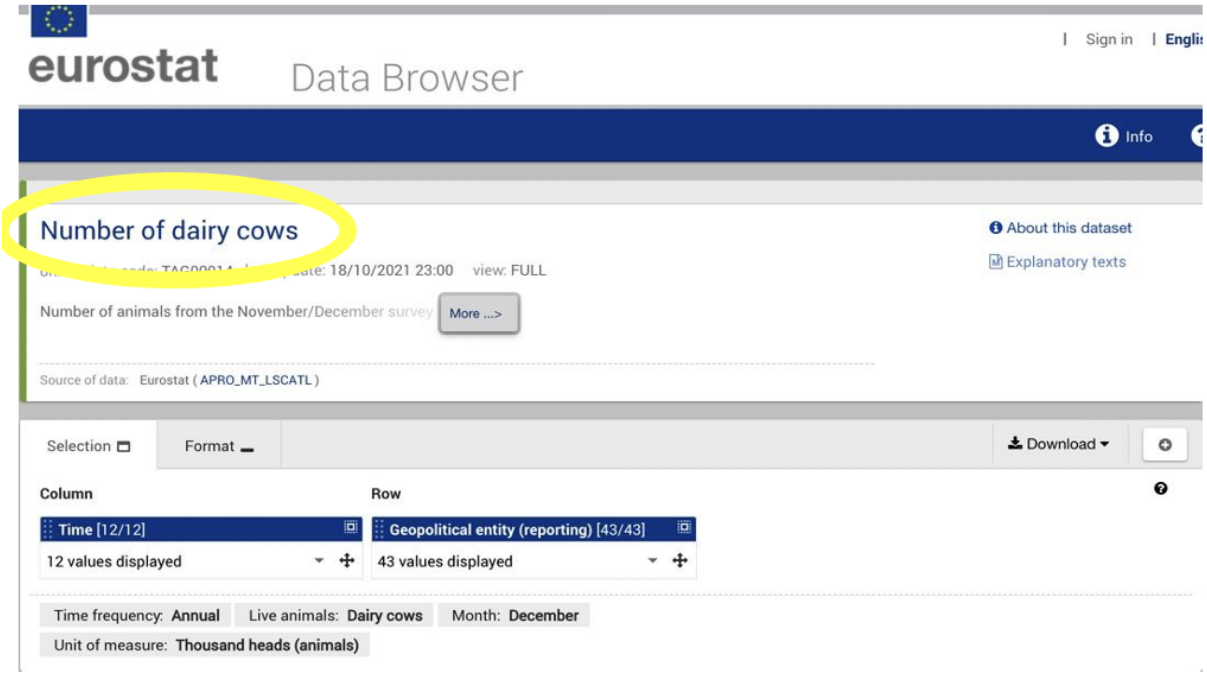
Once the dataset to characterize has been identified, the user can proceed with the data entry. Let's take as an example the dataset "Number of dairy cows":
https://ec.europa.eu/eurostat/databrowser/view/tag00014/default/table?lang=en



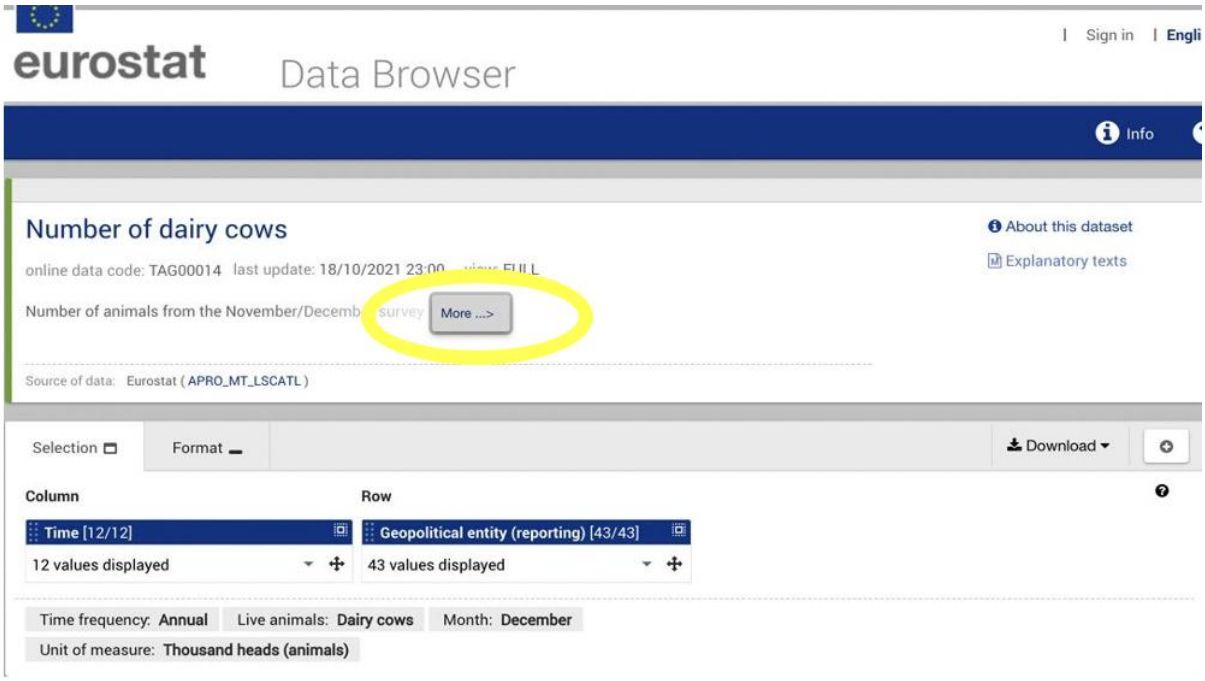
This page contains all the information we need to characterize this document.

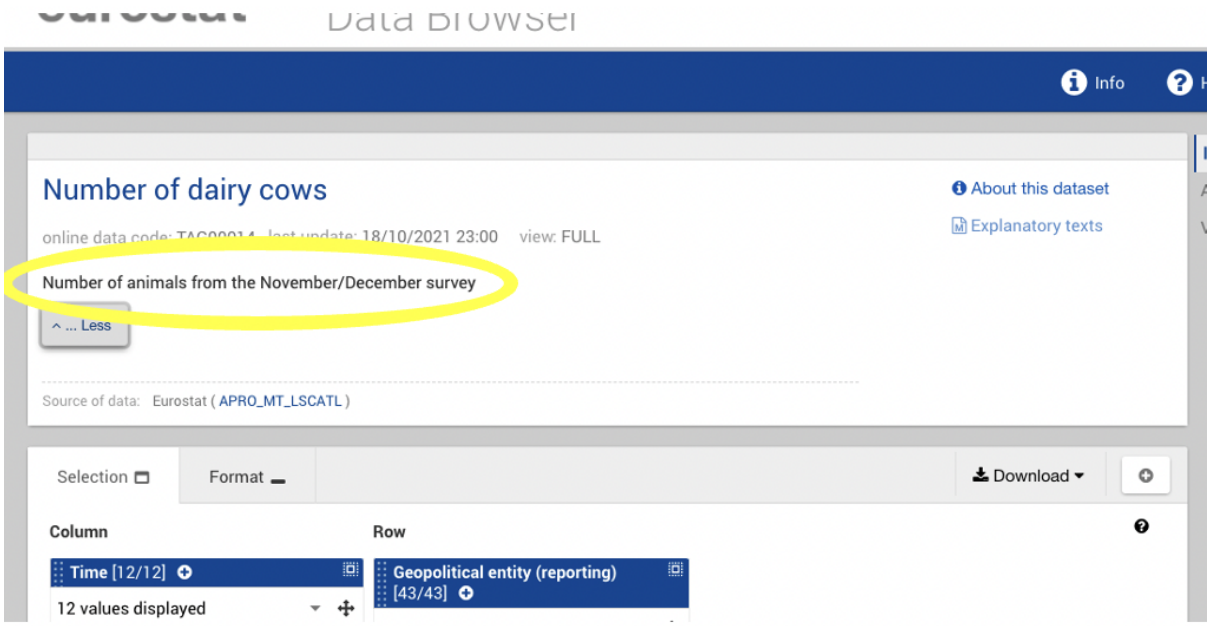
A1.2.1.1 General

- **Title**: Number of dairy cows

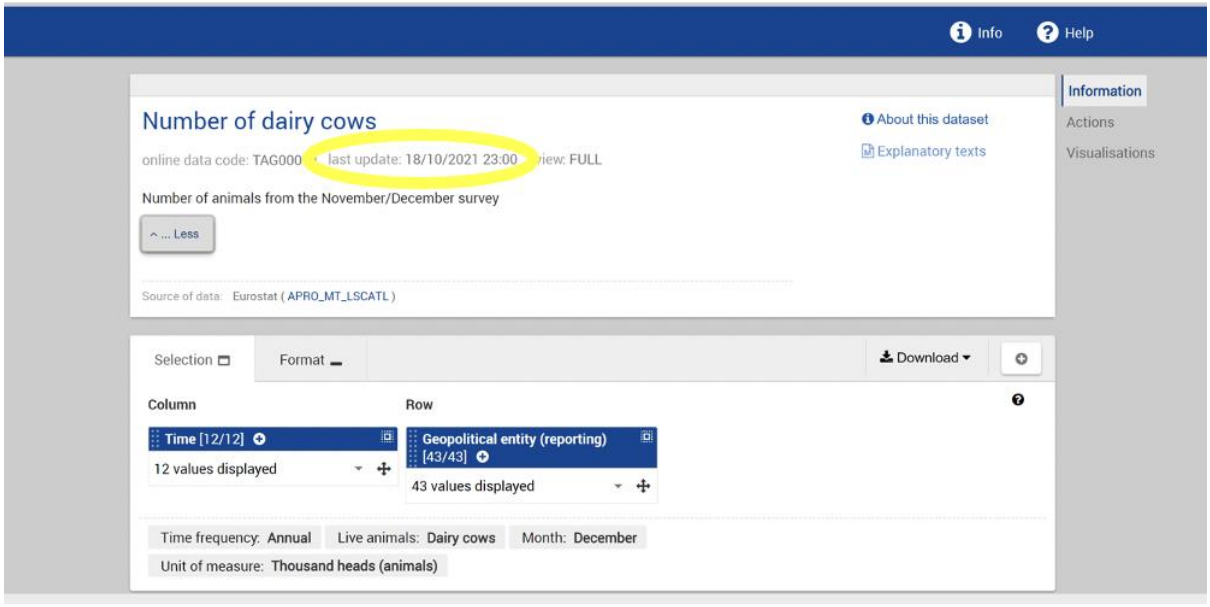


- **Description**: The description of the dataset can be found by clicking on "More" → Number of animals from the November/December survey.

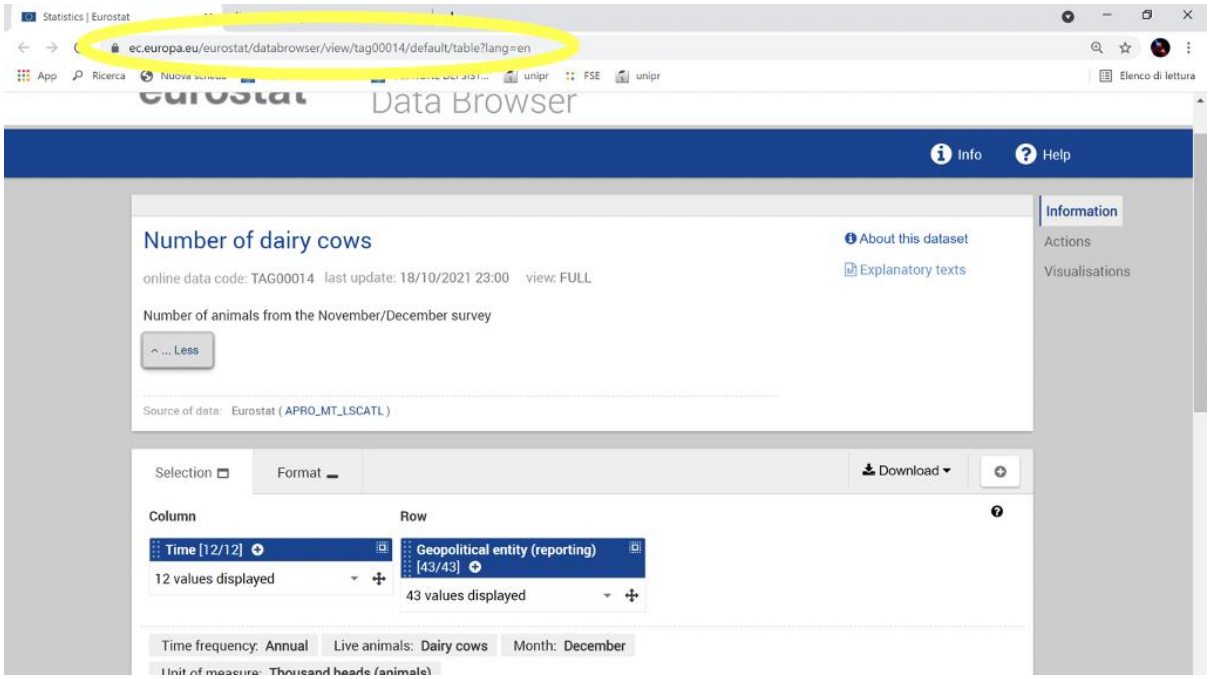




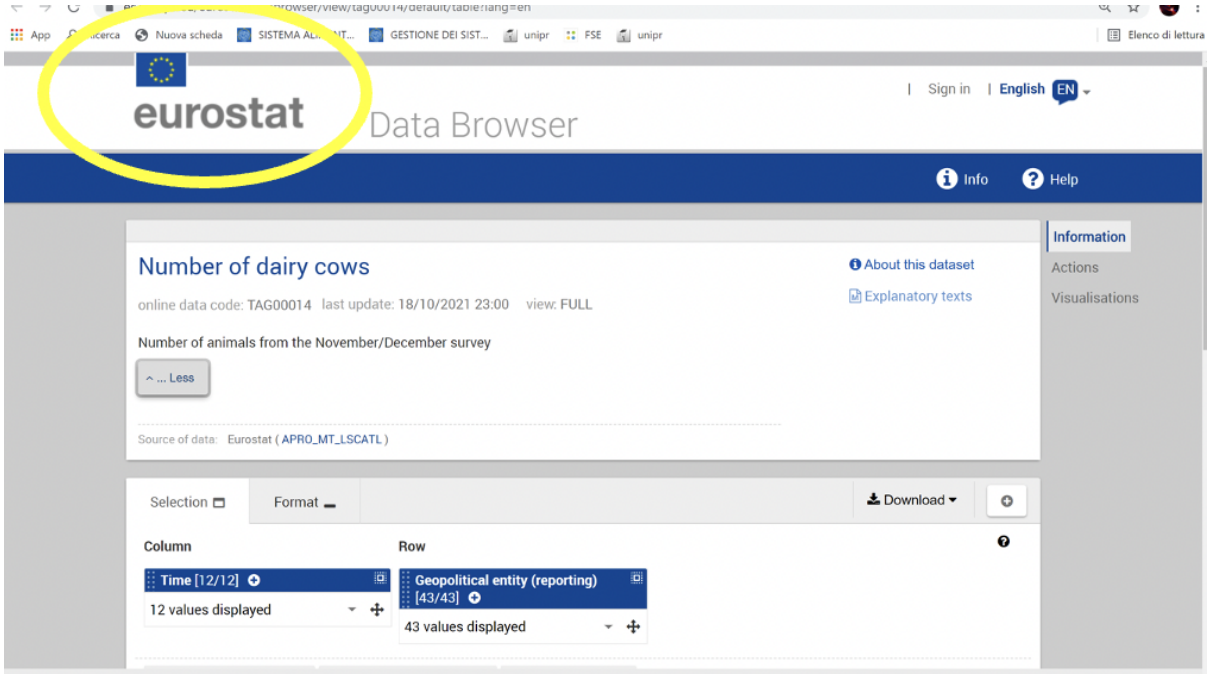
- **Last update**: 18/10/2021 23:00



- **Link**: details on the distributions can be found in this section.

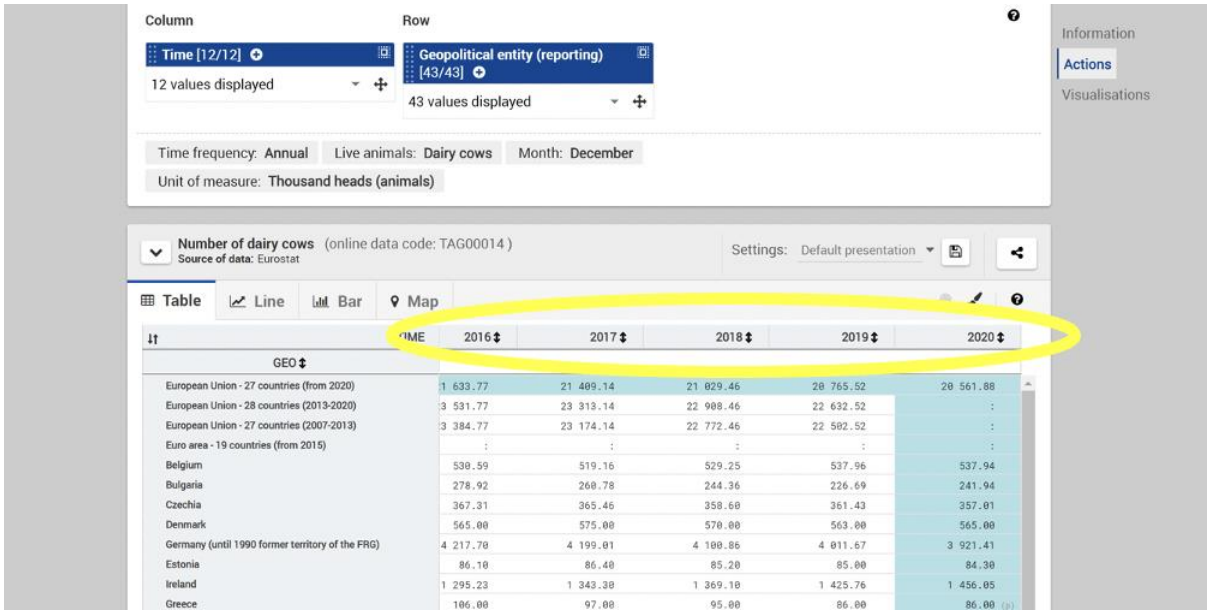


- **Producer:** Eurostat

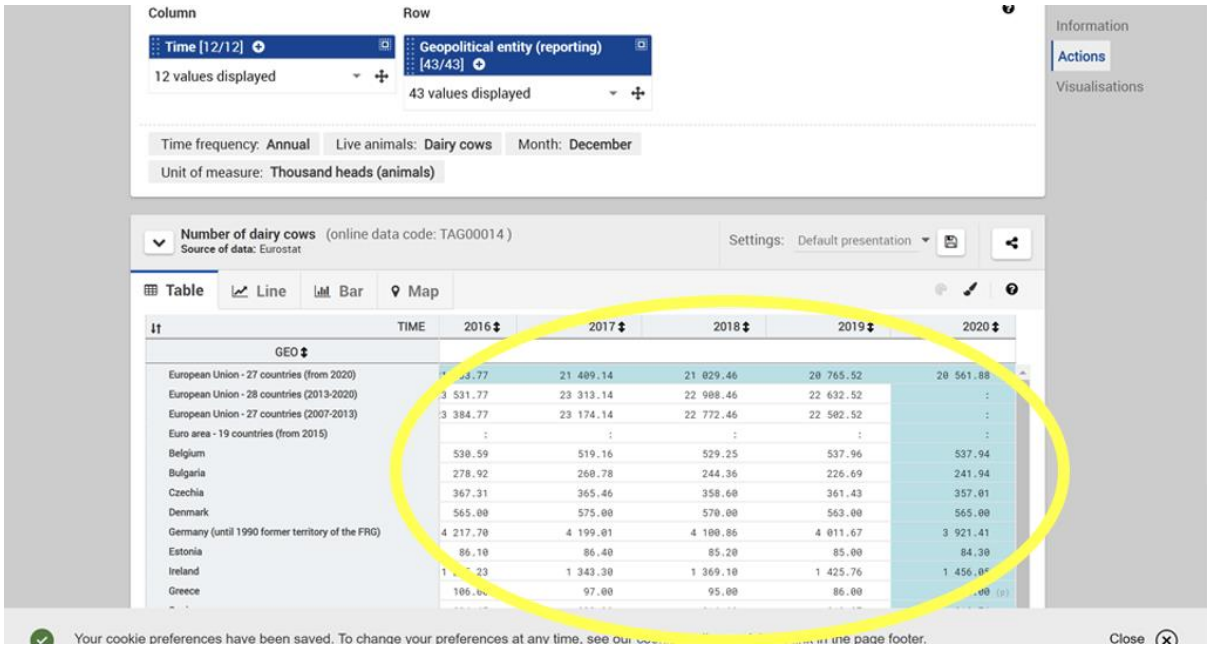


- **Dataset type:** socioeconomic (it depends on the dataset chosen; it can be georeferenced too).
- **Language:** English.
- **Periodicity of the publication:** annual.
- **Catalogue:** Eurostat- Agriculture, forestry and fisheries.

- **Temporal resolution:** annual

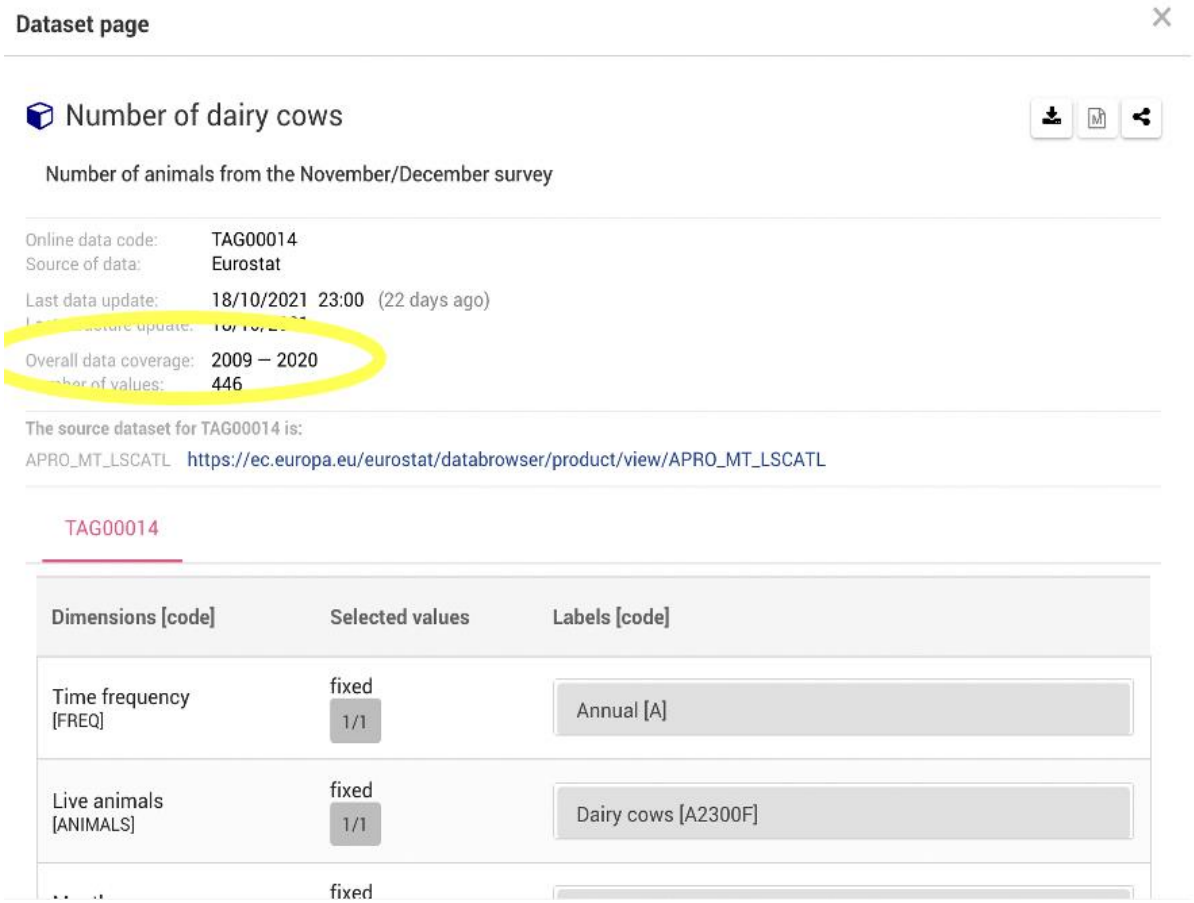


- **Resource type:** dataset.



A1.2.1.2 Purpose

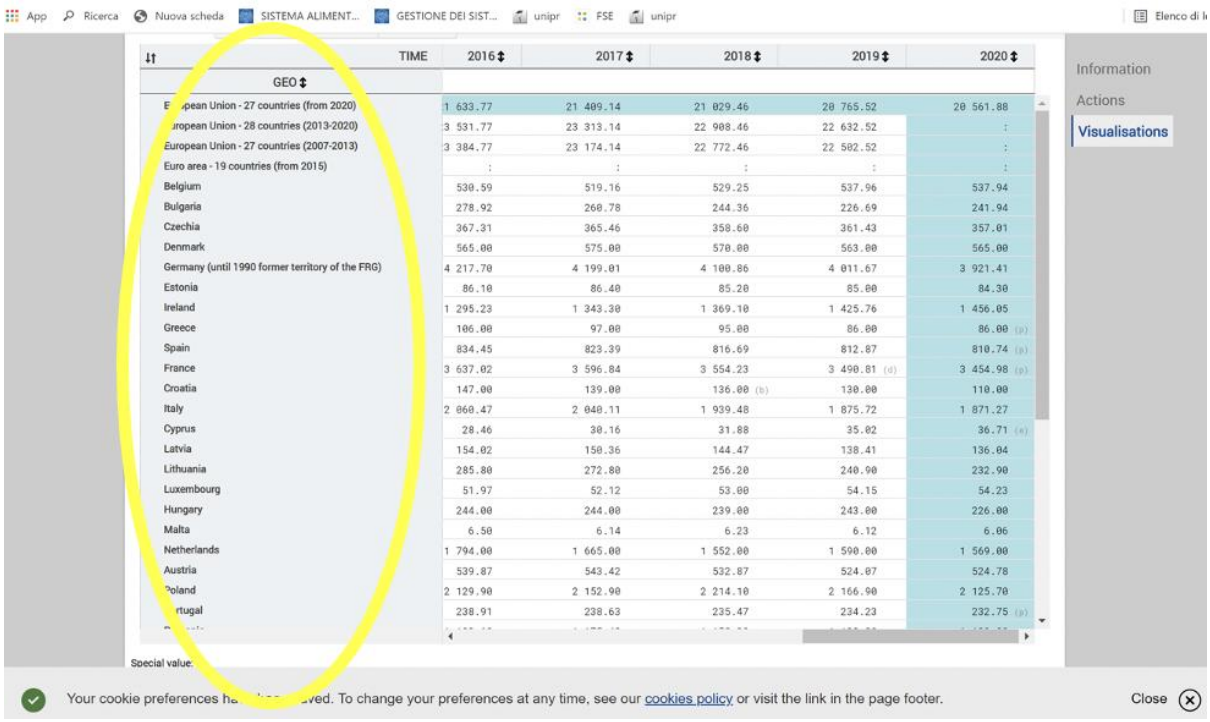
- **Temporal extent:** 2009/01 – 2020/12



- **Subject:** livestock (users can choose the most appropriate subject from the dropdown menu; the list in the menu is coming from a public vocabulary hence it's not changeable).
- **Useful for the analysis of (purpose):** technical efficiency analysis (users can add terms to the vocabulary).
- **Themes covered:** agriculture, fisheries, forestry and food.

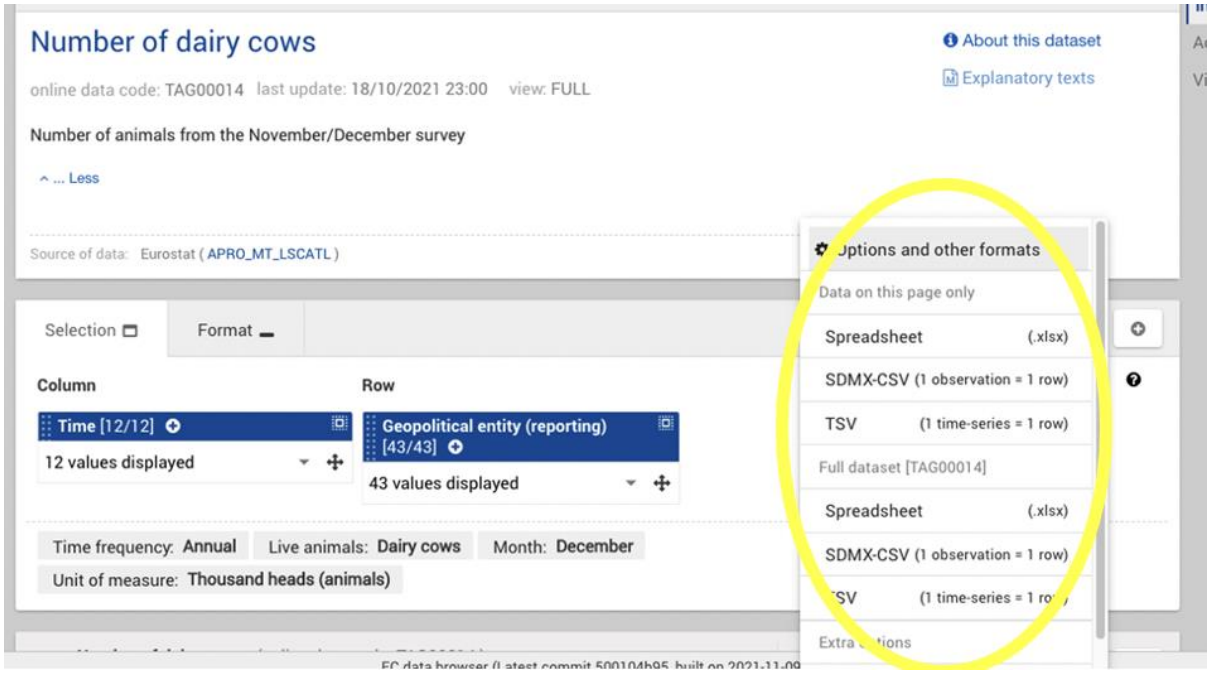
A1.2.1.3 Geographical coverage

- **Country coverage:**



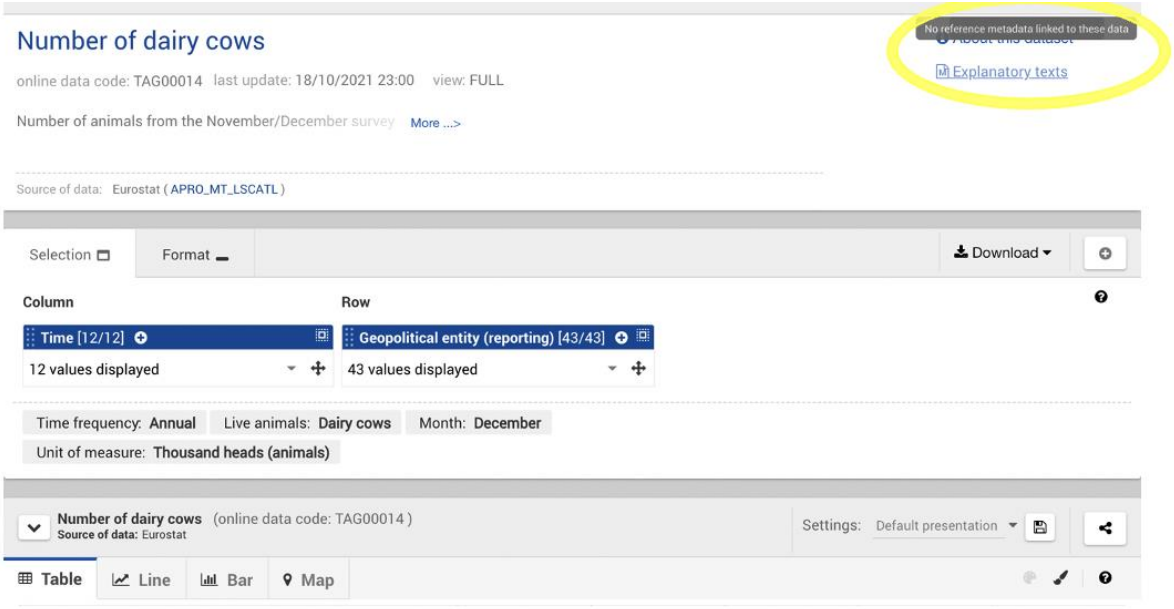
A1.2.1.4 Distribution

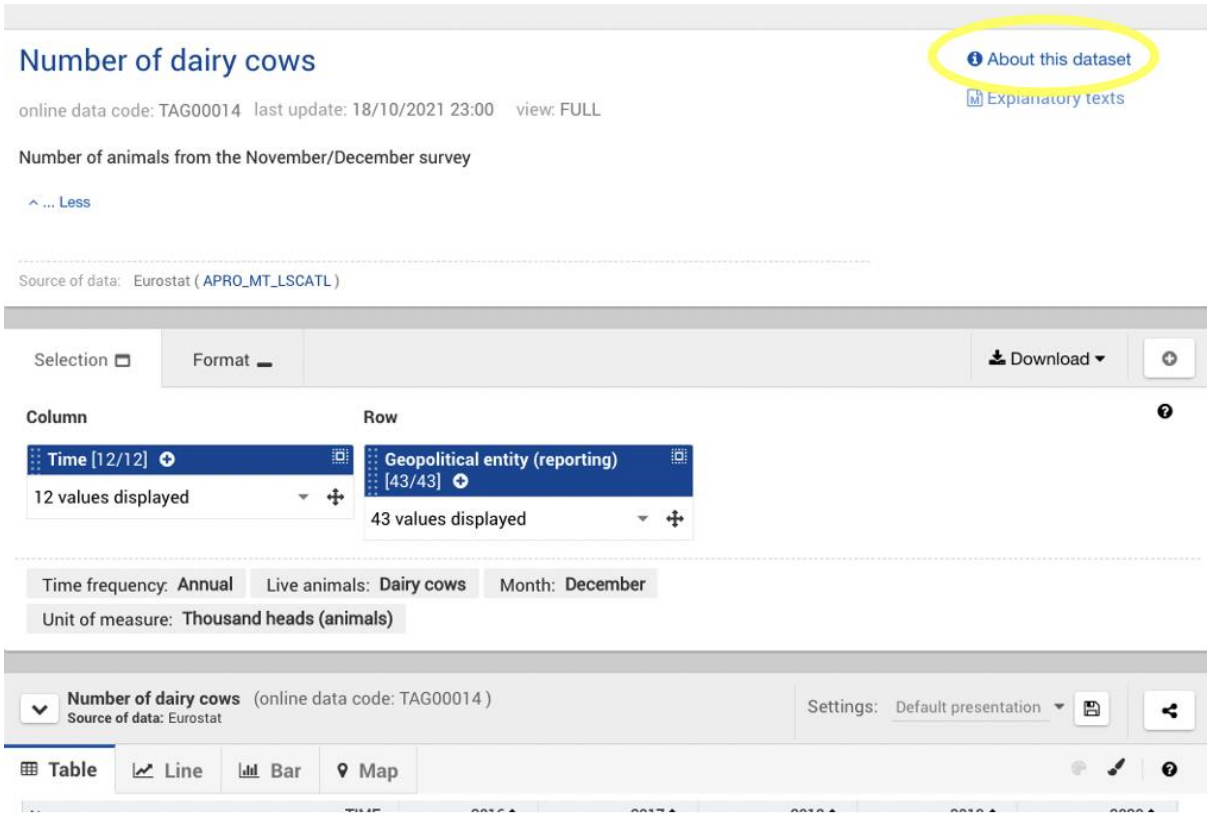
- SDMX-CSV
- Spreadsheet (xlsx)
- TSV



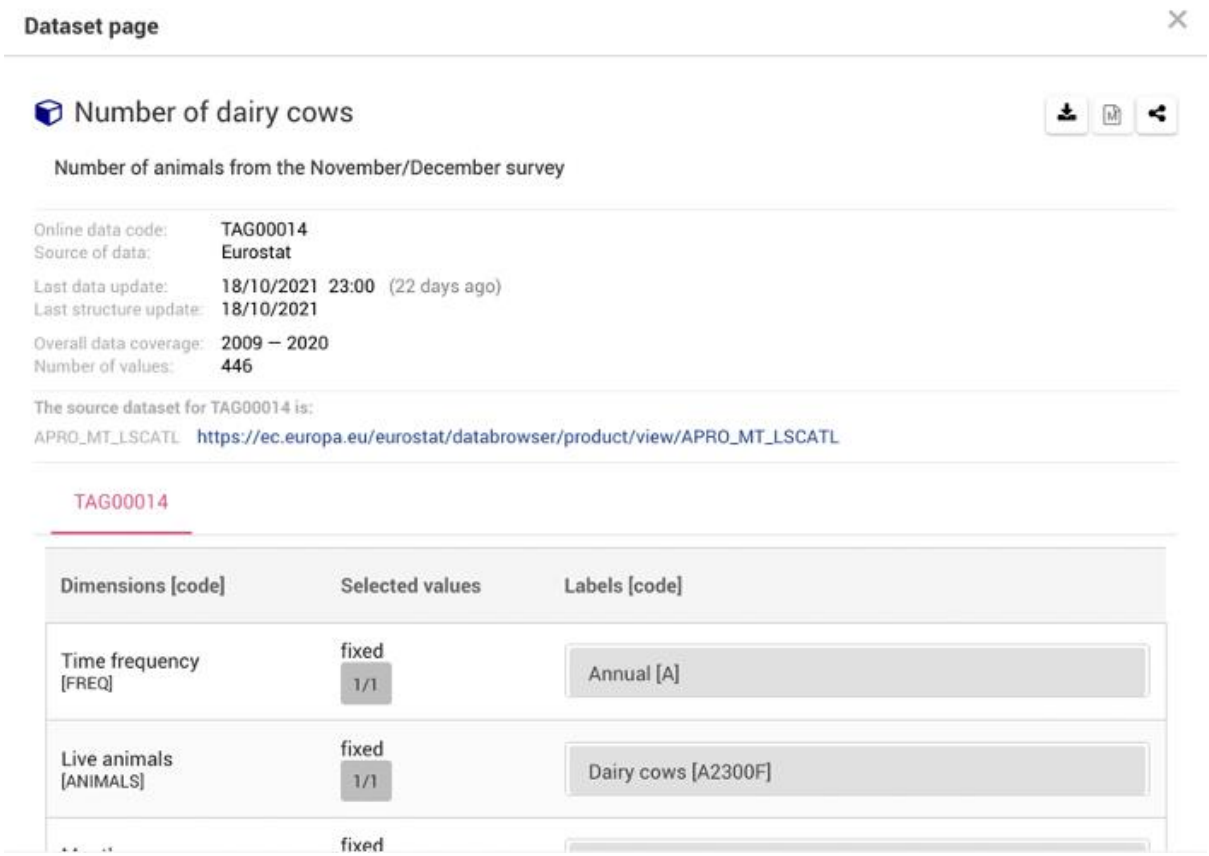
A1.2.1.5 Resolution and representativeness

- **Unit of analysis:** agriculture holding (users can find this information accessing metadata).In this case, metadata is not immediately available.

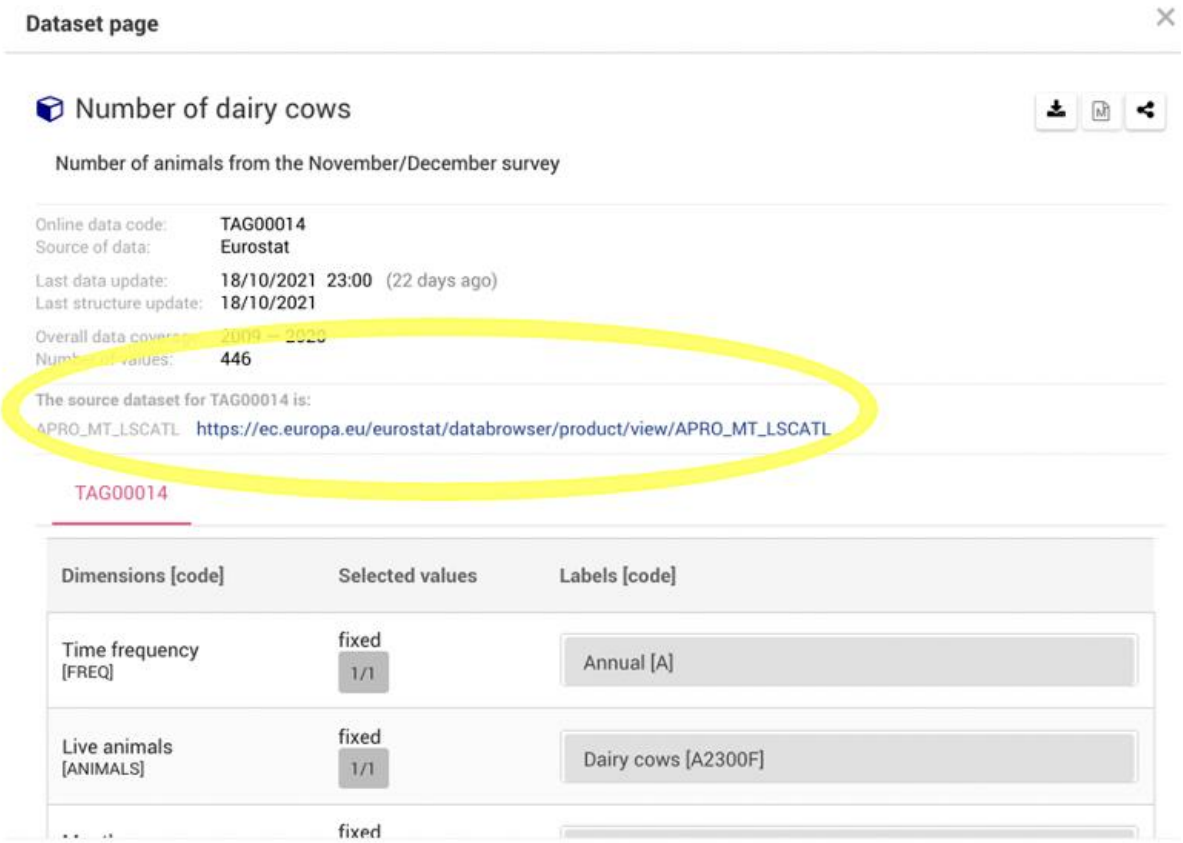




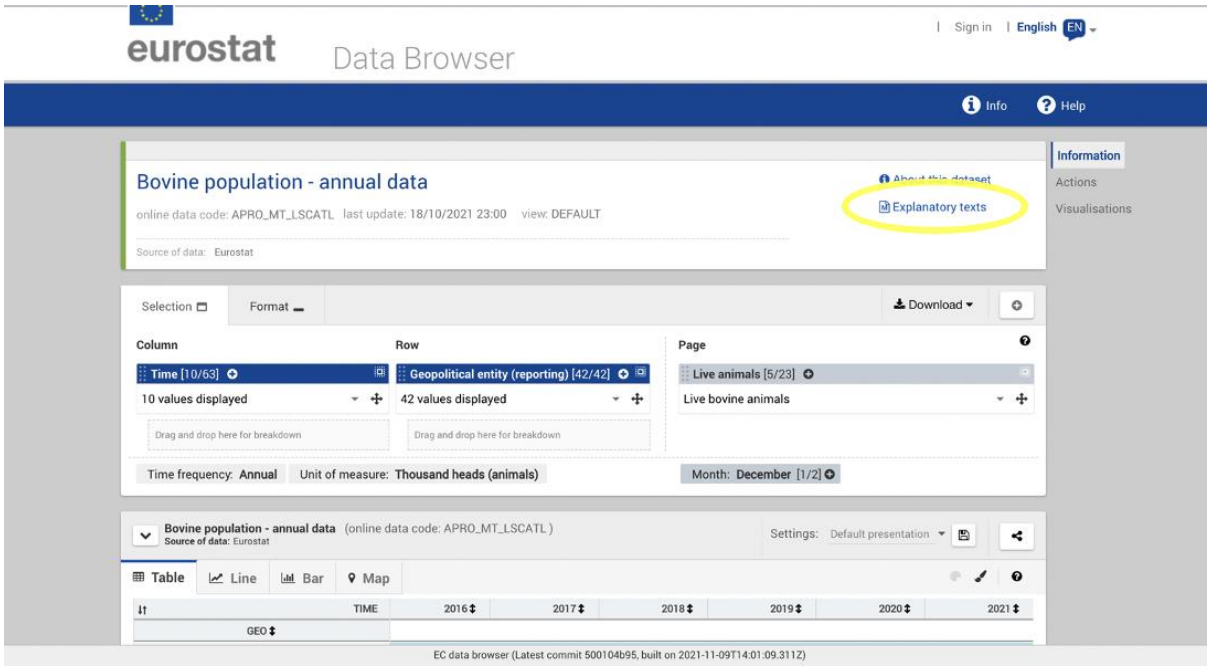
This table will then appear, from which we can get every variable needed.

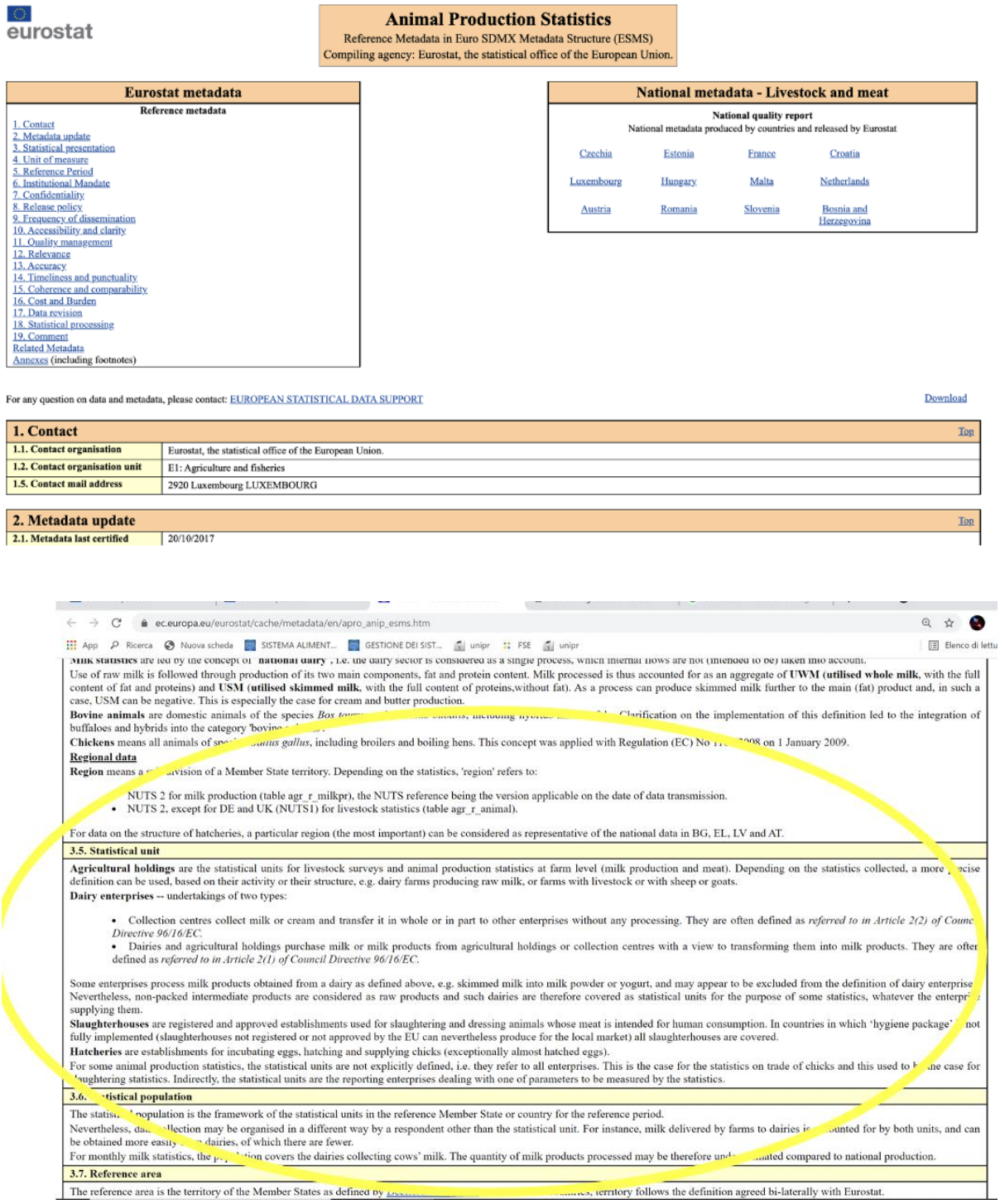


By clicking on the link below,



It's possible to access to the original metadata, which is useful for a further characterization.





eurostat

Animal Production Statistics
Reference Metadata in Euro SDMX Metadata Structure (ESMS)
Compiling agency: Eurostat, the statistical office of the European Union.

Eurostat metadata

Reference metadata

1. Contact
2. Metadata update
3. Statistical presentation
4. Unit of measure
5. Reference Period
6. Institutional Mandate
7. Confidentiality
8. Release policy
9. Frequency of dissemination
10. Accessibility and clarity
11. Quality management
12. Relevance
13. Accuracy
14. Timeliness and punctuality
15. Coherence and comparability
16. Cost and Burden
17. Data revision
18. Statistical processing
19. Comment

Related Metadata
Annexes (including footnotes)

National metadata - Livestock and meat

National quality report
National metadata produced by countries and released by Eurostat

Czechia, Estonia, France, Croatia
Luxembourg, Hungary, Malta, Netherlands
Austria, Romania, Slovenia, Bosnia and Herzegovina

For any question on data and metadata, please contact: EUROPEAN STATISTICAL DATA SUPPORT

Download

1. Contact Top

1.1. Contact organisation	Eurostat, the statistical office of the European Union.
1.2. Contact organisation unit	E1: Agriculture and fisheries
1.5. Contact mail address	2920 Luxembourg LUXEMBOURG

2. Metadata update Top

2.1. Metadata last certified	20/10/2017

ec.europa.eu/eurostat/cache/metadata/en/apro_anip_esms.htm

Use of raw milk is followed through production of its two main components, fat and protein content. Milk processed is thus accounted for as an aggregate of **UWM (utilised whole milk**, with the full content of fat and proteins) and **USM (utilised skimmed milk**, with the full content of proteins,without fat). As a process can produce skimmed milk further to the main (fat) product and, in such a case, USM can be negative. This is especially the case for cream and butter production.

Bovine animals are domestic animals of the species *Bos taurus*, including hybrids. Clarification on the implementation of this definition led to the integration of buffaloes and hybrids into the category 'bovine animals'.

Chickens means all animals of species *Gallus gallus*, including broilers and boiling hens. This concept was applied with Regulation (EC) No 1165/2008 on 1 January 2009.

Regional data

Region means a subdivision of a Member State territory. Depending on the statistics, 'region' refers to:

- NUTS 2 for milk production (table agr_r_milkpr), the NUTS reference being the version applicable on the date of data transmission.
- NUTS 2, except for DE and UK (NUTS1) for livestock statistics (table agr_r_animal).

For data on the structure of hatcheries, a particular region (the most important) can be considered as representative of the national data in BG, EL, LV and AT.

3.5. Statistical unit

Agricultural holdings are the statistical units for livestock surveys and animal production statistics at farm level (milk production and meat). Depending on the statistics collected, a more precise definition can be used, based on their activity or their structure, e.g. dairy farms producing raw milk, or farms with livestock or with sheep or goats.

Dairy enterprises -- undertakings of two types:

- Collection centres collect milk or cream and transfer it in whole or in part to other enterprises without any processing. They are often defined as *referred to in Article 2(2) of Council Directive 96/16/EC*.
- Dairies and agricultural holdings purchase milk or milk products from agricultural holdings or collection centres with a view to transforming them into milk products. They are often defined as *referred to in Article 2(1) of Council Directive 96/16/EC*.

Some enterprises process milk products obtained from a dairy as defined above, e.g. skimmed milk into milk powder or yogurt, and may appear to be excluded from the definition of dairy enterprise. Nevertheless, non-packed intermediate products are considered as raw products and such dairies are therefore covered as statistical units for the purpose of some statistics, whatever the enterprise supplying them.

Slaughterhouses are registered and approved establishments used for slaughtering and dressing animals whose meat is intended for human consumption. In countries in which 'hygiene package' is not fully implemented (slaughterhouses not registered or not approved by the EU can nevertheless produce for the local market) all slaughterhouses are covered.

Hatcheries are establishments for incubating eggs, hatching and supplying chicks (exceptionally almost hatched eggs).

For some animal production statistics, the statistical units are not explicitly defined, i.e. they refer to all enterprises. This is the case for the statistics on trade of chicks and this used to be the case for slaughtering statistics. Indirectly, the statistical units are the reporting enterprises dealing with one of parameters to be measured by the statistics.

3.6. Statistical population

The statistical population is the framework of the statistical units in the reference Member State or country for the reference period.

Nevertheless, data collection may be organised in a different way by a respondent other than the statistical unit. For instance, milk delivered by farms to dairies is accounted for by both units, and can be obtained more easily from dairies, of which there are fewer.

For monthly milk statistics, the population covers the dairies collecting cows' milk. The quantity of milk products processed may be therefore underestimated compared to national production.

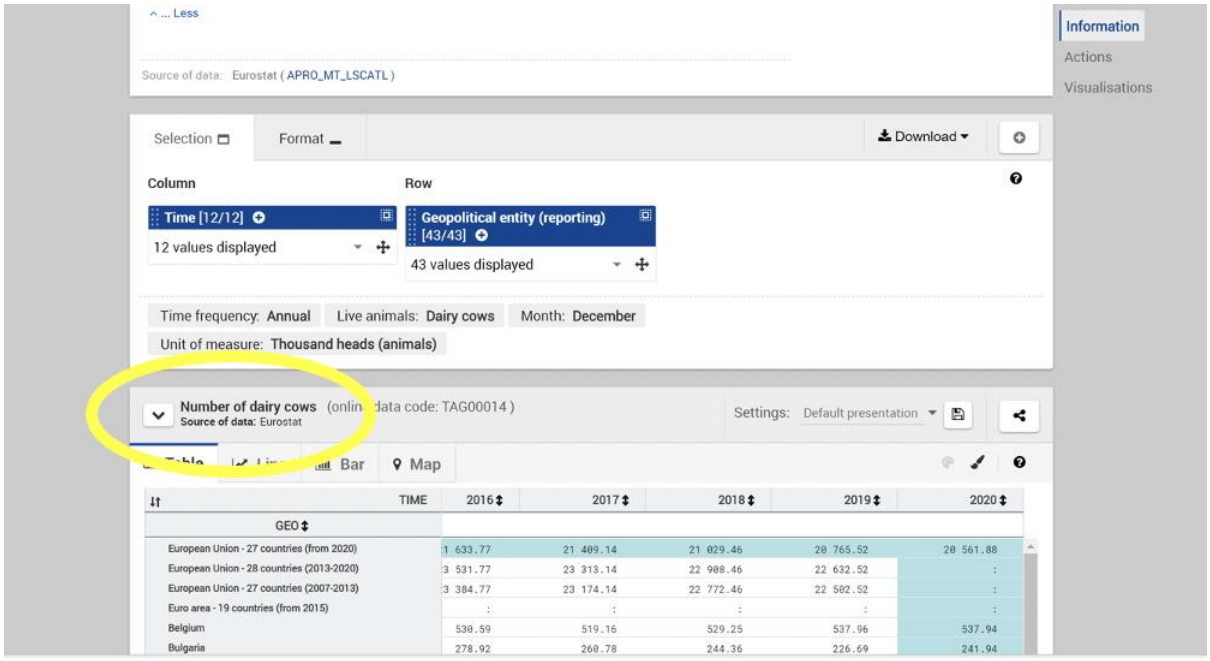
3.7. Reference area

The reference area is the territory of the Member States as defined by [illegible], territory follows the definition agreed bi-laterally with Eurostat.

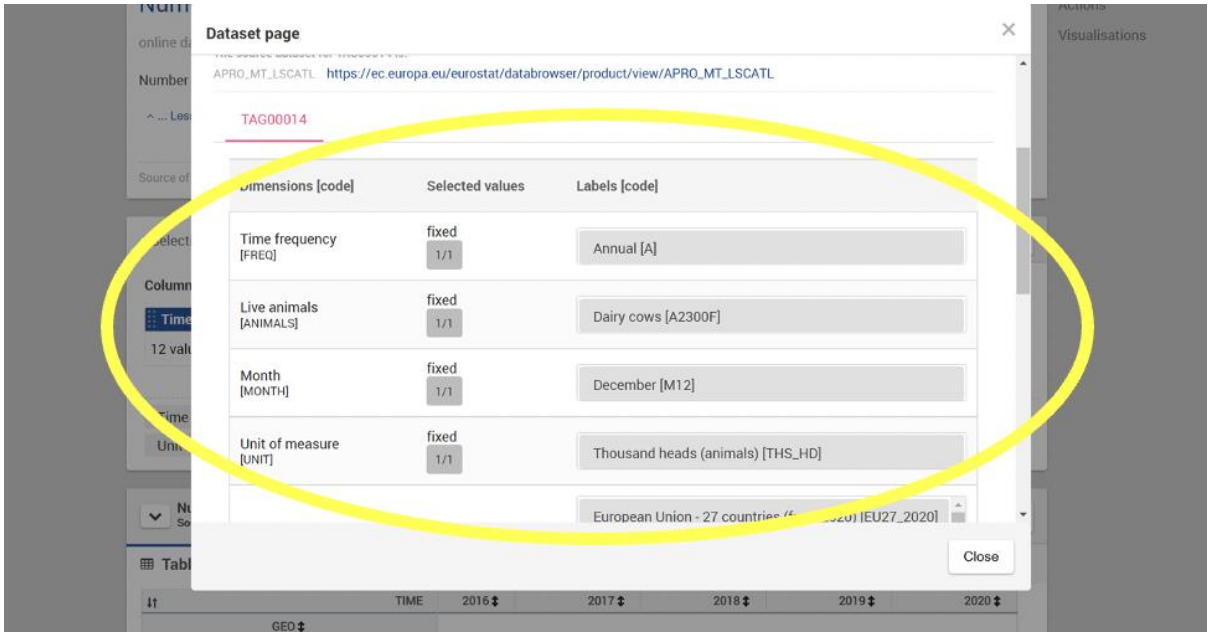
A1.2.1.6 Variables included

User has to click on 'Add a new variable'

- **Name:** Number of dairy cows

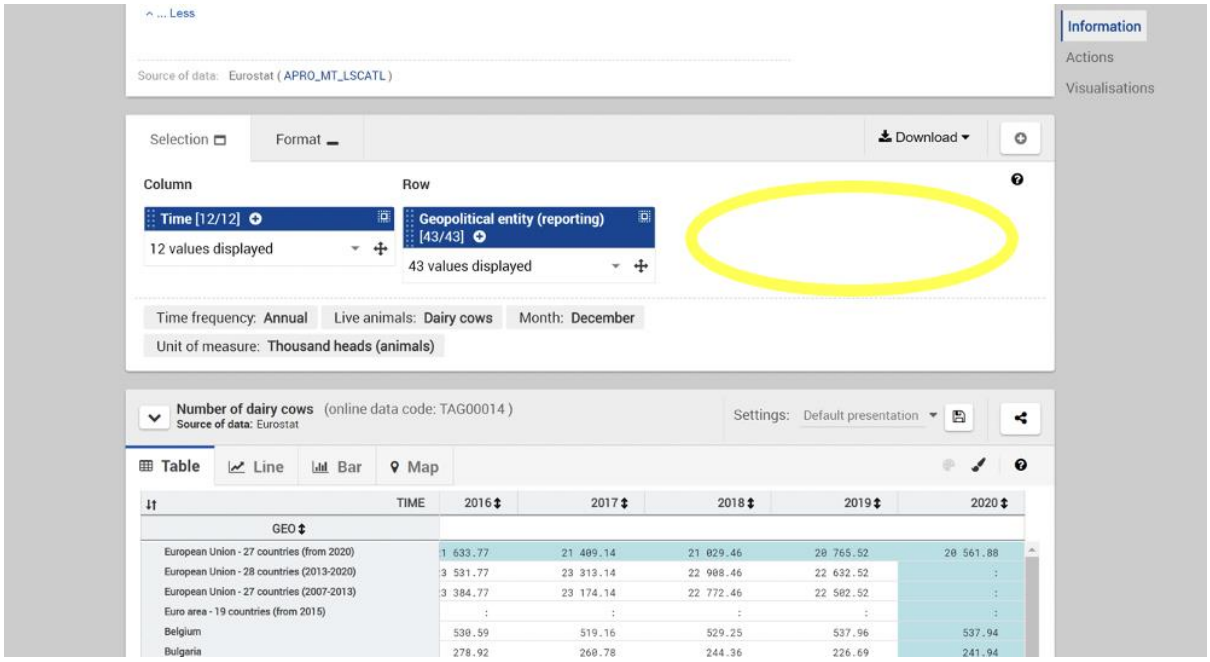


- **Unit of measurement:** thousand heads
- **Temporal extent:** 2009/01 – 2020/12
- **Mathematical representation:** instant value
- **Data frequency:** annual



- **Unit of analysis:** in this field, users can link the variable to the unit of analysis characterized before.

If there's more than one variable, the list of variables will appear on the right side of the screen.



A1.2.2 Example 2: OECD

The OECD database brings together information from various national and international databases, both within the OECD external organizations. The OECD brings together Member

countries and partners that collaborate closely on key global issues at national, regional and local level to find solutions to common challenges, develop global standards and identify best practices to promote better policies by measuring and benchmarking economies and data.

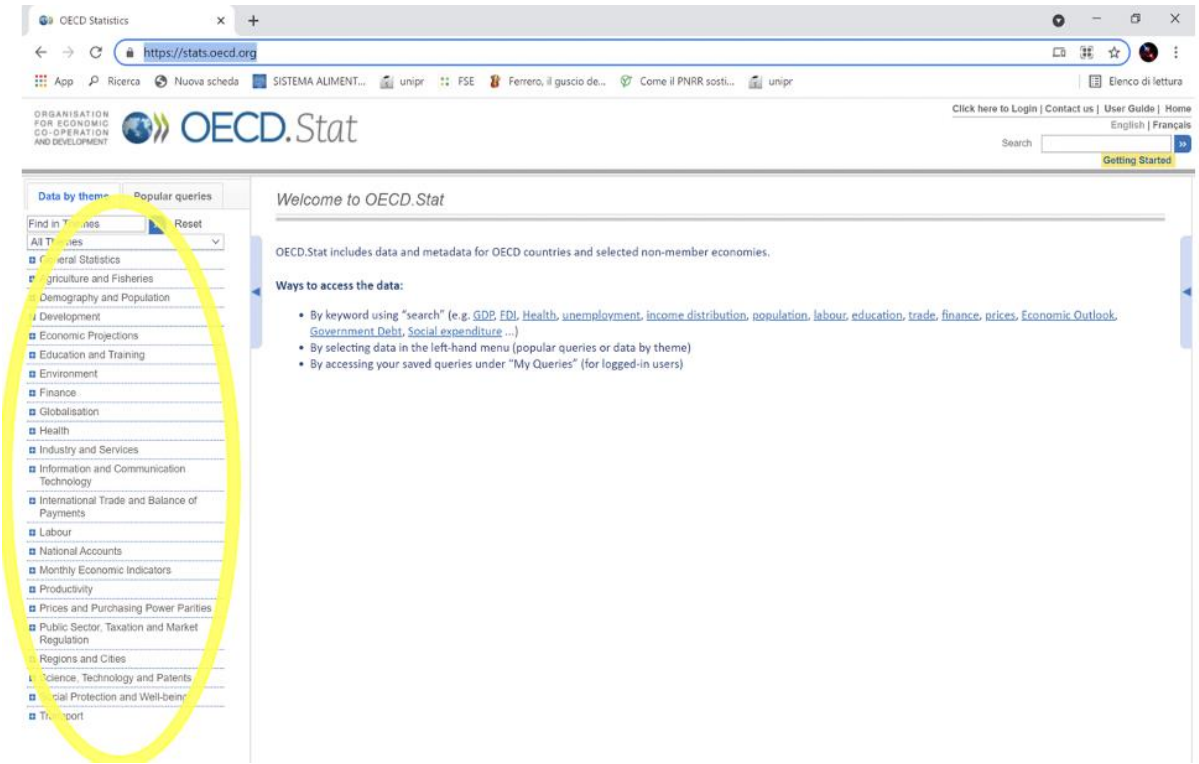
In the OECD case we have a subdivision by themes; for our project we decided to characterise the tables under the themes:

- Agriculture and Fisheries: where we find tables concerning Agricultural Outlook, Agricultural Policy Indicators and Environmental Indicators for Agriculture.
- Environment: where we find tables concerning Air and Climate, Water, Material Resources, Forest, Biodiversity, Land Resources, Innovation in environment.

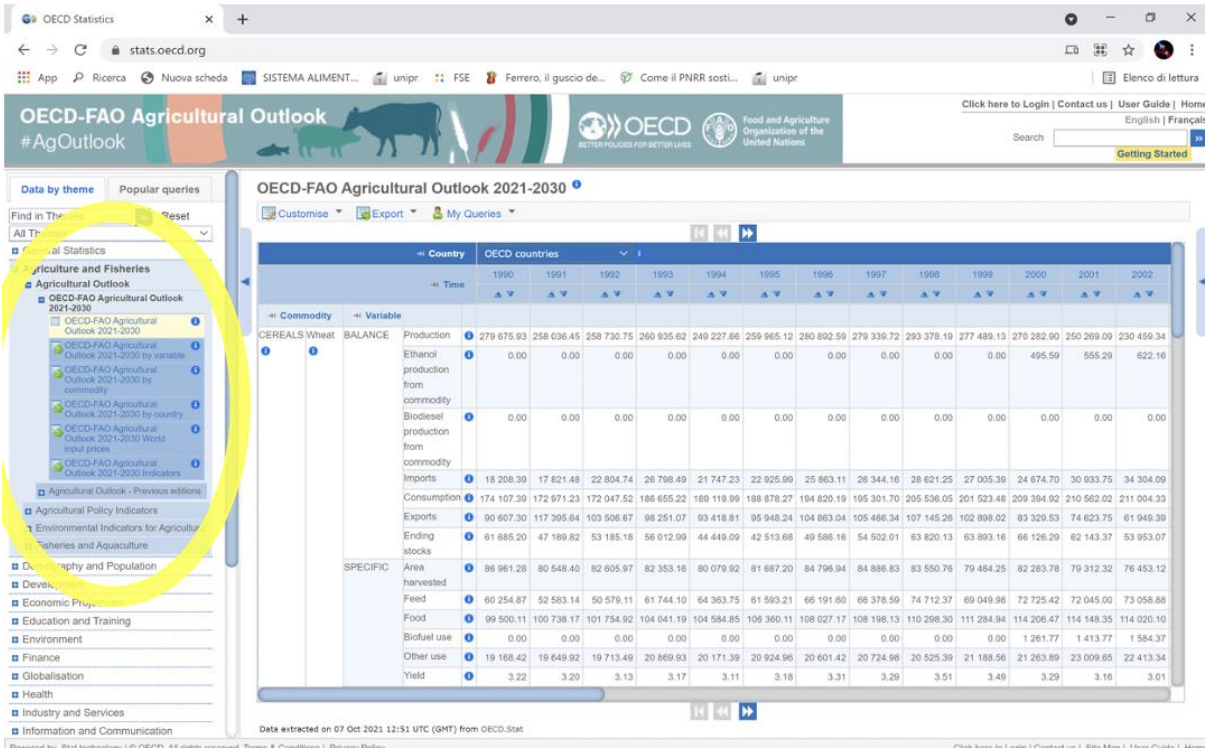
This is the link to access the OECD database:

https://stats.oecd.org/

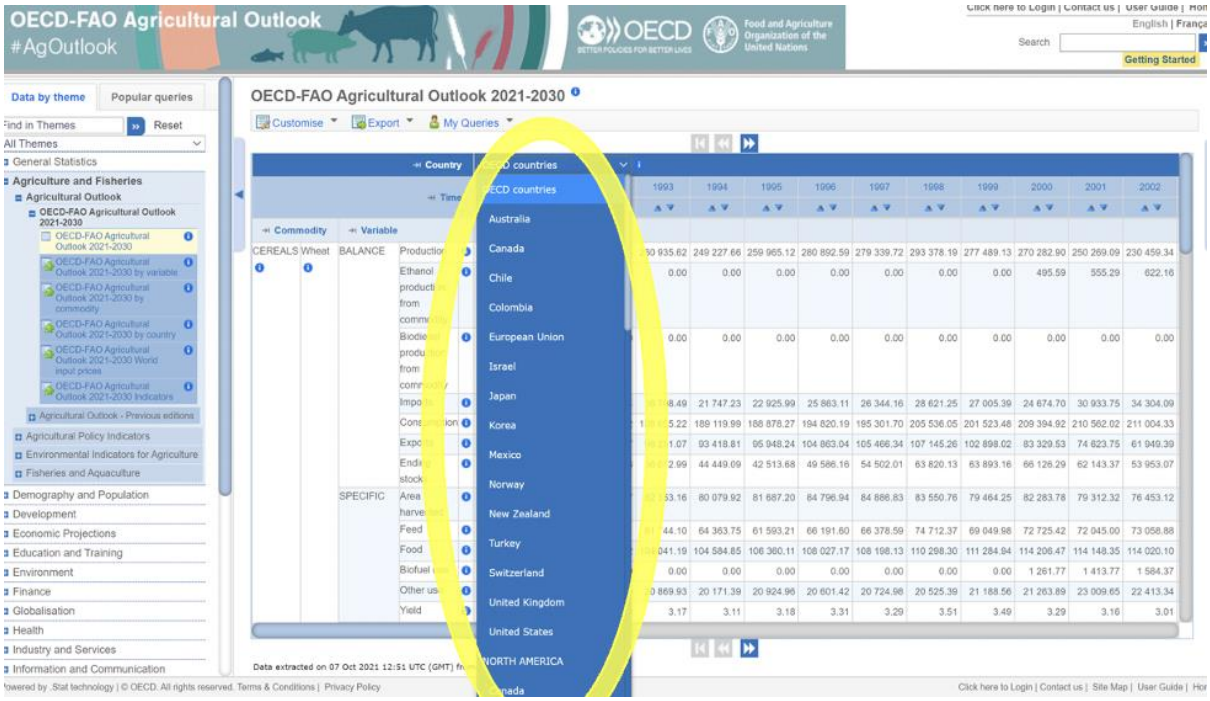
On the left side of this website there is a categorization of data by theme.



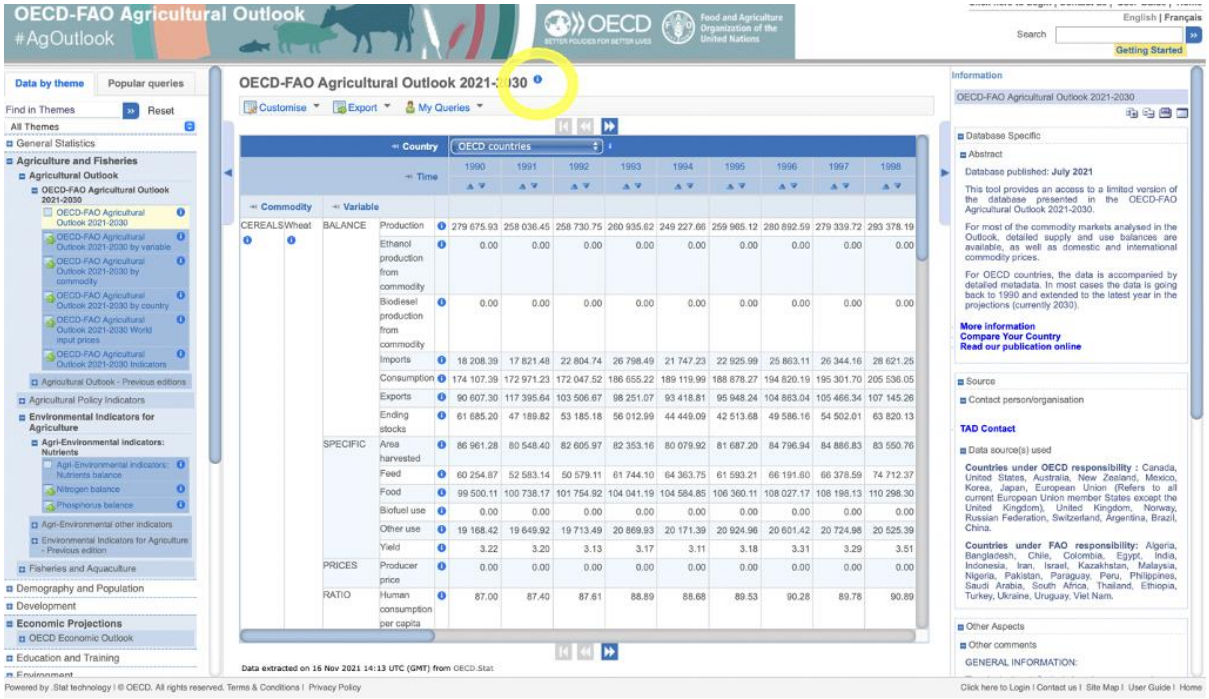
It's the user's responsibility to choose the appropriate theme to find the right document for characterization. Data will appear on the right side of the screen.



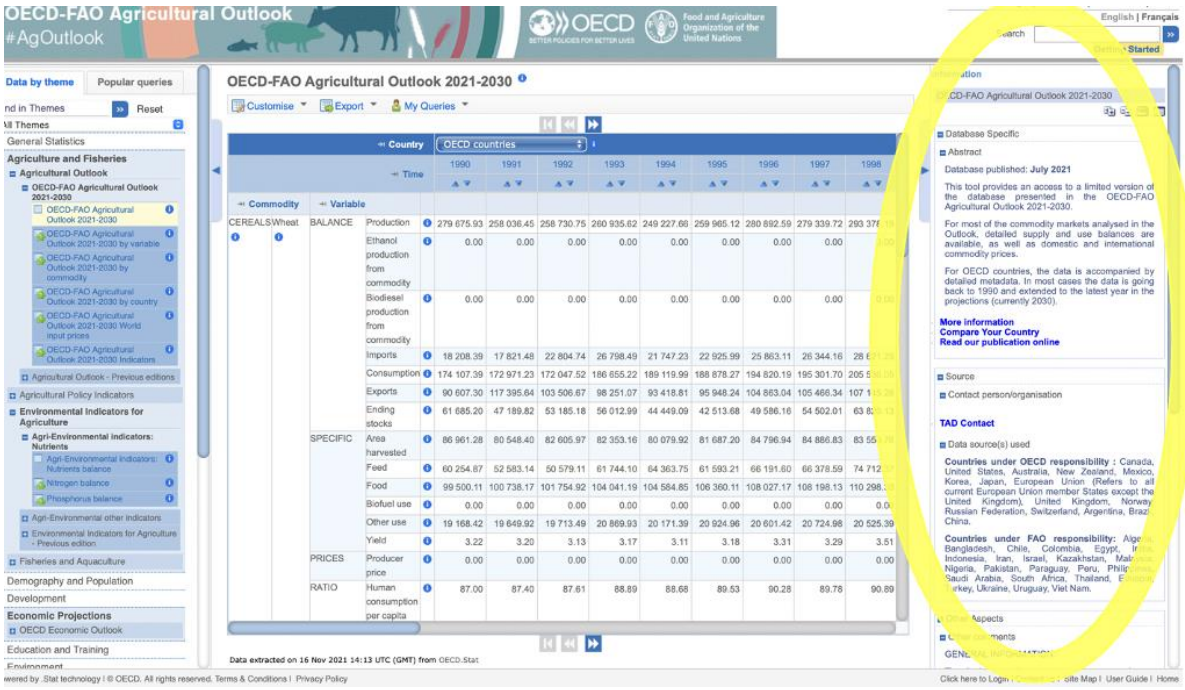
It's possible to select the information needed for a targeted search.



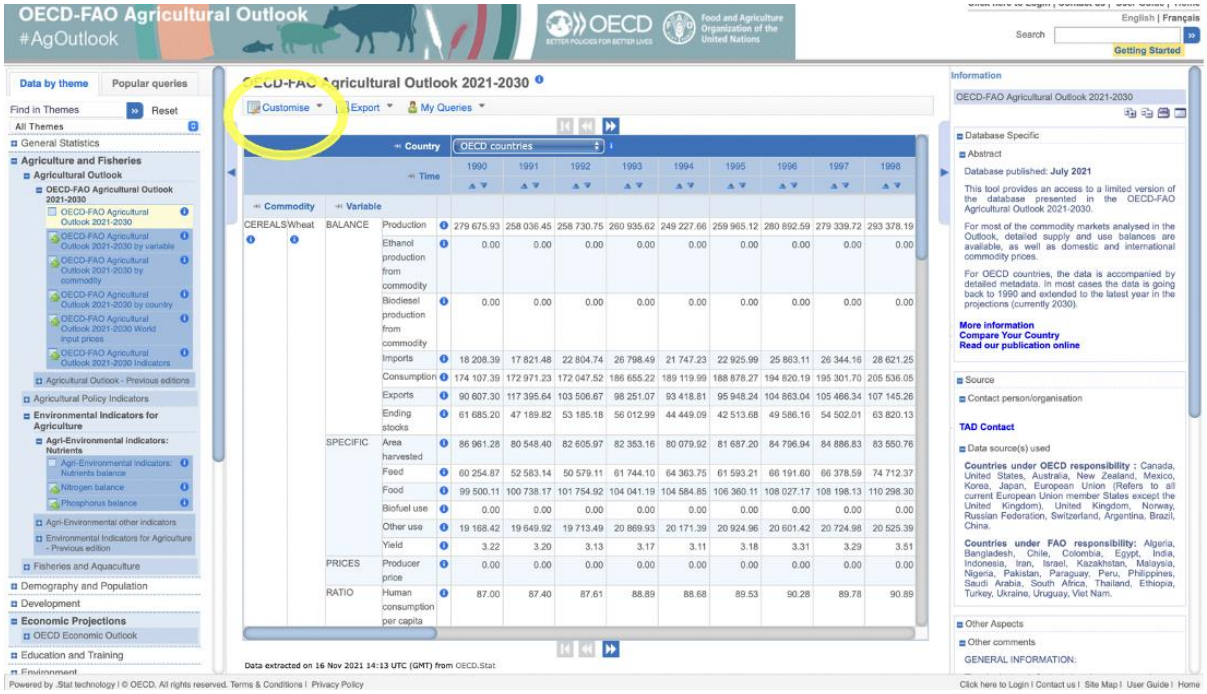
Users can find information for a better characterization by clicking on  .

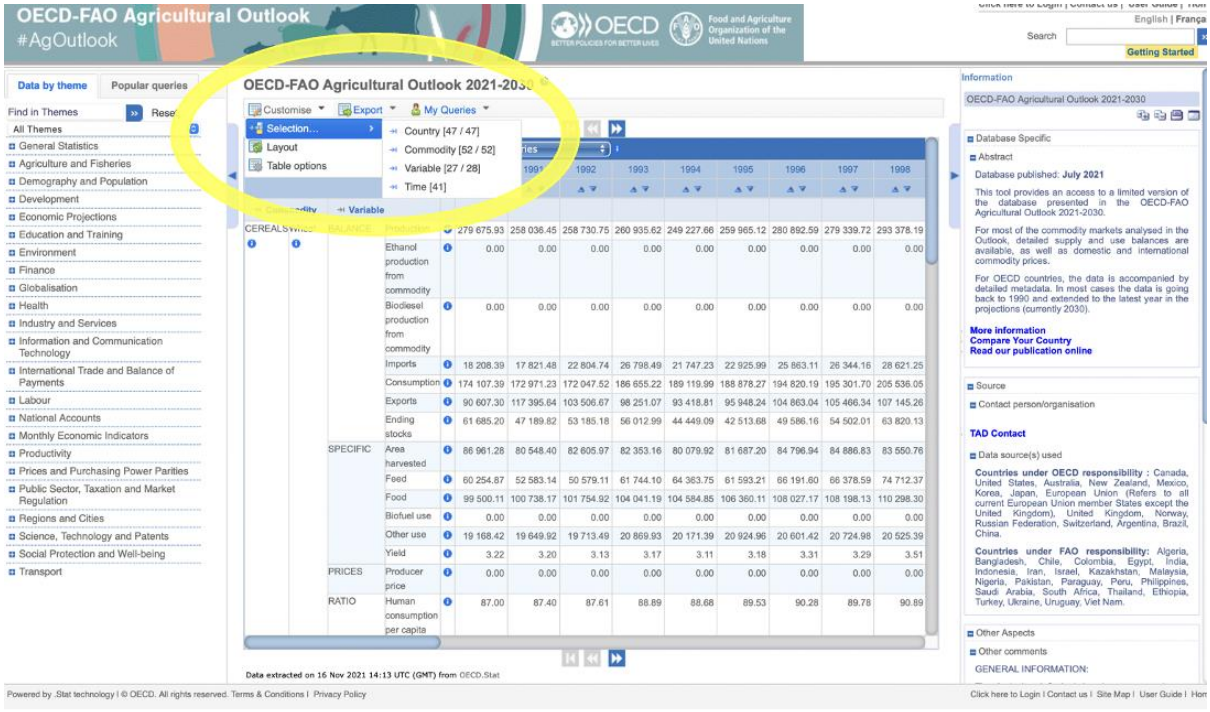


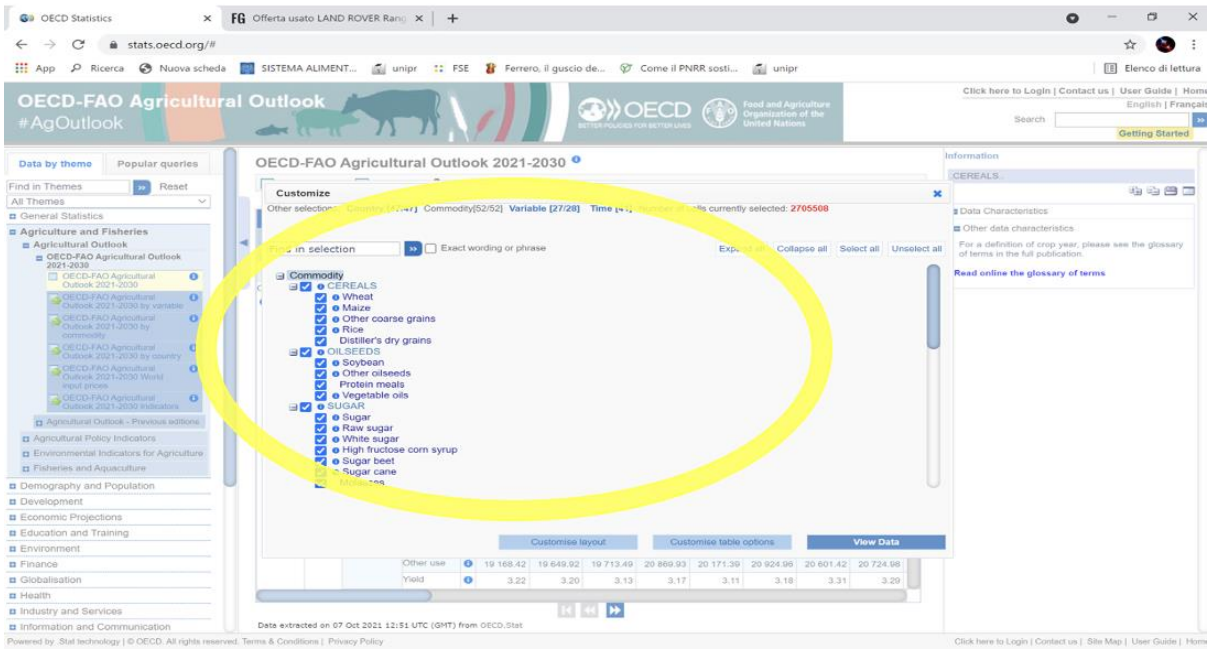
A column will appear on the right side of the screen with the required information.



Users can click on 'Commodity - variable - time' to show variables in a specific dataset. Those variables are useful for a subsequent characterization. It's up to the user choosing relevant elements for the dataset.


OECD-FAO Agricultural Outlook
#AgOutlook
OECD
Food and Agriculture Organization of the United Nations
English | Français
Search
Getting Started
Data by theme
Popular queries
Find in Themes
Reset
All Themes
General Statistics
Agriculture and Fisheries
Agricultural Outlook
OECD-FAO Agricultural Outlook 2021-2030
OECD-FAO Agricultural Outlook 2021-2030
OECD-FAO Agricultural Outlook 2021-2030 by variable
OECD-FAO Agricultural Outlook 2021-2030 by commodity
OECD-FAO Agricultural Outlook 2021-2030 by country
OECD-FAO Agricultural Outlook 2021-2030 World input prices
OECD-FAO Agricultural Outlook 2021-2030 Indicators
Agricultural Outlook - Previous editions
Agricultural Policy Indicators
Environmental Indicators for Agriculture
Agri-Environmental indicators: Nutrients
Agri-Environmental indicators: Nutrients balance
Nitrogen balance
Phosphorus balance
Agri-Environmental other indicators
Environmental Indicators for Agriculture - Previous edition
Fisheries and Aquaculture
Demography and Population
Development
Economic Projections
OECD Economic Outlook
Education and Training
Environment
gricultural Outlook 2021-2030
Customise
Export
My Queries
Country
OECD countries
Time
1990 1991 1992 1993 1994 1995 1996 1997 1998
Commodity
Variable
CEREALS
Wheat
BALANCE
Production 279 675.93 258 036.45 258 730.75 260 935.62 249 227.66 259 965.12 280 892.59 279 339.72 293 378.19
Ethanol production from commodity 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Biodiesel production from commodity 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Imports 18 208.39 17 821.48 22 804.74 26 798.49 21 747.23 22 925.99 25 863.11 26 344.16 28 621.25
Consumption 174 107.39 172 971.23 172 047.52 186 655.22 189 119.99 188 878.27 194 820.19 195 301.70 205 536.05
Exports 90 607.30 117 395.64 103 506.67 98 251.07 93 418.81 95 948.24 104 863.04 105 466.34 107 145.26
Ending stocks 61 685.20 47 189.82 53 185.18 56 012.99 44 449.09 42 513.68 49 586.16 54 502.01 63 820.13
SPECIFIC
Area harvested 86 961.28 80 548.40 82 605.97 82 353.16 80 079.92 81 687.20 84 796.94 84 886.83 83 550.76
Feed 60 254.87 52 583.14 50 579.11 61 744.10 64 363.75 61 593.21 66 191.60 66 378.59 74 712.37
Food 99 500.11 100 738.17 101 754.92 104 041.19 104 584.85 106 360.11 108 027.17 108 198.13 110 298.30
Biofuel use 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Other use 19 168.42 19 649.92 19 713.49 20 869.93 20 171.39 20 924.96 20 601.42 20 724.98 20 525.39
Yield 3.22 3.20 3.13 3.17 3.11 3.18 3.31 3.29 3.51
PRICES
Producer price 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
RATIO
Human consumption per capita 87.00 87.40 87.61 88.89 88.68 89.53 90.28 89.78 90.89
Data extracted on 16 Nov 2021 14:13 UTC (GMT) from OECD.Stat
Information
OECD-FAO Agricultural Outlook 2021-2030
Database Specific
Abstract
Database published: July 2021
This tool provides an access to a limited version of the database presented in the OECD-FAO Agricultural Outlook 2021-2030.
For most of the commodity markets analysed in the Outlook, detailed supply and use balances are available, as well as domestic and international commodity prices.
For OECD countries, the data is accompanied by detailed metadata. In most cases the data is going back to 1990 and extended to the latest year in the projections (currently 2030).
More information
Compare Your Country
Read our publication online
Source
Contact person/organisation
TAD Contact
Data source(s) used
Countries under OECD responsibility : Canada, United States, Australia, New Zealand, Mexico, Korea, Japan, European Union (Refers to all current European Union member States except the United Kingdom), United Kingdom, Norway, Russian Federation, Switzerland, Argentina, Brazil, China.
Countries under FAO responsibility: Algeria, Bangladesh, Chile, Colombia, Egypt, India, Indonesia, Iran, Israel, Kazakhstan, Malaysia, Nigeria, Pakistan, Paraguay, Peru, Philippines, Saudi Arabia, South Africa, Thailand, Ethiopia, Turkey, Ukraine, Uruguay, Viet Nam.
Other Aspects
Other comments
GENERAL INFORMATION:
Powered by .Stat technology | © OECD. All rights reserved. Terms & Conditions | Privacy Policy
Click here to Login | Contact us | Site Map | User Guide | Home





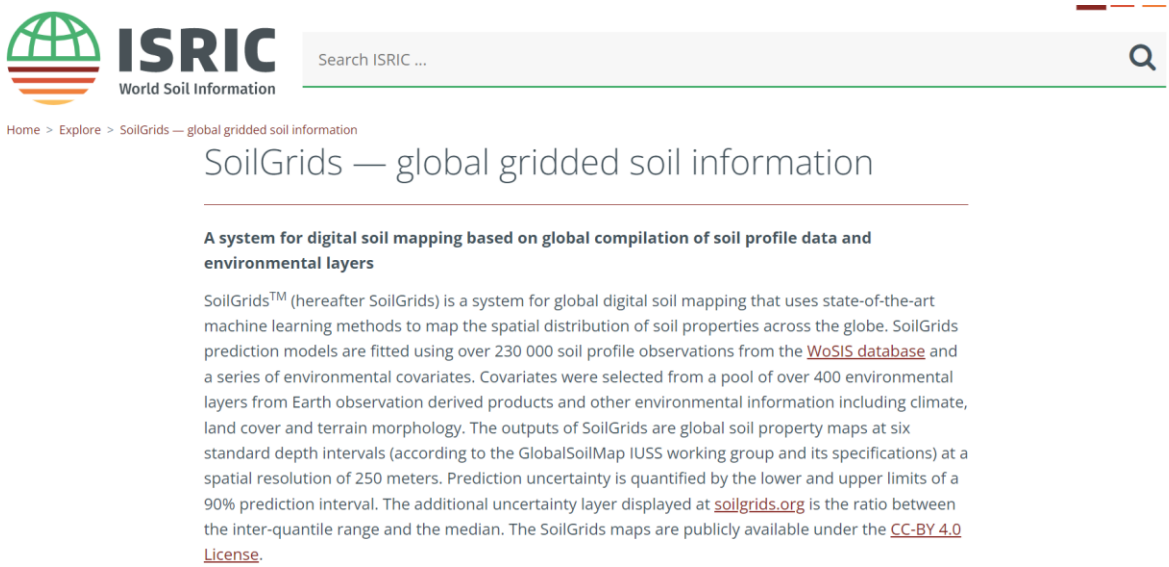
A1.3 Characterisation of 'Geo-referenced' Datasets

A1.3.1 Example 3: SoilGrids250m 2.0

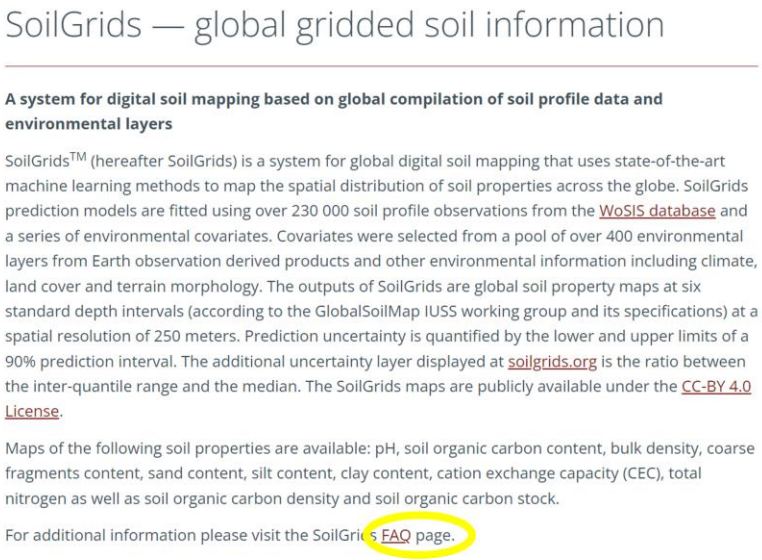
SoilGrids produces maps of soil properties for the entire globe at medium spatial resolution (250 m cell size) using state-of-the-art machine learning methods to generate the necessary models. It takes as inputs soil observations from about 240 000 locations worldwide and over 400 global environmental covariates describing vegetation, terrain morphology, climate, geology and hydrology. The aim of this work was the production of global maps of soil properties, with

cross-validation, hyper-parameter selection and quantification of spatially explicit uncertainty, as implemented in the SoilGrids version 2.0 product incorporating state-of-the-art practices and adapting them for global digital soil mapping with legacy data.

https://www.isric.org/explore/soilgrids



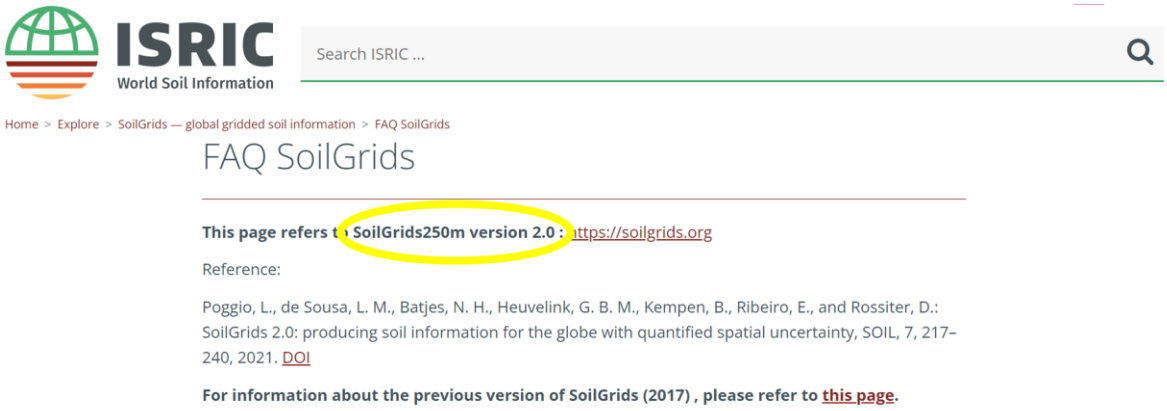
On this page it is possible to find most of the information needed to characterize the dataset, and the missing ones can be found on the FAQ page. For the link the user has to scroll down a little bit the page.



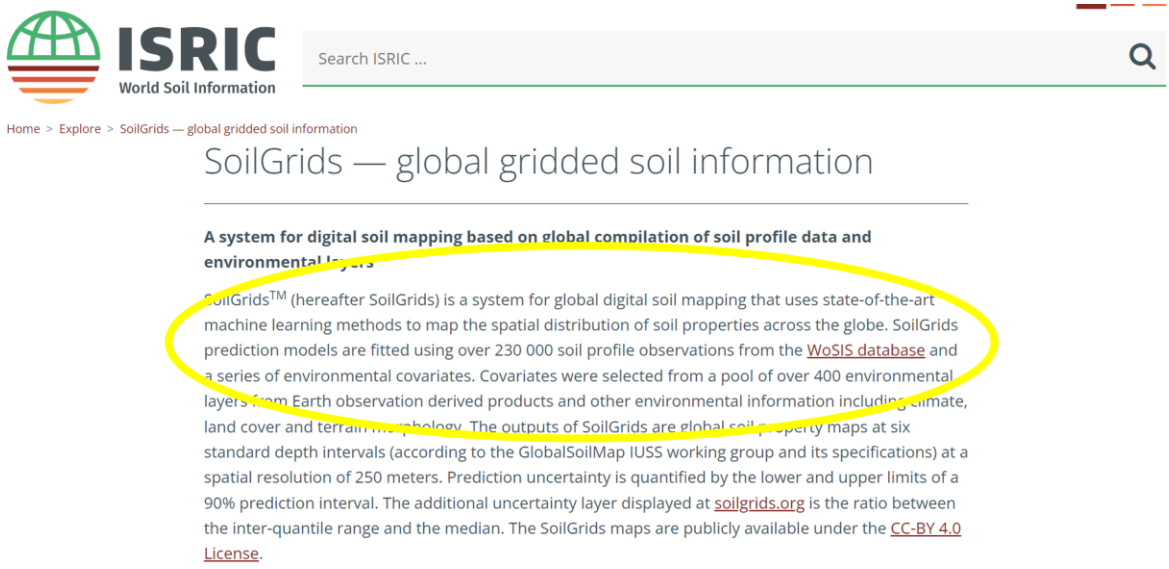
A1.3.1.1 General

- **Title**: SoilGrids250m 2.0.

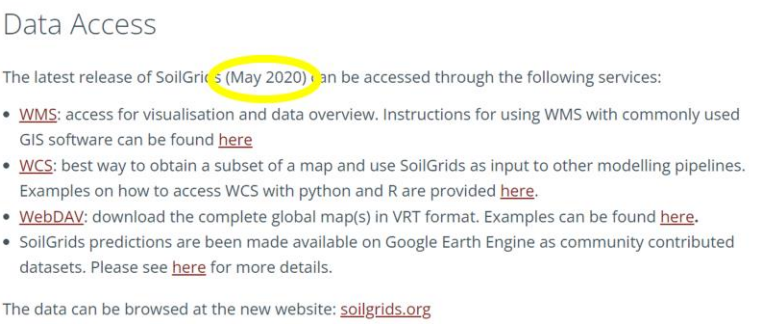
When more than one version is available, it is important to characterize the latest one.



- **Description**: The description can be found in the general page.

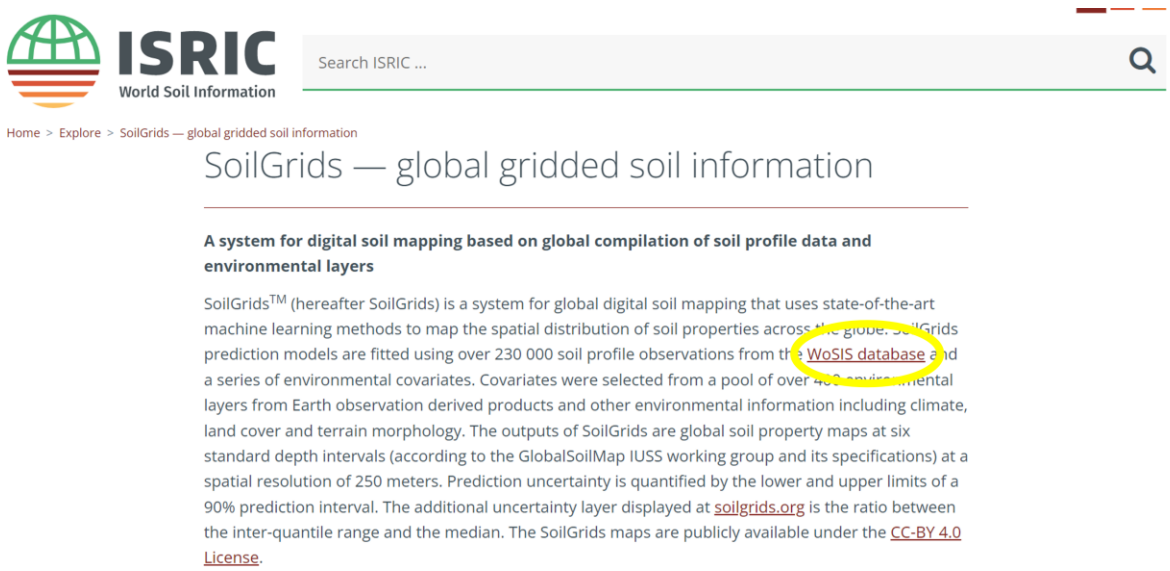


- **Last update**: May 2020. Also this information can be found in the general page.



- **Dataset type**: Georeferenced
- **Producer**: ISRIC

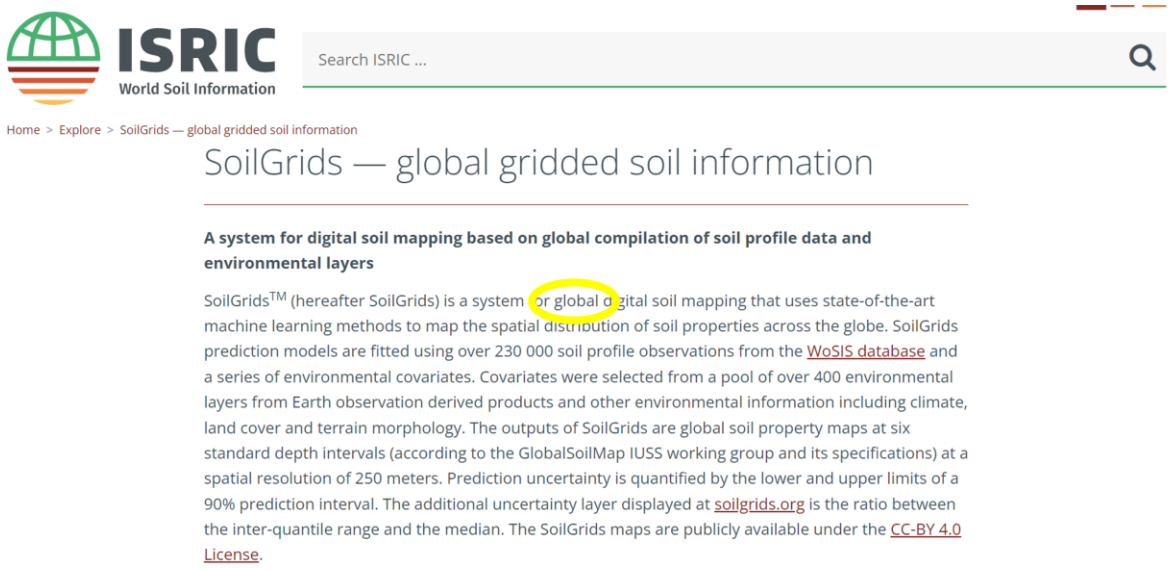
- **Link**: Here the link should refer to the "home" page of the dataset. So, in this case, the URL of the general page: https://www.isric.org/explore/soilgrids
- **Language:** English.
- **Periodicity of the publication:** Irregular (the previous version 1 was published in 2017, version 2 in 2020, and there is no foreseen date for the next publications).
- **Catalogue:** None (the dataset does not belong to a catalogue).
- **Spatial resolution (in meters):** 250. This field is crucial for georeferenced datasets, since it expresses how the data are spatially distributed.
- **Temporal resolution:** NA
- **Was generated by:** SoilGrids is generated by prediction models fitted on data from the WoSIS database. So here we put the link referring to the WoSIS dataset:- https://data.isric.org/geonetwork/srv/eng/catalog.search#/metadata/ca880bd4-cff8-11e9-8046-0cc47adaa92c



A1.3.1.2 Purpose

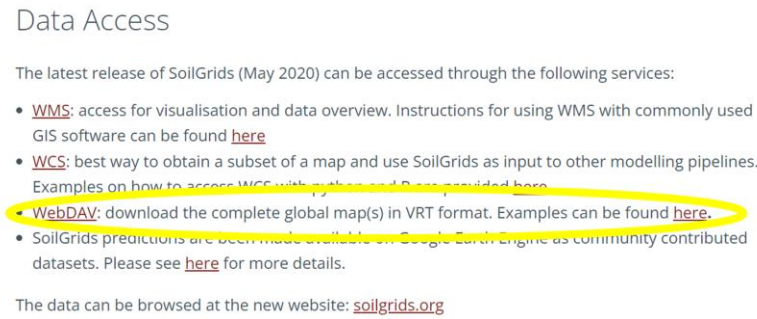
- **Temporal extent:** 05/2020 - 05/2020 (the temporal extent of this dataset is not specified, hence the user can refer to the data of publication).

- **Subject:** soil resources, sample survey, chemical compound (users can choose the most appropriate subject(s) from the drop down menu; the list in the menu is coming from a public vocabulary hence it's not changeable).
- **Useful for the analysis of (purpose):** Environmental policy, Agricultural policy (users can add terms to the vocabulary).
- **Themes covered:** Geographical grid systems, Soil.
- **Geographical coverage**
- **Continental coverage:** Europe, Asia, Africa, America, Antarctica, Oceania. As written in the description, SoilGrids maps spatial distribution of soil properties across the globe, hence the geographical coverage is global (select all continents).



A1.3.1.3 Distributions

Since WebDAV distribution is the only one that gives direct access to the data download, we characterized only this distribution. Clicking in the related "here" we obtain the Access URL.



- **Title:** WebDAV
- **Description:** Each map has three elements: a master VRT file; an OVR file with overviews for faster visualisation; a folder with the GeoTIFF tiles. This description can be found in the FAQ page, where more details are reported.
- **Access rights:** Public (the data download is available to everyone, no restrictions).
- **Access URL:** https://files.isric.org/soilgrids/latest/data/. By clicking on the "WebDAV" link, the user is directed to the menu where the user can select which data table to download. Here it is possible to select the soil property of interest:

Index of /soilgrids/latest/data

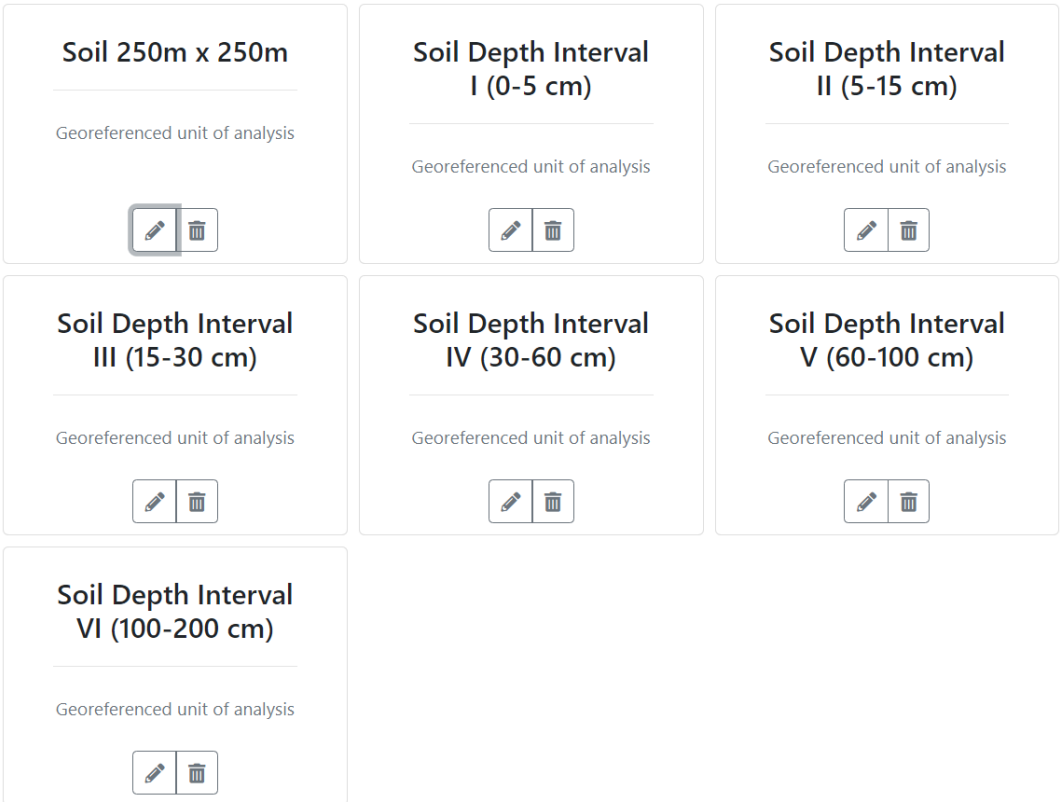
- Parent Directory
- README.md
- bdod/
- cec/
- cfvo/
- clay/
- landmask/
- nitrogen/
- ocd/
- ocs/
- phh2o/
- sand/
- silt/
- soc/
- wrb/

Apache/2.4 Server at files.isric.org Port 8080

- **Format:** XML (VRT is a kinf of XML data format).

A1.3.1.4 Unit of analysis

When the data of georeferenced dataset is related to a grid, the unit of analysis should be represented by the area of the cells of the grid. The outputs of SoilGrids are global soil properties mapped at six standard depth intervals at a spatial resolution of 250 meters, hence in this case the data are related to a 3D grid. In order to express this complex grid, a unit of analysis for each depth interval and one for the spatial resolution has to be created.



- **Area size:** it describes the magnitude of the unit of analysis. This information can be found on the FAQ page. E.g. Soil 250mx250m: 250 square metres; Soil Depth Interval I: 5

centimetres; Soil Depth Interval II: 10 centimetres; Soil Depth Interval III: 15 centimetres; and so on.

Depth intervals

SoilGrids predictions are made for the six standard depth intervals specified in the GlobalSoilMap IUSS working group and its specifications:

	Interval I	Interval II	Interval III	Interval IV	Interval V	Interval VI
Top depth (cm)	0	5	15	30	60	100
Bottom depth (cm	5	15	30	60	100	200

A1.3.1.5 Variables included

SoilGrids — global gridded soil information

A system for digital soil mapping based on global compilation of soil profile data and environmental layers

SoilGrids™ (hereafter SoilGrids) is a system for global digital soil mapping that uses state-of-the-art machine learning methods to map the spatial distribution of soil properties across the globe. SoilGrids prediction models are fitted using over 230 000 soil profile observations from the WoSIS database and a series of environmental covariates. Covariates were selected from a pool of over 400 environmental layers from Earth observation derived products and other environmental information including climate, land cover and terrain morphology. The outputs of SoilGrids are global soil property maps at six standard depth intervals (according to the GlobalSoilMap IUSS working group and its specifications) at a spatial resolution of 250 meters. Prediction uncertainty is quantified by the lower and upper limits of a 90% prediction interval. The additional uncertainty layer displayed at soilgrids.org is the ratio between the inter-quantile range and the median. The SoilGrids maps are publicly available under the CC-BY 4.0 License.

Maps of the following soil properties are available: pH, soil organic carbon content, bulk density, coarse fragments content, sand content, silt content, clay content, cation exchange capacity (CEC), total nitrogen as well as soil organic carbon density and soil organic carbon stock.

For additional information, please visit the SoilGrids FAQ page

The variables are the soil properties mapped in SoilGrids. On the FAQ page there is a table with name, description, mapped units, conversion factor and conventional units of each soil property (see figure below).

Name	Description	Mapped units	Conversion factor	Conventional units
bdod	Bulk density of the fine earth fraction	cg/cm³	100	kg/dm³
cec	Cation Exchange Capacity of the soil	mmol(c)/kg	10	cmol(c)/kg
cfvo	Volumetric fraction of coarse fragments (> 2 mm)	cm3/dm3 (vol‰)	10	cm3/100cm3 (vol%)
clay	Proportion of clay particles (< 0.002 mm) in the fine earth fraction	g/kg	10	g/100g (%)

- **Name:** in order to be searchable through e semantic research, as name it is important to utilise an understandable name. So, in this case, for each soil property the user can use the "Description" field of the table.
- **Unit of measurement:** the "mapped units" of the table, which represent the unit of measurement of the reported data.
- **Temporal extent:** also in this case, as for the Temporal extent of the dataset, it is not specified.
- **Mathematical representation:** not specified.
- **Data origin:** forecast. Indeed, SoilGrids' data are the result of machine learning methods.
- **Unit of analysis:** since the values of the soil properties are available for all the points in the 3D grid, all variables are related to all units of analysis.

ANNEX II: Frameworks programming languages, tools and libraries

A2.1 Backend side

A2.1.1 API endpoints

ARDIT platform uses the open source framework Swagger [20], which provides a set of rules, specifications and tools to document and support the consumption of the platform's API. Allowing users and developers to know in detail the defined endpoints and how HTTP requests are constructed to connect to them, helping non-expert users to access the resources provided by the API without the need for in-depth knowledge of REST or the HTTP protocol. Swagger UI, ARDIT API endpoints and all of the documentation is available through https://www.ardit-api.agricore-project.eu unlike the graphical user interface, which is through https://www.ardit.agricore-project.eu.

A2.1.2 Docker

Open source tool that enables the automation of application deployments within software containers on any machine, independently of its operating system or software. It simplifies deployments in highly distributed systems and also provides software projects with the advantage that all developers involved can work with the same versions of the tools, languages or libraries that compose the project both on their local computers and in the different development environments [21].

To enable its deployment, ARDIT uses Docker Compose [22], a tool that simplifies the use of Docker with YAML format files, where all the services and tools that compose the project are defined along with ports, volumes for persistence, connectivity between containers, etc. In this way, the entire deployment of the project can be automated through a simple command that creates and initialises all the services defined in the file.

ARDIT uses various YAML files for deployment depending on the development environment in which the project must be deployed (local, test, development, production).

A2.1.3 Spring Framework

Spring [23] is the open source framework par excellence for the development of enterprise applications using Java programming language. It has a modular structure and great flexibility when implementing different types of architectures, being especially useful in the development of complex, scalable and distributed applications, where aspects such as security, maintenance and reusability of components are essential. Some of the main Spring modules are:

- **Spring Framework**: module that provides the framework itself. It includes features like:
    - **Spring Core**: technologies that enable the reutilisation of code through dependency injection, validation, Aspect Oriented Programming (AOP), etc.
    - **Data access**: support for transactions, Data Access Objects (DAO), Object-Relational Mapping (ORM), Java Database Connectivity (JDBC), etc.
    - **Spring MVC**: support for web development and Model-View-Controller architectures.
- **Spring Boot**: provides dependencies to simplify project building configurations.

- **Spring Data**: provide a programming model for data access, making it easy to use data access technologies, relational and non-relational databases.
- **Spring Security**: provides authentication and authorization principles to Java applications and protection against different types of attacks.
- **Spring Cloud**: provides tools to build patterns in distributed systems.
- **Spring HATEOAS**: provides some APIs to ease creating REST representations that follow the HATEOAS principle.
- **Spring Batch**: enable the development of batch applications.

Figure 32 shows the modules that compose Spring Framework [24]:

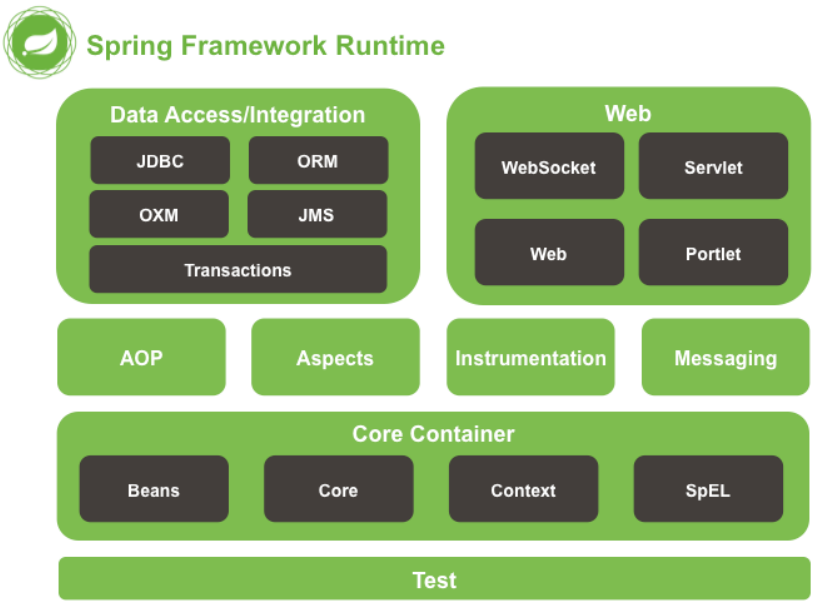


Figure 32 Spring framework overview

ARDIT platform makes use of some of these modules such as **Core**, **Boot**, **Data**, **Security**, **Web**, including **Spring LDAP**, which simplifies the development of applications that use LDAP, using Spring Framework 5 version with Spring Boot 2.3.0 version and Java 14 (Java SE 14).

A2.1.4 PostgreSQL

PostgreSQL [25] is an open source relational database management system (RDBMS), which is characterised by its high scalability, stability and robustness, complying with the ACID (Atomicity, Consistency, Isolation and Durability) rules:

- **Atomicity**: if a transaction consists of a block or sequence of operations, they must all be successfully completed for the results to be stored in the database. Otherwise, all operations will be undone (rollback).
- **Consistency**: a transaction will only be executed if it will not break the integrity rules of the database to always ensure that the data is accurate and consistent.

- **Isolation**: the execution of two transactions on the same information must be independent and not affect each other.
- **Durability**: the results of a completed transaction shall remain in the database even when failures of any type may occur.

ARDIT uses PostgreSQL in its version 13.2 to persist all information and data related to dataset characterisations.

A2.1.5 LDAP

The Lightweight Directory Access Protocol (LDAP), as its name suggests, is a lightweight protocol for accessing directory services, which are specialized databases optimized for read access, that contain attribute-based information and filtering capabilities. LDAP is a good solution when projects require data to be centrally managed, stored and accessible, being some of its most important uses machine and user authentication, user groups definition, organization representation, user resource management, etc.

One of the most important aspects of LDAP is the way it organises the information, which can be described as follows:

"*The LDAP information model is based on entries. An entry is a collection of attributes that has a globally-unique Distinguished Name (DN). The DN is used to refer to the entry unambiguously. Each of the entry's attributes has a type and one or more values. The types are typically mnemonic strings, like "cn" for common name, or "mail" for email address. The syntax of values depend on the attribute type. For example, a cn attribute might contain the value Babs Jensen. A mail attribute might contain the value "babs@example.com". A jpegPhoto attribute would contain a photograph in the JPEG (binary) format...In LDAP, directory entries are arranged in a hierarchical tree-like structure*" [26]

The following picture, Figure 33, shows an example LDAP directory tree where the information is classified in the following order:

1. Country: Great Britain and Unit States.
2. State: California.
3. Organisation: Acme.
4. Department: Sales and Marketing.
5. Employee: Barbara Jenson.

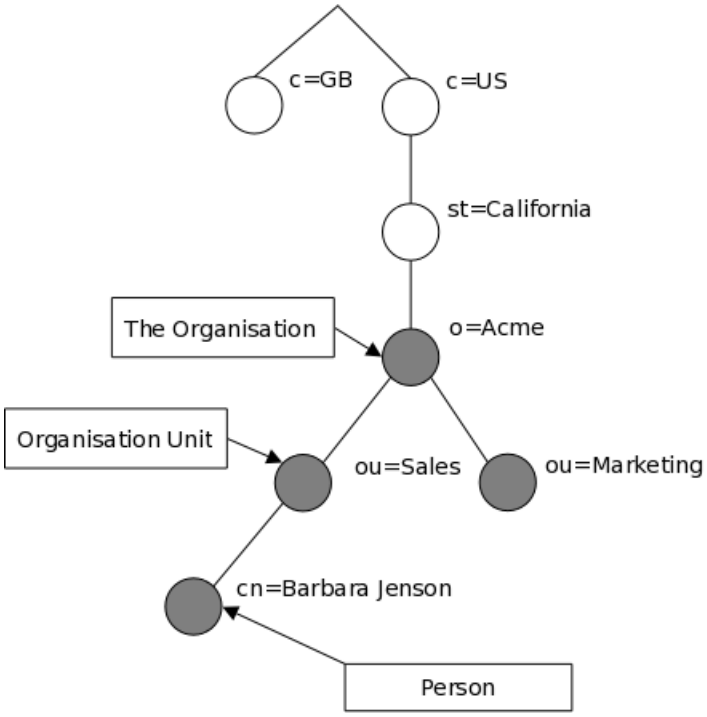


Figure 33 LDAP tree structure example

Each entry can be referenced and identified from others by its "distinguished name" (DN), which consists of a concatenation between the name of the entry and the names of its ancestor entries. In the example, a concatenation between the name (cn), the organisation unit/department (ou), the organisation name (o) and the rest of the entries. Through this identification, LDAP provides methods for accessing, creating, updating or deleting entries by search and modification capabilities.

The AGRICORE project and ARDIT use **OpenLDAP** [27] on its version 1.3.0, which is an open source implementation of the LDAP protocol developed by the Internet community and distributed using the OpenLDAP Public License, to store and manage user data, including credentials, in the project.

A2.2 Frontend side

A2.2.1 Angular

Open source framework developed by Google for webs written in TypeScript, used for complete separation of frontend and backend or to build single page applications (SPA). Angular allows to build the frontend side of a web application, in other words, the web client that consumes a REST service (backend). Being both two separate and independent entities, but interact with each other. It is also a modular and scalable framework, which, being based on the web component standard, facilitates code reutilisation [28].

In order to containerise Angular and deploy it, it is necessary to compile and build the project with Node [29] and use Nginx [30] to host the Angular build inside the Docker container. At the time of generating this documentation, ARDIT uses Angular version 10.0.14, Node 14.9 and Nginx 1.21.6.

A2.2.2 NG Bootstrap

Bootstrap [31] is an open source library for web design. It contains templates for different components and elements of modern websites, such as forms, menus, navigation bars, alerts or buttons. All of them with HTML and CSS based designs. ARDIT makes use of the NG Bootstrap library, an implementation of Bootstrap for Angular that allows developers to use all the components of this library without the need to use jQuery code. All the code necessary to control and manage the library's components and elements is done with TypeScript.

ARDIT uses the NG Bootstrap [32] library on its 8.0.1 version over the fourth version of Bootstrap (v4). In mid 2021, Bootstrap released the fifth version of its library (v5), which could be updated in the tool in future revisions.

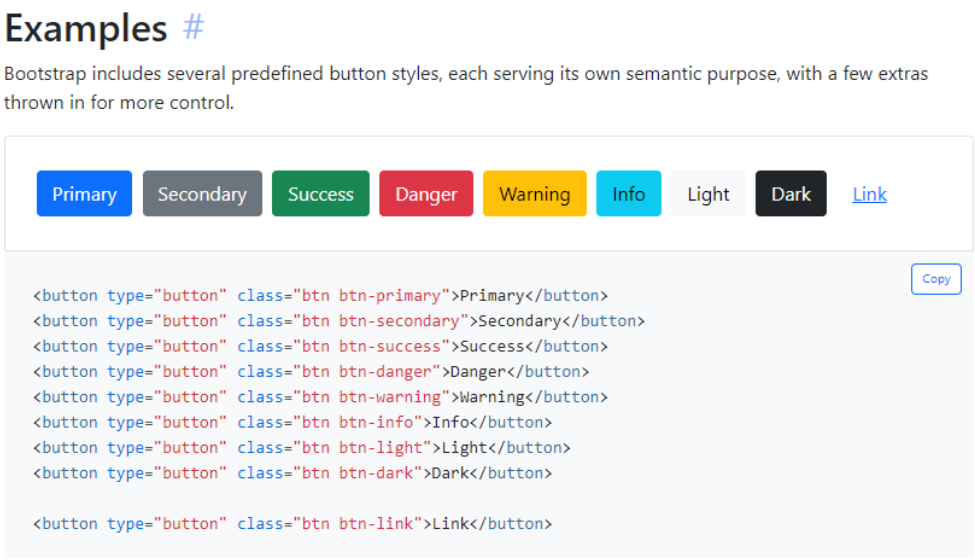


Figure 34 Example of Bootstrap buttons

A2.2.3 NG Select

It is a library for Angular distributed under MIT license, that allows to add to the projects a powerful drop-down menu for the selection of one or multiple options, with many customisation and configuration options [33]. Among its main features it includes:

- Custom binding to property or object.
- Scroll support with large data sets (>5000 items).
- Keyboard navigation.
- Multiselect.
- Custom search and filtering.
- Accessibility.
- Theme customization.

ARDIT uses NG Select 5.0.9 version at the time of writing this document but may be updated in the future if it is required.

Figure 35 Example of NG Select menu in ARDIT

A2.2.4 NGX Pagination

Library available for Angular 4 versions onwards, which is distributed under the MIT license [34]. It provides the necessary components to incorporate pagination to a website, that is, the organisation and separation of a set of elements in a certain number of pages, showing a fixed amount of elements in each one of them. ARDIT uses NGX Pagination 5.1.1 version at the time of writing this document but may be updated in the future if it is required.

Figure 36 ARDIT pagination component

A2.2.5 Font Awesome

Font Awesome [35] is a library of icons for web development that is distributed under different licenses. The free version, which is open source, provides a smaller but still sufficient set of icons than the paid versions. ARDIT uses the 5.15.1 free version of this library.

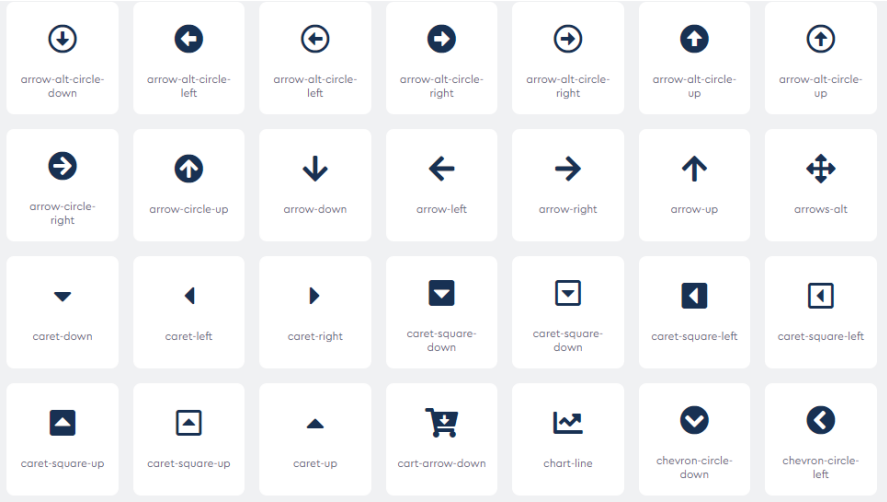


Figure 37 Font Awesome icons examples

ANNEX III: Guide to the ARDIT data model

The following tables represent the translation of the structure of classes and properties defined in the AGRICORE DCAT-AP 2.0 ontology to the ARDIT database scheme.

A3.1 Dataset

Table 2 Dataset properties

Name of the attribute	Corresponds to ClassName (ARDIT data model)	Corresponds to PropertyName (AGRICORE DCAT-AP Data Model)	Data Type	Controlled Vocabulary (if applicable)	Example
Title	Dataset: title	dcterms:title : rdfs:Literal	String		`raw_value = "Price indices of the means of agricultural production, input (2015 = 100) - annual data"` representation: Price indices of the means of agricultural production, input (2015 = 100) - annual data
Link	Dataset: link	dcat:landingPage : foaf:Document	String		`raw_value = "https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=apri_pi15_ina&lang=en"` representation: "https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=apri_pi15_ina&lang=en"
Creation date timestamp (Characterisation in the plataform)	Dataset: creationDateTime		Date		`raw_value = "2021-04-19T14:11:35.506+00:00"` **representation**: 2021-04-19T14:11:35.506+00:00
Last update timestamp (Characterisation in the platform)	Dataset: lastUpdateDateTime		Date		`raw_value = "2021-04-19T14:11:35.506+00:00"` **representation**: 2021-04-19T14:11:35.506+00:00
Issued	Catalogue: issued	dcterms:issued : xsd:dateTime	Date		`raw_value = "2021-04-19"` **representation**:

					2021-04-19
Modified	Catalogue: modified	dcterms:modified : xsd:dateTime	Date		**raw_value** = "2021-05-04" **representation**: 2021-05-04
Producer	Dataset: producer	dcterms:creator : foaf:Agent	String		raw_value = "European Commission" representation: European Commission
Catalogue	Dataset: catalogue	Agricore: catalogue	Catalogue		WIP
Type of data set (subjects)	Dataset: subjects		List<VocabularyValue>	https://op.europa.eu/en/web/eu-vocabularies/at-concept-scheme/-/resource/authority/det/?target=Browse&uri=http://data.europa.eu/uxp/det	raw_value = [{ "uid": "1010", "url": "http://data.europa.eu/uxp/1010", "label": "export financing" }, { "uid": "1788", "url": "http://data.europa.eu/uxp/1788", "label": "agricultural market" },] representation: export financing, agricultural market

Languages	Dataset: language	dcterms:language : dcterms:LinguisticSystem	List<VocabularyValue>	https://id.loc.gov/vocabulary/iso639-2.html	raw_value = [
					{
					"code": "eng",
					"url": "http://id.loc.gov/vocabulary/iso639-2/eng",
					"label": "English"
					},
					{
					"code": "spa",
					"url": "https://id.loc.gov/vocabulary/iso639-2/spa",
					"label": "Spanish"
					}
					]
					
					representation:
					English, Spanish

Periodicity of publications	Dataset: periodicity	dcat:accrualPeriodicity : dcterms:Frequency	VocabularyValue	https://www.dublincore.org/specifications/dublin-core/collection-description/frequency/	raw_value = { "uid": "annual", "url": "http://purl.org/cld/freq/annual", "label": "Annual" } representation: Annual
Temporal extent	• Dataset: tmpExtentFrom • Dataset: tmpExtentTo	dcterms:temporal : dcterms:PeriodOfTime	Date Date		raw_value = { "from": 2015/03/17, "to": 2019/11/03} representation: 2015 - 2019

Useful for the analysis of (purposes)	Dataset: purposes	datasetPurpose : DatasetPurpose	List<VocabularyValue>	Intenal vocabulary	raw_value= [
					{
					"uid": "ENVP",
					"url": "",
					"label": "Environmental policy"
					},
					{
					"uid": "EUIA",
					"url": "",
					"label": "Energy use in agriculture"
					}
					]
					representation:
					Environmental policy, Energy use in agriculture

Themes covered	Dataset: themes	dcat:theme : skos:Concept	List<VocabularyValue>	More than one vocabulary is used: • http://publications.europa.eu/resource/authority/data-theme • https://inspire.ec.europa.eu/theme/	raw_value = [{ "uid": "ECON", "url": "http://publications.europa.eu/resource/authority/data-theme/ECON", "label": "Economy and finance" }, { "uid": "ENVI", "url": "http://publications.europa.eu/resource/authority/data-theme/ENVI", "label": "Environment" },]

					representation: Economy and finance, Environment
Continental coverage (geocoverage)	Dataset: continentalCoverage	geoCoverages : dcterms:Location	List<VocabularyValue>	http://publications.europa.eu/resource/authority/continent	raw_value = { "uid": "EUROPE", "url": "http://publications.europa.eu/resource/authority/continent/EUROPE", "label": "Europe" } representation: Europe

Country coverage (geocoverage)	Dataset: countryCoverage	geoCoverages : dcterms:Location	List<VocabularyValue>	http://publications.europa.eu/resource/authority/country	raw_value = [
					{
					"uid": "ESP",
					"url":
					"http://publications.europa.eu/resource/authority/country/ESP",
					"label": "Spain"
					},
					{
					"uid": "ITA",
					"url":
					"http://publications.europa.eu/resource/authority/country/ITA",
					"label": "Italy"
					}
					]
					
					representation:
					Spain, Italy

NUTS1 (geocoverage)	Dataset: nuts1	geoCoverages : dcterms:Location	List<VocabularyValue>	http://nuts.geovocab.org/id/nuts1.html	raw_value = [
					{
					"uid": "ES6",
					"url":
					"http://nuts.geovocab.org/id/ES6",
					"label": "SUR"
					},
					{
					"uid": "ITD",
					"url":
					"http://nuts.geovocab.org/id/ITD",
					"label": "NORD-EST"
					}, {
					"uid": "ITC",
					"url":

					"http://nuts.geovocab.org/id/ITC", "label": "NORD-OVEST" }] representation: ES6, ITD, ITC
NUTS2 (geocoverage)	Dataset: nuts2	geoCoverages : dcterms:Location	List<VocabularyValue>	http://nuts.geovocab.org/id/nuts2.html	raw_value = [{ "uid": "ES61", "url": "http://nuts.geovocab.org/id/ES61", "label": "Andalucía" }, { "uid": "ITD5", "url": "http://nuts.geovocab.org/id/ITD5", "label": "Emilia-Romagna" },] representation: ES61, ITD5
NUTS3 (geocoverage)	Dataset: nuts3	geoCoverages : dcterms:Location	List<VocabularyValue>	http://nuts.geovocab.org/id/nuts3.html	raw_value = [{ "uid": "ES618", "url": "http://nuts.geovocab.org/id/ES618", "label": "Sevilla" }, { "uid": "ITD52", "url": "http://nuts.geovocab.org/id/ITD52", "label": "Parma" },] representation: ES618, ITD52
Geonames ADM1(geocoverage)	Dataset: adm1	geoCoverages : dcterms:Location	List<VocabularyValue>	• Source (allCountries.zip file): (https://download.geonames.org/export/dump/)	raw_value = [{ "uid": "11071623", "url":

				• ADM1 values: https://www.geonames.org/advanced-search.html?q=ADM1&country=&featureClass=&continentCode=EU	"https://www.geonames.org/11071623/occitanie.html", "label": "Occitanie" }, { "uid": "11071621", "url": "https://www.geonames.org/11071621/normandie.html", "label": "Normandie" },] representation: Occitanie, Normadie
Geonames ADM2(geocoverage)	Dataset: adm2	geoCoverages : dcterms:Location	List<VocabularyValue>	• Source (allCountries.zip file): (https://download.geonames.org/export/dump/) • ADM2 values: https://www.geonames.org/advanced-search.html?q=ADM2&country=&featureClass=&continentCode=EU	raw_value = [{ "uid": "2516394", "url": "https://www.geonames.org/2516394/provincia-de-jaen.html", "label": "Provincia de Jaén" }, { "uid": "2510910", "url": "https://www.geonames.org/2510910/provincia-de-sevilla.html", "label": "Provincia de Sevilla" },] representation: Provincia de Jaén, Provincia de Sevilla
Statistical representativeness **(HIDDEN at user level)**	Dataset: statsRepresentativeness	Agricore-dcatap: statsRepresentativeness	Integer		raw_value = 100 representation: 100
Aggregation level **(HIDDEN at user level)**	Dataset: aggregationDistance	aggregationLevel : AggregationLevel	Integer		raw_value = 1000 representation: 1000

Aggregation level unit **(HIDDEN at user level)**	Dataset: aggregationGeoReference	aggregationLevel : AggregationLevel	VocabularyValue	• http://publications.europa.eu/resource/authority/measurement-unit
				• Internal vocabulary using the following list of values:
				○ Continental
				○ Country
				○ NUTS 1, NUTS 2, NUTS 3
				○ Geonames
					raw_value = {
					"uid": "COU",
					"url": ",
					"label": "Country"
					}
					representation:
					Country

Aggregation scale **(HIDDEN at user level)**	Dataset: aggregationScale	aggregationLevel : AggregationLevel	Integer		raw_value = 1000 representation: 1000
Unit of analysis	Dataset: analysisUnits		List<AnalysisUnit> • *See 'Analysis unit class' table*		• *See 'Unit of analysis class' table*
Dataset variables	Dataset: variables		List<DatasetVariable> • *See 'Dataset variable class' table*		• *See 'Dataset variable class' table*
Distribution	Dataset: distributions	dcat:distribution : dcat:Distribution	List<Distribution> • *See 'Distribution class' table*		• *See 'Distribution class' table*
Spatial resolution in meters	Dataset: spatialResolutionInMeters	dcat:spatialResolutionInMeters : xsd:decimal	Float		raw_value = 1000 representation: 1000
Temporal resolution	Dataset: temporalResolution	dcat:temporalResolution : xsd:duration	String		raw_value = "PT15M" representation: PT15M
Was generated by	Dataset: wasGeneratedBy	prov:wasGeneratedBy : prov:Activity	List<DatasetGenerationActivity> • *DatasetGenerationActivity: new class representing an activity that*		raw_value = [{ "id": 1, "label": "https://www.example.com" }, { "id": 2, "label": "https://www.example-2.com"

			generated the dataset		}] representation: https://www.example.com, https://www.example-2.com
Keywords	Dataset: keywords	dcat:keyword : rdfs:Literal	List<Keyword>		raw_value = [{ "id": 12, "label": "land cover" }, { "id": 13, "label": "land use" }] representation: land cover, land use
Description	Dataset: description	dcterms:description : rdfs:Literal	String		raw_value = { "description": "LUCAS stands for the Land Use and Coverage Area frame Survey. Eurostat has carried out this survey every 3 years since 2006 to identify changes in the European Union in land use and land cover" } representation: LUCAS stands for the Land Use and Coverage Area frame Survey. Eurostat has carried out this survey every 3 years since 2006 to identify changes in the European in Union land use and land cover
Is referenced by	Dataset: isReferencedBy	dcterms:isReferencedBy : dcat:Resource	List<DatasetReferencedResource> • *DatasetReferencedResource: new class representing reosurces that point to the dataset*		raw_value = [{ "id": 1, "label": "https://www.example.com" }, { "id": 2, "label": "https://www.example-2.com" }] representation: https://www.example.com, https://www.example-2.com

Resource type	Dataset: resourceType	dcterms:type : rdfs:Resource	VocabularyValue	ttps://www.dublincore.org/specifications/dublin-core/dcmi-terms/	raw_value = {
					"uid": "DAT",
					"url": "http://purl.org/dc/dcmitype/Dataset",
					"label": "Dataset"
					}
					representation:
					Dataset

A3.2 Distribution

Table 3 Distribution properties

Name of the attribute	Corresponds to ClassName (ARDIT data model)	Corresponds to PropertyName (AGRICORE DCAT-AP Data Model)	Data Type	Controlled Vocabulary (if applicable)	Example
Title	Distribution: title	dct: title	string		raw_value = "JSON" representation: JSON
Description	Distribution: description	dct: description	string(1000)		raw_value = "JSON format distribution of the dataset" representation: JSON format distribution of the dataset
Create at	DataService: createAt	dct: createAt	Date		raw_value = "01-16-2018" representation: 01/16/2018
Last update	DataService: lastUpdate	dct: lastUpdate	Date		raw_value = "08-31-2021" representation: 08/31/2021
License	Distribution: license	dct: license	string (URI)		raw_value = "https://creativecommons.org/publicdomain/zero/1.0/legalcode" representation: https://creativecommons.org/publicdomain/zero/1.0/legalcode

Access rights	Distribution: accessRights	dct: accessRights	VocabularyValue	http://publications.europa.eu/resource/authority/access-right	raw_value = { "uid": "PUBLIC", "url": "http://publications.europa.eu/resource/authority/access-right/PUBLIC", "label": "public" } representation: public
Procedures to access	Distribution: accessProcedures		string(1000)		raw_value = "On request after registration (online)" representation: On request after registration (online)
Access URL	Distribution: accessURL	dcat: accessURL	string		raw_value = "https://www.example.com" representation: https://www.example.com
Access service	Distribution: accessService	dcat: accessService	DataService • *See 'Data service class' table*		• *See 'Data service class' table*
Download URL	Distribution: downloadURL	dcat: downloadURL	string		raw_value = "https://www.example.com" representation: https://www.example.com
Byte size	Distribution: byteSize	dcat: byteSize	float		raw_value = 2000.0 representation: 2000.0
Format	Distribution: format	dct: format	VocabularyValue	http://publications.europa.eu/resource/authority/file-type	raw_value = { "uid": "XML", "url": "http://publications.europa.eu/resource/authority/file-type/XML", "label": "XML" } representation: JSON
Compress format	Distribution: compressFormat	dcat: compressFormat	VocabularyValue	http://publications.europa.eu/resource/authority/file-type	raw_value = { "uid": "ZIP", "url": "http://publications.europa.eu/resource/authority/f

					ile-type/ZIP", "label": "ZIP" } representation: ZIP

Packaging format	Distribution: packagingFormat	dcat: packagingFormat	VocabularyValue	http://publications.europa.eu/resource/authority/file-type	raw_value = {
					"uid": "TAR",
					"url":
					"http://publications.europa.eu/resource/authority/file-type/TAR",
					"label": "TAR"
					}
					representation:
					TAR

A3.3 Data service

Table 4 Data service properties

Name of the attribute	Corresponds to ClassName (ARDIT data model)	Corresponds to PropertyName (AGRICORE DCAT-AP Data Model)	Data Type	Controlled Vocabulary (if applicable)	Example
Title	DataService: title	dct: title	String		raw_value = "Data service example title" representation: Data service example title
Description	DataService: description	dct: description	String(1000)		raw_value = "Data service example description" representation: Data service example description
Create at	DataService: createAt	dct: createAt	Date		raw_value = "01-16-2018" representation: 01/16/2018
Last update	DataService: lastUpdate	dct: lastUpdate	Date		raw_value = "08-31-2021" representation: 08/31/2021

Creator	DataService: creator	dct: creator	String		raw_value = "EUROSTAT" representation: EUROSTAT
Publisher	DataService: publisher	dct: publisher	String		raw_value = "EUROSTAT" representation: EUROSTAT
AccessRights	DataService: accessRights	dct: accessRights	VocabularyValue	http://publications.europa.eu/resource/authority/file-type	raw_value = { "uid": "PUBLIC", "url": "http://publications.europa.eu/resource/authority/access-right/PUBLIC", "label": "public" } representation: public
Endpoint URL	DataService: endpointURL	dcat: endpointURL	String		raw_value = "https://www.example.com" representation: https://www.example.com
Endpoint description	DataService: endpointDescription	dcat: endpointDescription	String		raw_value = "https://www.example.com" representation: https://www.example.com
Served datasets	DataService: servedDatasets	dcat: servesDataset	List<String>		raw_value = [{ https://ardit-agricore.idener.es/datasets/45 }, { https://ardit-agricore.idener.es/datasets/97 }] representation: https://ardit-agricore.idener.es/datasets/45, https://ardit-agricore.idener.es/datasets/97

A3.4 Socio-economic unit of analysis

Table 5 Socio-economic unit of analysis properties

Name of the attribute	Corresponds to ClassName (ARDIT data model)	Corresponds to PropertyName (AGRICORE DCAT-AP Data Model)	Data Type	Controlled Vocabulary (if applicable)	Example

Unit of reference	AnalysisUnit: unitReference	unitReference : AnalysisUnitReference	String*	• TBD	`raw_value = "5 arc-minutes grid cell"`
					representation:
					`5 arc-minutes grid cell`

Temporal extent	AnalysisUnit: tmpExtentFrom AnalysisUnit: tmpExtentTo	dcterms:temporal : dcterms:PeriodOfTime	Date Date		`raw_value = { "tmpExtentFrom": 2015/03/17, "tmpExntentTo": 2019/11/03}` **representation**: `March,2015` - November, 2019

Area size	• AnalysisUnit: areaSizeValue
	• AnalysisUnit: areaSizeUnit
		sizeUnit: SizeUnit	• Integer
			• VocabularyValue
				http://publications.europa.eu/resource/authority/measurement-unit	`raw_value = { "areaSizeUnit": 100 "areaSizeUnit": { "uid": "KMK", "url": "http://publications.europa.eu/resource/authority/measurement-unit/KMK", "label": "square kilometre" } }`
					representation:
					`100, square kilometre(s)`

Aggregation level	AnalysisUnit: aggregationDistance	aggregationLevel : AggregationLevel	Integer		raw_value = 1000 representation: 1000

Aggregation level unit	AnalysisUnit: aggregationGeoReference	aggregationLevel : AggregationLevel	VocabularyValue	• http://publications.europa.eu/resource/authority/measurement-unit
				• Internal vocabulary using the following list of values:
				○ Continental
				○ Country
				○ NUTS 1, NUTS 2, NUTS 3
				○ Geonames
					raw_value = {
					"uid": "COU",
					"url": ",
					"label": "Country"
					}
					representation:
					Country

Aggregation scale	AnalysisUnit: aggregationScale	aggregationLevel : AggregationLevel	Integer		raw_value = 1000 representation: 1000

Population coverage	AnalysisUnit: populationCoverage		Integer (%)		raw_value = {
					"populationCoverage": 80
					}
					
					representation:
					80%

Census	AnalysisUnit: census		Boolean		raw_value = {
					"census": True
					}
					
					representation:
					True

Unit of analysis number	AnalysisUnit: unitAnalysisNumber	unitAnalysisNumber : integer	Integer		raw_value = 9331200 **representation**: 9331200
Statistical representativeness	AnalysisUnit: statsRepresentativeness	statsRepresentativeness : string	String(1000)		raw_value = "Lore ipsum..." **representation**: Lore ipsum...
Available stratification criteria & suggested downscaling methodologies	AnalysisUnit: downscalingMethodology		String(1000)		raw_value = "Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris" **representation**: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris

A3.5 Geo-referenced unit of analysis

Same properties as the socio-economic unit of analysis, but in addition, it includes the following:

Table 6 Geo-referenced unit of analysis properties

Name of the attribute	Corresponds to ClassName (ARDIT data model)	Corresponds to PropertyName (AGRICORE DCAT-AP Data Model)	Data Type	Controlled Vocabulary (if applyable)	Example

Area size	• AnalysisUnit: areaSizeValue
	• AnalysisUnit: areaSizeUnit
		sizeUnit: Size Unit	• Integer
			• Vocabulary Value
				http://publications.europa.eu/resource/authority/measurement-unit	raw_value = {
					"areaSizeUnit": 100
					"areaSizeUnit": {
					"uid": "KMK",
					"url": "http://publications.europa.eu/resource/authority/measurement-unit/KMK",
					"label": "square kilometre"
					}
					}
					representation:
					100, square kilometre(s)

A3.6 Socio-economic and geo-referenced variable

Both types of variables share the same properties.

Table 7 Variable properties

Name of the attribute	Corresponds to ClassName (ARDIT data model)	Corresponds to PropertyName (AGRICORE DCAT-AP Data Model)	Data Type	Controlled Vocabulary (if applicable)	Example
Name	DatasetVariable: name	variableSubject : string	String		raw_value = "Monthly growing area" **representation**: Monthly growing area

Unit of measurement	DatasetVariable: measureUnit	measureUnit : MeasureUnit	String* (*Needs to be changed to a list of strings*)		`raw_value = "hectare"`
					representation:
					`hectare`

Temporal extent	• DatasetVariable: tmpExtentFrom • DatasetVariable: tmpExtentTo	dcterms:temporal : dcterms:PeriodOfTime	Date Date		`raw_value = { "tmpExtentFrom": 2015/03/17, "tmpExntentTo": 2019/11/03}` **representation**: `March,2015` - November, 2019

Mathematical representation	DatasetVariable: mathRepresentation	mathRepresentation : string	VocabularyValue		**raw_value** = {
					"uid": "AVG",
					"url": "",
					"label": "Average"
					}
					representation:
					Average

Data frequency	DatasetVariable: dataFrequency	dataFrequencyElaboration : DataFrequencyElaboration	VocabularyValue	Frequency: http://publications.europa.eu/resource/dataset/frequency (English terms: http://publications.europa.eu/resource/cellar/87532442-261d-11e8-ac73-01aa75ed71a1.0001.04/DOC_6)	**raw_value** = {
					"uid": "ANNUAL",
					"url": "http://publications.europa.eu/resource/cellar/87532442-261d-11e8-ac73-01aa75ed71a1.0001.04/DOC_6",
					"label": "annual"
					}
					representation:
					Annual

Data frequency mathematical representation	DatasetVariable: dataFrequencyMathRep	dataFrequencyElaboration : DataFrequencyElaboration	VocabularyValue	Internal vocabulary	**raw_value** = {
					"uid": "AVG",
					"url": "",
					"label": "Average"
					}
					representation:
					Average

Aggregation level	DatasetVariable: aggregationLevel	aggregationLevel : AggregationLevel	Integer		**raw_value** = 1000 **representation**: `1000`
Aggregation level unit	DatasetVariable: aggregationUnit	aggregationLevel : AggregationLevel	VocabularyValue	• http://publications.europa.eu/resource/authority/measurement-unit • Internal vocabulary composed by: o Continental	**raw_value** = { "uid": "KMK",

				○ Country ○ NUTS 1, NUTS 2, NUTS 3 ○ Geonames	"url": "http://publications.europa.eu/resource/authority/measurement-unit/KMK", "label": "square kilometre } **representation**: square kilometre
Aggregation scale	DatasetVariable: aggregationScale	aggregationLevel : AggregationLevel	Integer		**raw_value** = 1000 **representation**: 1000
Statistical representativeness	DatasetVariable: statsRepresentativeness	statsRepresentativeness : string	String(1000)		raw_value = "Lore ipsum..." **representation**: Lore ipsum...

Data origin	DatasetVariable: dataOrigin	dataOrigin : DataOrigins	VocabularyValue	Internal vocabulary composed by:
				• Forecast
				• Observed
				• Measured
					raw_value = {
					"uid": "FOR",
					"url": "Internal vocabulary",
					"label": "Forecast"
					}
					representation:
					Forecast

Reference values	DatasetVariable: referenceValues	referenceValues : string	List<String>		**raw_value** = [{ refereceValue: "Rice" }, { refereceValue: "Potatoes" },] **representation**: Rice, Potatoes

Unit of analysis	DatasetVariable: analysisUnits	Agricore: AnalysisUnit	List<AnalysisUnit>		**raw_value:**
					[
					{
					"unitReference": "5 arc-minutes grid cell"
					"unitAnalysisNumber": 9331200,
					"statsRepresentative": 0,
					"tmpExtentFrom": 1998/01/01,
					"tmpExtentTo": 2002/12/01,
					"downscalingMethodology" = ""
					}
					]
					representation:

					5 arc-minutes grid cell, 9331200, 0, January,1998 - December, 2002, ""
Downscaling methodology suggestions	DatasetVariable: downscalingMethodology		String(1000)		`raw_value = "`Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris**"** **representation**: `Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris`

A3.7 Price variable

It is a socio-economic variable subclass. It shares the same properties, but also includes:

Table 8 Price variable properties

Name of the attribute	Corresponds to ClassName (ARDIT data model)	Corresponds to PropertyName (AGRICORE DCAT-AP Data Model)	Data Type	Controlled Vocabulary (if applicable)	Example

Currency	DatasetVariable: currency	currency : Currency	VocabularyValue	Currency vocabulary: http://publications.europa.eu/resource/authority/currency	**raw_value** = {
					"uid": "ESP",
					"url": "http://publications.europa.eu/resource/authority/currency/ESP",
					"label": "Peseta"
					}
					representation:
					Peseta

Price type	DatasetVariable: priceType	priceType : String	String	Internal vocabulary ("Purchase", "Selling", ...)	**raw_value** = { "uid": "PUR",

					"url": "", "label": "Purchase" } **representation**: Purchase
Size unit	• DatasetVariable: sizeUnitAmount • DatasetVariable: sizeUnit	sizeUnit : SizeUnit	• Integer • Vocabulary Value	A new vocabulary for measurement units: http://dd.eionet.europa.eu/vocabulary/eurostat/unit/	raw_value = { "uid": "BN_EUR", "url": "http://dd.eionet.europa.eu/vocabularyconcept/eurostat/unit/BN_EUR", "label": "billion euro" } **representation**: 100 billion euro

A3.8 Catalogue

Table 9 Catalogue properties

Name of the attribute	Corresponds to ClassName (ARDIT data model)	Corresponds to PropertyName (AGRICORE DCAT-AP Data Model)	Data Type	Controlled Vocabulary (if applicable)	Example
Title	Catalogue: title	dcterms:title : rdfs:Literal	String		raw_value = "Land Use and Coverage Area frame Survey" **representation**: Land Use and Coverage Area frame Survey
Description	Catalogue: description	dcterms:description : rdfs:Literal	String		raw_value = "Accurately characterizing land surface changes with Earth Observation requires geo-localized ground truth. In the European Union (EU), a tri-annual surveyed sample of land cover and land ..." **representation**: Accurately characterizing land surface changes with Earth Observation requires geo-localized ground truth. In the European Union (EU), a tri-annual surveyed sample of land cover and land ...
Publisher	Catalogue: publisher	dcterms:creator : foaf:Agent	String		raw_value = "Eurostat" **representation**:

					Eurostat
Link	Catalogue: link	• foaf:homepage : foaf:Document • dcat:landingPage : foaf:Document	String		`raw_value = "https://ec.europa.eu/eurostat/web/lucas/overview"` **representation**: https://ec.europa.eu/eurostat/web/lucas/overview
Issued	Catalogue: issued	dcterms:issued : xsd:dateTime	Date		`raw_value = "2021-04-19T14:11:35.506+00:00"` **representation**: 2021-04-19T14:11:35.506+00:00
Modified	Catalogue: modified	dcterms:modified : xsd:dateTime	Date		**raw_value** = "2021-05-04T14:16:42.885+00:00" **representation**: 2021-05-04T14:16:42.885+00:00
Temporal extent	Catalogue: tmpExtentFrom Catalogue: tmpExtentTo	dcterms:temporal : dcterms:PeriodOfTime	Date Date		`raw_value = { "tmpExtentFrom": 2015/03/17, "tmpExntentTo": 2019/11/03}` **representation**: March, 2015 - November, 2019
Spatial resolution In Meters	Catalogue: spatialResolutionInMeters	dcat:spatialResolutionInMeters : xsd:decimal	Float		`raw_value = "100.0"` **representation**: 100.0
Temporal Resolution	Catalogue: temporalResolution	dcat:temporalResolution : xsd:duration	String		`raw_value = "PT1H"` **representation**: PT1H
Periodicity	Catalogue: periodicity	dcat:accrualPeriodicity : dcterms:Frequency	VocabularyValue	https://www.dublincore.org/specifications/dublin-core/collection-description/frequency/	`raw_value = { "code": "annual", "url": "http://purl.org/cld/freq/annual", "label": "Annual" }` **representation**: Annual
Languages	Catalogue: languages	dcterms:language : dcterms:LinguisticSystem	List<VocabularyValue>	https://id.loc.gov/vocabulary/iso639-2.html	`raw_value = [ { "code": "eng", "url": "http://id.loc.gov/vocabulary/iso639-2/eng", "label": "English" },`

					<pre>{
					"code": "spa",
					"url": "https://id.loc.gov/vocabulary/iso639-2/spa",
					"label": "Spanish"
					}
					]</pre>**representation**:
					English, Spanish

Themes	Catalogue: themes	dcat:theme : skos:Concept	List<Vocabulary Value>	This property uses two vocabularies:
				• http://publications.europa.eu/resource/authority/data-theme
				• https://inspire.ec.europa.eu/theme/
					<pre>raw_value = [
					{
					"code": "gg",
					"url": "https://inspire.ec.europa.eu/theme/gg",
					"label": "Geographical grid systems"
					},
					{
					"code": "ENVI",
					"url": "http://publications.europa.eu/resource/authority/data-theme/ENVI",
					"label": "Environment"
					}
					]</pre>**representation**:
					Environment, Geographical grid systems