D1.1 Standardised Methodology and Set of Ontologies for the Characterisation of Data Sources

agricore

Deliverable Number	D1.1
Lead Beneficiary	UNIPR
Authors	Mario Veneziani, Federico Antonioli, Giorgia Eranio, Carlos Leyva, Pablo Báez, Álvaro Fernández
Work package	WP1
Delivery Date	M09 → M11
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)

Version History

Version	Description	Organisation	Date
0.1	First template version	UNIPR	31 March 2020
0.2	Revision of document structure	IDENER	7 April 2020
0.3	Background information added	UNIPR	16 April 2020
0.4	Description of Ontologies and previous efforts	UNIPR, STAM	5 May 2020
0.5	Restructuring of the document. Inclusion of DCAT-AP extension	STAM	8 May 2020
0.6	Content completion	UNIPR	28 May 2020
0.7	DCAT-AP extension details	STAM	15 June 2020
0.8	Methodology description	UNIPR, STAM	22 June 2020
0.9	Coordinator review, suggestions for updates	IDENER	30 June 2020
0.10	Further explanation of the methodology	IDENER, UNIPR	10 July 2020
0.11	Extension of DCAT-AP AGRICORE extension	STAM	15 July 2020
0.12	Intensive review and extension of the content and format	IDENER, UNIPR, STAM	10-July-2020 - 30 July-2020
1.0	First release version - Delivered to the EC	UNIPR, IDENER	31 July 2020

Executive Summary

This deliverable presents the methodology defined within the AGRICORE project to characterise data sources useful for performing agricultural research analysis. This methodology has been developed as part of the first of the work packages defined in the AGRICORE project. AGRICORE is a research project proposing an innovative way to apply agent-based modelling to improve the capacities of policymakers to evaluate the impact of agricultural-related measurements under and outside the framework of the Common Agricultural Policy. This project was funded by the European Commission as a result of the RUR-04-2018 call, part of the H2020 programme.

The present document starts providing an introduction to the reader about the framework of this development, covering the AGRICORE project and focusing on the relevant parts of it mostly related to the use of data for performing impact assessment analysis. Then, the usage of ontologies as part of the characterisation methodology is explained and an analysis of the already existing research work in this area provided. After that, the proposed methodology is detailed including also the process followed to design it. As part of this methodology, the AGRICORE partners have developed an extension to the Data Catalogue Application Profile (DCAT-AP) standard which will serve as a basis to compile the required information during the characterisation process. This section is available as part of this deliverable but also released as a separate document. Finally, some conclusions regarding the characterisation and mapping (and the needs for it) of data sources is provided.

It is important to remark that although this deliverable has been developed in the framework of the AGRICORE project, the participating partners have aimed for broader usage of the proposed methodology. As the final goal of this work package, the proposed EU Index Tool (now renamed as Agricultural Research Data Index Tool (ARDIT) aims to serve as a central entry point for locating useful datasets useful for agricultural research. Accordingly, the methodology here presented will be used to characterise the set of datasets identified by the AGRICORE consortium, not limiting the analysis to those used in the project.

Abbreviations

Abbreviation	Full name
ABM(s)	Agent-Based Model(s)
AFFRIS	Aquaculture Feed and Fertilizer Resources Information System
AGLINK-COSIMO	AGLINK-COmmodity SImulation MOdel
AGMEMOD	Agricultural Member State Modelling
AI	Artificial Intelligence
AP	Application Profile
API	Application Programming Interface
ARDIT	Agricultural Research Data Index Tool
BIODIVTHES	Biodiversity Thesaurus
BioMa	Biophysical Model Applications
CAP	Common Agricultural Policy
CAPRI	Common Agricultural Policy Regional Impact
CAT	Chinese Agricultural Thesaurus
CIARD	Coherence in Information for Agricultural Research for Development
CO	Crop Ontology
CREA	Consiglio per la Ricerca e la Sperimentazione in Agricoltura
CSV	comma-separated values
DCAT	Data Catalogue
DCAT-AP	Data Catalogue Application Profile
DG CONNECT	Directorate-General for Communications Networks, Content and Technology
DP	Direct Payments
DWH	Data Warehouse
EC	European Commission
EEA	European Economic Area
EGDIP	European Green Deal Investment Plan
EP	European Parliament
ESYRCE	Spanish Survey on Crop Surfaces and Yields
ETL(s)	Extraction Transformation and Loading script(s)
EU	European Union
FADN	Farm Accountancy Data Network
FAO	Food and Agriculture Organization of the United Nations
FSS	Farm Structure Survey
FTP	File Transfer Protocol
GeoDCAT-AP	A geospatial extension for the DCAT application profile for data portals in Europe
GFAR	Global Forum on Agricultural Research
GODAN Action	Global Open Data for Agriculture & Nutrition
GUI	Graphic User Interface
IACS	Integrated Administration and Control System
IAM	Integrated Assessment and Modelling
ICT	Information and Communication Technologies
IFM-CAP	EU-Wide Individual Farm Model for Common Agricultural Policy Analysis
INSPIRE	Infrastructure for Spatial Information in the European Community
INRA	Institut National de la Recherche Agronomique
ISA^2	Interoperability Solutions for public Administrations, businesses and citizens (programme)

ISTAT	(Italian) National Institute for Statistics
JRC	Joint Research Centre
KPIs	Key Performance Indicators
LDAP	Lightweight Directory Access Protocol
LIPS	Land Parcel Identification System
LUCAS	Land Use/Land Cover Area Frame Survey
MAGNET	Modular Applied GeNeral Equilibrium Tool
MFF	Multiannual Financial Framework
MIDAS	Modelling Inventory and Knowledge Management System
MS(s)	Member State(s)
NALT	National Agricultural Library Thesaurus
NASA	National Aeronautics and Space Administration of the United States of America
NCBO	National Center For Biomedical Ontology
NetCDF	Network Common Data Form
NUTS	Nomenclature of Territorial Units for Statistics
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OBO	Open Biological and Biomedical Ontologies
OECD	Organization for Economic Cooperation and Development
OWL	Web Ontology Language
PWT	Penn World Table
RDA	Research Data Alliance
RDF	Resource Description Framework
REST	Representational State Transfer
RING	Routemap to Information Nodes and Gateways
ROMA	Official Registers of Agricultural Machinery
RSS	RDF Site Summary
SeTA	Semantic Text Analysis
SIFR	Semantic Indexing of French biomedical Resources
SKOS	Simple Knowledge Organisation System
SOA	Service-Oriented Architecture
SOIL	agINFRA Soil Vocabulary
SPARQL	SPARQL Protocol and RDF Query Language
StatDCAT-AP	DCAT Application Profile for description of statistical datasets
TOP	Thesaurus of Plant Characteristics
UC(s)	Use Case(s)
UNFCCC	United Nations Framework Convention on Climate Change
URL	Uniform Resource Locator
USDA	United States Department for Agriculture
VEST	Vocabularies, mEtadata Sets and Tools
WP	Work Package
WTO	World Trade Organisation
W3C	World Wide Web Consortium
XML	eXtensible Markup Language

List of Figures

List of Tables

Table of Contents

1 Introduction - Framework of this methodology

The AGRICORE project proposes a novel tool for improving the current capacity to model policies dealing with agriculture by taking advantage of the latest progress in modelling approaches and Information and Communication Technologies (ICT). Specifically, the AGRICORE tool will be built as an Agent-Based Model (ABM) in which each farm will be modelled as an autonomous decision-making entity that individually assesses its context and makes decisions based on its current situation and expectations. This modelling approach will allow simulating the interaction between farms and their context (which will account for the natural environment, rural integration, ecosystem services, land use, input and output markets organisation and dynamics) at various geographic scales - from regional to global. The AGRICORE tool will fill a few gaps in the characteristics of existing modelling frameworks. In particular, it will be one of the few individual farms-based models going beyond the limitations imposed by other approaches centred on "average" farms. The latter are farms whose characteristics are obtained as averages - across the farms belonging to each farm type or specialisation and in each defined area (i. e., administrative or Nomenclature of Territorial Units for Statistics (NUTS) regions) - of the variables of every farm belonging to that farm type and area. For instance, for the whole of Europe, the Common Agricultural Policy Regional Impact (CAPRI) model [1] is built considering 2,000 farm types (i. e., different organisations of the production activity) and 280 NUTS2 regions. Moreover, the AGRICORE tool will mark a clear advancement on previous modelling efforts since it will develop a particular focus on the analyses of Pillar II Measures of the Common Agricultural Policy (CAP). Furthermore, the AGRICORE tool will benefit from innovative, advanced and open-sourceable ICT tools and procedures which will facilitate significantly the otherwise very resource-intensive and time-consuming nature of the model calibration phase, brought about by the complex nature of the modelling undertaken, which often characterises extant suites.

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. This includes the conditions and expectations regarding the future states of the input and output markets, the natural conditions related to the nutrients in the soil and its moisture due to the incidence of rainfalls and temperature (changes), and the existing and foreseen (agricultural) policy scenario, among others. To tackle this, Work Package (WP) 1 of the AGRICORE project is tasked with the characterisation and access retrieval of several databases that can provide the information instrumental to the development of the AGRICORE tool. The preliminary characterisation of the datasets will allow model developers and quantitative researchers to plan their data requirements, methodological choices and the actual specification of the variables employed ahead of starting acquiring the datasets, which may be a time-consuming activity. Indeed, the dataset characterisation work done in WP1 of the project, especially in Tasks 1.1 to 1.6, aims to provide a methodology for characterising (Task 1.1), and the actual characterisation according to the selected methodology (Tasks 1.2 through to Task 1.6) of, several datasets which may be of interest to the research community when undertaking agricultural policy analysis and impact assessment.

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. Indeed, the scope of this Deliverable is to present such methodology. 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

[1] The CAPRI model is also known as a regionalised agricultural sector model using an activity-based approach.

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). To provide this assistance, semantic services will be provided and will exploit the defined ontologies detailed below. The ARDIT will initially be populated with the datasets characterised within the project, gathered during the execution of Tasks 1.2 to 1.6. These include the European Union (EU) statistics datasets (e. g., the Farm Accountancy Data Network (FADN)), geo-referenced datasets (e. g., the Land Use/Land Cover Area Frame Survey (LUCAS), national and regional information sources (e. g., the Italian or Spanish FADN; the statistics of land prices in Andalusia) and previous research results (e. g., the EU-Wide Individual Farm Model for Common Agricultural Policy Analysis (IFM-CAP); the Biophysical Model Applications (BioMa)) for the modelling of land use, policy, biophysical, social, economic and environmental aspects related to farming activities.

The characterisation here defined will go a step beyond other current characterisation efforts providing technical details on the information contained in the sources mentioned above up to the level of the variables included on them. This 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).

1.1 Motivation and AGRICORE presentation

Over the last decade, the CAP has focused on the reinforcement of the support for rural development across the EU, the improvement of the integration of environmental requirements and the increase of market orientation of agriculture. Based on two linked pillars, the CAP (2000-2020) looks for strengthening rural development and enhancing the competitiveness and sustainability of the EU agricultural sector. While Pillar I includes measures mainly targeting income support (i. e., Direct Payments (DP), either decoupled, roughly 90% of the budget, or coupled), Pillar II entails targeted and more farm-specific components (e. g., fostering knowledge transfer, enhancing farm viability and competitiveness, promotion of food chain organisations, preservation and restoration of ecosystems, promotion of resource efficiency and social inclusion). The CAP is evolving (around €4 billion have been transferred from Pillar I to Pillar II over the period 2014-2019), and the European Commission (EC) has recently presented specific legislative proposals to further target the CAP goals for the period after 2020. Among the additional most relevant domains of the CAP after the 2020 reform, and current discussions, it is possible to highlight: the issue of the continued (from the 2013 CAP reform) external convergence of the CAP per hectare payments (i. e., the harmonisation of CAP payments per hectare across EU Member States (MSs)); the issue of internal convergence (i. e., the equalisation of the value of the decoupled DP within each MS or region); the pursuit of a "truly" greener and more sustainable agriculture through the reform of the CAP in compliance with other EU policy frameworks (e. g., the European Green Deal Investment Plan (EGDIP) and - for what concerns the agri-food industry - the included Farm to Fork Strategy, the EU Biodiversity Strategy); the role of domestic support, as foreseen in the CAP after 2020, in the realm of the EU membership of the World Trade

Organisation (WTO) and - obviously - how much of the next Multiannual Financial Framework (MFF) will be earmarked for the CAP (i. e., which will be the CAP share of the next MFF). [2]

Current applied agricultural models (e. g., AGLINK-COmmodity SImulation MOdel (AGLINK-COSIMO), CAPRI, Agricultural Member State Modelling (AGMEMOD), AROPAj, Modular Applied GeNeral Equilibrium Tool (MAGNET)) were developed for modelling early CAP instruments such as those included in the Pillar I, and to capture their impact on markets, prices and trade. As a consequence, these models are not so suitable for representing many of the new policy instruments, to capture farm heterogeneity and to address a smaller geographical scale than the regional level. In response to these needs, agent-based modelling represents a powerful framework able to tackle these challenges, modelling the system as a collection of autonomous decision-making entities (i. e., the agents). Each agent individually assesses its situation and makes decisions based on a set of rules[1]. The main advantages of ABMs in the agricultural domain are the explicit modelling of the interactions among farmers and the consideration of the spatial dimension of agricultural activities[2].

In this framework, the AGRICORE tool (please see the figure below) will be built as an ABM in which each farm will be modelled as an autonomous decision-making entity which individually assesses its context and makes decisions based on its current situation and expectations (WP3 of the project). This modelling approach will allow simulating the interaction between farms and their context (which will account for the natural environment, rural integration, ecosystem services, land use, input and output markets organisation and dynamics through the additional model modules developed in the WP5 of the project) at various geographic scales - from regional to global.

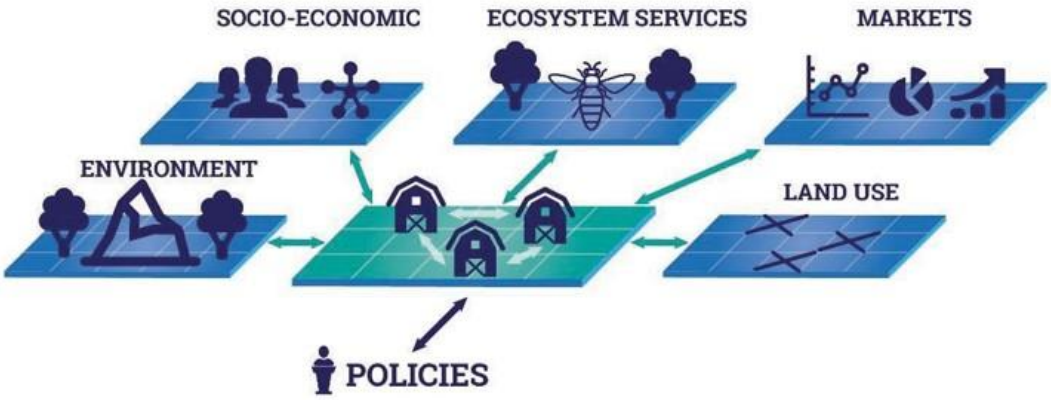

Figure 1 The AGRICORE Project Framework

For the AGRICORE tool to be fully functioning and, to address the modelling needs of the stakeholders of the project, suitable data should be identified and gathered both by relying on existing datasets and research outcomes and participatory research purposely undertaken in the realm of the AGRICORE project (WP1). A comprehensive survey of EU statistics and geo-referenced datasets, national and regional information sources and previous research results for the modelling of land use, policy, biophysical, social, economic and environmental aspects related to farming activities, will allow the preparation of the ARDIT. Building on the latest progress in ICT, the AGRICORE project will produce a synthetic population generator capable of generating realistic synthetic populations mimicking the distribution and characteristics of the real farmers'

[2] For an up-to-date analysis of, among other things, the most pressing issues being discussed in the area of EU (agricultural) policy reform, Prof. Alan Matthews' blog CAP Reform (http://capreform.eu/) is a recommended reading.

population of interest as captured in the datasets identified, characterised and acquired in WP1 (WP2). This will allow minimising the time and user efforts currently required for the parameterisation and calibration of ABMs. WP3 will develop an evolved ABM with improved capacity to model policies dealing with agriculture. Partners will elaborate on a dynamic quadratic model explicitly accounting for agents' interactions, whose computation is to be enabled by recent advancements in the capacities of mathematical solvers and ICT. The ABM is further expanded relying on modules evaluating the social, economic and environmental impact of agricultural policies at farm, sector and global levels (WP5). The ABM (WP3), the synthetic population generator and the data warehouse (DWH) infrastructure (WP2); allowing the AGRICORE tool to be operational for agricultural policy analysis and the impact assessment modules (WP5); will be packaged in an AGRICORE suite (WP6) capable of meeting the standard requirements mandated by the usability analysis and design (i. e., the presence of a user interface and big data visualisation module to facilitate input procedures and output gathering/presentation) (WP4). The flexible and integrated AGRICORE simulation suite (WP6) will constitute a simulation environment ready to use either for ex-ante (for policy design) or ex-post (for monitoring) analysis, allowing interoperability. The AGRICORE suite will be demonstrated in three Use Case(s) (UC(s)) aimed at evaluating the impact of measures of Pillar II of the CAP in Andalusia (Spain), Poland and Greece. In particular, UC1 will evaluate the M11 measure "Ecologic agriculture" for assessing the environmental impacts of the olive sector in Andalusia. UC2 will concern the M10.1 measure "Agri-environment-climate commitments" in Poland, focusing mainly on the provision of ecosystem services and the environmental and climate impacts of agriculture (transformation). Finally, UC3 will analyse M6.1 measure "Start-up aid for young farmers" and its effects on Greek agriculture, focusing on the socio-economic aspects of the integration of agriculture in rural society. Therefore, UC1 will employ the AGRICORE suite at a regional level, while UC2 and UC3 at the national level. Having verified the capabilities of the AGRICORE suite within the framework of the UCs, it will be packaged for open-source release to the research, practice and policymaking community for further use and improvement (WP8). Throughout the lifetime of the project, and across all WPs, interaction with all the stakeholders of the project through communication and dissemination (WP9), as well as research activities (WP1 and WP8), will make sure the project results address the needs of a community of users of the AGRICORE suite and its adoption is as widespread as possible, given the novelty and advances of this suite over existing ones (WP9).

The AGRICORE project hinges on four main development pillars:

1. An advanced population concept to efficiently parameterise and calibrate the ABMs, taking into account of the combination of multiple agricultural-related sources;
2. An ABM integrating Artificial Intelligence (AI) for overcoming the main drawbacks of the current modelling techniques;
3. A user-friendly interface allowing users without a strong computational and/or scientific background to build case studies and obtain meaningful outcomes;
4. A highly modular and customisable ICT architecture to handle agent-based simulations at various geographic scales (from regional to global).

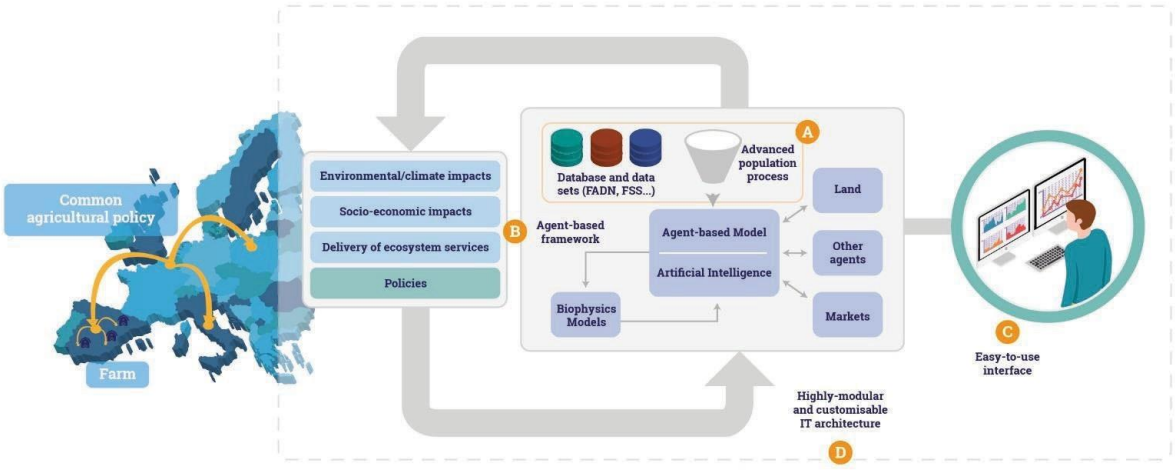

Figure 2 The AGRICORE Project Overall Approach

Regarding the ABM structure conceptualised in the AGRICORE model, the following figure offers a synthetic glance at it.

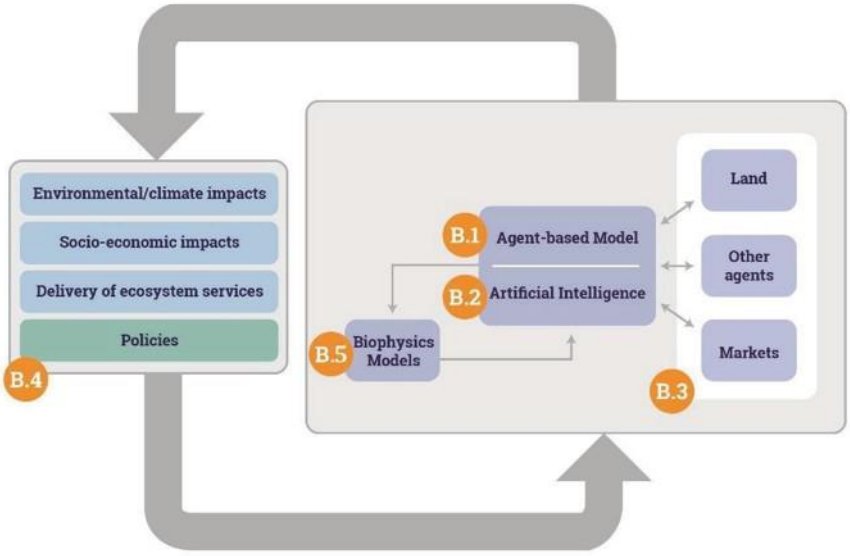

Figure 3 The AGRICORE Project Conceptualisation

Each farm interacts with the other components of the agricultural structure, which in the case of AGRICORE include other farmers, input and output markets - of which the land one is especially detailed. Additionally, the farms are embedded within their context, which is considered to account for the environmental, climatic, socio-economic characteristics (rural integration) they are exposed to and impact on, as well as the delivery of ecosystem services to which they contribute. The AGRICORE conceptual framework has been translated into a model structure composed of five main elements which include: (*B.1*) the non-linear dynamic model of the farm (agent); (*B.2*) the AI-based farmers' behavioural foundation; (*B.3*) the model interactions; (*B.4*) the context relationships and (*B.5*) the links with biophysical models from the BioMa platform.

Farming is deeply connected to what the AGRICORE project refers to as context. To assess the impacts of agriculture in its context (as a response to policies), and also to determine how changes

in such a context affect farming, the AGRICORE project will establish links with the following dedicated modules:

- the policy environment. This module will translate the policy schemes of interest into the AGRICORE simulation environment. For instance, policies related to price support would modify the agent's objective function whereas policies establishing production quotas would modify the agent's production constraints. This flexibility will allow the AGRICORE suite to simulate both CAP pillar I policies (e. g., coupled and decoupled DP, price support, set-aside and production quotas) and more targeted and potentially complex CAP pillar II and post-2020 policies (e. g., subsidies for organic farming, animal welfare payments, advisory services, a mix of different and progressive schemes).
- the environmental and climate impacts of agriculture. The relationship between farming and the environment in the AGRICORE suite will be bi-directional: agriculture has significant impacts on the environment and climate while climate change affects how much agricultural output can be produced and where. In parallel, policy (intervention) is an effective means to help farmers adapting to climate change as well as incentivising sustainable agricultural practices. The AGRICORE project addresses the modelling of these aspects twofold. On the one hand, this dedicated module will provide regional climatic patterns as an input to the ABM. On the other hand, the module will compute main Key Performance Indicators (KPIs) related to the environmental and climatic impact assessment of policies (e. g., land conversion and habitat loss, wasteful water consumption, soil erosion and degradation, pollution, genetic erosion, climate change).
- the delivery of ecosystem services. In line with the Millennium Ecosystem Assessment guidelines, this dedicated AGRICORE suite module will model and provide ecosystems services KPIs.
- the socio-economic aspects of the integration of agriculture in rural society. Maintaining viable rural communities is one of the strategic aims of the CAP as set out in the Commission's Communication "The CAP towards 2020: Meeting the food, natural resources and territorial challenges of the future" (COM(2010) 672). Accordingly, the AGRICORE suite will include a dedicated module aiming to assess the relationship between policy incentives and KPIs related to the integration of agriculture in rural systems. Among them rural employment, the viability of local micro, small and medium-sized enterprises within the agricultural value chain, young people's startup initiatives in rural areas and gross value added of agriculture can be listed.

1.2 Need for a unique entry point to identify relevant data sources for agricultural researchers

Agriculture policy analysis in Europe, and especially the impact assessment of the CAP policies, mainly relies on the FADN and the Farm Structure Survey (FSS) databases. The FADN is an annual farm-level survey that gathers detailed EU-wide accounting data from a sample of agricultural holdings. The FSS provides harmonised data on the structure of farm holdings regarding land use and livestock, farm labour force, machinery and equipment, as well as the participation in rural development programs, although it lacks in economic variables. The basic unit underlying the FSS is the agricultural holding and a complete agricultural census is updated every 10 years (with intermediate sample surveys).

Despite the information available in the FADN and FSS has allowed the impact analysis of previous CAP versions, it seems these data sources are not sufficient to assess more recent and targeted policies. Indeed, policymakers and researchers are completing the information provided by the FADN and FSS with an extended list of datasets. This includes datasets from sources such

as Aquastat, FaoStat and the Aquaculture Feed and Fertilizer Resources Information System (AFFRIS) provided by the Food and Agriculture Organization of the United Nations (FAO); Feedipedia; FeedPrint; Penn World Table (PWT); the statistics provided under the United Nations Framework Convention on Climate Change (UNFCCC) and by Eurostat and the Organization for Economic Cooperation and Development (OECD). Moreover, the new analysis needs the EC to perform more detailed research and impact evaluations with a stronger focus on local effects. To provide this, geo-referenced data sources such as LUCAS provide valuable information. LUCAS is released by Eurostat every 3 years since 2006 and allows identifying changes in land use (meaning the socioeconomic use of land, e. g., agriculture, forestry, recreation or residential use) and land cover (e. g., crops, grass, broad-leaved forest, or built-up area) in the EU. At the same time, the integrated Administration and Control System (IACS), which is a European network of databases used for the management and control of CAP payments disbursed by the MSs is also a valuable resource for detailed analysis. This system consists of a network of databases (including the Land Parcel Identification System (LIPS)) that are generated and controlled at the national level by the respective governments. Access to these databases is valuable because they include spatially explicit and even field-level data (e. g., geo-localisation of the farm, arable land, permanent grassland, permanent crops) that would be relevant to many analytical and policy issues. However, being authorised to use these datasets poses significant challenges related to ensuring the anonymity of the farms in these spatially-explicit databases. Furthermore, it is fair to acknowledge that it is not easy to link the IACS data to other farm-level databases like the FADN.

At a more granular level, local information (covering also regional and country levels) is of crucial importance to perform micro-level policy analysis. Despite the availability of the above mentioned EU-level data sources, more localised ones are available and can provide additional information which would be crucial to accurately model the behaviour of farmers and their reactions to potential policy changes enacted. The Bin database for The Netherlands; the IACS-AGEA database (i. e., the Italian version of the IASC) and the data provided by the Italian National Institute for Statistics (ISTAT) for Italy; the GRIA, the Official Registers of Agricultural Machinery (ROMA) and the Spanish Survey on Crop Surfaces and Yields (ESYRCE) for Spain are among the many others datasets of this type.

Besides, agricultural researchers have been working in Europe for more than 20 years evaluating the impact of different policies and, in this process, they have generated a lot of knowledge. This covers both models (that can be used to generate information for the analyses) or even already processed data sets. This additional source may provide a key benefit not only by saving efforts and avoiding repeating already-performed studies but also by enabling increasing the consistency between different research activities.

Finally, and despite the vast availability of data sources for conducting agricultural policy analysis, the data needs for doing this for the new planned policies are even greater; which means that the current data acquisition methods and datasets lack critical information instrumental in performing detailed analysis such as detailed social information on farms.

Within this framework, and as already introduced in previous sections, the AGRICORE project proposes creating an advanced index tool providing easy access to agricultural data sets. The ARDIT aims to become the first stop for agricultural researchers in their process of identifying useful data sources that can help them answer the specifics questions they would like to address, especially those related to the impact of new agricultural policies and schemes. In addition to including all the technical information about the characterised data sources, ARDIT will also focus on providing the required information to get access to the data sources. Indeed, European researchers experience problems in accessing these databases due to privacy or administrative reasons[3] which may be a stumbling block to further extending their research. To avoid this situation, the ARDIT will include detailed information on how to gain such access.

2 Ontologies

2.1 What is an ontology

An ontology can be defined as a model that represents a set of concepts and their relationships within a knowledge domain such that it can be seen as a model of knowledge and the tool for its management. Within an organisation, an ontology represents a structure created for users to answer complex questions that they address to the information systems.

Users need daily access to large quantities of information, typically collected in different formats and for which it is often impossible to understand the relationships among them. To deal with these challenges, ontologies have been developed. They can also be defined as a logical model that allows data properties and their relationships to become visible. More in detail, an ontology is a formal and explicit description of concepts dedicated to a particular knowledge field or domain, of their properties and characteristics and the relationships among them. It is created through a formal naming and definition of the categories, properties and relationships among concepts, data and entities that link one, many or all knowledge domains. From a graphical point of view, ontologies are visualised and thought of as semantic networks: nodes correspond to concepts and links identify the connections among the concepts[4].

Ontologies have become an important topic in several fields of computer science and they are crucial for the development of the semantic web, to which they provide the semantic vocabulary used to annotate websites in a way meaningful for machine interpretation[5]. The semantic approach is currently useful not only on the web but also in other knowledge domains. The complexity of a large number of knowledge domains is quickly growing and the information systems are no longer characterised by simple data that can be easily managed with a set of programming languages but are now abundant in documents in natural language and different formats. Therefore, as mentioned above, relational database technologies may no longer be useful and efficient for storing, managing and querying these types of documents.

Ontologies can be built *ex-novo* or by reusing other available ontologies and the final result depends on the different building processes which are used. Due to the increasing importance of ontologies for the construction of the semantic, the building process has become more refined but the existing models of an ontology can be summarised as follows[6],[7]:

1. Representation: define the representation primitives of a knowledge representation system. An ontology based on a representation system object-centred includes the definition of class, instance, of the relation between a class and its superclass.
2. General or upper-level: define very general concepts that are highly reusable across several domains and applications. An ontology based on time focuses on time points or intervals and their relations.
3. Domain: define concepts from a given domain. An ontology on the elements of a domain, defines concepts, such as the class of the features of elements.

In the ontology creation process, there is no correct way to model a knowledge domain, because there are always a few viable alternatives. The best solution to choose from is to define in detail what type of application the ontology creator has in mind, which currents needs it must satisfy and which future extensions it may accommodate[8]. Irrespective of the philosophy chosen for the construction of the ontology, the next steps to follow for developing it are quite common across approaches. Five main stages can be identified:

- Specification: identify the reason, needs for building the ontology and its potential users

- Conceptualisation: describe the ontology to be developed. The conceptual model of the ontology is composed of concepts in the domain of interest and relationships among those concepts
- Formalisation: move from the conceptual to the formal model, by defining axioms to shrink the possible interpretations of the meaning of those concepts
- Implementation: choose a representation language to write the formal model of the ontology
- Maintenance: update of, and fix bugs in, the implemented ontology.

The specification, conceptualisation and formalisation phases of the construction of an ontology can be thought of being preparatory phases. While they can be undertaken in non-machine-readable/interpretable languages, this is not the case for the implementation phase; which possibly requires a recognised standard. One of the most used standard ontology languages is the Web Ontology Language (OWL) from the World Wide Web Consortium (W3C). OWL is a markup language employed to describe concepts and to explicitly represent the meaning and semantics of terms according to the vocabularies used and the relationships among them. OWL is based on a logical model that makes it possible for concepts to be defined as well as to describe how complex concepts can be built up from the management of a set of simpler concepts. Several software packages for preparing an ontology using the OWL markup language exist and EntryScape and Protégé are two of them. EntryScape is a web application developed for the collection of work material, allowing for collaboration, which is used in some contexts by the Joint Research Centre (JRC) of the EC[9],[10]. The platform is reasonably user friendly and it has support for handling metadata beyond title and description, relations between resources, linking to web material and defining groups that can form communities and be used for access control. Unfortunately, the free version of EntryScape is a hosted data management platform with a lot of limitations, especially for large and complex projects. The free version of the platform has restrictions on the number of data catalogues or datasets it can manage, the possibility to upload files to datasets, the possibility of adapting the Data Catalogue Application Profile (DCAT-AP) to the user needs and the possibility of harvesting from any data portal supporting DCAT-AP. Protégé is a software package based on the OWL language, developed at Stanford University, which is free and open-source and allows for the creation and management of ontologies. This software is a tool created to support the development of an ontology for the semantic web using a graphic user interface. It is part of the Protégé ontological development platform and it allows you to create and edit ontologies in the OWL while using description logic classifiers to maintain the consistency of their ontologies[11]. The tool accommodates the manual, semi-automatic or automatic ontology creation or management starting *ex-novo* or modifying and reusing existing ontologies. It also includes deductive classifiers to validate that models are consistent and to infer new information based on the analysis of an ontology. Protégé has a large user community of some 300,000 registered users which has elevated it to possibly the preferred tool for preparing an ontology[12]. [3]

Ontologies can be developed both manually and automatically. The manual approach has the drawback of requiring a lot of design time and high-level expertise. Automatic methods are more parsimonious from the point of view of human commitment since they define an ontology model by defining a domain in the form of the metadata that can characterise the domain and apply rules to the metadata[13].

For both the manual and automatic construction of an ontology there are two techniques which differ in the starting point of the design:

[3] The AGRICORE project ontology has been developed in Protégé after having compared its features to the ones of the free version of EntryScape, evaluated the needs of the project, the expertise in building ontologies within the Consortium and the size of its community of users and developers.

- **Top-down approach** in which the core concepts of the ontology are the starting point of construction and are used as an upper-level structure for the specific ontology of the domain. The domain or knowledge of the domains is then analysed and the relationships between the concepts are identified. The last step is to add the concepts and relationships in the ontology building tool.
- **Bottom-up approach** in which the available data sources or the data dictionary inform attempt to extract all the concepts using domain corpus and removing non-domain related concepts while manually adding the concepts and their relationship in the ontology building tool.

Both approaches have advantages and drawbacks. More in detail, a top-down ontology, being close to the upper-level type, is very rich from a semantic point of view, it represents knowledge very well and it ensures interoperability. On the other side, a bottom-up approach is close to the idea of an experience-based ontology meaning that it can capture the experience, is close to the application and is dynamic although less structured. The possibility of combining the two approaches has already been shown in other domains[14]. These two strategies for building ontologies can be seen as complementary using the top-down definition for the development of the conceptual structure of the domain, thanks to the contribution of domain experts. This structure can be harmonized with other ontologies concerning similar or complementary knowledge to expand the conceptual basis. On the other hand, the bottom-up strategy based on the semantics of available information and data sources, managed with the automatic method, brings improvements and extensions to the conceptual structure by maximizing the completeness and specificity of the resulting knowledge domain.

Despite the approach followed for developing an ontology, some questions need to be answered to plan appropriately for its development. These questions are:

- What is the domain that the ontology will cover?
- What is the purpose of the ontology?
- Which types of questions the information in the ontology should provide answers to?
- Who will use and maintain the ontology?

2.1.1 Need for an Ontology for the ARDIT

As already introduced at the beginning of this document, one of the goals of the AGRICORE project is to develop the ARDIT to facilitate the task of identifying relevant and useful data for performing agricultural policy analysis. To do so, the AGRICORE partners devised a characterisation methodology (detailed in the next section) for describing the available data sources and their content to enable proper mapping and searching capabilities over the gathered information. To design such a methodology, one of the key elements is the definition or adoption of an ontology (or a set of them), which allows:

- To share a **common understanding** of the structure of information among people or software agents.
- To enable the **reuse** of domain knowledge.
- To make domain **assumptions explicit.**
- To **separate domain** knowledge from the **operational** knowledge
- To **analyse** domain knowledge
- To secure the interoperability of datasets

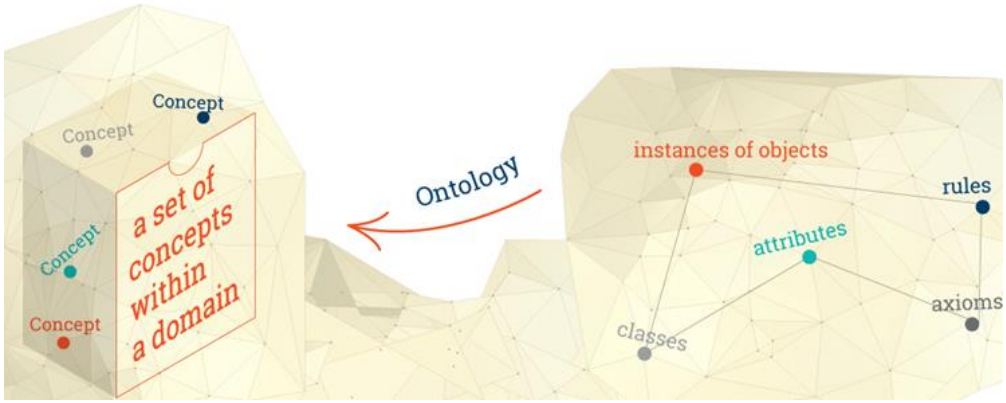

Figure 4 A Graphical Representation an Ontology at Work

All these motivations are present in the AGRICORE project. Indeed, to ensure the long-term usefulness of the ARDIT platform well beyond the scope of the AGRICORE project, it is critical developing it in a way that it is easy to upgrade and maintain as well as user-friendliness for the research community as a whole. Accordingly, consortium partners performed an analysis of existing ontologies that could be used for developing the ARDIT. This analysis is presented below and its findings have been used to devise the methodology proposed in this document.

2.1.2 Previous ontologies and related work in the agricultural knowledge domain

The domain of agriculture, due to initiatives from organisations such as the FAO, has several substantial semantic resources and data interchange standards at its disposal. However, the application of semantic web technologies in agriculture remains infrequent.

The largest and most comprehensive semantic resource, AGROVOC, was developed by the FAO and the EC in the early 1980s to identify documents and other information resources for indexing and searching in twenty-seven languages[15]. AGROVOC is a standardised and controlled vocabulary that contains 35,000 concepts and 40,000 terms in the domains of agriculture, forestry, fisheries, food and related areas like the environment. AGROVOC is a sizeable monolithic resource which includes explicit semantics of a hierarchical structure between terms representing agricultural concepts when compared to ordinary word-lists or glossaries. It also involves generic associative relationships that imply a semantic relationship between two entities and can be further developed in more complex relationships[16].

Additional examples of vocabulary/thesauri include the Chinese Agricultural Thesaurus (CAT), the Cab Thesaurus, the National Agricultural Library Thesaurus (NALT), the agINFRA Soil Vocabulary (SOIL), Biodiversity Thesaurus (BIODIVTHES) and the Thesaurus of Plant Characteristics (TOP). The CAT was created as an information management tool in the domains of agriculture, forestry and biology. It is China's second most comprehensive multidisciplinary thesaurus. CAT can be deemed a translated thesaurus, where definitions are simply introduced in Chinese and are translated in English. CAT has become a requirement to build upon for any document retrieval system in the area of agriculture and for archiving administrative and scientific research outcomes of the Ministry of Agriculture in China. The Taiwan Agricultural Science Information Centre has expanded the CAT [17]. The CAB Thesaurus is the primary search tool available to users of the CAB Abstracts and Global Health databases and related products, offering nearly 2.9 million descriptive terms in applied sciences. It has been in use since 1983 and it is regularly updated. It includes specific terminology for all subjects covered with about

265,900 names of plants, animals and microorganisms. It is multi-lingual, with Dutch, Portuguese and Spanish equivalents for most English terms, but less content in Danish, Finnish, French, German, Italian, Norwegian and Swedish. The SOIL vocabulary is a first formalisation of the Infrastructure for Spatial Information in the European Community (INSPIRE) data model into a Resource Description Framework (RDF) vocabulary. [4]

It has been developed, among others, by the Italian Consiglio per la Ricerca e la Sperimentazione in Agricoltura (CREA) in collaboration with the Global Forum on Agricultural Research (GFAR) and the FAO in the realm of the EU funded FP7-INFRASTRUCTURES project "A data infrastructure to support agricultural scientific communities" (agINFRA, Grant agreement ID: 283770). The SOIL vocabulary and the agINFRA project outcomes were built upon by the recently completed EU funded H2020-EINFRA project "Accelerating user-driven e-infrastructure innovation in Food Agriculture" (AGINFRA PLUS (AGINFRA+), Grant agreement ID: 731001) and H2020-INFRASUPP project "Towards an e-infrastructure Roadmap for Open Science in Agriculture" (e-ROSA, Grant agreement ID: 730988). [5]

Ontologies may be interpreted as evolutions of vocabularies and thesauri having become a significant resource for the representation of domain knowledge, and a central component of many information management, decision support and other smart systems also in the domain of agriculture[18]. The use of ontologies in agriculture is increasing for various purposes, including the possibility of sharing the knowledge built up in the agricultural sector among farmers all over the world, irrespective of the language they speak and/or read (AGROVOC Thesaurus;[19]). Ontologies in agriculture also facilitate farmers' decisions[20] and create semantic interoperability of agricultural systems[21].

Ontologies, as well as other semantic resources, can be found and are often accessible through repositories hosted on the public Internet. Examples of these repositories include the AgroPortal[22], available at http://agroportal.lirmm.fr/, the Crop Ontology (CO) curation and annotation tool on germplasm and traits[23], accessible at http://www.cropontology.org/, the Coherence in Information for Agricultural Research for Development (CIARD) Routemap to Information Nodes and Gateways (RING), accessible at https://ring.ciard.net/, and the VEST/AgroPortal Map of Standards, available at http://vest.agrisemantics.org/.

The AgroPortal, reusing the National Center For Biomedical Ontology (NCBO) BioPortal technology, aims to offer a reference ontology repository for agronomy. Building on the scientific outcomes and the experience of the biomedical domain, partners of the Semantic Indexing of French biomedical Resources (SIFR) project - which include e. g., the NCBO, the Research Data Alliance (RDA), the FAO, Global Open Data for Agriculture & Nutrition (GODAN Action) and the Institut National de la Recherche Agronomique (INRA) - have worked on the agronomy domain focusing on plants, food, environment and, possibly, animal sciences. AgroPortal features include ontology hosting, searching, versioning, visualising, commenting, recommending, semantic annotating as well as storing and exploiting ontology alignments while relying on a fully semantic web compliant infrastructure. In particular, the AgroPortal is built upon the requirements of the agronomic community such as the Simple Knowledge Organisation System (SKOS) RDF Schema and of the five agronomic use cases which have shaped the construction of the repository. These include the AgroLD on rice, the RDA Wheat Data Interoperability Working Group on wheat, the Open Vocabularies @ INRAE collecting all the vocabularies produced by INRA researchers and scientists, the CO and the GODAN Action which is is a map of standards in use for the exchange of agriculture and nutrition data.

[4] INSPIRE was established by Directive 2007/2/EC of the European Parliament (EP) and of the Council of 14 March 2007.

[5] The evidence and outcomes of the agINFRA, AGINFRA+ and e-ROSA projects will be reviewed in Task 1.6 of the AGRICORE project.

The CO comprises logically defined relationships on, among others, crop phenotype, breeding, germplasm, pedigree and traits allowing for computational reasoning on data annotated with a structured vocabulary. The use of ontology terms to describe agronomic phenotypes and the accurate mapping of these descriptions into databases is important in comparative phenotypic and genotypic studies across species and gene-discovery experiments since it provides a harmonized description of the data and therefore facilitates the retrieval of information. The CO is built using the Open Biological and Biomedical Ontologies (OBO) Format Syntax and Semantics and can be edited using the open-source, platform-independent application OBO-Edit. Furthermore, the CO is accessible via an Application Programming Interface (API), is hosted on Google App Engine and its versioned code is hosted on GitHub.

The VEST/AgroPortal Map of Standards is an online repository of standards and vocabularies which are used in the exchange of agriculture and nutrition data, which is promoted by both the FAO and GODAN Action. It builds on the FAO Vocabularies, mEtadata Sets and Tools (VEST) Directory and includes the AgroPortal ontology repository. It comprises 398 resources, as well as a graphical overview of the alignment of the semantic resources. In addition to the list of semantic resources, the VEST/AgroPortal Map of Standards also has an RDF query interface where all of the aforementioned semantic resources, can be queried using the SPARQL Protocol and RDF Query Language (SPARQL) through a webpage or via a Representational State Transfer (REST) Web API.[6]

The interoperability of semantic resources through the use of Linked Data is often referred to as agrisemantics (http://agrisemantics.org) and Linked Data Hubs can be seen as the first step taken towards the aim of the agrisemantics movement, which is the interoperability between semantic resources for agriculture. In this regard, the EU funded FP6 "System for Environmental and Agricultural Modelling; Linking European Science and Society" (SEAMLESS project, Grant agreement ID: 773786) made a great effort at trying to facilitate translating policy questions into alternative scenarios via a set of indicators capturing the key economic, environmental, social and institutional issues of the produced questions. SEAMLESS provided a smooth linkage between different scales (i. e., point or field scale, farm, region, EU and the world) allowing an Integrated Assessment and Modelling (IAM) approach. This approach enables smoother management of complex systems, improving integrated assessment and balancing the integration of the biophysical, economic, social and institutional aspects. To do so, linkages between micro- and macro-levels, balancing-methods for diverse disciplines (i. e., economic, social, biophysical) and institutional constraints were generated. Moreover, the project recognised the existence of numerous models and databases which are case-specific. The SEAMLESS outputs consisted of a combination of the SeamFrame server with the SEAMLESS database and knowledge base to facilitate the proposed IAM. Both the databases and the knowledge base depend extensively on the ontologies developed during the project, which plays a central role in harmonising and relating different concepts from diverse sources. The set of ontologies the SEAMLESS project developed (see[24] for more detailed information) aimed at facilitating the integration "of a variety of combinatorial, simulation and optimisation models related to agriculture […] by using them to specify data communication across the models and with a relational database."[24, p.1]. The ontologies characterising SEAMLESS are model-focussed and propose a common data schema developed accounting for structural and semantic interoperability. Similarly to the process employed to create the ontology in the AGRICORE project, SEAMLESS adopted a community process for knowledge elicitation (see[25]): the first step involved the researchers working on the project, who were asked to produce a list of concepts they considered relevant (for all the selected datasets in AGRICORE, for "model coupling" in SEAMLESS). These concepts were later populated with examples and comments to ensure easy readability and a comprehensive final list was set serving as the lexicon. Iterative discussions among project partners modified the full list, clearing out unclear and conflicting concepts as well as generating

[6] SPARQL is a query language, designed by W3C in an attempt to standardise querying of RDF data sources.

the common ontology to be used (see also[26] for further details on this approach). Eventually, SEAMLESS developed 11 small ontologies, each one of them focused on a specific aspect of the project, while granular ontologies share common concepts and relationships (see[27] for a detailed discussion on data-related ontologies within SEAMLESS).

2.1.2.1 Online data set repositories and index tools

Similarly to the objective of the AGRICORE project of creating the ARDIT, other initiatives have already tried to generate repositories of datasets related to the agricultural knowledge domain. One of the most extensive ontology and most comprehensive global repository of agri-food datasets and data services is the CIARD RING, promoted by GFAR to allow information providers to register their services and datasets to facilitate the discovery of sources of agriculture-related information across the world. It is one of the outcomes of the agINFRA project (Grant agreement ID: 283770) and currently hosts some 34 million records, of which some 6 million ones are fully available resources. The CIARD RING employs the W3C Government Linked Data Working Group vocabulary and indexes both "datasets" and "data services". According to the relevant vocabulary, the former are "a collection of data, published or curated by a single source, and available for access or download in one or more formats" while the latter are any type of data service on the web, from a simple website to a search engine to an API to a data dump. Examples of datasets indexed in the CIARD RING include a RDF Site Summary (RSS) feed reachable at a Uniform Resource Locator (URL), an eXtensible Markup Language (XML) dump downloadable via File Transfer Protocol (FTP) or reachable at a URL, a comma-separated values (csv) or Network Common Data Form (NetCDF) file available at a URL and an API call already parametrised to retrieve a specific dataset. A particular class of datasets available in the CIARD RING comprises the dynamic dataset endpoints such as SPARQL engines that respond to a query with an RDF response that represents a dataset, Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) targets that respond to a verb call with an XML response and any web service or API endpoint whose response is a dataset.

The EC, through the Open Data Portal, provides access to an increasingly wide range of data from the EU and other EU bodies. The data can be used and reused for commercial or non-commercial purposes. An easy and free access to data has the goal of contributing to their innovative use and realising the economic potential that can derive from them. At the same time, the portal aims to make EU institutions and other bodies more transparent and accountable. The EU Open Data Portal was established in 2012 and all EU institutions are invited to make their data available to the public whenever possible through this tool. The domains of data provided include:

- geographic, geopolitical and financial data,
- statistics,
- election results,
- legal acts,
- data on crime, health, the environment, transport and scientific research.

All these data, as already pointed out, are available for free but the source must be cited and they can be reused in databases, reports or projects. Only a small part of this data is subject to specific conditions on reuse, most of which have to do with the protection of third party intellectual property rights. The portal offers:

- a standardised catalogue, which provides easier access to open EU data,
- a list of ICT applications and web tools that reuse such data,
- a SPARQL endpoint query editor,
- a REST API access,

- advice on how to best use the site.

The EU Open Data Portal provides access to 860,000 public datasets from 35 countries (EU MSs, the European Economic Area (EEA), Switzerland and countries in the EU Neighbourhood Policy programme). Data resources are available in six languages and are indexed by the EC from national, regional, local and domain-specific public data providers. The data can be extracted using an easy tool which facilitates the discovery of public open data. Data can be retrieved by geographical location and by time period. The time series contain up to five decades of statistical observations. Coverage varies depending on national, regional, local or domain-specific public data providers.

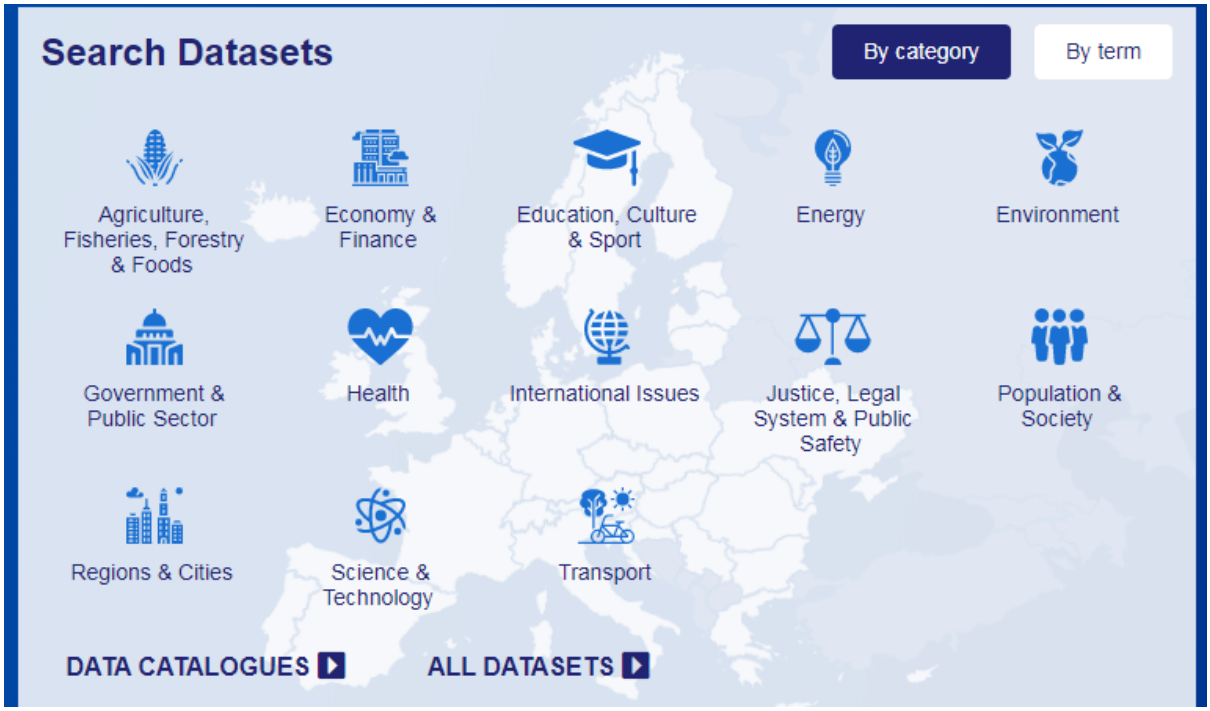

Figure 5 Datasets by Category in the EU Open Data Portal

The EU Open Data Portal contains a large number of datasets in the knowledge domain of agri-food. For the "agri" part of the domain, it relates to farming as the process of producing food, feed, fibre, and other desired products that can be obtained from cultivating selected plants and raising domesticated animals (livestock) and provides benchmark data of the agricultural sector. The EU Open Data Portal hosts a collection of open data to facilitate the transformation of agriculture and ensures food security around the world. These datasets include, among others, weather data, data on seed genetics, data on environmental conditions and soil data to help agriculture face its modern challenges applying data-driven (evidence-based) strategies.

The Modelling Inventory and Knowledge Management System (MIDAS) of the EC is a Commission-wide knowledge management tool designed, in a Service-Oriented Architecture (SOA), to relate data, models, scientific publications and policy actions associated with the EC and its dedicated services like the JRC, policy impact assessment efforts. By enhancing the transparency and traceability of the models used by the EC to inform its policymaking, MIDAS contributes to the EC Better Regulation Agenda. Currently, the EC uses more than 150 models applied on a broad set of domains to pursue evidence-based policymaking. The majority of these models are run in combination with other models forming complex networks of interactions and dependencies. The proliferation of models, and their combinations, required the development of a tool for maintaining an overview of ongoing modelling activities for transparent and coherent use of models in support of the policy cycle. MIDAS is based on high quality consistent and updated data written in such a way that both experts and non-experts can benefit from it. This

has been achieved thanks to the following key principles: capture the relevant, update & check regularly, retrieve information stored and maintained elsewhere, use of permanent identifiers for consistency and data access. In particular, updating and checking - occurring at least once a year - has been relying recently on the Semantic Text Analysis (SeTA) word embedding neural network. The latter, also developed at the JRC, processes approximately 500,000 documents from sources like the EU Bookshop, EUR-Lex, CORDIS, the EU Open Data Portal and the JRC PUBSY to update the relationships among model acronyms, the data the models used and the relevant terms identifying the relevant policy measures and impacts. MIDAS is characterised by good usability thanks to the implemented data visualisation techniques capable of communicating the resulting complex network of relationships to a non-technical audience, revealing the bigger picture and allowing users to identify patterns about model use they might not have been aware of. MIDAS is an online platform accessible from within the EC Network to all its staff and services, as well as to those of the EP. Additional knowledge sharing (and accountability) would be brought about by opening the tool up to the general public, but this is still a matter for discussion.

2.1.3 Metadata and its relation with ontologies

As already described and quoting from the literature, "an ontology is a model language that can build models, which support the conceptual integration of the distributed domain data and the inference of relationships among the concepts as a result of activities such as concept analysis and domain modelling using the standard methodology"[13, p.1]. Therefore, ontology development requires developers to discuss domain concepts, relationships and constraints with experts in the relevant knowledge domain. Because of this interaction process consumes a lot of human resources and time, advances in ontology development are relying on methods to automatically define an ontology model by casting a domain in the form of the metadata that can characterise the domain and apply rules to the metadata[13]. Metadata is structured and coded data that describes the characteristics of media objects by facilitating their identification, detection, evaluation and management. Metadata is used to describe the meaning and properties of information to understand, classify, manage and exploit the data better. Accordingly, metadata is employed to facilitate interoperability and integrating resources between humans and machines. Methodologies that automatically generate an ontology from metadata must pre-process them to create a template and then apply an ontology-generating rule. The generated ontology model does not focus on how to manage the presence of many individuals. Individual inputs into the generated ontology model can be stored in a table in one of the datasets in a triple form that consists of a subject, a predicate and an object. Using this approach, it is possible to provide efficient management and query functions for individuals of the corresponding schema. The individual is the basic component of an ontology. The role of individuals in an ontology is to classify objects according to their class, which is the concept of a domain. Individuals in OWL correspond to constants in first-order logic and instances in the RDF[13].

In the process of defining the AGRICORE ontology, the above-described methodology was taken into consideration but appeared not viable. It was preferred to choose a data model and use the related metadata specifications to construct the ontology. The metadata specification chosen to describe AGRICORE related dataset is the DCAT-AP. The DCAT-AP for data portals in Europe is a specification based on W3C's Data Catalogue (DCAT) vocabulary for describing public sector datasets in Europe. Its primary use case is to enable searching for datasets across data portals and make public sector data better searchable across borders and sectors. This can be achieved by exchanging the descriptions of datasets among data portals. The specification of the DCAT-AP was a joint initiative of Directorate-General for Communications Networks, Content and Technology (DG CONNECT) of the EC, the EU Publications Office and the Interoperability Solutions for public Administrations, businesses and citizens (ISA^2) programme of the EC. The specification was elaborated by a multi-disciplinary Working Group with representatives from 16 EU MSs, some European Institutions and the United States. The first version (1.0) of the DCAT-

AP was published in September 2013. In 2015, a revised version (1.1) was developed and released in November 2015 with changes based on the feedback received from users. The description of the updated DCAT-AP (2.0) is presented in the section AGRICORE DCAT-AP extension, including the description of classes (mandatory, recommended and optional).

2.1.4 Why then a new ontology is needed

As described in the previous section, during the extensive research activities carried out to design the ARDIT and the related AGRICORE ontology, semantic web resources were surveyed including controlled vocabularies, taxonomies, thesauri and ontologies. The agricultural knowledge domain is well served by freely available resources because, starting from the nineties, there has been a concerted effort to develop semantic resources for agriculture by various national agencies. The already mentioned semantic resource AGROVOC by the FAO is an example of the most comprehensive and extensive vocabulary of the agricultural knowledge domain available. Semantic resources for the knowledge domain of agriculture are typically one of two types: they are either concerned with agriculture as a general concept or are specialised in (one or more of) its sub-domains, while commonly allowing for some overlapping between the resources.

Exploiting the results of an existing review, an insight in the available resources confirmed the overlapping but at the same time the gaps in the representation of the general agricultural knowledge domain.

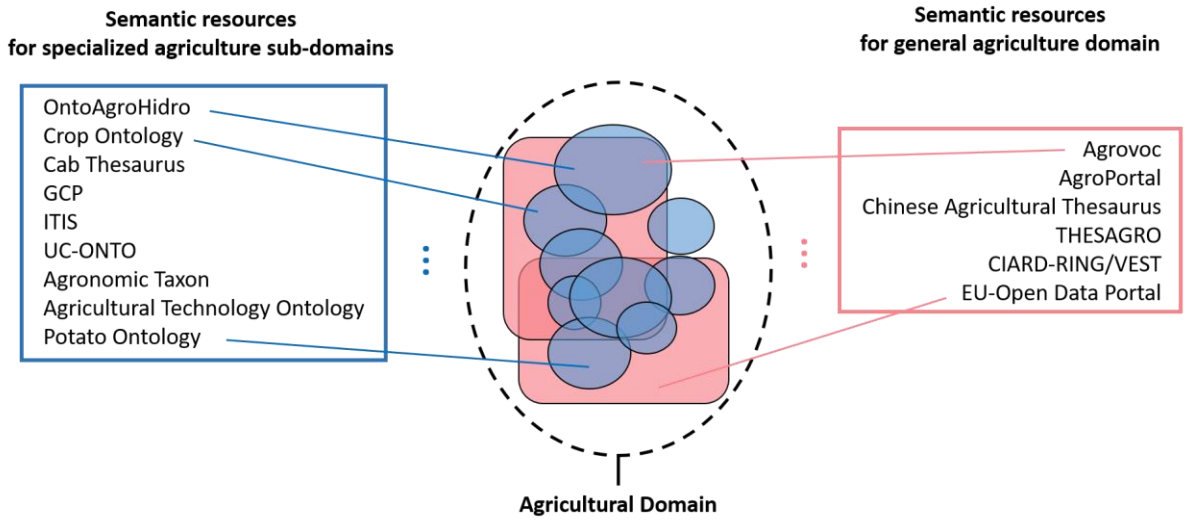

Figure 6 Specialised and General Semantic Resources in the Domain of Agriculture

Existing semantic resources include research targeted available ontologies, ontologies repositories and dataset repositories in the knowledge domain of agriculture. For instance, AgroPortal agriculture ontologies repository indicates that there are 124 domain-specific ontologies in the agricultural area. However, the ontology needs of AGRICORE (and the ARDIT platform) relate on a general domain ontology which provides a general view of the agricultural domain which has not been identified within the analysed sources.

A different perspective is embodied in the CIARD RING, a federated and curated metadata registry of agri-food datasets and data services which is an index of vocabularies and a repository of semantic web services. The RDF data model behind the CIARD RING is DCAT-AP and, through the portal, it is possible to register an information service or dataset.

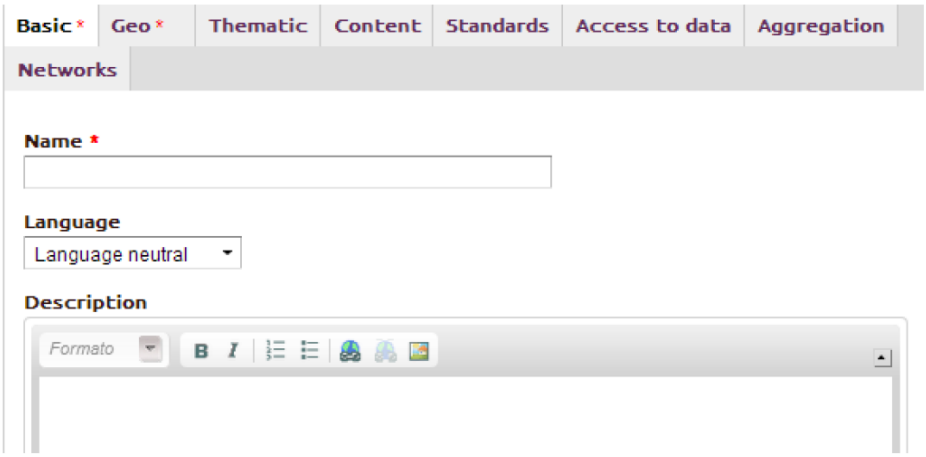

Figure 7 CIARD Ring GUI to Register a Dataset or Data Service

The CIARD RING is also an RDF store. An RDF store is a way of storing data using a machine-readable "grammar" (the RDF) and documented semantics (RDF vocabularies).

Similarly, the EU Open Data Portal is a dataset repository acting as a point of access to public data published by the EU institutions, agencies and other bodies. The data model behind the EU Open Data Portal is again the DCAT-AP RDF vocabulary.

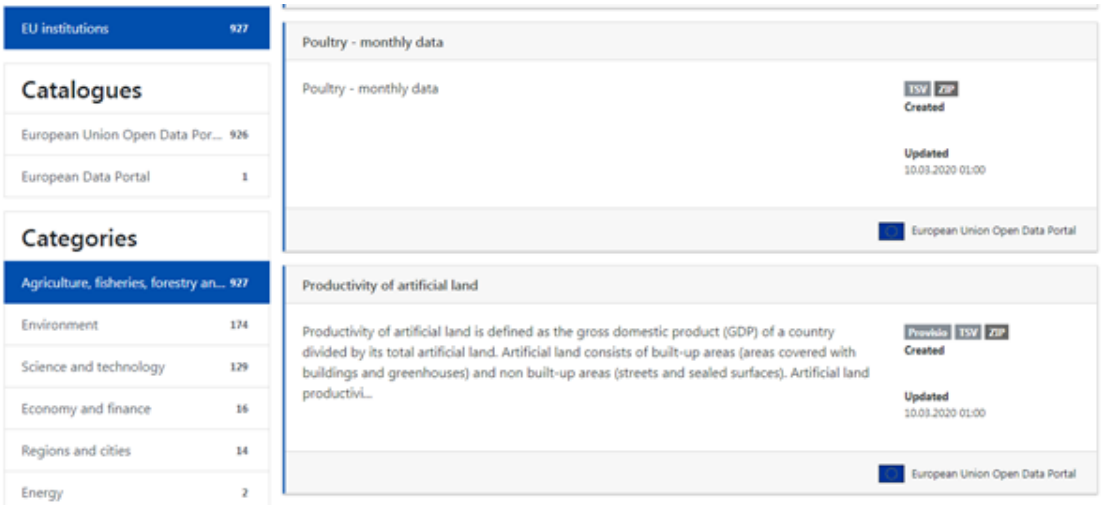

Figure 8 Results of Dataset Search on EU Open Data Portal

Likewise, datasets registration and retrieval follow the same CIARD Ring approach, with different options for data publication and consultation available.

However, the capabilities of searching for datasets characteristics of interest are limited to providing high-level information while not allowing to perform a search at the level of (a) dataset variable(s), which is the gap that the ARDIT functionalities - built upon the ontology developed in the AGRICORE project - is aiming to fill in. The possibility of performing semantic searches capable of retrieving information also about individual and selected variables in a dataset would be a significant advancement in the search functionalities of a public data index or repository, allowing for more effective data exploitation. This ambitious objective can be achieved by building an ontology upon a selected data model (e. g., DCAT-AP), inheriting all its classes and properties, but extending it to obtain a more detailed characterisation of datasets through class extension and properties inheritance.

As mentioned in the previous sections, an ontology is an explicit formal specification of a shared conceptualisation. An ontology provides concept definitions, hierarchies, and relationships between concepts in a knowledge domain. Ontologies enhance the performance of information retrieval systems and offer solutions to the effective management of extensive collections of data by modern information systems. An ontology-based semantic representation of agricultural data sources enables semantic concept-based data processing and retrieval. With this perspective, the AGRICORE DCAT-AP 2.0 ontology was developed to allow users to overcome the impossibility of retrieving data referring to (a) specific variable(s) contained in a given dataset from existing open data portals.

The process enabled expressing the information relevant for researchers through the ontology schema, keeping track of the properties and relations of the imported data model and the newly created classes. The data model chosen for the ARDIT ontology is the DCAT-AP RDF vocabulary, which constituted a solid basis for data description, and the data schema provided by the newly created AGRICORE DCAT-AP 2.0, with new extensions and classes, have proven satisfactory to map the required dataset properties, as described in details in the AGRICORE DCAT-AP 2.0 Technical Documentation, annexed to this document.

Find below an example of the extension of the DCAT-AP class DATASET as done in the AGRICORE DCAT-AP 2.0 ontology, showing how an Environmental Variable might be described in the model.

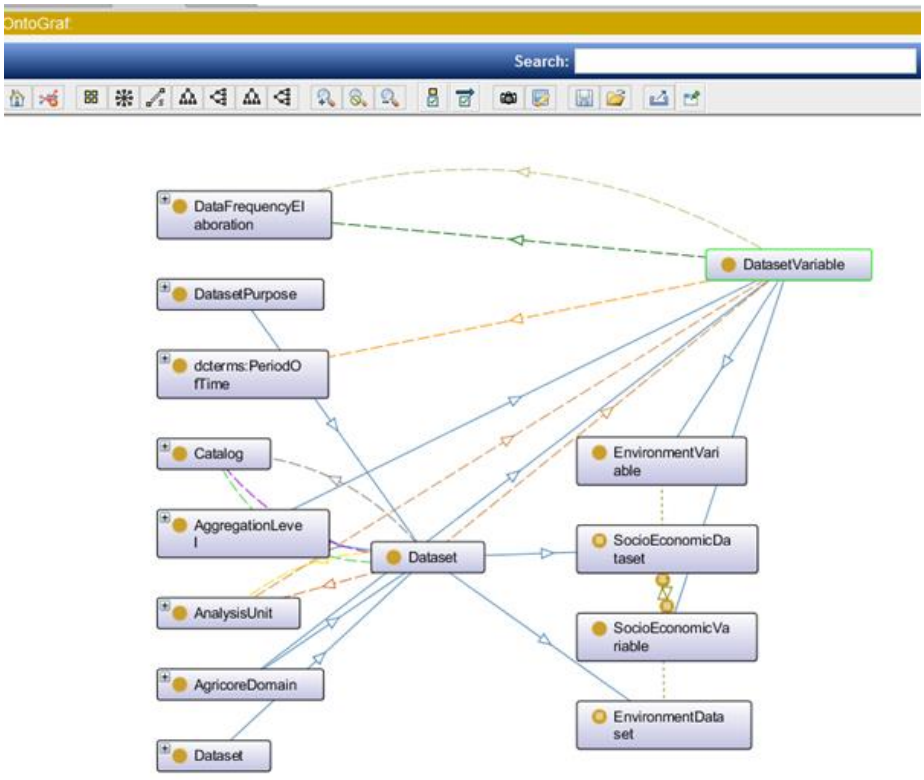

Figure 9 Environmental Variable Description in the AGRICORE DCAT-AP 2.0 Ontology for the ARDIT

3 ARDIT (AGRICORE) Datasets characterisation methodology

3.1 Introduction

Studying the impacts of policy measures on agricultural systems relies on a wide range of datasets from diverse knowledge domains which, for the execution of the activities of the AGRICORE project as well as for the operation of large-scale and sophisticated models often combining different platforms, include environmental, economic, social, biophysical, policy and climate data. The datasets, in each knowledge domain, most suitable to be employed in each modelling effort first need identification based on model requirements and research questions. Often, this is impossible without preliminarily access to the data themselves, resulting in a possible wasteful effort of the researcher should the data reveal themselves inadequate to the modelling job at hand. This could be because the unit of analysis of the dataset is different from the level at which the model operates (i. e., data available for every region vs. individual (agent/farm) level model(s)) or because the data are available for only one year when the model is dynamic in nature and could require repeated observations of the variables over time. Most commonly though, a dataset may not fit the requirements of the model and/or the researcher because of the lack of one or more key variables necessary to run the model to respond to the research question currently investigated. In turn, the researcher will have to look for another data source to either replace or integrate the one(s) already obtained and examined. Furthermore, researchers may find retrieving the actual datasets from the original provider or other dataset repositories particularly challenging. This could be due to data being available only upon having complied with the access requirement imposed by the provider. Because fulfilling these requirements could be a time-consuming activity which may delay the acquisition of the data, early and complete knowledge of the procedure for accessing the data could be valuable for the researcher. Likewise, the amount of data necessary to use large-scale models, or combinations of multiple models, is likely to be significant. Therefore, properly organising the relevant datasets to be ready for use, by selecting or manipulating parts of them, could be a limiting factor for (some) researchers.

Given all these needs of (agricultural) researchers, the AGRICORE project has addressed them in Task 1.1 by defining a methodology for characterising the datasets relevant for the analysis of policy impacts in the domain of agriculture, without necessarily having to obtain the data in advance. Therefore, this activity has relied on having access to good and complete descriptions (i. e., metadata) of the contents of the dataset, especially on the characteristics of the variables contained in them. This methodology has been developed to collect all the relevant information on datasets, such that this knowledge could be stored, manipulated and displayed employing the AGRICORE ontology. Lastly, this information will be gathered and hosted in the ARDIT (previously EU Index Tool), which will be then used to allow researchers to identify useful data sources for them.

Next sections describe how this methodology has been devised and in what it consists of. The methodology will be used first during the AGRICORE project characterisation efforts (undertaken in Task 1.2 to 1.6) and then, after the project, to keep updating the index content by external partners. Please notice that the design of the methodology is intrinsically linked to the development of the ARDIT ontology, as the former describes how to capture the information which is structured following the latter.

3.2 A standardised process for the characterisation

The proposed methodology to characterise agricultural datasets is described through the three main elements that it is composed of:

1. The proposed ontology: The AGRICORE DCAT-AP 2.0 Extension
2. The ARDIT tool
3. The data governance process

This process will be applied within the project to characterise an extensive set of data sources covering EU-wide and local datasets (especially those relevant to the three use cases covered in the AGRICORE project). Moreover, this methodology also describes the process that should be followed after the project to provide more content in the proposed ARDIT. The next points describe in detail each of these three elements.

3.2.1 Ontology for ARDIT

3.2.1.1 Introduction

The approach followed to implement the ontology for the ARDIT followed an iterative process involving experts' knowledge of agricultural, geographic and statistical datasets and ICT developers exploiting cross-cutting competencies on data standards and controlled vocabularies.

The different steps which led to the AGRICORE extension of the DCAT-AP 2.0 ontology, instrumental to developing the ARDIT, are hereby summarised and described in the following "Process description" section, according to this general outline:

1. Analysing the tools available for editing the ontology;
2. Reviewing the existing ontologies in the domain of agriculture to match the AGRICORE project needs;
3. Reviewing the experts' needs in terms of datasets of interest and elements required to characterise them and develop the ontology;
4. Choosing the data model suitable to develop the ontology and create the backbone of the data schema;
5. Choosing the ontology approach, bottom-up, starting from needed data, top-down starting from a standardised data model or mixed approach;
6. Develop the ontology extending the base data model with new classes, including all needed variables and themes to describe the data sources, making use of properties and controlled vocabularies.

3.2.1.2 Process description

3.2.1.2.1 Step 1 - Analysis of available tools

As already described in section 4.1 "What is an ontology" of the present document, different editing tools were taken into consideration but Protégé 5 http://protegeproject.github.io/protege/getting-started/, was selected due to the large community deeming it a reliable and flexible cooperative tool available as an open-source product both in a desktop and web version. As reported by previous studies[28], it has a suite of tools to construct domain models and knowledge-based applications with ontologies. It implements a rich set of knowledge-modelling structures and actions that support the creation, visualisation and manipulation of ontologies in various representation formats. It can be customised to provide domain-friendly support to creating knowledge models and entering data.

Also, it can be extended by a plug-in architecture and Java-based API for building knowledge-base tools and applications. Protégé allows the definition of classes, class hierarchy's variables, variable-value restrictions, and the relationships between classes and the properties of these relationships.

The significant advantage of Protégé is its scalability and extensibility. Protégé allows building and processing large ontologies efficiently. Through its extensibility, Protégé might be adopted and customised to suit users' requirements and needs. The most popular type of plug-ins is tab plug-ins; currently available tabs provide capabilities for - among others - advanced visualisation, ontology merging, version management and inference. The OntoViz and Jambalaya tabs, for example, present different graphical views of a knowledge base, with the Jambalaya tab allowing interactive navigation, zooming in on particular elements of the structure and different layouts of nodes in a graph to highlight connections between clusters of data. Protégé supports collaborative ontology editing as well as annotation of both ontology components and ontology changes, hierarchy's variables, variable-value restrictions, and the relationships between classes and the properties of these relationships. Many of the Protégé features mentioned above were used during the development of the AGRICORE ontology.

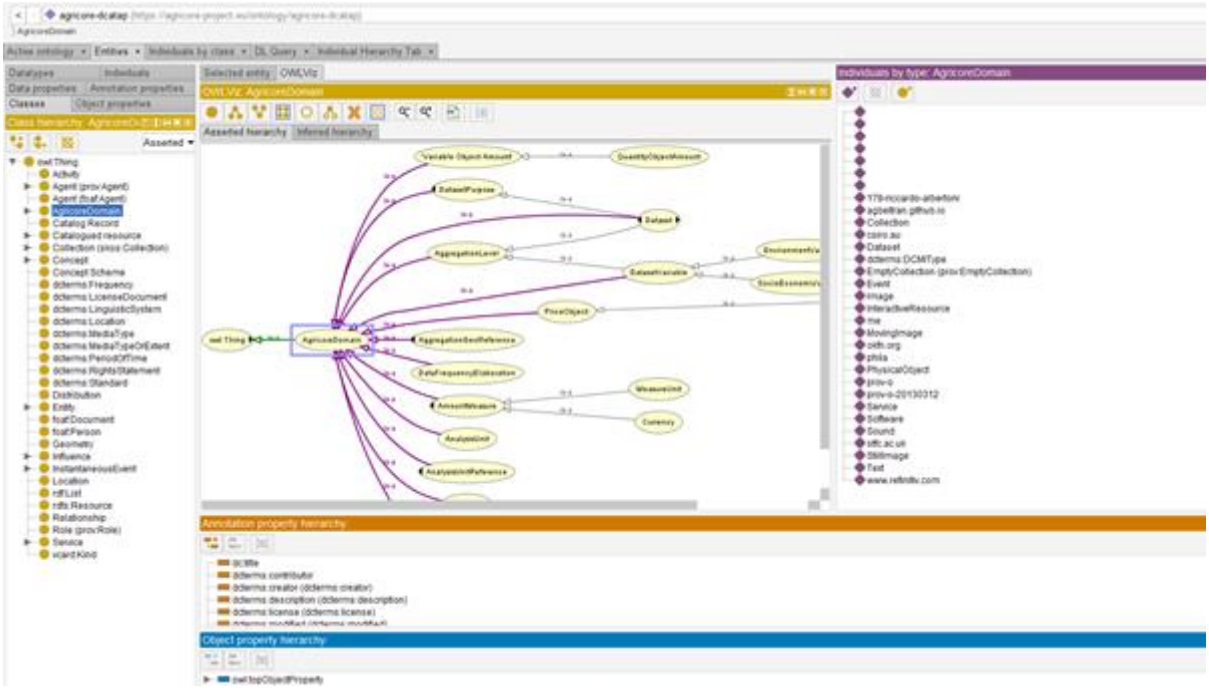

Figure 10 Screenshot of the Functionalities of the Protégé Editor

3.2.1.2.2 Step 2 - Review of existing ontologies

The second step in the development of the ontology was the review of the existing ontologies, as described in sections 4.1.2 "Previous ontologies and related work in the agricultural knowledge domain" of the present document. As stated in section 4.1.4 "Why then a new ontology is needed", the decision taken was to focus more on a standardised data model than on a pre-existing ontology, which would have been difficult to be adapted to the needs of characterising datasets in the AGRICORE project.

This decision was taken mainly due to the requirements of step 3, Analysing the experts' needs in terms of datasets of interest and elements required to characterise them and develop the ontology. Namely, the requirements are listed here below:

- Ontology Domain: Datasets to be included in ARDIT;

- Ontology Use: Datasets consultation, retrieval and addition of the characterisation of a new dataset in the ARDIT;
- Specific Competency Questions: Semi-natural language users' queries to retrieve the datasets in the ARDIT, also variables-based;
- Users: AGRICORE users and agricultural researchers outside the project;
- Updates: AGRICORE developers (during the project) and agricultural researchers after it.

3.2.1.2.3 Step 3 - Assess user needs

The assessment of the actual needs for both the methodology and the being-defined ontology has been one of the more time-consuming stages. This process has been done collaboratively between all project partners and especially by WP1 contributors.

First, partners were requested to extend the list of data sources already identified at the proposal stage. This new list would allow to have a clear image of the datasets that would be characterised within the project. This list should include both EU-level datasets but also those specific and needed for the execution of the three planned UCs. Furthermore, to provide the research community with a complete and noteworthy search tool, partners were asked to consider - already from the start of the project and, more importantly, over the whole duration of Tasks 1.2 to 1.6 - the possibility of characterising any dataset which might be relevant for the analysis of the impacts of (agricultural) policy (reforms). The number of datasets to be characterised by the end of Tasks 1.2 to 1.6 is not pre-defined and will be continuously updated.

Second, this list was split into two sublists identifying those most likely to be used within the project. The "probability" of its use was based both, on the knowledge of project experts and their opinion on such data sources utility for the project goals as well as the easiness to get hold of the data sources.

Third, the project partners were requested to start characterising the data sources present in this sub-list to produce a first image of the information that would need to be captured by the ARDIT ontology. Accordingly, the initial dataset characterisation effort and AGRICORE/ARDIT knowledge domain construction have relied on the information on some 90 datasets, of which a subset is presented in the following table.

Database Characterised	Type of Dataset
FEGA	Socioeconomic (Policy)
(Spanish) Statistics of meterophenological variables	Env & Climate
Statistics of land prices in Andalusia	Socioeconomic
Organic Farming in Spain	Socioeconomic
MARM. Household consumption database	Socioeconomic
BDSICE. Cost and price index	Socioeconomic
BDSICE. National production and demand indicators	Socioeconomic
BDSICE. Price and costs. Agricultural wage index	Socioeconomic
BDSICE. Price and costs. Salary increases in agreement and salary increases registered in agriculture	Socioeconomic
(Spanish) Agrifood Foreign Trade Statistics	Socioeconomic
(Spanish) Monthly production, movement and stock data (AICA)	Socioeconomic
ELSTAT - Livestock Surveys (for Greece)	Socioeconomic and Env & Climate
ELSTAT - Annual Agricultural Statistical Survey (for Greece)	Socioeconomic
ELSTAT - Farm Structure Survey (FSS) (for Greece)	Socioeconomic
ELSTAT - Crops Survey (for Greece)	Socioeconomic

Hellenic National Meteorological Service	Env & Climate
ELSTAT - Census of Agricultural and Livestock Holdings 2009	Socioeconomic and Env & Climate
EU FADN	Socioeconomic
AgMERRA Climate Forcing Dataset for Agricultural Modeling	Env & Climate
AgCFSR Climate Forcing Dataset for Agricultural Modeling	Env & Climate
Global Summary of the Day (GSOD)	Env & Climate
WorldClim Version 2.1	Env & Climate
Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2)	Env & Climate
Climate-related dataset for Poland	Env & Climate
Database on mineral nitrogen content in Poland	Soil/Land/Quality/Biodiversity
Climatic Research Unit Time-series (CRU TS) dataset v. 4.04	Env & Climate
SoilGrids	Soil/Land/Quality/Biodiversity
SoilHydroGrids	Soil/Land/Quality/Biodiversity
BISE Biodiversity Information System for Europe	Env & Climate
European Climate Assessment & Dataset	Env & Climate
MIRCA 2000	Soil/Land/Quality/Biodiversity
NASA Prediction of Worldwide Energy Resources (POWER)	Env & Climate
ARMA - RDP 2014-20 (Poland) Implementation reports	Socioeconomic (Policy)
Statistics Poland - Agricultural and horticultural crops	Socioeconomic
Statistics Poland - Local Data Bank	Socioeconomic and Env & Climate
Poland - Animal production, Farm animals	Socioeconomic
ADMS - Agricultural Drought Monitoring System	Env & Climate
GIOŚ – datasets (Inspectorate of Environmental Protection Reports)	Env & Climate

Table 1 Selection of Datasets Characterised so Far

For this characterisation process, the following initial knowledge table was defined and filled in for each dataset.

Characteristic to be captured and Presented in ARDIT	Explanation/Example (When missing or irrelevant for the specific dataset, leave blank)
Name	*Please add the name of the dataset*
Type of dataset (Objective of analysis)	*Please add a brief description of the general scope of the dataset (e. g.: Absolute and index prices for agricultural input and output products, per MS*
Number of farms covered	*Please add the number of farms surveyed in the dataset. Leave blank for non-farm-specific datasets*
Includes all farms in the covered area	*Yes/No*
Number of grid cells	*Please indicate the number of grid cells, if this is relevant for the dataset type surveyed (e. g., environment, climate)*
Number of variables included	*Please indicate the number of variables in the dataset*
Type of variables included	*Please provide the type of variables included in the dataset (e. g.: nutrients, temperature, input prices)*
Link	*Please provide the direct link to the specific database*
Data contained	*Please provide the list of data (tables) contained in the dataset E. g.:* *Selling prices of agricultural products (absolute prices).*

	• *Selling prices of crop products (absolute prices) - annual price (from 2000 onwards).* • *Selling prices of animal products (absolute prices) - annual price (from 2000 onwards).* • *Purchase prices of the means of agricultural production (absolute prices) - annual price (from 2000 onwards).* • *Selling prices of crop products (absolute prices) - annual - old codes - data from 1969 to 2005.* • *Selling prices of crop products (absolute prices) - monthly - old codes - data from 1969 to 2006.* • *Selling prices of animal products (absolute prices) - annual - old codes - data from 1969 to 2005.* • *Selling prices of animal products (absolute prices) - monthly - old code - data from 1969 to 2006.* • *Purchase prices of the means of agricultural production (absolute prices) - annual - old codes - data from 1969 to 2005.* • *Purchase prices of the means of agricultural production (absolute prices) - monthly - old codes - data from 1969 to 2006.*
Data link	*Please provide the direct link to the specific database*
Number of crops	*Please indicate the number of crops for which the variables are available*
Activity sector	*Please indicate the activity sector for which the dataset provides information on*
Dataset structure	*Please describe the structure of the dataset*
Dataset format	*Please provide the format in which the dataset can be downloaded*
Data source access	*Please provide where and how the data are made available to the public (e.g. statistical yearbooks, reports)*
Geographic scope	*Please provide the geographic area(s) to which the dataset refers to*
Geo-referenced dataset	*Please indicate whether the dataset is a geo-referenced dataset or not. In case it is, please indicate the resolution of the data*
Anonymous	*Please indicate the confidentiality of the data retrieved*
Linked to individual farms	*Please indicate whether the dataset is linked to individual farms*
Source	*Please indicate the source of the raw data*
Process followed to gather	*Please indicate the process followed to gather the raw data*
Author	*Please indicate the data provider*
Maintainer	*Please indicate the data maintainer*
Last actualisation	*Please indicate the last actualisation of the dataset*
Update frequency	*Please indicate the frequency with which the dataset is updated*
Periods covered	*Please indicate the time span covered by the data*
Release date (past and expected future ones)	*Please indicate the dates of last and next release of the dataset*
Measurement units	*Please indicate the measurement units employed in the dataset*
Additional information	*Please provide any additional relevant information*

Table 2 Template for Detailed Dataset Characterisation

3.2.1.2.4 Step 4 - Choosing the data model for developing the ontology

In order to properly define a data model to be used as a basis for developing the ontology, the project partners analysed the results of the characterisations done in the previous step. In parallel, several consultation sessions and discussion panels were organised to evolve the above-presented table into a more mature characterisation template, which would then be used as a basis to define the ontology classes. The meetings resulted in the following improved characterisation template.

Characteristic Captured and Presented in ARDIT	*Explanation/Example*
Geographical coverage	*E. g., EU-28; World; Spain; Poland; Greece. It is also possible to use a coordinates box for this*
Type of data set (Object of analysis)	*E. g., Agricultural dataset, Survey on the structure of the agricultural holdings (Agricultural dataset), Administrative*
Unit of analysis	*In case of environmental, biophysical and environmental datasets this specifies the definition of the covered units (parcels, grid cells, climatic zones, etc.). Some adaptation from partners is required. In other cases: e. g., Single farm; NUTS# regions; Prices; Quantities; Policy measures; Meteorological stations*
Name	*Please indicate the name of the dataset*
Distribution	*Please indicate the distribution details*
Data Service	*If present, please indicate the data service (API) for access details*
Link to the dataset information	*Please indicate the hyperlink to the dataset*
Producer	*Please indicate the institution which publishes/maintains it*
Language of the dataset	*Please fill in with the language in which the data and metadata are available*
Type of access	*E. g., Publicly available; Access request required; Registration required*
Description of procedures to access the data	*If "Type of access" is "access request required", please describe the main characteristic of the procedures to access the data*
Statistically representative	*Yes/No; Please specify what it is representative of and the weighting factor(s)*
Aggregation level	*Please specify the spatial units of the data (i. e., 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 data set/ 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)*
Temporal extent	*Please indicate the temporal extent of the dataset (e. g., From 2008-01-01T11:45:30 to 2008-12-31T09:10:00)*
Periodicity of publications	*Please include how frequently the data are published (e. g., Yearly, Quarterly, Monthly, Weekly, Daily, Intraday)*
Data frequency	*Please indicate the frequency of the data in the dataset (e.. g., Annual, Quarterly, Monthly, Weekly, Daily, Intraday, Hourly)*
Mathematical representation of the data	*Please indicate here if the data appear in the dataset as e. g., Average, Instant value, Max, Median, Min, Mode, Sum, Variance, Standard deviation of other data*
Dataset format	*Please indicate the dataset format (e. g., xls, csv, html, pc axis, spss, tsv, gdx)*
Useful for the analysis of	*Please indicate which type of policy is more commonly evaluated using this dataset (e. g., Environmental policy; Income and distributional policies; Technical efficiency levels; Technical efficiency gains; Economic efficiency levels; Economic efficiency gains; Trade policy; Climate change policy; Insurance policy; Greenhouse gas emissions policy)*
Values	*Please indicate the values with which the variables are recorded (e. g., Index(es), Absolute, Percentage, ShareIndex(es), absolute, percentage, share)*

Themes covered	*E. g., General information on the holding; Type of occupation; Labour; Assets; Quotas and Other Rights*
Variables included	*Please list the variables included in the dataset and whether they are socioeconomic or environmental variables*

Table 3 Template for the ARDIT Dataset Characterisation (Methodological Grid)

To properly assess the linking of this information to the one already existing in DCAT-AP 2.0, the DCAT-AP 2.0 data model (please check sections 5.3.3 DCAT-AP extension and 6 AGRICORE DCAT-AP 2.0 Technical Documentation for a detailed description) was imported in Protégé, namely the DCAT-AP 2.0 RDF vocabulary was imported (https://www.w3.org/ns/dcat2.rdf), and the required relationships among classes implemented.

3.2.1.2.5 Step 5 - Choosing the approach for designing the ontology

Step 5 of the approach consisted in the development of the ARDIT ontology itself, adopting a mixed top-down/bottom-up approach, allowing the developer to adapt the DCAT-AP data model to the characterisation template in an iterative and collaborative relationship with datasets experts.

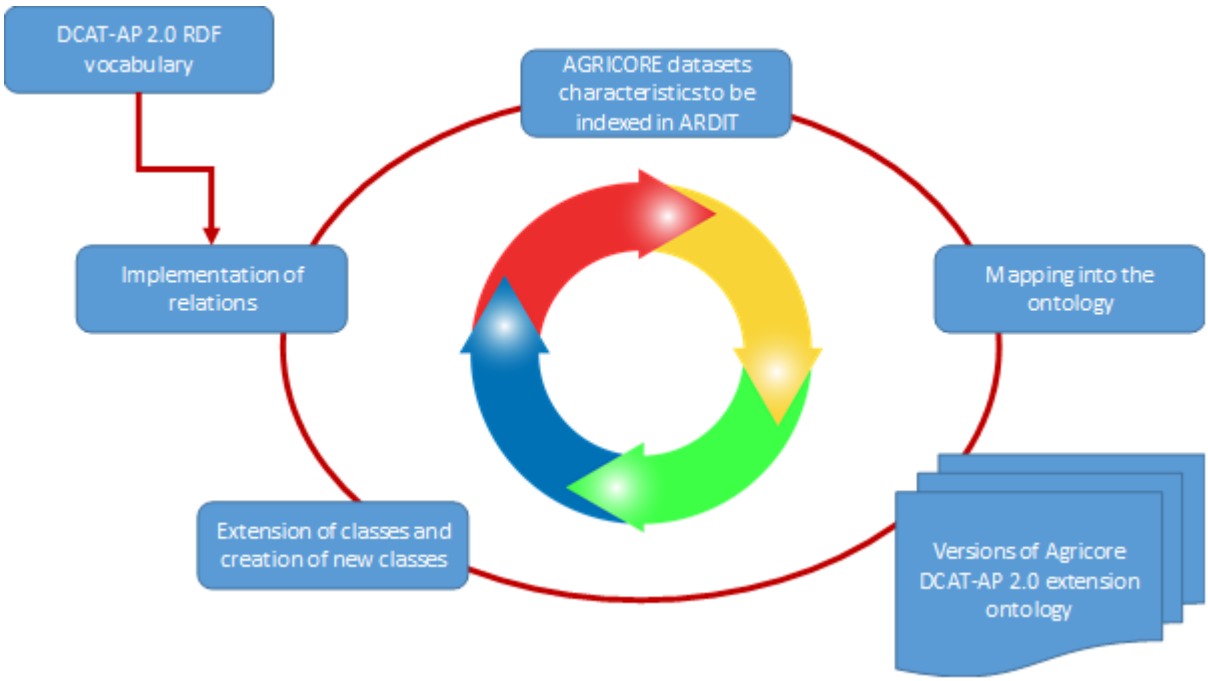

Figure 11 The Iterative Process of Creating the AGRICORE DCAT-AP 2.0 Extension Ontology

3.2.1.2.6 Step 6 - Development of the ontology

Finally, Step 6 consisted of the implementation of the ontology. This process ended up in a technical definition of the AGRICORE DCATA-AP 2.0 extension which is deeply described in the technical documentation attached at the end of this Deliverable.

3.2.1.2.7 Continuation of the process

The overall success of the ARDIT strongly depends on both the content generation process (the addition of the characterisation of new data sources) and the continuous upgrade of the AGRICORE DCAT-AP 2.0 extension ontology. The former will allow the expansion of the knowledge including in the tool, which will enable attracting more users that can consult it. The latter will enable increasing the quality of the information contained and iterating the ontology to reflect new researcher needs for fast dataset identification. Hence, the ontology (with its data

model and relations) will be iteratively upgraded as the characterisation of new data sources is subsequently added to the ARDIT tool. The proposed workflow is depicted in the following figure.

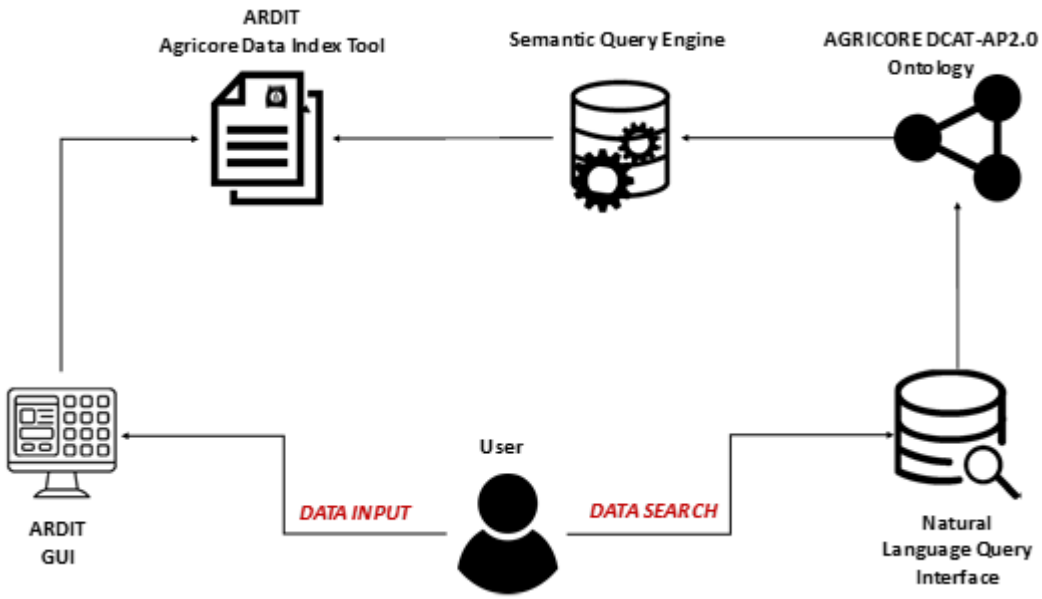

Figure 12 The Workflow for Data Input into the ARDIT and Search Through the AGRICORE Ontology

As described in the figure, the ontology will allow the application of semantic queries in ARDIT for searching datasets information and characterisations. Likewise, it will allow a continuous enrichment of the number and quality of datasets characterisations through the ARDIT Graphic User Interface (GUI) tailored to fit the data model structure described by the ontology.

3.2.1.2.8 DCAT- AP extension implementation details

The DCAT-AP is based on the specification of the DCAT developed under the responsibility of the Government Linked Data Working Group at W3C. DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web. Additional classes and properties from other well-known vocabularies are reused where necessary.

In its COM(2011) 882 of 12 December 2011, the EC stated that the availability of information in a machine-readable format as well as a thin layer of commonly agreed metadata could facilitate data cross-referencing and interoperability and, therefore, it would considerably enhance its value for reuse. [7]

Much of the public sector information that would benefit from interoperability is published as datasets in data portals. Therefore, an agreement on a common format for data exchange would support the sharing, discovery and re-use of this data.

The primary use case for DCAT-AP is to enable performing searches of datasets across portals and make public sector data more easily searchable across borders and sectors. This can be achieved exchanging the descriptions of datasets among data portals. From the start, the DCAT-AP had the purpose of adapting DCAT to facilitate the reuse of data, which implies that:

- It proposes mandatory, recommended or optional classes and properties to be used for a particular application;

[7] COM(2011) 882 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions "Open data An engine for innovation, growth and transparent governance" can be accessed at https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF.

- It identifies requirements to control vocabularies for this particular application;
- It gathers other elements to be considered as priorities or requirements for an application such as conformance statement, agent roles or cardinalities.

DCAT-AP has been implemented by over 15 open data portals in the EU, including the EU Open Data Portal. Moreover, some EU MSs have created extensions for the DCAT-AP such as the DCATAP_IT. To facilitate the implementation of DCAT-AP, many users have developed tools such as validators, harvesters and exporters of DCAT-AP metadata. An overview of those tools is available via Joinup (https://joinup.ec.europa.eu/). To better respond to the information requirements of the statistical and geospatial domains, while ensuring consistency with DCAT-AP, the ISA^2 Programme of the EC has created two extensions to DCAT-AP: the DCAT Application Profile for description of statistical datasets (StatDCAT-AP) and a geospatial extension for the DCAT application profile for data portals in Europe (GeoDCAT-AP). The former, developed in close collaboration with Eurostat, brings the statistical and open data communities closer by enhancing the visibility and facilitating the inclusion of statistical data sets in open government data portals. The latter, describes geospatial datasets, dataset series and services. It provides an RDF syntax binding together metadata elements defined in the core profile of ISO 19115:2003 and those defined in the framework of the INSPIRE Directive. Its basic use case is to make spatial datasets, data series and services searchable on general data portals; thereby making geospatial information better searchable across borders and sectors. A specific working group from the JRC and the ISA^2 programme was responsible for this extension[29].

During the process of preparing the AGRICORE/ARDIT ontology, the need for building upon a solid standardised data model became immediately clear. Because the semantic search was a top priority for the functionality of the ARDIT and ontology, adopting the DCAT-AP data model was a natural consequence. This required exerting some effort adapting the template for the characterisation of datasets to the data model itself. The DCAT-AP RDF vocabulary (DCAT-AP for data portals in Europe Version 2.0.0 https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe/distribution/dcat-ap-200-pdf) was imported in Protégé. The classes with their related properties were examined and compared with the dataset characteristics listed in the Template for the ARDIT Dataset Characterisation (Methodological Grid) table.

A new AP AGRICORE DCAT AP 2.0 extension was created and implemented following the detailed description available in the AGRICORE DCAT-AP 2.0 Extension section.

The five Mandatory classes of DCAT (i. e., Agent; Catalogue; Dataset; Literal; Resource), the Recommended classes "Distribution" and "License Document" and other 12 Optional classes were retained.

Moreover, 22 new classes and sub-classes were created as indicated in the next figure (left). As detailed in the second figure (right), the DatasetVariable has been the most significant addition as it compiles all the details about the different variables contained in the dataset.

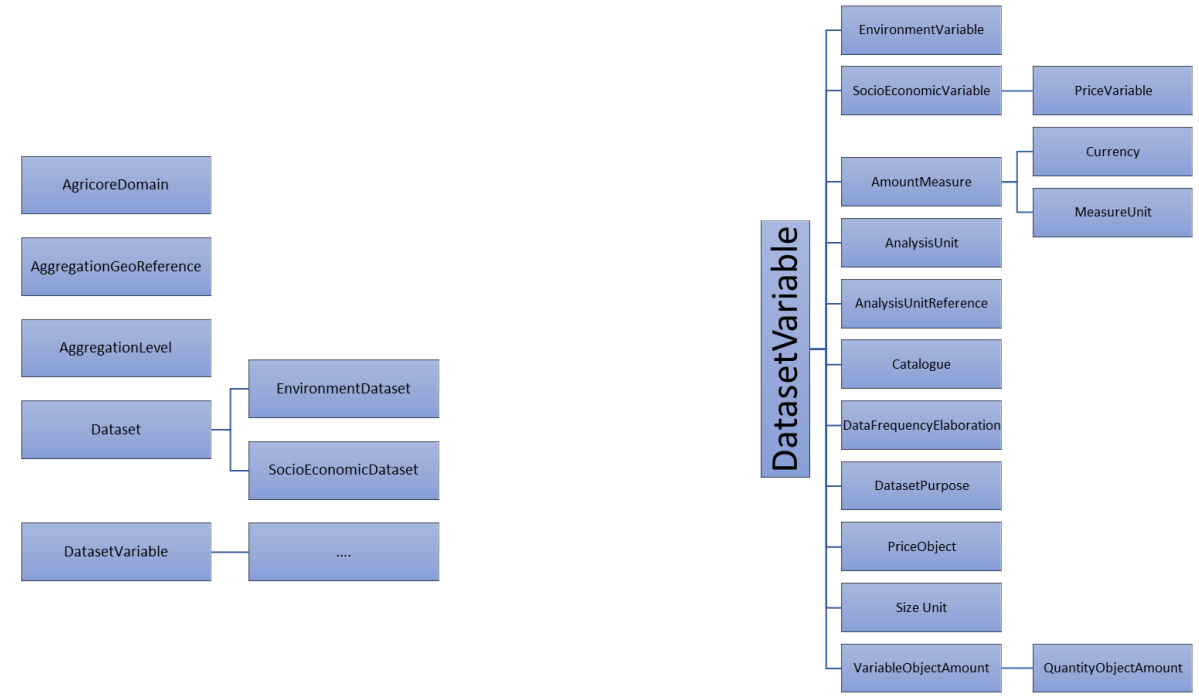

Figure 13 AGRICORE DCAT-AP 2.0 Extension, New Classes (Left), Detail of DatasetVariable Class (Right)

The new classes of concepts have properties linked to controlled vocabularies to allow selecting the required categories browsing standardised linked lists. The Class:Dataset has a property dct:spatial linked to a controlled vocabulary: http://dd.eionet.europa.eu/vocabulary/common/nuts. This property covers the Dataset characteristic Geographical coverage. A similar approach was followed to describe other required Datasets characteristics, such as Type of data set (Object of analysis), through the Dataset property dct:subject, which can take any value in the Digital Europa Thesaurus.

Figure 14 Graphic Representation of the AGRICORE DCAT-AP 2.0 Domain Classes in Protégé

Other suggested controlled vocabularies are:

- http://dd.eionet.europa.eu/vocabulary/eurostat/ind_farm/ (covered unit of a dataset analysis)
- Country lists: e. g., EU MSs, Switzerland, Albania, Montenegro, Turkey; Country groupings: e. g., EU, EU 15, EU25, EU27, EU28, EU_2020, Euro Area 11, Euro Area 12, Euro Area 16, Euro Area 19
- NUTS0, NUTS1, NUTS2 codes and labels
- Policy Measure(s) (in Agriculture and/or Biofuels)
- Grid cells
- Meteorological stations

The following figure portrays the Protégé Ontograf representation with Agricore Domain classes, subclasses and annotations

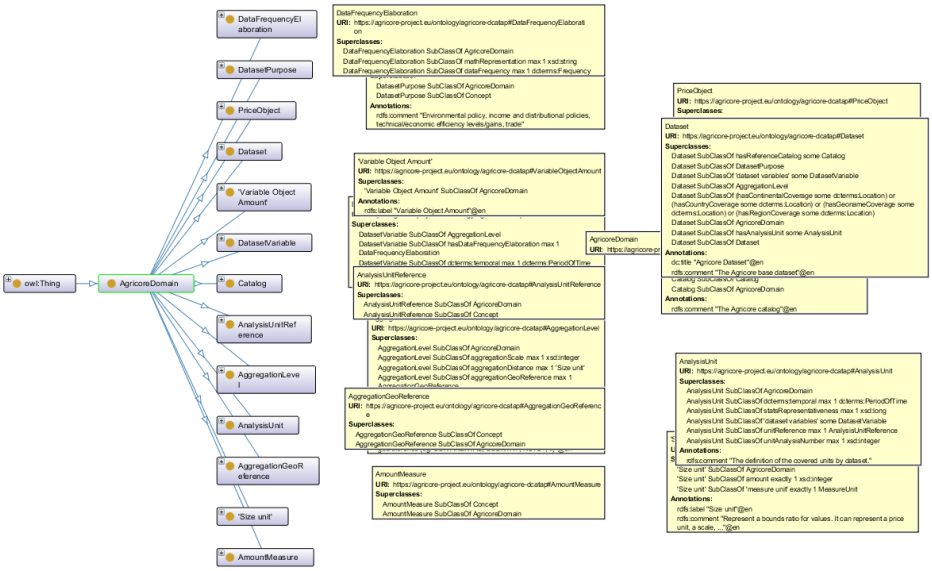

Figure 15 AGRICORE DCAT-AP2.0 Agricore Domain Ontograf in Protégé

3.2.2 Data governance process

3.2.2.1 Within the project duration

Although project partners already performed an initial characterisation of a wide set of datasets (listed in the table in Selection of Datasets Characterised so Far), this characterisation will be repeated in order to follow the final data model established within the ontology definition process. Nonetheless, this process will take place once the first version (for internal use) of the ARDIT is available, as its GUI will be used to ensure and enforce meeting the requirements established in the ontology. The list of datasets which will be characterised within the project will be continuously updated to address any additional need identified by the project modellers. However, a quite mature list of aimed datasets is already provided in this document in the Planned Datasets section. All these analyses will take place in the framework of project's WP1, specifically within Tasks 1.2 to Task 1.6. During this extensive characterisation activity, the partners involved in Task 1.1 will remain available to revise both the characterisation template and the ontology to suit any relevant, yet currently unforeseen, feature of the dataset and/or of all the variables therein. This will be possible also because STAM, UNIPR and IDENER are also involved in the effort of characterising the datasets.

To ensure a consistent quality of the characterisation of the different datasets and - in turn - of the information provided to the research community through the ARDIT, a governance process will be defined and improved within this timeline. This iterative approach will allow maturing the governance structure aiming for its maintenance after the project finalisation. In the current stage, the AGRICORE partners have established that for each dataset to be included in the platform, a reviewer of the quality of its characterisation will be defined. The reviewers of the characterisation will be identified within the consortium.

In a similar way to ensuring the quality of the characterisation effort, it is important to establish a procedure to define which datasets should be characterised. During the project, this process will be easy to follow as, in the current stage, an extensive list of the required datasets has been already produced. Nonetheless, as coordinator of the project, IDENER will be in charge of monitoring any additional data characterisation need identified within the execution of the UCs. Any additional need will be notified by the UC responsible to the coordinator and the actions required to tackle it.

Finally, a special committee formed by representatives of STAM, UNIPR and IDENER will be formed. This Committee will be tasked with the continuous monitoring of the characterisation process, identifying the relevant aspects that may require a modification of the current ontology. At the same time, this Committee will be in charge of producing a new version of such an ontology and to communicate the required changes and adaptations both to the ARDIT developers and to the data providers (people performing the characterisations).

3.2.2.2 Characterisation after the project

Upon completing the activities of the AGRICORE project and having published the ARDIT on the public internet, researchers interested in using and contributing to the tool with the characterisation of new datasets will be able to do so by means of the same functionality employed by project partners in the lifetime of the project. It is obvious that for ensuring the survival of the developed portal, a continuous upgrade and maintenance of the data included should be promoted. To do so, a clear governance structure should be defined.

Although the final version of such governance will be established later in the project (as this relies on the activities pertaining to other WPs and tasks), an initial potential governance structure is already under discussion and presented next.

The ARDIT tool will be developed to include a Datasource life cycle management system. This system, among other things, will establish a set of roles within the platform which will be linked to specific responsibilities. In this way, three main roles have been already identified:

- The platform administrator role will be played either by representatives of AAT (Ayesa) or IDENER, as the main developer of the ARDIT tool (the former) and project coordinator (the latter). The administrator will be tasked with the monitoring, control and management of the rest of users permissions as well as for the other typical tasks of an IT administrator.
- The figure of "Characterisation Reviewer" will be defined and a specific role defined for it. The people entitled with this role will be the ones in charge of deeply reviewing a new characterisation requested to be added to the platform.
- The role of Moderator will be established. The people (up to 3) entitled with this role will need to coordinate the activities of the Characterisation Reviewers. To do so, they will monitor the new characterisation addition requests (submitted within the ARDIT tool) and will assign such requests to specific Characterisation Reviewers.
- The role of Characterisation Volunteer will be also established. This role can be requested by any registered user on the platform. Once a person requests this role, the platform moderators will decide if granting such a role based on the previous contributions of the user or the experience of such a person. The people entitled with this role volunteer to perform a

new characterisation upon request. These requests will be managed by the Moderators, which will receive some characterisation requests from the own ARDIT tool (submitted by registered users). At the same time, Moderators and platform administrators will meet at least twice a year to define new roadmaps, tackling both development and data characterisation needs.

The above-presented roles cover mainly the data governance process, but not in detail the development one. The corresponding roles will be defined within the project by both AAT and IDENER and the people assigned with them will be identified; initially within the personnel of such companies. However, project activities already include efforts to increase the adoption and interest on the project tools (including ARDIT) and external requests from potential contributors are expected.

3.2.3 ARDIT tool, preliminary architecture and functionality

The ARDIT will be the place where all the information will be gathered. It will allow any user (registering will not be mandatory) to search for useful datasets taking advantage of the ontology defined in the AGRICORE project. It is worth noting that the ARDIT will not store real data belonging to any data source, only its references. The main functionality of the tool (besides the obvious data discoverability capabilities) is described in the next points:

- **Datasource life cycle management**. Creation, categorisation, publication, ETL edition activities are possible by means of the ARDIT web interface. Each activity requires that the respective role is granted (by an administrator).
- **Ontology categorisation**. The ARDIT counts on an initial ontology configuration. Notwithstanding, this ontology is open to new versions and releases.
- **Advanced searching**. The ARDIT includes advanced search functionalities for experienced researchers.
- **Semantic searching.** This is a valuable feature which allows the ARDIT to make global and specific queries in natural language to recover all the available datasets indexed by the platform. This semantic searching mechanism will use English as the official language. Multi-language search capabilities could be implemented in a possible future extension of the project.
- **Global and local index tool**. A global ARDIT tool will be deployed in the cloud for centralised and public usage of dataset references. Nonetheless, private users are allowed to deploy local versions of the indexer, which will be synchronised with the public one but that can also manage the connection of the indexer to the DWH (specifically, to the one designed within the AGRICORE project (WP2)).
- **Content syndication**. The global ARDIT will publish content upgrades by RSS allowing people to receive change notifications by the web browser without the need for a local ARDIT to be installed.
- **Isolated Local Indexer**. The Local Indexer can extend the searches to the Global Indexer database. Nonetheless, because the Local Indexer must be working separately, a database synchronisation will be needed.
- **ETL DWH**. The instructions to extract the information from the official source (i. e., public folder, URL, database) and load it into a DWH will be also stored in the ARDIT database allowing for the easy population of the DWH by people who are not familiar with the technology.

- **Loader tool library**. Several utilities will be developed to access the information stored in the ARDIT, in order to facilitate the development of the ETL. This feature implies the standardisation and unification of the ETL development.
- **DWH index**. Thanks to the loader tool library, local ARDIT deployments will be able to know if a dataset has been successfully loaded in the DWH. This is a remarkable feature which will allow the local user to look up which datasets have been loaded, their name, version as well as their destination (i. e., Hadoop or Hive database). This traceability is guaranteed by the loader tool library.
- **Security**. Both the global and local ARDIT instances count on their respective and separated Lightweight Directory Access Protocol (LDAP) databases, where user credentials are stored for authentication purposes. No replication is planned among them. Currently, there are not requirements to store any external source credentials due to this type of information being in the public internet. Nevertheless, when needed, this is the best place to store third party credentials.

The following figure depicts the Global and Local ARDIT architecture together with the AGRICORE DWH. This diagram shows these components to achieve an insightful and seamless explanation beyond the sole ARDIT.

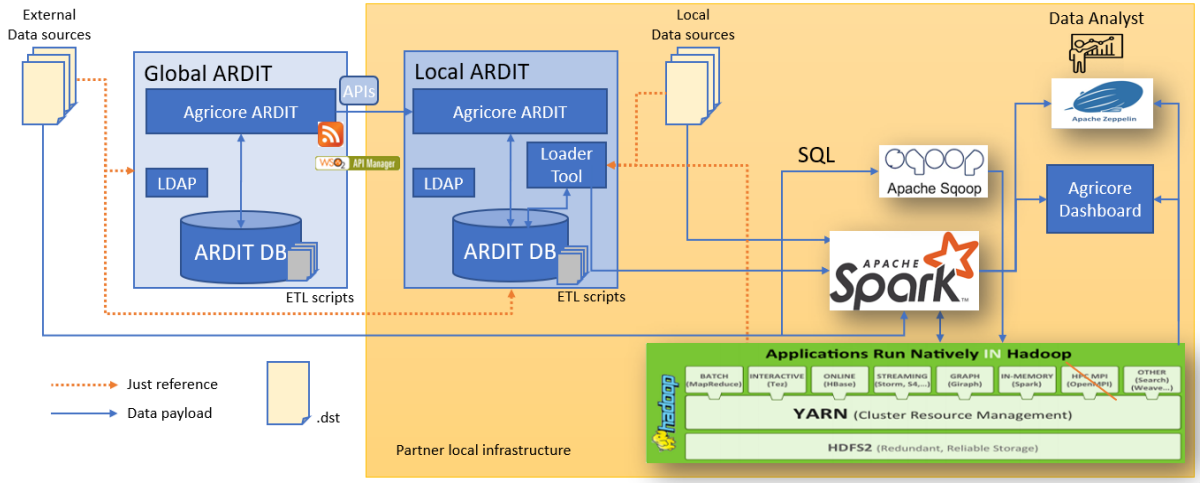

Figure 16 The Global and Local ARDIT Architecture Together with the DWH

It is important to remark that this architecture is being discussed within the AGRICORE consortium and that it extends, notably, the planned characteristics defined in the project Grant Agreement. Indeed, the capabilities of managing ETLs within the index tool is a new feature that has been added to expand the usefulness of the ARDIT further. Although the main goal is still increasing data discoverability (allowing researchers to identify useful datasets taking advantage of advanced search capabilities), the ARDIT tool will go a step beyond including details on how this data can be imported (and or curated) in a DWH. These ETLs will be defined within the project to fit in with the AGRICORE DWH but can provide valuable information to any external researcher on how to import the information in other systems. In the future, if specific applications (e. g., the inclusion of the data on existing JRC repositories) is identified, the ARDIT ETL management system will be upgraded to allow defining ETLs for different target systems.

4 Planned datasets

Tasks 1.2 to 1.6 will undertake the characterisation of several datasets, many of which are available online from institutional repositories which group them together. Besides the aforementioned, and partly already characterised LUCAS, FADN and FSS datasets, many more are provided by Eurostat, the FAO and the OECD. For instance, the FAO maintains the Aquastat, FaoStat and AFFRIS dataset repositories while the OECD and Eurostat have a unique entry point to dataset access. Due to the large number of datasets which might provide relevant information for the analysis of the impact of policy (reforms) on agriculture, a preliminary selection of the datasets which may be characterised in the AGRICORE project has been undertaken on the basis of the sole name of the dataset. The table below contains the datasets which have been identified as of potential interest in this manner, clarifying which is the provider and the location of the dataset in the institutional repository, in terms of the folder and sub-folder it is necessary to reach to find the associated dataset(s). Moreover, the table below includes the datasets already characterised according to the initial template and whose characteristics informed the preparation of the final template on which the AGIRCORE DCAT-AP 2.0 ontology has been developed.

Distributing the dataset characterisation effort among partners will spur a more detailed examination of which of the following datasets might be of real interest to the AGRICORE project and to the researcher interested in assessing the impact of policy (reform(s)) on agriculture. This will determine which datasets will be actually characterised and whose metadata will populate the ARDIT.

Dataset Provider	Folder	Sub-Folder/Dataset Name
OECD	Agriculture and Fisheries	Agricultural Outlook
OECD	Agriculture and Fisheries	Agricultural Policy Indicators
OECD	Agriculture and Fisheries	Environmental Indicators for Agriculture
OECD	Environment	Air and Climate
OECD	Environment	Water
OECD	Environment	Environmental risks and health
OECD	Environment	Waste
OECD	Environment	Material Resources
OECD	Environment	Forest
OECD	Environment	Biodiversity
OECD	Environment	Land Resources
OECD	Environment	Innovation in environment-related technologies
OECD	Environment	Environmentally Adjusted Multifactor Productivity
OECD	Environment	Environmental Expenditures and Revenues
OECD	Environment	Agri-Environmental indicators: Nutrients
OECD	Environment	Environmental policy
OECD	Environment	Agri-Environmental other indicators
OECD	Environment	Green Growth
OECD	Environment	Mineral and Energy Resource Accounts

OECD	Environment	Policy Indicators on Trade and Environment
OECD	Globalisation	Maritime Transport Costs
OECD	Health	Health Status
OECD	Health	Non-Medical Determinants of Health
OECD	Labour	Earnings
OECD	Labour	Employment Protection
OECD	Labour	Labour Force Statistics
OECD	Labour	Labour Market Programmes
OECD	Labour	Trade Unions and Collective Bargaining
OECD	Labour	World Indicators of Skills for Employment
OECD	Labour	ILOSTAT Database
OECD	Labour	Job quality
OECD	Labour	Skills for Jobs
OECD	Monthly Economic Indicators	Main Economic Indicators
OECD	Productivity	Productivity and ULC – Annual, Total Economy
OECD	Productivity	Productivity and ULC by industry, Annual
OECD	Productivity	Productivity and ULC, Total economy, Quarterly early estimates
OECD	Productivity	Productivity Archives
OECD	Productivity	Prices and Purchasing Power Parities
OECD	Productivity	Consumer and Producer Price Indices
OECD	Productivity	Purchasing Power Parities (PPP) Statistics
OECD	Public Sector, Taxation and Market Regulation	Government at a Glance
OECD	Public Sector, Taxation and Market Regulation	Taxation
OECD	Public Sector, Taxation and Market Regulation	Fiscal decentralisation
OECD	Public Sector, Taxation and Market Regulation	Market Regulation
OECD	Public Sector, Taxation and Market Regulation	Going for Growth 2019.
OECD	Social Protection and Well-being	Income distribution and poverty
Eurostat (DB by themes)	General and regional statistics	
Eurostat (DB by themes)	General and regional statistics	European and national indicators for short-term analysis (euroind)
Eurostat (DB by themes)	General and regional statistics	Regional statistics by NUTS classification (reg)
Eurostat (DB by themes)	General and regional statistics	Regional statistics by typology (reg_typ)

Eurostat (DB by themes)	General and regional statistics	Degree of urbanisation (degurb)
Eurostat (DB by themes)	General and regional statistics	City statistics (urb)
Eurostat (DB by themes)	General and regional statistics	Other sub-national statistics (reg_nat)
Eurostat (DB by themes)	General and regional statistics	Land cover and land use, landscape (LUCAS) (lan)
Eurostat (DB by themes)	Economy and finance	National accounts (ESA 2010) (na10)
Eurostat (DB by themes)	Economy and finance	Government statistics (gov)
Eurostat (DB by themes)	Economy and finance	Exchange rates (ert)
Eurostat (DB by themes)	Economy and finance	Interest rates (irt)
Eurostat (DB by themes)	Economy and finance	Prices (prc)
Eurostat (DB by themes)	Economy and finance	Balance of payments - International transactions (bop)
Eurostat (DB by themes)	Economy and finance	Balance of payments - International transactions (BPM6) (bop_6)
Eurostat (DB by themes)	Population and social conditions	Demography and migration (demo)
Eurostat (DB by themes)	Population and social conditions	Asylum and managed migration (migr)
Eurostat (DB by themes)	Population and social conditions	Population projections (proj)
Eurostat (DB by themes)	Population and social conditions	Health (hlth)
Eurostat (DB by themes)	Population and social conditions	Labour market (labour)
Eurostat (DB by themes)	Population and social conditions	Living conditions and welfare (livcon)
Eurostat (DB by themes)	Population and social conditions	Income, consumption and wealth - experimental statistics (icw)
Eurostat (DB by themes)	Industry, trade and services	Statistics on the production of manufactured goods (prom)
Eurostat (DB by themes)	Agriculture, forestry and fisheries	Agriculture (agr)
Eurostat (DB by themes)	Agriculture, forestry and fisheries	Forestry (for)
Eurostat (DB by themes)	International trade	International trade in goods (ext_go)
Eurostat (DB by themes)	Environment and energy	Environment (env)
Eurostat (DB by themes)	Environment and energy	Energy (nrg)
PWT	RGDPNA	Real GDP using national-accounts growth rates, for studies comparing (output-based) growth rates across countries
PWT	CGDPe	Expenditure-side real GDP at current PPPs, to compare relative living standards across countries at a single point in time
PWT	CGDPo	Output-side real GDP at current PPPs, to compare relative productive capacity across countries at a single point in time
PWT	RGDPe	Expenditure-side real GDP at chained PPPs, to compare relative

		living standards across countries and over time
PWT	RGDPo	Output-side real GDP at chained PPPs, to compare relative productive capacity across countries and over time
PWT	DA	Development accounting, the sources of differences in living standards at a point in time
PWT	GA	Growth accounting, the sources of economic growth over time
FAO	Production	Crops
FAO	Production	Crops processed
FAO	Production	Live Animals
FAO	Production	Livestock Primary
FAO	Production	Livestock Processed
FAO	Production	Production Indices
FAO	Production	Value of Agricultural Production
FAO	Trade	Crops and livestock products
FAO	Trade	Live animals
FAO	Trade	Detailed trade matrix
FAO	Trade	Trade Indices
FAO	Food Balance	New Food Balances
FAO	Food Balance	Food Balances (old methodology and population)
FAO	Food Balance	Commodity Balances - Crops Primary Equivalent
FAO	Food Balance	Commodity Balances - Livestock and Fish Primary Equivalent
FAO	Food Balance	Food Supply - Crops Primary Equivalent
FAO	Food Balance	Food Supply - Livestock and Fish Primary Equivalent
FAO	Food Security	Indicators from Household Surveys (gender, area, socioeconomics)
FAO	Food Security	Suite of Food Security Indicators
FAO	Prices	Producer Prices - Annual
FAO	Prices	Producer Prices - Monthly
FAO	Prices	Producer Price Indices - Annual
FAO	Prices	Producer Prices - Archive
FAO	Prices	Consumer Price Indices
FAO	Prices	Deflators
FAO	Prices	Exchange rates - Annual
FAO	Inputs	Fertilizers by Nutrient
FAO	Inputs	Fertilizers by Product
FAO	Inputs	Fertilizers archive
FAO	Inputs	Pesticides Use
FAO	Inputs	Pesticides Trade

FAO	Inputs	Land Use
FAO	Inputs	Employment Indicators
FAO	Population	Annual population
FAO	Investment	Machinery
FAO	Investment	Machinery Archive
FAO	Investment	Government Expenditure
FAO	Investment	Credit to Agriculture
FAO	Investment	Development Flows to Agriculture
FAO	Investment	Foreign Direct Investment (FDI)
FAO	Investment	Country Investment Statistics Profile
FAO	Macro-Statistics	Capital Stock
FAO	Macro-Statistics	Macro Indicators
FAO	Agri-Environmental Indicators	Fertilizers indicators
FAO	Agri-Environmental Indicators	Land use indicators
FAO	Agri-Environmental Indicators	Land Cover
FAO	Agri-Environmental Indicators	Livestock Patterns
FAO	Agri-Environmental Indicators	Livestock Manure
FAO	Agri-Environmental Indicators	Pesticides indicators
FAO	Agri-Environmental Indicators	Emissions by sector
FAO	Agri-Environmental Indicators	Emissions intensities
FAO	Agri-Environmental Indicators	Temperature change
FAO	Emissions - Agriculture	Agriculture Total
FAO	Emissions - Agriculture	Enteric Fermentation
FAO	Emissions - Agriculture	Manure Management
FAO	Emissions - Agriculture	Rice Cultivation
FAO	Emissions - Agriculture	Synthetic Fertilizers
FAO	Emissions - Agriculture	Manure applied to Soils
FAO	Emissions - Agriculture	Manure left on Pasture
FAO	Emissions - Agriculture	Crop Residues
FAO	Emissions - Agriculture	Cultivation of Organic Soils
FAO	Emissions - Agriculture	Burning - Savanna
FAO	Emissions - Agriculture	Burning - Crop Residues
FAO	Emissions - Agriculture	Energy Use
FAO	Emissions - Land Use	Land Use Total
FAO	Emissions - Land Use	Forest Land
FAO	Emissions - Land Use	Cropland
FAO	Emissions - Land Use	Grassland
FAO	Emissions - Land Use	Burning - Biomass
FAO	Forestry	Forestry Production and Trade
FAO	Forestry	Forestry Trade Flows
FAO	ASTI R&D Indicators	ASTI-Researchers
FAO	ASTI R&D Indicators	ASTI-Expenditures
Eurostat (Tables on EU Policy)	Euro indicators / PEEIs	Balance of payments (teieuro_bp)
Eurostat (Tables on EU Policy)	Euro indicators / PEEIs	Business and consumer surveys (teieuro_bs)
Eurostat (Tables on EU Policy)	Euro indicators / PEEIs	International trade (teieuro_et)

Eurostat (Tables on EU Policy)	Euro indicators / PEEIs	Industry, trade and services (teieuro_is)
Eurostat (Tables on EU Policy)	Euro indicators / PEEIs	Labour market (teieuro_lm)
Eurostat (Tables on EU Policy)	Euro indicators / PEEIs	Monetary and financial indicators (teieuro_mf)
Eurostat (Tables on EU Policy)	Euro indicators / PEEIs	National accounts (teieuro_na)
Eurostat (Tables on EU Policy)	Euro indicators / PEEIs	Consumer prices (teieuro_cp)
Eurostat (Tables on EU Policy)	Europe 2020 indicators	Headline indicators (t2020_h)
Eurostat (Tables on EU Policy)	Europe 2020 indicators	Resource efficiency indicators (t2020_r)
Eurostat (Tables on EU Policy)	Circular economy indicators	Production and consumption (cei_pc)
Eurostat (Tables on EU Policy)	Circular economy indicators	Waste management (cei_wm)
Eurostat (Tables on EU Policy)	Circular economy indicators	Secondary raw materials (cei_srm)
Eurostat (Tables on EU Policy)	Circular economy indicators	Competitiveness and innovation (cei_cie)
Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 1 - No poverty (sdg_01)
Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 2 - Zero hunger (sdg_02)
Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 3 - Good health and well-being (sdg_03)
Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 4 - Quality education (sdg_04)
Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 5 - Gender equality (sdg_05)
Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 6 - Clean water and sanitation (sdg_06)
Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 7 - Affordable and clean energy (sdg_07)
Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 8 - Decent work and economic growth (sdg_08)
Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 9 - Industry, innovation and infrastructure (sdg_09)
Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 10 - Reduced inequalities (sdg_10)
Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 11 - Sustainable cities and communities (sdg_11)
Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 12 - Responsible consumption and production (sdg_12)
Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 13 - Climate action (sdg_13)
Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 14 - Life below water (sdg_14)
Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 15 - Life on land (sdg_15)
Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 16 - Peace, justice and strong institutions (sdg_16)

Eurostat (Tables on EU Policy)	Sustainable development indicators	Goal 17 - Partnerships for the goals (sdg_17)
Eurostat (Cross-cuting Topics)	Migrant integration and children in migration	Migrant integration (mii)
Eurostat (Cross-cuting Topics)	Economic globalisation indicators	International trade (egi_tr)
Eurostat (Cross-cuting Topics)	Economic globalisation indicators	Foreign direct investment (egi_fi)
Eurostat (Cross-cuting Topics)	Economic globalisation indicators	Employment (egi_em)
Eurostat (Cross-cuting Topics)	Economic globalisation indicators	Research and development (egi_rd)
Eurostat (Cross-cuting Topics)	Economic globalisation indicators	Value added (egi_va)
Eurostat (Cross-cuting Topics)	Quality of employment	Safety and ethics of employment (qoe_saet)
Eurostat (Cross-cuting Topics)	Quality of employment	Income and benefits from employment (qoe_inbe)
Eurostat (Cross-cuting Topics)	Quality of employment	Working time and work-life balance (qoe_woli)
Eurostat (Cross-cuting Topics)	Quality of employment	Security of employment and social protection (qoe_soe)
Eurostat (Cross-cuting Topics)	Quality of employment	Social dialogue (qoe_sod)
Eurostat (Cross-cuting Topics)	Quality of employment	Skills development and training (qoe_trsk)
Eurostat (Cross-cuting Topics)	Quality of employment	Employment-related relationships and work motivation (qoe_relmot)
Eurostat (Cross-cuting Topics)	Climate change	Greenhouse gas emissions (cli_gge)
Eurostat (Cross-cuting Topics)	Climate change	Drivers (cli_dri)
Eurostat (Cross-cuting Topics)	Climate change	Mitigation (cli_mit)
Eurostat (Cross-cuting Topics)	Climate change	Impact and adaptation (cli_iad)
Eurostat (Cross-cuting Topics)	Climate change	Climate action initiatives (cli_act)
National Statistics Institute of Spain	NA	SSAO2016 Survey on the Structure of Agricultural Holdings 2016
Ministry of Agriculture, Fisheries and Alimentation	NA	Spanish Survey on Crop Surfaces and Yields (ESYRCE)
Ministry of Agriculture, Fisheries and Alimentation	Spanish Statistics on Means of Production	Use of Fertilizers
Ministry of Agriculture/Junta de Andalucía	NA	Sistema de Identificación Geográfica de Identificación de Parcelas Agrícolas (SIGPAC)
Ministry of Agriculture, Fisheries and Alimentation	NA	ESYRCE (for Spain)
Ministry of Agriculture	NA	FADN for Spain
Ministry of Agriculture, Fisheries and Alimentation	NA	National Agrarian Accounting Network (RECAN) (for Spain)
National Statistics Institute of Spain	NA	Agricultural Census (2009) (for Spain)
National Statistics Institute of Spain	NA	Survey on Production Methods in Agricultural Operations (2009)
Ministry of Agriculture, Fisheries and Alimentation	Spanish Statistics on Means of Production	Commercialization of Phytosanitary Products
Ministry of Agriculture, Fisheries and Alimentation	Spanish Statistics on Means of Production	Utilization of Phytosanitary Products

Ministry of Agriculture, Fisheries and Alimentation	Spanish Statistics on Means of Production	Registration of New Machinery
Ministry of Agriculture, Fisheries and Alimentation	Spanish Statistics on Means of Production	Use of Fertilizers
Ministery of Agriculture and Fishing, Alimentation and Environment	National Agricultural Economic Statistics	Short-term prices of agricultural products (for Spain)
Ministery of Agriculture and Fishing, Alimentation and Environment	National Agricultural Economic Statistics	Agricultural rates and salaries (for Spain)
Ministery of Agriculture and Fishing, Alimentation and Environment	National Agricultural Economic Statistics	Agricultural rates and prices received (for Spain)
Ministery of Agriculture and Fishing, Alimentation and Environment	National Agricultural Economic Statistics	Agricultural rates and prices paid (for Spain)
Ministery of Agriculture and Fishing, Alimentation and Environment	National Agricultural Economic Statistics	Average land prices for agricultural use (for Spain)
Ministry of Agriculture, Fisheries and Alimentation	NA	Surfaces and annual crops productions (for Spain)
Ministry of Agriculture, Fisheries and Alimentation	NA	Surface advances and crop productions (for Spain)
Ministry of Environment, Rural and Marine Environment	NA	National Soil Erosion Inventory (INES) (for Spain)
Institute of Statistics and Cartography of Andalusia	NA	Sistema de Información Multiterritorial de Andalucía
Ministry for Ecological Transition and Demographic Challenge	NA	National Hydrological Report (for Spain)
Ministry of Agriculture, Fisheries and Rural Development	NA	Climate data obtained by the Agroclimatic Stations Net (for Spain)
Eurostat	Agriculture	Farm indicators by agricultural area, type of farm, standard output, legal form and NUTS 2 regions
Eurostat	Agriculture	Organic farming
Eurostat	Agriculture	Agricultural production
Eurostat	Agriculture	Estimated soil erosion by water, by erosion level, land cover and NUTS 3 regions
Eurostat	Agriculture	Farm structure
Eurostat	Agriculture	Economic accounts for agriculture
Eurostat	Agriculture	Agricultural prices and price indices
Eurostat	Agriculture	Agriculture and environment. Gross nutrient balance
Eurostat	Agriclture	Share of main land types in utilised agricultural area
Eurostat	Agriculture	Share of irrigable and irrigated areas in utilised agricultural area by NUTS 2 regions

Ministry of Environment, Rural and Marine Environment	NA	Survey on fruit tree plantations, olive groves and table grapes. (for Spain)
Regional Ministry of Agriculture, Livestock, Fisheries and Sustainable Development, Junta de Andalucía	NA	Evolution of provincial agricultural macromagnitudes 2005-2014 (for Spain)
Regional Ministry of Agriculture, Livestock, Fisheries and Sustainable Development, Junta de Andalucía	NA	Olive. Data obtained from the monitoring of pests and diseases in the biological control stations (for Spain)
Regional Ministry of Agriculture, Livestock, Fisheries and Sustainable Development, Junta de Andalucía	NA	Rice. Data obtained from the monitoring of pests and diseases in the biological control stations (for Spain)
Ministry of Agriculture, Alimentation and Environment	NA	Survey on fruit tree plantations, olive groves and table grapes (for Spain)
Regional Ministry of Agriculture, Livestock, Fisheries and Sustainable Development, Junta de Andalucía	NA	Winter cereals. Data obtained from the monitoring of pests and diseases in the biological control stations (for Spain)
Ministry of Agriculture, Fisheries and Alimentation	NA	Cuentas Económicas de la Agricultura (Renta Agraria: Macromagnitudes Agrarias)
Ministry of Agriculture, Fisheries and Alimentation	NA	Fondo Español de Garantía Agraria
Meteorology Statal Agency	NA	Statistics of meterophenological variables (for Spain)
Junta de Andalucía, Ministry of Agriculture, Fisheries and Sustainable Development	NA	Statistics of land prices in Andalusia
Ministry of Agriculture, Fisheries and Alimentation	NA	Organic Farming in Spain
Junta de Andalucía, Ministry of Agriculture, Fisheries and Sustainable Development	NA	Information system for organic production in Andalusia (SIPEA) (for Spain)
Junta de Andalucía, Ministry of Agriculture, Fisheries and Sustainable Development	NA	Livestock Management and Information System (SIGGAN) (for Spain)
Ministry of Agriculture, Fisheries and Alimentation	NA	MARM. Household consumption database (for Spain)
Ministry of Agriculture, Fisheries and Alimentation	NA	Official Agricultural Machinery Register (ROMA) (for Spain)
Subdirectorate General of Short-term Analysis and Economic Forecasts of the General Directorate of Macroeconomic Analysis	NA	BDSICE Costs and prices Agricultural price index (for Spain)
Subdirectorate General of Short-term Analysis and Economic Forecasts of the General Directorate of Macroeconomic Analysis	NA	BDSICE National production and demand indicators Agriculture (for Spain)

Subdirectorate General of Short-term Analysis and Economic Forecasts of the General Directorate of Macroeconomic Analysis	NA	BDSICE Price and costs Agricultural wage index (for Spain)
Subdirectorate General of Short-term Analysis and Economic Forecasts of the General Directorate of Macroeconomic Analysis	NA	BDSICE Prices and costs Salary increases in agreement and salary increases registered in agriculture (for Spain)
Junta de Andalucía. Counseling of Agriculture, Fisheries and Sustainable Development	NA	Agrifood Foreign Trade Statistics
Ministry of Agriculture, Fisheries and Alimentation	NA	Monthly production, movement and stock data (AICA)
ELSTAT	Agriculture, Livestock, Fishery	Livestock Surveys (for Greece)
ELSTAT	Agriculture, Livestock, Fishery	Annual Agricultural Statistical Survey (for Greece)
ELSTAT	Agriculture, Livestock, Fishery	FSS (for Greece)
ELSTAT	Agriculture, Livestock, Fishery	Crops Survey (for Greece)
Hellenic National Meteorological Service	NA	Hellenic National Meteorological Service
ELSTAT	NA	Census of Agricultural and Livestock Holdings 2009 (for Greece)
EU Commision. Directorate General for Agriclture	NA	EU FADN
Eurostat	NA	EU Farm Structure Survey (FSS)
NASA	NA	AgMERRA Climate Forcing Dataset for Agricultural Modeling
NASA	NA	AgCFSR Climate Forcing Dataset for Agricultural Modeling
National Climatic Data Center	NA	Global Summary of the Day (GSOD)
WorldClim	NA	WorldClim Version 2.1
NASA	NA	Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2)
IMGW	NA	Climate-related dataset for Poland
Polish National Chemical-Agricultural Station	NA	Database on mineral nitrogen content in Poland
University of East Anglia	NA	Climatic Research Unit Time-series (CRU TS) dataset v. 4.04
ISRIC — World Soil Information	NA	SoilGrids
JRC	NA	SoilHydroGrids
The European Environment Agency (EEA)	NA	Biodiversity Information System for Europe (BISE)
The European Environment Agency (EEA)	NA	European Climate Assessment & Dataset
Goethe-University	NA	MIRCA 2000
NASA	NA	NASA Prediction of Worldwide Energy Resources (POWER)

Agency for Restructuring and modernisation of Agriculture	NA	ARMA - RDP 2014-20 (PL) Implementation reports (for Poland)
Statistics Poland	NA	Agricultural and horticultural crops
Statistics Poland	NA	Local Data Bank
Statistics Poland	NA	Animal production, Farm animals
Chief Inspectorate of Environmental Protection	NA	Agricultural Drought Monitoring System (ADMS)
Chief Inspectorate of Environmental Protection	NA	Inspectorate of Environmental Protection Reports datasets (for Poland)

Table 4 List of datasets to be characterised within AGRICORE

5 AGRICORE DCAT-AP 2.0 Technical Documentation

5.1 Background

The ARDIT will be a publicly accessible index of data sources available for agricultural policy assessment. It will be an important resource available to all stakeholders (from data analysts to policymakers and researchers) and it will contain detailed information about each relevant dataset such as fields, spatial scope and resolution, aggregation level, update frequency, last update available, privacy level of the data and accessibility.

The objective of this activity in Task 1.1 is the definition of an ontology to enable semantic searching capabilities on the platform. The idea behind the AGRICORE ontology for the ARDIT is to follow the path of the Open Data Portals for the publication, organisation, and retrieval of published data. A thorough assessment has been carried out to determine the appropriate standardised vocabularies to describe the characteristics of datasets employed by the AGRICORE suite and, more generally, by the stakeholders involved in the evaluation of the impacts of policy (reform) on agriculture. The analysis carried out confirmed the consolidated usage of the DCAT vocabulary to describe the datasets and to publish them on Open Data Portals. DCAT is an RDF vocabulary developed by the W3C designed to facilitate interoperability between different data catalogues published on the Web. It enables applications to easily consume metadata from multiple catalogues and it is what makes initiatives like the EU Open Data Portal possible. The DCAT-AP is a specification based on W3C DCAT for describing metadata of public sector datasets in Europe. The benefits of DCAT-AP are that by using a common metadata schema to describe datasets:
1. Data publishers increase the discoverability of the data and thus re-use;
2. Data re-users can search across platforms without facing difficulties caused by the use of separate models or language differences.

While DCAT-AP is not a mandatory standard (e. g., by national or EU law), it is widely accepted as the standard way for describing a dataset; therefore, it has been adopted by portal owners (https://www.europeandataportal.eu/sites/default/files/edp_s3wp4_sustainability_recommendations.pdf).

The DCAT-AP most updated version (DCAT-AP 2.0) was imported and the contained relationships were used to create the AGRICORE-DCAT-AP 2.0 Extension, now enlisting new classes suitable to describe the characteristics of the datasets characterised in the AGRICORE Project. In the following sections, such classes and related properties will be described.

5.2 The DCAT-AP Data Structure

The basis of the DCAT-AP is the specification of the DCAT Vocabulary. DCAT was developed in the period from June 2011 to December 2013 by the Government Linked Data Working Group. The specification was published as a W3C Recommendation in January 2014. DCAT is an RDF vocabulary designed to facilitate the interoperability between data catalogues published on the Web. By using DCAT to describe datasets in data catalogues, publishers increase discoverability and enable applications to use metadata from multiple catalogues easily. It further enables the decentralised publishing of catalogues and facilitates federated dataset search across sites. Aggregated DCAT metadata can serve as a manifest file to facilitate digital preservation. The specification defines RDF Classes and Properties in a model that has four main entities:

- Catalogue (dcat:Catalog), defined as a curated collection of datasets' metadata;

- Catalogue Record (dcat:CatalogRecord), defined as a record in a data catalogue describing a single dataset;
- Dataset (dcat:Dataset), defined as a collection of data, published or curated by a single agent, and available for access or download in one or more formats;
- Distribution (dcat:Distribution), defined as representing a specific available form of a dataset. Each dataset might be available in different forms, these forms might represent different formats of the dataset or different endpoints.

The data model of DCAT is presented in the figure below.

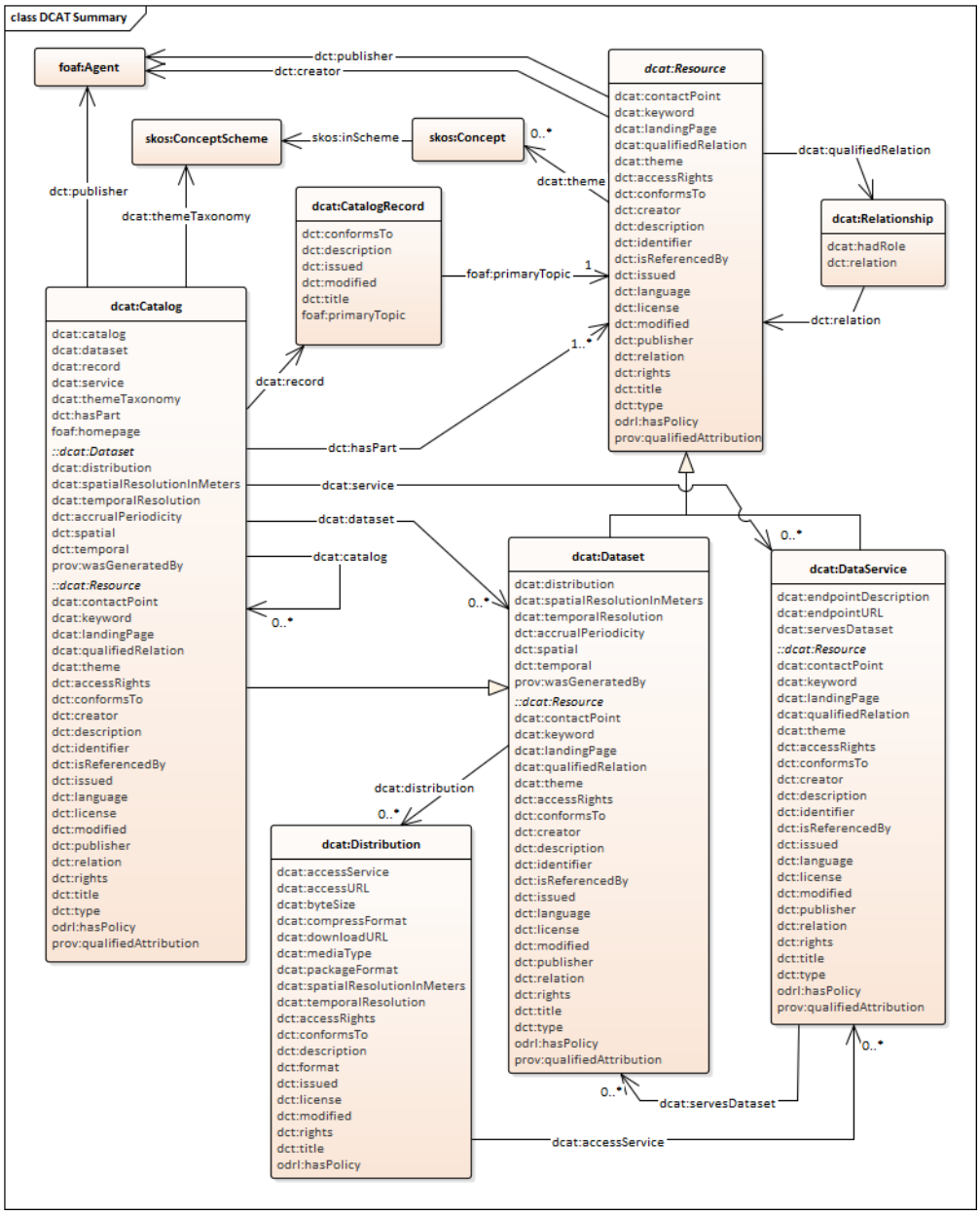

Figure 17 DCAT-AP Data Model

5.3 The AGRICORE DCAT-AP Data Model

The DCAT-AP is intended as a common layer for the exchange of metadata for a wide range of dataset types. The availability of such a common layer creates the opportunity for a wide range of professional communities to look onto the emerging landscape of interoperable portals by aligning with the common exchange format. In addition to the basic DCAT-AP, specific communities can extend the basic Application Profile to support description elements which are specific to their particular data.
Developing a DCAT-AP extension for the exchange datasets metadata, named AGRICORE DCAT-AP, is in line with that approach, firstly by determining which description elements in agricultural data standards can be exposed in the DCAT-AP format and second by extending the DCAT-AP with descriptive elements that can further help in the discovery and use of datasets for the analysis of the impacts of policy (reform) on agriculture.

Under these premises, the AGRICORE DCAT-AP is an extension of DCAT-AP version 2.0 (https://www.w3.org/TR/vocab-dcat-2/) created to describe the datasets employed in the AGRICORE Project and all that are included in the ARDIT, irrespective of the format they are in, such as those published in SDMX, Data Cube, CSV and other formats. Its purpose is to provide a specification that is fully conformant with DCAT-AP version 2.0 as it meets all obligations of the DCAT-AP Conformance Statement. As a result, data portals that comply with the DCAT-AP will be able to understand the core of the AGRICORE DCAT-AP. In addition, the AGRICORE DCAT-AP defines a small number of additions to the DCAT-AP model that are particularly relevant for agricultural datasets. Considering the high number of agricultural datasets that are available on data portals and are of interest to their users, it is likely that recognising and exposing the additions to the DCAT-AP proposed by the AGRICORE DCAT-AP will benefit the general data portals which will then be able to provide enhanced services to collections of statistical data.
The AGRICORE DCAT-AP data model includes the four main entities that are also present in DCAT-AP:

- The **Catalogue**: it represents a collection of Datasets. It is defined in the DCAT Recommendation as "a curated collection of metadata about datasets". The description of the Catalogue includes links to the metadata for each of the Datasets that are in the Catalogue.
- The **Catalogue Record**: defined by DCAT as "a record in a data catalog, describing a single dataset". The Catalogue Record enables statements about the description of a Dataset rather than about the Dataset itself. Catalogue Records may not be used by all implementations. It is optional in DCAT-AP and mostly used by aggregators to keep track of harvesting history.
- The **Dataset**: it represents the published information. It is defined as "a collection of data, published or curated by a single agent, and available for access or download in one or more formats". The description of a Dataset includes links to each of its Distributions, if they are available. Nonetheless, a Dataset is not required to have a Distribution; examples are Datasets that are described before the associated data is collected, Datasets for which the data has been removed and Datasets that are only accessible through a landing page.
- The **Distribution**: according to DCAT, it "represents a specific available form of a dataset. Each dataset might be available in different forms, these forms might represent different formats of the dataset or different endpoints. Examples of distributions include a downloadable CSV file, an API or an RSS feed". The description of a Distribution contains information about the location of the data files or access point and about the file format and licence for use or reuse. In the case of statistical datasets, Distributions may be available in specific formats like SDMX-ML or using the Data Cube vocabulary.

5.3.1 Overview of the model

In the following sections, classes and properties are grouped under headings 'mandatory', 'recommended' and 'optional'. These terms have the following meaning.

- Mandatory class: a receiver of data MUST be able to process information about instances of the class; a sender of data MUST provide information about instances of the class.
- Recommended class: a receiver of data MUST be able to process information about instances of the class; a sender of data SHOULD provide information about instances of the class. However, if information about the instances of a class is available, then the sender of data MUST provide this information.
- Optional class: a receiver MUST be able to process information about instances of the class; a sender MAY provide the information but is not obliged to do so.
- Mandatory property: a receiver MUST be able to process the information for that property; a sender MUST provide the information for that property.
- Recommended property: a receiver MUST be able to process the information for that property; a sender SHOULD provide the information for that property if it is available.
- Optional property: a receiver MUST be able to process the information for that property; a sender MAY provide the information for that property but is not obliged to do so.

The meaning of the terms MUST, MUST NOT, SHOULD and MAY in this section and in the following sections are as defined in RFC 211945. In this context, the term "processing" means that receivers must accept incoming data and transparently provide these data to applications and services. It does neither imply nor prescribe what applications and services finally do with the data (e. g., parse, convert, store, make searchable, display to users).

5.3.2 Namespaces

The AP reuses terms from various existing specifications. Classes and properties specified in the next sections have been taken from the following namespaces:

Prefix	Namespace URI
adms	http://www.w3.org/ns/adms#
dcat	http://www.w3.org/ns/dcat#
dct	http://purl.org/dc/terms/
dqv	http://www.w3.org/ns/dqv#
foaf	http://xmlns.com/foaf/0.1/ or http://www.w3.org/ns/oa#
qb	http://purl.org/linked-data/cube#
rdfs	http://www.w3.org/2000/01/rdf-schema#
schema	http://schema.org/
skos	http://www.w3.org/2004/02/skos/core#
spdx	http://spdx.org/rdf/terms# stat http://data.europa.eu/(xyz)/statdcat-ap/46
vcard	http://www.w3.org/2006/vcard/ns#
xsd	http://www.w3.org/2001/XMLSchema#

Table 5 Specifications reused by DCAT-AP

5.4 Description of classes

5.4.1 Mandatory classes of DCAT present in the AGRICORE DCAT-AP

Class name	Usage note for the AP	URI	Reference
Agent	An entity that is associated with Catalogues and/or Datasets. If the Agent is an organisation, the use of the Organization Ontology47 is recommended.	foaf:Agent	http://xmlns.com/foaf/spec/#term_Agent , http://www.w3.org/TR/vocab-org/
Catalogue	A catalogue or repository that hosts the Datasets being described.	dcat:Catalog	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-catalog
Dataset	A conceptual entity that represents the information published.	dcat:Dataset	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-dataset
Literal	A literal value such as a string or integer; Literals may be typed, e. g., as a date according to xsd:date. Literals that contain human-readable text have an optional language tag as defined by BCP 4748.	rdfs:Literal	http://www.w3.org/TR/rdf-concepts/#section-Literals
Resource	Anything described by RDF.	rdfs:Resource	http://www.w3.org/TR/rdf-schema/#ch_resource

Table 6 Mandatory classes of DCAT present in the AGRICORE DCAT-AP

5.4.2 Recommended classes of DCAT present in the AGRICORE DCAT-AP

Class name	Usage note for the AP	URI	Reference
Distribution	A physical embodiment of the Dataset in a particular format, including visualisations of the data.	dcat:Distribution	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-distribution
Licence document	A legal document giving official permission to do something with a resource.	dct:LicenseDocument	http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=terms#LicenseDocument

Table 7 Recommended classes of DCAT present in the AGRICORE DCAT-AP

The class 'Distribution' is classified as 'Recommended' to allow for cases in which a particular Dataset does not have a downloadable Distribution. Therefore, the sender of data would not be able to provide this information. However, it can be expected that the vast majority of Datasets do have downloadable Distributions, and in these instances the provision of information on the Distribution is mandatory.

5.4.3 Optional classes of DCAT present in the AGRICORE DCAT-AP

Class name	Usage note for the AP	URI	Reference
Catalogue Record	A description of a Dataset's entry in the Catalogue.	dcat:CatalogRecord	http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-catalog-record
Document	A textual resource intended for human consumption that contains information, e. g., a web page about a Dataset.	foaf:Document	http://xmlns.com/foaf/spec/#term_Document
Frequency	A rate at which something recurs, e. g., the publication of a Dataset.	dct:Frequency	http://dublincore.org/documents/dcmi-terms/#terms-Frequency
Kind	A description following the vCard specification, e. g., to provide a telephone number and an e-mail address for a contact point. Note that the class Kind is the parent class for the four explicit types of vCard (Individual, Organization, Location, Group).	vcard:Kind	http://www.w3.org/TR/2014/NOTE-vcard-rdf-20140522/#d4e181
Linguistic system	A system of signs, symbols, sounds, gestures, or rules used in communication, e. g., a language.	dct:LinguisticSystem	http://dublincore.org/documents/dcmi-terms/#terms-LinguisticSystem
Location	A spatial region or named place. It can be represented using a controlled vocabulary or with geographic coordinates. In the latter case, the use of the Core Location Vocabulary49 is recommended, following the approach	dct:Location	http://dublincore.org/documents/dcmi-terms/#terms-Location

	described in the GeoDCAT-AP specification.		
Media type or extent	A media type or extent, e. g., the format of a computer file.	dct:MediaTypeOrExtent	http://dublincore.org/documents/dcmi-terms/#terms-MediaTypeOrExtent
Period of time	An interval of time that is named or defined by its start and end dates.	dct:PeriodOfTime	http://dublincore.org/documents/dcmi-terms/#terms-PeriodOfTime
Rights statement	A statement about the intellectual property rights held in or over a resource, a legal document giving official permission to do something with a resource, or a statement about access rights.	dct:RightsStatement	http://dublincore.org/documents/dcmi-terms/#terms-RightsStatement
Standard	A standard or other specification to which a Dataset or Distribution conforms.	dct:Standard	http://dublincore.org/documents/dcmi-terms/#terms-Standard
Provenance Statement	A statement of any changes in ownership and custody of a resource since its creation that are significant for its authenticity, integrity, and interpretation.	dct:ProvenanceStatement	http://dublincore.org/documents/dcmi-terms/#terms-ProvenanceStatement

Table 8 Optional classes of DCAT present in the AGRICORE DCAT-AP

5.5 Extensions of DCAT-AP 2.0 and specific usage in the AGRICORE Project or in the ARDIT

Discussions during the development of the AGRICORE ontology specifications surfaced a number of requirements for the description of the datasets employed in the AGRICORE Project and/or characterised in the ARDIT that were not met by existing properties in the DCAT-AP. The following sections present the extensions that have been included in the AGRICORE DCAT-AP to meet these requirements. Some of the extensions are re-used from existing RDF vocabularies, others are defined in a new namespace specific for the AGRICORE DCAT-AP. The URI for this AGRICORE DCAT-AP dedicated namespace is AGRICORE DCATAP, the URI can be assigned following the W3.org rules and result similarly to: https://agricore-project.eu/ontology/agricore-dcatap

5.5.1 New classes created in the AGRICORE DCAT-AP

Class name	Usage note for the Application Profile	URI	Reference
AgricoreDomain	New class without properties to frame the newly created classes and allow a clearer representation of the ontology	Agricore-dcatap:AgricoreDomain	AGRICORE project http://www.agricore-projet.eu/ontology/agricore-dcatap# AgricoreDomain
Dataset	New class Dataset which extends the DCAT-AP class Dataset.	Agricore-dcatap:dataset	AGRICORE project http://www.agricore-project.eu/ontology/agricore-dcatap#Dataset
AnalysisUnit	New class which represents the definition of the units covered by a specific dataset.	Agricore-dcatap:AnalysisUnit	AGRICORE project http://www.agricore-projet.eu/ontology/agricore-dcatap#AnalysisUnit
AnalysisUnitReference	To define the structure of the vocabulary, we created a specific class in the ontology, AnalysisUnitReference, which is an extension of skos: concept.	Agricore-dcatap:AnalysisUnitReference	AGRICORE project http://www.agricore-projet.eu/ontology/agricore-dcatap# AnalysisUnitReference
AggregationLevel	A new class created to represent the level of aggregation of the data represented. It could be: - equivalent scale (e. g., 1/1000 -1000) integer. - distance (e. g., 1 km) class. - georeference (e. g., CONTINENTAL, COUNTRY, NUTS1).	Agricore-dcatap:AggregationLevel	AGRICORE project http://www.agricore-projet.eu/ontology/agricore-dcatap# AggregationLevel
DataFrequencyElaboration	New class, with two properties: the frequency and a mathematical representation.	Agricore-dcatap:DataFrequencyElaboration	AGRICORE project http://www.agricore-projet.eu/ontology/agricore-dcatap# DataFrequencyElaboration
DatasetVariable	A new class which is an Aggregation Level subclass which could have more specific properties depending on the dataset (e. g., statistic variable, geo variable).	Agricore-dcatap:DatasetVariable	AGRICORE project http://www.agricore-projet.eu/ontology/agricore-dcatap# DatasetVariable
Size Unit	A new class representing a bounds ratio for values. It can represent a price unit, a scale, etc.	Agricore-dcatap:SizeUnit	http://www.agricore-projet.eu/ontology/agricore-dcatap#SizeUnit

AggregationGeoReference	A new class, Concept subclass. Concept is a skos class (reference to an existing vocabulary).	Agricore-dcatap:AggregationGeoReference	http://www.agricore-projet.eu/ontology/agricore-dcatap#AggregationGeoReference
AmountMeasure	A new class with has two subclasses: Currency and MeasureUnit (both are skos concepts, referring to existing vocabularies).	Agricore-dcatap:AmountMeasure	http://www.agricore-projet.eu/ontology/agricore-dcatap#AmountMeasure
Catalog	A new class Catalog which extends the DCAT-AP class Catalog.	Agricore-dcatap:Catalog	http://www.agricore-projet.eu/ontology/agricore-dcatap#Catalog
Currency	See AmountMeasure. The currency of an amount. It assumes a specific value as defined by the controlled vocabulary available at http://publications.europa.eu/resource/authority/currency.	Agricore-dcatap:Currency	http://www.agricore-projet.eu/ontology/agricore-dcatap#Currency
DatasetPurpose	A new class, subclass of Concept (skos concept, referring to existing vocabulary), it refers to different purposes of datasets such as: Environmental policy analysis, income and distributional policies analysis, analysis of the technical/economic efficiency levels/gains, trade policy analysis.	Agricore-dcatap:DatasetPurpose	http://www.agricore-projet.eu/ontology/agricore-dcatap#DatasetPurpose
EnvironmentDataset	A new class created to indicate an AGRICORE/ARDIT geo-referenced dataset, subclass of Dataset.	Agricore-dcatap:EnvironmentDataset	http://www.agricore-projet.eu/ontology/agricore-dcatap#EnvironmentDataset
EnvironmentVariable	A new class created to indicate an Environmental Variable, subclass of DatasetVariable	Agricore-dcatap:EnvironmentVariable	http://www.agricore-projet.eu/ontology/agricore-dcatap#EnvironmentVariable
MeasureUnit	A new AmountMeasure subclass. It represents the measuring unit of an amount. It assumes a specific value as defined by the controlled vocabulary available at http://publications.europa.eu/resource/authority/measurement-unit.	Agricore-dcatap:MeasureUnit	http://www.agricore-projet.eu/ontology/agricore-dcatap#MeasureUnit
PriceObject	It is a new class, used to represent information about a price variable	Agricore-dcatap:PriceObject	http://www.agricore-projet.eu/ontology/agricore-dcatap#PriceObject

PriceVariable	It is a new class, PriceObject and SocioEconomic Variable subclass and it inherits their properties.	Agricore-dcatap:PriceVariable	http://www.agricore-projet.eu/ontology/agricore-dcatap#PriceVariable
QuantityObjectAmount	It is a Variable Object Amount subclass.	Agricore-dcatap:QuantityObjectAmount	http://www.agricore-projet.eu/ontology/agricore-dcatap#QuantityObjectAmount
SocioEconomicDataset	It is a Dataset subclass (inheriting all its properties), and it has all the SocioEconomicVariable of the Dataset.	Agricore-dcatap:SocioEconomicDataset	http://www.agricore-projet.eu/ontology/agricore-dcatap#SocioEconomicDataset
SocioEconomicVariable	It is a DatasetVariable subclass.	Agricore-dcatap:SocioEconomicVariable	http://www.agricore-projet.eu/ontology/agricore-dcatap#SocioEconomicVariable

Table 9 New classes created in the AGRICORE DCAT-AP

5.5.2 Description of properties of the new AGROCRE DCAT-AP classes

Property	URI	Range	Card.
ContinentalCoverage	Agricore-dcatap:ContinentalCoverage	dcterms:Location	some
CountryCoverage	Agricore-dcatap:CountryCoverage	dcterms:Location	some
GeonameCoverage	Agricore-dcatap:GeonameCoverage	dcterms:Location	some
RegionCoverage	Agricore-dcatap:RegionCoverage	dcterms:Location	some
AnalysisUnit	Agricore-dcatap:AnalysisUnit	AnalysisUnit	some AnalysisUnit
ReferenceCatalog	Agricore-dcatap:ReferenceCatalog	Catalog	some Catalog
dataset variables	Agricore-dcatap:dataset variables	DatasetVariable	Some DatasetVariable

Table 10 Properties for DATASET (AGRICORE)

Property	URI	Range	Card.
dataset variables'	Agricore-dcatap:dataset variables'	DatasetVariable	some
temporal	Agricore-dcatap:dcterms:temporal	dcterms:PeriodOfTime	max 1
unitReference	Agricore-dcatap:unitReference	AnalysisUnitReference	max 1
statsRepresentativeness	Agricore-dcatap:statsRepresentativeness	xsd:long	max 1
unitAnalysisNumber	Agricore-dcatap:unitAnalysisNumber	xsd:long	max 1

Table 11 Properties for AnalysisUnit (AGRICORE)

Property	URI	Range	Card.
AnalysisUnit	Agricore-dcatap:AnalysisUnit	AnalysisUnitReference	max 1
unitReference	Agricore-dcatap:unitReference	AnalysisUnitReference	

Table 12 Properties for AnalysisUnitReference (AGRICORE)

Property	URI	Range	Card.
aggregationDistance	Agricore-dcatap:aggregationDistance	Size unit	
aggregationGeoReference	Agricore-dcatap:aggregationGeoReference	AggregationGeoReference	max 1
aggregationScale	Agricore-dcatap:aggregationScale	xsd:integer	max 1

Table 13 Properties for AggregationLevel (AGRICORE)

Property	URI	Range	Card.
dataFrequency	Agricore-dcatap:dataFrequency	dcterms:Frequency	max 1
aggregationScale	Agricore-dcatap:aggregationScale	xsd:string	max 1

Table 14 Properties for DataFrequencyElaboration (AGRICORE)

Property	URI	Range	Card.
temporal	dcterms:temporal	dcterms:PeriodOfTime	max 1
DataFrequencyElaboration	Agricore-dcatap:DataFrequencyElaboration	DataFrequencyElaboration	max 1
mathRepresentation	Agricore-dcatap:mathRepresentation	xsd:string	max 1

Table 15 Properties for DatasetVariable (AGRICORE)

Property	URI	Range	Card.
measure unit	Agricore-dcatap:measure unit	MeasureUnit	exactly 1
amount	Agricore-dcatap:amount	xsd:integer	exactly 1

Table 16 Properties for Size Unit (AGRICORE)

Property	URI	Range	Card.
Dataset	Agricore-dcatap:Dataset	Dataset	some

Table 17 Properties for Catalog (AGRICORE)

Property	URI	Range	Card.
currency	Agricore-dcatap:currency	Currency	Exactly 1
size unit	Agricore-dcatap:size unit	Size unit	Exactly 1
priceType	Agricore-dcatap:priceType	"PURCHASE", "SELLING"	Exactly 1

Table 18 Properties for PriceObject (AGRICORE)

Property	URI	Range	Card.
measure unit	Agricore-dcatap:measure unit	MeasureUnit	exactly1
amount	Agricore-dcatap:amount	xsd:decimal	exactly1

Table 19 Properties for QuantityObjectAmount (AGRICORE)

Property	URI	Range	Card.
measure unit	Agricore-dcatap:measure unit	MeasureUnit	exactly1
amount	Agricore-dcatap:amount	xsd:decimal	exactly1

Table 20 Properties for QuantityObjectAmount (AGRICORE)

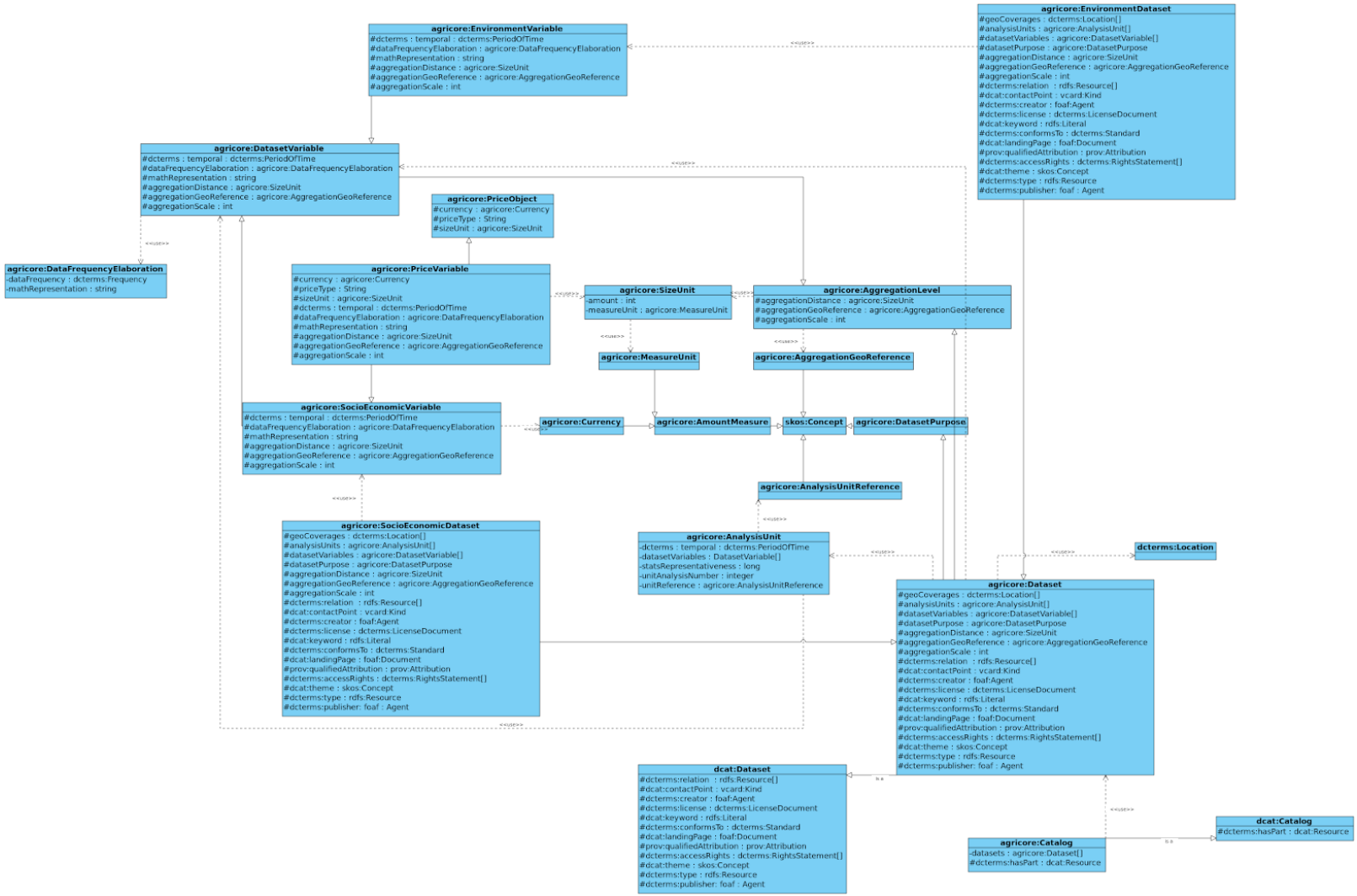

Figure 18 The AGRICORE DCAT-AP Data Model Representation

Characteristic Captured and Presented in the ARDIT	Expression into the AGRICORE Ontology
Geographical coverage	Property **dct:spatial**. A list of spatial regions or named places may be represented using a controlled vocabulary (e.g. http://dd.eionet.europa.eu/vocabulary/common/nuts): or geographic coordinates. A prioritised methodology is needed to select the Geographical coverage. E. g., Use DCAT-AP list of vocabularies (3EU ones + geonames), provide a prioritised selection (first continent, then country, then region (all using EU vocabulary) and if not available, geonames.
Type of data set (Object of analysis)	Property **dct:subject**. Each dataset may have one or more subjects. The values may be any value present in the Digital Europa Thesaurus.
Unit of analysis	A new class in the ontology was created: **AnalysisUnit,** with 3 properties (list): • unitReference (a skos:Concept), referring to a vocabulary that defines the covered unit of a dataset (e. g.: http://dd.eionet.europa.eu/vocabulary/eurostat/ind_farm/); • (optional) unitAnalysisNumber (a xsd:integer); • AnalysisUnit will have a list of children (which are dataset Variables). • Temporal extent/coverage (check DCAT-AP) may be used at 3 levels, at the dataset level, at Analysis Unit (level) and at the Dataset Variable (level). In order to define the structure of the vocabulary, a specific class in the ontology was created: AnalysisUnitReference. It is an extension of skos: concept, and some related individuals, these lists shall be extended, linking to existing vocabularies.
Name	Property **dct:title**.
Distribution	Property **dcat:distribution**.
Data Service	Property **dcat:dataservice**.
Link to the dataset information	Property **dcat:landingPage**.
Producer	Property **dct:publisher**.
Language of the dataset	Property **dct:language** (list of languages also supported already in DCAT). It also includes regional information on how the data is represented (I. e., decimal points, dates).
Type of access	Property **dcat:accessRights** should be referred to RightsStatement. Other information like API endpoint shall be provided by the **DataService** class.
Statistically representative	Class: AnalysisUnit. A property Representativeness containing only a field (represented units) indicates if the dataset is a Census, or a Statistical Representation.
Aggregation level	A specific class AggregationLevel has been created, with these properties: • equivalent scale (e. g., 1/1000 -> 1000) integer. • distance (e. g., 1 km) class. • geo reference name (specific class with this example list of individuals: CONTINENTAL, COUNTRY, NUTS1.
Temporal extent	Property **dct:temporal**.
Periodicity of publications	Property **dct:accrualPeriodicity**.
Data frequency	A new class: DataFrequencyElaboration with 2 properties has been created: Frequency (vocabulary http://publications.europa.eu/resource/dataset/frequency); Mathematical representation, which, currently is a range of literal constants, defined in the ontology e. g.: • AVERAGE,

	• INSTANT_VALUE, • MAX, • MEDIAN, • MIN, • MODE, • SUM, • VARIANCE, • STANDARD DEVIATION
Dataset format	Property **dcat:distribution** that refers to one or more Distribution object.
Useful for the analysis of	DatasetPurpose, a new class, extension of skos: concept has been created, with some related individuals, e. g.: • Environmental policy. • Greenhouse gas emissions. • Energy use in agriculture. • Climate change. • Income and distributional policies. • Insurance policy. • Technical efficiency analysis.
Values	The same mathematical representation used by DataFrequencyElaboration shall be used.
Themes covered	The property **dcat:theme** shall be used. Each dataset can have one or more themes. The values can be the URI of the Eurovoc controlled vocabulary or Eionet Data Dictionary.
Variables included	The class **DatasetVariable** has been created, which could have more specific properties depending on the dataset (e. g., statistic variable, geo variable).

Table 21 Mapping of the ARDIT Datasets Characteristics into the AGRICORE DCAT-AP 2.0 Extension Ontology

6 Conclusions

Deliverable 1.1 has provided evidence on the use of ontologies to capture and systematise rich domains of knowledge, such as agriculture. Using an ontology for the efficient management of a lot of information may be more important when large-scale/complex models require making use of more than one data source. This is especially true when it is necessary to ensure that one/a few variables are available to the researcher to operate the model(s).

Due to the lack of extant ontologies capable of identifying the relevant information on key variables contained in (a) dataset(s), and the relationships among them, this Deliverable has documented the development of the AGRICORE DCAT-AP 2.0 ontology. It is an extension of the DCAT-AP 2.0 data model which has been undertaken mainly by adding new classes - and relationships - capable of capturing the needs to know the characteristics (of the variables contained in) the datasets which could be employed for the analysis of the impacts of policy (reform(s)) on agriculture. Modellers involved in the AGRICORE project expressed these knowledge requirements during the execution of Task 1.1 helping to prepare a template for dataset characterisation, which could capture the characteristics of statistical and geo-referenced datasets alike. The template will allow collecting the relevant information about the datasets of interest to a researcher, without having access to the data. Because of the information collected by means of the characterisation template will be manipulated with and managed by the AGRICORE DCAT-AP 2.0 ontology, the two have been developed synergically. The characterisation template will constitute the tool for enacting the characterisation of the datasets to be undertaken in Task 1.2 to 1.6. Nonetheless, upon verifying that the template may fall short of capturing important characteristics of datasets to be characterised, partners of the AGRICORE project will update both the template and the ontology which organises and manages the information.

Researchers' awareness and knowledge of which variables, and their characteristics, are available in which datasets are crucial in making research efforts effective. Therefore, the simultaneously developed characterisation template and AGRICORE DCAT-AP 2.0 ontology will allow capturing many details about a large number of datasets of potential interest to the research community. This information will be stored in the ARDIT (to be Delivered in Task 1.8) and will be searchable - also by means of semantic services (to be provided in Task 4.4) - on the public internet, thanks to the AGRICORE DCAT-AP 2.0 ontology. Search queries will return information regarding the datasets, and - most importantly - the variables contained therein, which may be employed for running models for the analysis of the impacts of policy (reform(s)) on agriculture. Hopefully, the ARDIT will become a reference tool for identifying the datasets relevant to the modelling efforts of the research community.

7 References

1. ^ E. Bonabeau, “Agent-based modeling: Methods and techniques for simulating human systems,” Proceedings of the national academy of sciences, vol. 99, no. suppl 3, pp. 7280–7287, 2002.
2. ^ K. Happe, Agricultural policies and farm structures-Agent-based modelling and application to EU-policy reform. 2004.
3. ^ S. Shrestha, A. Barnes, and B. V. Ahmadi, Farm-level modelling: Techniques, applications and policy. CABI, 2016.
4. ^ B. Frédérick, C. Hanachi, M. Lauras, P. Couget, and V. Chapurlat, “A metamodel and its ontology to guide crisis characterization and its collaborative management,” in Proceedings of the 5th International Conference on Information Systems for Crisis Response and Management (ISCRAM), Washington, DC, USA, May, 2008, pp. 4–7.
5. ^ J. Domingue, D. Fensel, and J. A. Hendler, Handbook of semantic web technologies. Springer Science & Business Media, 2011.
6. ^ G. Van Heijst, A. T. Schreiber, and B. J. Wielinga, “Using explicit ontologies in KBS development,” International journal of human-computer studies, vol. 46, no. 2–3, pp. 183–292, 1997.
7. ^ N. Guarino, “Understanding, building and using ontologies,” International Journal of Human-Computer Studies, vol. 46, no. 2–3, pp. 293–310, 1997.
8. ^ I. S. Bajwa, “A framework for ontology creation and management for semantic web,” International Journal of Innovation, Management and Technology, vol. 2, no. 2, pp. 116–118, 2011.
9. ^ E. S. Alatrish, “Comparison Some of Ontology Editors,” Management Information Systems, vol. 8, no. 2, pp. 18–24, 2013.
10. ^ T. Slimani, “Ontology Development: A Comparing Study on Tools, Languages and Formalisms,” Indian Journal of Science and Technology, vol. 8, no. 24, 2015.
11. ^ D. L. Rubin, H. Knublauch, R. W. Fergerson, O. Dameron, and M. A. Musen, “Protege-owl: Creating ontology-driven reasoning applications with the web ontology language,” in AMIA Annual Symposium Proceedings, 2005, vol. 2005, p. 1179.
12. ^ Gaševic, Dragan, D. Djuric, and Devedžic, Vladan, “Ontologies,” in Model Driven Engineering and Ontology Development, Springer, 2009, pp. 45–80.
13. ^ [1][2][3][4] J. An and Y. B. Park, “Methodology for automatic ontology generation using database schema information,” Mobile Information Systems, vol. 2018, 2018.
14. ^ E. Francesconi, S. Montemagni, W. Peters, and D. Tiscornia, “Integrating a bottom–up and top–down methodology for building semantic resources for the multilingual legal domain,” in Semantic Processing of Legal Texts, Springer, 2010, pp. 95–121.
15. ^ D. Soergel, B. Lauser, A. Liang, F. Fisseha, J. Keizer, and S. Katz, “Reengineering thesauri for new applications: The AGROVOC example,” Journal of Digital Information, vol. 4, no. 4, 2004.
16. ^ B. Lauser, M. Sini, A. Liang, J. Keizer, and S. K. BorisLauser, From AGROVOC to the Agricultural Ontology Service / Concept Server An OWL model for creating ontologies in the agricultural domain. .

17. ^ A. Liang et al., “The Mapping Schema from Chinese Agricultural Thesaurus to AGROVOC.,” in 6th Agricultural Ontology Service (AOS) Workshop on Ontologies, 2005, pp. 1–6.

18. ^ A. Goldstein, L. Fink, and G. Ravid, “A Framework for Evaluating Agricultural Ontologies,” Jun. 2019, [Online]. Available: http://arxiv.org/abs/1906.10450.

19. ^ M. T. Maliappis, Using agricultural ontologies. Springer US, 2009, pp. 493–498.

20. ^ R. Gaire, L. Lefort, M. Compton, G. Falzon, D. Lamb, and K. Taylor, “Demonstration: Semantic Web Enabled Smart Farm with GSN,” 2013, [Online]. Available: http://smartfarm-ict.it.csiro.au.

21. ^ C. Goumopoulos, A. D. Kameas, and A. Cassells, “An ontology-driven system architecture for precision agriculture applications,” International Journal of Metadata, Semantics and Ontologies, vol. 4, no. 1–2, pp. 72–84, 2009, doi: 10.1504/IJMSO.2009.026256.

22. ^ C. Jonquet et al., “AgroPortal: A vocabulary and ontology repository for agronomy,” Computers and Electronics in Agriculture, vol. 144, pp. 126–143, Jan. 2018, doi: 10.1016/j.compag.2017.10.012.

23. ^ R. Shrestha et al., “Bridging the phenotypic and genetic data useful for integrated breeding through a data annotation using the Crop Ontology developed by the crop communities of practice,” Frontiers in Physiology, vol. 3 AUG, 2012, doi: 10.3389/fphys.2012.00326.

24. ^ 1 2 I. N. Athanasiadis, A.-E. Rizzoli, S. Janssen, E. Andersen, and F. Villa, “Ontology for seamless integration of agricultural data and models,” in Research conference on metadata and semantic research, 2009, pp. 282–293.

25. ^ M. A. Musen, “Dimensions of knowledge sharing and reuse,” Computers and biomedical research, vol. 25, no. 5, pp. 435–467, 1992.

26. ^ C. W. Holsapple and K. D. Joshi, “A collaborative approach to ontology design,” Communications of the ACM, vol. 45, no. 2, pp. 42–47, 2002.

27. ^ S. Janssen, E. Andersen, I. N. Athanasiadis, and M. K. van Ittersum, “A database for integrated assessment of European agricultural systems,” Environmental Science & Policy, vol. 12, no. 5, pp. 573–587, 2009.

28. ^ E. Alatrish, “Comparison Some of Ontology Editors,” Management Information Systems, vol. 8, no. 2, pp. 18–24, 2013, [Online]. Available: https://pdfs.semanticscholar.org/bdf8/056feec90b60c338d3818a20e17c4b199458.pdf?_ga=2.36105125.422640853.1595319726-1671690332.1587569242.

29. ^ B. Wyns, M. Dekkers, N. Loutas, V. Peristeras, and A. Karalopoulos, DCAT Application Profile for Data Portals in Europe. 2016.

For preparing this report, the following deliverables have been taken into consideration:

Deliverable Number	Deliverable Title	Lead beneficiary	Type	Dissemination Level	Due date
D10.1	Project Management Handbook	IDENER	Report	Confidential	M01