D9.3 Final Data Management Plan

agricore

Deliverable Number	D9.3
Lead Beneficiary	AAT
Authors	AAT, IDE
Work package	WP9
Delivery Date	M58
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	Initial version from D9.2	IDE	11-jul-2024
0.2	Adding datasets descriptions	IDE	22-jul-2024
0.3	Revision	IDE	29-ago-2024
1.0	Final version	IDE	31-ago-2024

Disclaimer

All the contributors to this deliverable declare that they:

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

Executive Summary

This document focuses on providing AGRICORE's consortium with information on how to deal with data generated during the execution of the project AGRICORE and how it should be managed. Specifically, this document provides information about the type of data is generated, the standards and the accessibility to the data for verification and re-use, and the curation and preservation of data procedures. At the same time, being a public document, it will allow the easy identification of the datasets generated within the project by other researchers who may be interested in reusing or validating the information produced.

This deliverable details all the data management procedures within AGRICORE in response to article 29.3 of the grant agreement, specifically regarding open access to research data. Accordingly, it presents information for partners in the consortium as well as for external parties about the different datasets generated and used within the project. They will be categorised and technically detailed in terms of data collection, processing and generation in order to manage the data that is associated and generated by the work of the AGRICORE consortium. This topic requires a significant level of detail posing for the need of an autonomous deliverable, rather than being a part of D10.1 (Project Management Handbook) where all of the other day-to-day procedures for the running of the project are presented. The monitoring of the AGRICORE project with regard to this area will be the subject of continuous reviews within the context of WP9 at each consortium meeting.

Moreover, the current data management plan (DMP) is the final draft of this document, which is based on the initial draft presented in D9.2 (M06). This means that the DMP has been a living document, continuously updated by its responsible with the information provided by all partners in the project.

This deliverable is based on the Guidelines on Data Management in Horizon 2020 document provided by the European Commission [1], the information provided by the UK Digital Curation Center and the DMPtool [2].

Abbreviations

Abbreviation	Full name
AAT	Ayesa Advanced Technologies
DCC	UK Digital Curation Centre
DMP	Data Management Plan
DPO of the project	Data Protection Officer of the project
EC	European Commission
FAIR	Findable, Accessible, Interoperable and Reusable
GDPR	General Data Protection Regulation
ID	Identifier
IDE	IDENER
IPR	Intellectual Property Rights
TBD	To Be Determined
WP	Work Package

List of Figures

List of Tables

Table of Contents

1 Introduction

This document presents the final version of the Data Management Plan (DMP) for the AGRICORE project. This is based on the initial version of the DMP (D9.2), which was partially prepared following the guidance of the UK Digital Curation Centre (DCC) (http://www.dcc.ac.uk), an internationally recognised centre of expertise in digital curation with a focus on building capability and skills for research data management. The DCC provides expert advice and practical help to research organisations wanting to store, manage, protect and share digital research data. This DMP for AGRICORE details the public datasets that the project:

- has generated,
- whether and how it will be exploited or made accessible for verification and re-use,
- how it will be curated and preserved.

The academic papers produced within the project have been made available as open access for several years (following the green or gold open-access approach). In a similar way, the data generated as a result of the project's research activities has also been put online and freely available in an open-access repository. The DMP includes the details on the datasets generated, providing detailed information in such datasets. Specifically, this DMP details the following aspects:

- metadata generation,
- data preservation,
- data storage beyond the end of the project.

In particular, the consortium partners acknowledge their responsibilities for fulfilling and updating this DMP, which includes:

- the first version of the DMP must be operative in the first 6 months of the project,
- it will be a live document continuously updated within the project timeframe,
- data identified (generated within the project) in DMP must be shared in an online repository by the producers of the data, appropriate support will be provided to those involved in the project from both the DMP responsible and the project coordinator.

At a deeper level of detail, the aim of the DMP is that its execution and update will lead to:

- a better understanding of the data produced as output from the project,
- clarity on how the data is actually used within the project and outside of it,
- continuity in the work of the consortium in the event of staff leaving or entering the project during its lifecycle or, equivalently, staff changing roles within the project during its lifecycle. This includes such areas as:
    - avoiding duplication of effort i.e., re-collecting or re-working data,
    - enabling validation of results,
    - contributing to collaboration through data sharing,
    - increasing visibility of output and thereby leading to greater impact. In particular, enabling other researchers to cite and use the datasets generated within the project.

AGRICORE embraces the open-access approach incentivised by the EC. This means that the project aims to improve and maximise access to and re-use of research data generated by it and considers the need to balance openness and protection of scientific information, commercialisation and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservations. Within the AGRICORE project, most of the deliverables produced are done as public documents. A similar approach will be undertaken with the generated data.

1.1 Update of the DMP

As said before, the DMP is a live document that must be updated during the project execution. Thus, based on the initial version of the DMP (M06), which settled the baseline to collect, process, analyse and store project data, the data-related activities were performed. During the project execution, the main source of data has been use cases execution from participatory research activities to simulation results.

Under the frame of WP1, a list of public datasets was searched as part of participatory research. These datasets were characterised and used to feed the ARDIT tool. In total, almost 300 datasets were characterised following an extended version of the characterisation template presented in D9.2 (see Section 7).

Moreover, for the simulation of use cases, several datasets were produced based on the collected data and non-public datasets. These datasets are described in Section 4.2, achieving 23 produced datasets. The main ones are the synthetic populations generated based on the FADN microdata. Despite the relevance of those datasets, it was decided to not publish them as open-source datasets given the data privacy and anonymity issues of the data from they derived, the FADN.

2 Data Management Plan

The DMP aims to regulate all the processes related to the data life cycle.

At this stage of the project, the consortium has worked on the whole lifecycle. Figure 1 shows the whole process developed in the data lifecycle, composed of mainly 4 steps:

- Data storage. This step is composed of the next 4 internal procedures: collection, description/characterization, analysis and recollection. This was done in the first stage of the project, concluding with the publication of D1.3, D1.4, D1.5 and D1.6.
- Archive. Datasets characterisations were stored in the project servers, feeding the ARDIT tool.
- Publication. Datasets generated during the project execution should have been published in the Zenodo repository, but this was not done due to data privacy issues.
- Data re-use. The implementation of the ARDIT tool makes data more accessible and facilitates their re-use for future research activities.

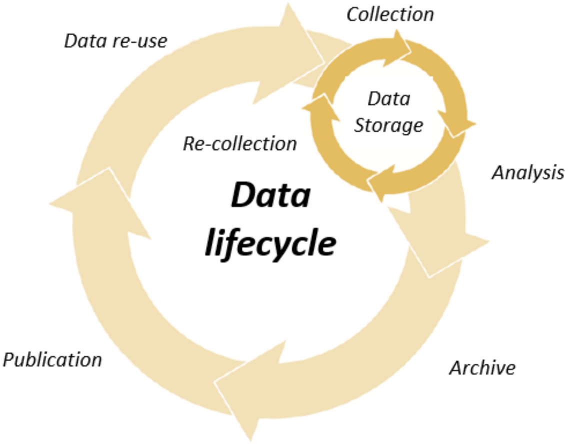

Figure 1 Data lifecycle

In addition, the usage of the data in the consortium has also been considered and monitored.

The initial DMP provided templates to characterise the datasets generated or employed within the project (which consists mainly of primary, derived and simulation data). During the project, there was no identified need for collecting personal data. Indeed, the participatory research activities limited the collection of sensitive information. Should it be the case, any data collection process was done considering the anonymization of the data, target that was assessed and guided by the recommendations of the DPO of the project.

Moreover, as part of the project work plan, the partners participating in AGRICORE have generated an Index tool of the data sources available at the regional, national and EU levels that can be used for conducting agricultural research. As part of this work, the consortium has identifyed and characterised these databases. The information contained in such data sources was not stored in any repository but rather characterised in the Confluence platform first and later in the Index tool (which is the scope of Task 1.8). The characterisation and mapping of this information are separated from the other project activities, meaning that the data sources analysis goes further than the specific needs for executing the rest of the project's tasks. This approach aims to produce a more extensive mapping of data sources, promoting and facilitating future research on related topics.

3 Data summary

This section includes in the final DMP the types and formats of data generated and/or collected during the project. With the aim of defining an effective DMP and following previous DMP schemes, the datasets utilised have been categorised according to different data-related features, source, format, stability and volume.

Source

Regarding the source, the data can be categorised into 4 types: observational, experimental, simulation and derived/compiled. For each category, the datasets utilised in AGRICORE will be analysed.

- *Observational*

Observational data is the data category that has been used to build and calibrate the models. In this case, the main source of observational data has been the FADN database. This database provides a comprehensive information source regarding the economic, management, and agricultural activity of agricultural holdings according to the techno-economic orientation, their economic level, and their geolocalisation. In addition to FADN, the Eurostat database has also been used to complete some required information fields not available in FADN.

Other observational datasets have been generated during the project by performing participatory research activities that involve different stakeholders. The aim of using these datasets is to fill the existing gaps in other observational datasets, especially when filling the information not available in the above mentioned datasets and that is required to generate the synthetic populations and to execute the activities of the AGRICORE project.

- *Experimental*

No experimental datasets were utilised during the project duration. None of the activities carried out during the project encompassed a controlled manipulation of variables in an experimental environment to assess the effect of such manipulations.

- *Simulation*

Several datasets were generated during the project activity that can be categorised as simulation datasets. Basically, these are the datasets that contain the results of the simulations of the models built within the project. For each simulation performed a new dataset is generated. Simulation datasets are reproducible as long as the simulation parameters, inputs and simulation configuration (including synthetic population) do not change.

- *Derived/compiled*

These kinds of datasets are generated by existing datasets and are reproducible. In this scope, all synthetic populations generated are derived datasets from a dataset composed of several observational datasets, including the FADN dataset, Eurostat datasets, etc. These primary datasets are processed and transformed following a set of rules to create a new dataset, which ultimately is the synthetic population. Synthetic populations are framed within this category because they are not a direct observation of the original dataset but a dataset that emulates the characteristics of the original.

Reports on primary datasets are also included in this category: different data analyses were carried out in the scope of crop representativeness, subsidies, economic compensations, crops associated and years of activity for each subsidy. These reports were generated for each use country included in the project at the use case level.

Format

When describing the format categorisation, the data can be generated in different forms: text, numeric, audiovisual, models, computer code, discipline-specific and instrument-specific. The AGRICORE project will generate data in all these forms. Two major categories cover the data generated in the project: numeric and text data. Numeric data, mainly in the shape of tabular data, is the result of simulations or derived datasets. With a lower presence than numeric data, text data is also used to populate the data warehouse needed and employed in the project.

Stability

Regarding the stability of the data, the data can be immutable or change over the course of the project. This feature will affect the way datasets are managed to preserve such immutability. The common categories of a dataset are:

- Fixed datasets: never change after being generated or collected.
- Growing datasets: new data may be added while old data are never changed or deleted.
- Revisable data: new data may be added while the old data may be changed or deleted.

In this context, partners have used all of the three types listed above during the execution of the AGRICORE project. For that, keeping track of data versions has been required.

Volume

Finally, some remarks regarding the volume of data managed during the project. As was forecasted at the project start, a high volume of data was utilised, including primary datasets and generated datasets.

3.1 Produced datasets

This section contains the template that all users have filled out for each of the used and produced datasets. As part of the continuous updating process of this document, the newly identified or generated datasets have been included both in this document and in the internal portal of the project (Confluence).

The following is the template that has been used to fill in the information relative to the datasets produced during the project.

Table 1 Template to characterise the produced datasets

Partner: (Partner name)	
Identifier	The identifier will have the following format (where internalID is an incremental ID starting from 1 for each dataset produced in the project): AGRICORE_shortDescription_MainAuthor_internalID
Dataset description	A detailed description of the dataset
Purpose of the data	Explanation of the purpose of this dataset
Type of data	Numeric, text, etc.
Form of the data	The way the information is provided and generated
Format of the data	The format in which the data are released
Origin of the data	The source generating this data. It should also include details on reused data
Dataset stability	Details on the stability (as defined in the previous section) of the data

Size of data	Estimation (or measurement when already produced) of the data size

3.2 Reused data sources

As indicated in the previous section, the AGRICORE project includes dedicated activities to map and characterise available data sources useful for agricultural modelling research. The findings of such activities will be released as part of the Index tool generated within the project. Nonetheless, a list of the data sources identified is included in this document. Throughout the course of the project, the consortium has identified and characterised the following datasets:

Table 2 Datasets selected to be characterised for the ARDIT tool

Region covered	Main topic covered	Dataset name
EU level	Climate-Related	European Climate Assessment & Dataset NASA Prediction Of Worldwide Energy Resources (POWER) Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) ADMS AgMERRA Climate Forcing Dataset for Agricultural Modeling AgCFSR Climate Forcing Dataset for Agricultural Modeling CRU TS (Climatic Research Unit Time-series) dataset Global Summary of the Day (GSOD)3 WorldClim Version2
	Soil/Land/Quality/Biodiversity-related	GIOŚ LUCAS - Land Cover and Land Use Landscape European Soil Database v2 Raster Library MIRCA2000 SoilGrids SoilHydroGrids
	General Agriculture data	FADN FSS (Farm Structure Survey) Eurostat OECD FaoStat IACS - Integrated Administration and Control System Feedipedia Feedprint AFFRIS - Aquaculture Feed and Fertilizer Resources Information System
	Geo-referenced datasets	CORINE Land Cover LUCAS Earthstat Gridded Livestock of the World

		WorldClim - Global Climate Data MOD16A3 v006 MODIS/Terra Net Evapotranspiration Yearly L4 Global 500 m SIN Grid Rainfall Erosivity Database on the European Scale (REDES) RUSLE High-Resolution Layer: Imperviousness Density (IMD) 2015
National and regional datasets	Farm characterization datasets	(Italian) RICA(Spanish) SSAO2016 Survey on the Structure of Agricultural Holdings (Polish) ARMA - RDP 2014-20 (PL) Implementation reports (Polish) Statistics Poland - Agricultural and horticultural crops (Polish) Statistics Poland - Animal production, Farm animals (Polish) Statistics Poland - LOCAL DATA BANK (Polish) ARMA - RDP 2014-20 (PL) Implementation reports (Greece) ELSTAT - Annual Agricultural Statistical Survey (Greece) ELSTAT - Crops Survey (Greece) ELSTAT - Farm Structure Survey (FSS) (Greece) ELSTAT - Livestock Surveys (Greece) Hellenic Statistical Authority on Agriculture and Livestock (Spanish) Statistics on Means of Production: Use of Fertilizers (Spanish) Statistics on Means of Production: Commercialization of Phytosanitary Products (Spanish) Statistics on Means of Production: Registration of New Machinery (Spanish) International Penn TablesSSAO2016 (Spanish) Agrifood Foreign Trade Statistics (Spanish) BDSICE. Cost and price index (Spanish) BDSICE. National production and demand indicators (Spanish) BDSICE. Price and costs. Agricultural wage index (Spanish) BDSICE. Price and costs. Salary increases in agreement and salary increases registered in agriculture (Spanish) Climate data obtained by the Agroclimatic Stations Net (Spanish) Evolution of provincial agricultural macromagnitudes 2005-2014 (Spanish) FEGA

		(Spanish) MARM. Household consumption database (Spanish) Monthly production, movement and stock data (AICA) (Spanish) Multi-territorial Information System of Andalusia (SIMA) (Spanish) National Agrarian Accounting Network RECAN (Spanish) National Agricultural Economic Statistics: Agricultural rates and prices paid (Spanish) National Agricultural Economic Statistics: Agricultural rates and prices received (Spanish) National Agricultural Economic Statistics: Agricultural rates and salaries (Spanish) National Agricultural Economic Statistics: Average land prices for agricultural use (Spanish) National Agricultural Economic Statistics: Short-term prices of agricultural products (Spanish) National Statistics Institute: Agricultural Census (2009) (Spanish) National Statistics Institute: Survey on Production Methods in Agricultural Operations (2009) (Spanish) Olive. Data obtained from the monitoring of pests and diseases in the biological control stations (Spanish) Organic Farming in Spain (Spanish) Rice. Data obtained from the monitoring of pests and diseases in the biological control stations (Spanish) SIGGAN
	Land Quality and Characteristics Datasets	(Polish) Database on mineral nitrogen content in Poland (3 depths, 2 times per year) (Spanish) SIGPAC (Spanish) ESYRCE (Spanish Survey on Crop Surfaces and Yields) (Spanish) National Soil Erosion Inventory (INES) (Spanish) National Hydrological Report
	Climate-Related Datasets	(Polish) Climate-related dataset for Poland (Polish) ADMS (Polish) Data set on mineral nitrogen content in Poland (3 depths, 2 times per year) (Greece) Climate-related dataset

4 Source datasets

This section includes a description of the main datasets used to develop and simulate project use cases. These are classified by source, indicating the name of the source in the header of each characterisation table in the case of observational datasets and a descriptive name in the case of generated datasets. These descriptions are at the metadata level, but more detailed descriptions of the observational datasets and other datasets were done for the characterisation and population of the ARDIT tool. For these datasets, identifiers were not assigned because they were not produced by the project, and some of them also contain non-public data.

4.1 Observational datasets

Table 3 FADN dataset

FADN	
Identifier	N/A
Dataset description	FADN (Farm Accountancy Data Network) dataset is a dataset that provides information on economic, farm management and overall farm, crop and livestock management decisions of European farms. The data contained in this dataset comes from real-world agricultural holdings across various countries in the European Union. All the data has been anonymised to protect the privacy of the individual farmers and ensure confidentiality. The dataset contains different variables with information relative to size, crop type, number of animals, incomes, expenditures, subsidies, money received from subsidies, number of workers in the holding, and demographical data. The dataset is representative of all the farmers contained in a specific region. This means that the dataset only contains a representative sample of the total farmers included in the represented region. The selection criteria ensure a wide range of farm sizes, types and regional distributions to fully capture the diversity and heterogeneity of the EU agricultural sector. Only farms considered professional due to their economic size are included in the dataset. This dataset has been used for Greece, and Poland use cases. The data request covered years from 2014 to 2020.
Purpose of the data	The purpose of this dataset is to serve as a foundational source of information on the economic, social, typological, and management-related characteristics of farmers. This data will be used to generate synthetic populations that accurately replicate the real-world attributes of farmers.
Type of data	Numeric and categorical data.
Form of the data	Dataset is organised in a tabular format. Each row represents the information relative to an anonymised farm. Each column contains the information for a specific farm attribute.
Format of the data	Dataset files have been provided in CSV format. There is one single table for each country and year.
Origin of the data	The data comes from real-world agricultural holdings across various countries in the EU. Data is collected annually from a representative sample of farms in all EU member states.
Dataset stability	This dataset has maintained its original format throughout the project to avoid altering or disturbing the original information. Subsequent modules using the dataset load the raw data and apply the required transformations and variables selection specified.
Size of data	Estimation (or measurement when already produced) of the data size

	Greece: 75,3MB; Poland: 378MB

Table 4 RECAN dataset

RECAN	
Identifier	N/A
Dataset description	RECAN (Red Contable Agraria Nacional) dataset is a dataset that contains accountancy agricultural data for Spain. It is a network similar to FADN but at the national level. The data contained in this dataset comes from real-world agricultural holdings at the Spanish level. All data is anonymised to protect the privacy of the individual farmers and ensure confidentiality. The dataset contains different variables with information relative to size, crop type, number of animals, incomes, expenditures, subsidies, money received from subsidies, number of workers in the holding, and demographical data. The dataset is representative of all the farmers contained in a specific region. This means that the dataset only contains a representative sample of the total farmers included in the represented region. The selection criteria ensure a wide range of farm sizes, types and regional distributions to fully capture the diversity and heterogeneity of the EU agricultural sector. Only farms considered professional due to their economic size are included in the dataset. This dataset has been used for the Andalusian use case. The data request covered years from 2014 to 2020.
Purpose of the data	The purpose of this dataset is to serve as a foundational source of information on the economic, social, typological, and management-related characteristics of farmers. This data will be used to generate synthetic populations that accurately replicate the real-world attributes of farmers.
Type of data	Numeric and categorical data.
Form of the data	Dataset is organised in a tabular format. Each row represents the information relative to an anonymised farm. Each column contains the information for a specific farm attribute.
Format of the data	Dataset files have been provided in CSV format. For each year, there are several topic-oriented CSV files containing all the variables related to that topic.
Origin of the data	The data comes from real-world agricultural holdings for all the NUTS3 regions included in Spain. The data origin is annual surveys for representative samples of farms.
Dataset stability	This dataset has maintained its original format throughout the project to avoid altering or disturbing the original information. Subsequent modules using the dataset load the raw data and apply the required transformations and variables selection specified.
Size of data	34,1MB

Table 5 Italian RICA dataset

ITALIAN RICA	
Identifier	N/A
Dataset description	Italian RICA (Rete di Informazione Contabile Agricola) is an annual sample survey established by the European Economic Commission. It has been applied in Italy since 1968, and it follows a similar approach to FADN in other Member States of the EU.

	It is a comprehensive and harmonised source of structural, microeconomic, social and financial data and the evolution of such indicators. The data contained in this dataset comes from real-world agricultural holdings at the Italian level. All data is anonymised to protect the privacy of the individual farmers and ensure confidentiality. The dataset is representative of all the farmers contained in a specific region. This means that the dataset only contains a representative sample of the total farmers included in the represented region. The selection criteria ensure a wide range of farm sizes, types and regional distributions to fully capture the diversity and heterogeneity of the EU agricultural sector. Only farms considered professional due to their economic size are included in the dataset. This dataset has been used for the Italian use case. The data available covered years from 2018 to 2020.
Purpose of the data	The purpose of this dataset is to serve as a foundational source of information on the economic, social, typological, and management-related characteristics of farmers. This data will be used to generate synthetic populations that accurately replicate the real-world attributes of farmers.
Type of data	Numeric and categorical data.
Form of the data	Dataset is organised in a tabular format. Each row represents the information relative to an anonymised farm. Each column contains the information for a specific farm attribute.
Format of the data	Dataset files have been provided in CSV format. For each year, there are several topic-oriented CSV files containing all the variables related to that topic.
Origin of the data	The data comes from real-world agricultural holdings in Italy. The data origin is annual surveys for representative samples of farms.
Dataset stability	This dataset has maintained its original format throughout the project to avoid altering or disturbing the original information. Subsequent modules using the dataset load the raw data and apply the required transformations and variables selection specified.
Size of data	4,29 MB

Table 6 Land value survey dataset

Land value survey dataset	
Identifier	N/A
Dataset description	The land value survey dataset is a dataset that contains land value in euros per hectare categorised by land type. Data differentiates between dry land and irrigated land. Additionally, there is another breakdown level based on the type of crops cultivated on the land. The breakdown is done for specific crops or groups of crops, and it reflects the value added by the crop to the land or the land suitability for cultivating specific crops. The crop resolution level is not very high, but it includes the main categories of crops according to land usage. Additionally, the information includes the evolution in prices with regard to previous years in absolute and relative differences.
Purpose of the data	This dataset is used to fill the agent attribute “landValue”. This is a relevant parameter used by the simulation model, and it is not present in none of the datasets FADN and RECAN.
Type of data	Numeric and categorical data.
Form of the data	Data was organised in tabular format. The layout of the information is one row per crop or crop category and one column per information field.

Format of the data	Data was initially contained in pdf format. Then, tables of interest were extracted and converted into CSV format.
Origin of the data	Information comes from surveys performed by the Government of Spanish, Ministry of Agriculture, Fisheries and Food about Land prices.
Dataset stability	The dataset has been transformed to modify its original format. Data was extracted from the original pdf and then converted into CSV.
Size of data	830 kB

Table 7 Agrarian census dataset

Agrarian Census	
Identifier	N/A
Dataset description	The Agrarian Census is a Spanish statistical operation performed periodically in Spain. It gathers detailed information about the structural features of agricultural holdings. The information contained in the Spanish Agrarian Census is typically used to analyse and elaborate agricultural and rural policies. The information contained covers a varied range of fields, including structural, crop and livestock, technology and agricultural practices, economic and social.
Purpose of the data	The utilisation of this dataset within AGRICORE is to infer geospatial regions below the NUTS3 level. Given specific structural, crop and livestock distributions and other holding features, a higher geospatial resolution level is assigned to synthetic farmers.
Type of data	Numeric and categorical data.
Form of the data	Dataset is organised in tabular format. Different possible configurations are allowed according to the variables of interest.
Format of the data	Dataset was downloaded in CSV format.
Origin of the data	The data recompilation for Agrarias Census is fundamented on the initiative promoted by Eurostat to harmonise agricultural statistics across all Member States of the EU. The Census is conducted by the INE (Instituto Nacional de Estadística), which develops the methodology, coordinates data-gathering campaigns, and processes and publishes the information. Data is gathered through structured surveys conducted online, on paper or via personal interviews.
Dataset stability	This dataset has maintained its original format throughout the project to avoid altering or disturbing the original information. Subsequent modules using the dataset load the raw data and apply the required transformations and variables selection specified.
Size of data	100 KB

Table 8 Statistical Yearbook of Agriculture dataset

Statistical Yearbook of Agriculture	
Identifier	N/A
Dataset description	The Statistical Yearbook of Agriculture is a comprehensive publication in the scope of agriculture statistics for Poland. This data recompilation has been published annually since 2005. It is organised in different agricultural-related themes, which cover a wide range of information on agricultural production results, production and economic conditions in agriculture, production balance for major agricultural

	products, supply and use of production goods, and information on income situation in agriculture.
Purpose of the data	The utilisation of this dataset within AGRICORE is to fill different agent attributes not found on FADN data for the Polish use case. Land information has been used to infer NUTS3 geospatial resolution according to the number of farms, surface distribution and crop distribution.
Type of data	Numeric and categorical data.
Form of the data	Dataset is organised in tabular format. The complete dataset contains different tables organised by themes.
Format of the data	The complete original report is in PDF format.
Origin of the data	The publication has been prepared by the Agriculture Department, Central Statistical Office of Poland.
Dataset stability	The dataset has been processed for two reasons: 1) to extract the relevant information and 2) to convert data format into CSV for easy dataframe conversion.
Size of data	9,29 MB

Table 9 Dataset of Regions and Regional Units for Greece

Regions and Regional Units for Greece	
Identifier	N/A
Dataset description	This dataset provides information about the agricultural and crop distributions in Greece. It contains information by NUTS2 and NUTS3 levels regarding the total area cultivated in each region differentiating into different agricultural groups of crops, including fallow land, arable land, garden area, vines areas under trees, fallow land, and land eligible for the payment of subsidies.
Purpose of the data	The utilisation of this dataset within AGRICORE is to fill geospatial attributes of the synthetic population not available on FADN data for the Greek use case. Using this dataset allows for a balanced allocation of the not available geospatial information according to real data.
Type of data	Numeric and categorical data.
Form of the data	Dataset is organised in tabular format. Each row contains the information for a single farm and each column is an attribute of the agent.
Format of the data	The complete original report is in .xlsx format.
Origin of the data	The data has been generated by the Greek Statistical Authority.
Dataset stability	The dataset has been maintained in its original format and distribution. No modifications have been required to use the information contained.
Size of data	80 KB

4.2 Produced datasets

Table 10 Synthetic population for the Andalusian use case

Synthetic population Andalusia	
Identifier	AGRICORE_SyntheticPopulationAndalusia_IDE_1
Dataset description	The synthetic population for the use case in Andalusia is a dataset containing all required information to initialise the agents used during the agent-based modelling scenario. In this case, the synthetic population represents the entire

	region of Andalusia and includes all agricultural holdings that are considered professional entities due to their economic size. The dataset has been created from a representative sample of data containing specific agent attributes through the use of a Bayesian Network algorithm. Each row in the dataset represents one holding, and it contains several individual variables indicating the attributes of the synthetic agent. The size of the dataset is defined by several factors: 1) the number of rows is equal to the number of agents to be simulated; 2) the number of variables depends on the product and livestock grouping performed according to the crop and livestock representativeness and grouping rules specific for each use case. Among all the columns present in the dataset, it is possible to differentiate some major sets of variables grouped by themes: a) Economic variables b) Holder social features c) Holding geospatial features d) Crops-related data e) Livestock-related data f) Subsidies related data There are several versions of the synthetic population as each new generation produces a unique result due to the stochastic nature of some generation methods used during the generation process.
Purpose of the data	In a practical scenario, this dataset is used in the AGRICORE project to initialise the agent-based modelling scenario. The dataset provides the simulation orchestrator with the true size and accurate representation of the population to be generated, thus creating a realistic simulation environment that will consequently yield realistic results.
Type of data	Numeric and categorical data.
Form of the data	Dataset is organised in tabular format. Each row contains the information for a single farm and each column is an attribute of the agent.
Format of the data	Synthetic populations are exported as CSV files.
Origin of the data	Synthetic population generation module from AGRICORE tool.
Dataset stability	Once a synthetic population has been created, no further modifications are performed.
Size of data	112 MB

Table 11 Synthetic population for the Italian use case

Synthetic population Italy	
Identifier	AGRICORE_SyntheticPopulationItaly_IDE_1
Dataset description	The synthetic population for the use case in Italy is a dataset containing all required information to initialise the agents used during the agent-based modelling scenario. In this case, the synthetic population represents the entire region of Emilia-Romagna and includes all agricultural holdings that are considered professional entities due to their economic size. The dataset has been created from a representative sample of data containing specific agent attributes through the use of a Bayesian Network algorithm. Each row in the dataset represents one holding, and it contains several individual variables indicating the attributes of the synthetic agent.

	The size of the dataset is defined by several factors: 1) the number of rows is equal to the number of agents to be simulated; 2) the number of variables depends on the product and livestock grouping performed according to the crop and livestock representativeness and grouping rules specific for each use case. Among all the columns present in the dataset, it is possible to differentiate some major sets of variables grouped by themes: a) Economic variables b) Holder social features c) Holding geospatial features d) Crops-related data e) Livestock-related data f) Subsidies related data There are several versions of the synthetic population as each new generation produces a unique result due to the stochastic nature of some generation methods used during the generation process.
Purpose of the data	In a practical scenario, this dataset is used in the AGRICORE project to initialise the agent-based modelling scenario. The dataset provides the simulation orchestrator with the true size and accurate representation of the population to be generated, thus creating a realistic simulation environment that will consequently yield realistic results.
Type of data	Numeric and categorical data.
Form of the data	Dataset is organised in tabular format. Each row contains the information for a single farm and each column is an attribute of the agent.
Format of the data	Synthetic populations are exported as CSV files.
Origin of the data	Synthetic population generation module from AGRICORE tool.
Dataset stability	Once a synthetic population has been created, no further modifications are performed.
Size of data	123 MB

Table 12 Synthetic population for the Greek use case

Synthetic population Greece	
Identifier	AGRICORE_SyntheticPopulationGreece_IDE_3
Dataset description	The synthetic population for the use case in Greece is a dataset containing all required information to initialise the agents used during the agent-based modelling scenario. In this case, the synthetic population represents the entire region of Central Macedonia and includes all agricultural holdings that are considered professional entities due to their economic size. The dataset has been created from a representative sample of data containing specific agent attributes through the use of a Bayesian Network algorithm. Each row in the dataset represents one holding, and it contains several individual variables indicating the attributes of the synthetic agent. The size of the dataset is defined by several factors: 1) the number of rows is equal to the number of agents to be simulated; 2) the number of variables depends on the product and livestock grouping performed according to the crop and livestock representativeness and grouping rules specific for each use case. Among all the columns present in the dataset, it is possible to differentiate some major sets of variables grouped by themes:

	a) Economic variables b) Holder social features c) Holding geospatial features d) Crops-related data e) Livestock-related data f) Subsidies related data There are several versions of the synthetic population as each new generation produces a unique result due to the stochastic nature of some generation methods used during the generation process.
Purpose of the data	In a practical scenario, this dataset is used in the AGRICORE project to initialise the agent-based modelling scenario. The dataset provides the simulation orchestrator with the true size and accurate representation of the population to be generated, thus creating a realistic simulation environment that will consequently yield realistic results.
Type of data	Numeric and categorical data.
Form of the data	Dataset is organised in tabular format. Each row contains the information for a single farm and each column is an attribute of the agent.
Format of the data	Synthetic populations are exported as CSV files.
Origin of the data	Synthetic population generation module from AGRICORE tool.
Dataset stability	Once a synthetic population has been created, no further modifications are performed.
Size of data	79 MB

Table 13 Synthetic population for the Polish use case

Synthetic population Poland	
Identifier	AGRICORE_SyntheticPopulationPoland_IDE_4
Dataset description	The synthetic population for the use case Poland is a dataset containing all required information to initialise the agents used during the agent-based modelling scenario. In this case, the synthetic population represents the entire region of Lubelskie and includes all agricultural holdings that are considered professional entities due to their economic size. The dataset has been created from a representative sample of data containing specific agent attributes through the use of a Bayesian Network algorithm. Each row in the dataset represents one holding, and it contains several individual variables indicating the attributes of the synthetic agent. The size of the dataset is defined by several factors: 1) the number of rows is equal to the number of agents to be simulated; 2) the number of variables depends on the product and livestock grouping performed according to the crop and livestock representativeness and grouping rules specific for each use case. Among all the columns present in the dataset, it is possible to differentiate some major sets of variables grouped by themes: a) Economic variables b) Holder social features c) Holding geospatial features d) Crops-related data e) Livestock-related data f) Subsidies related data

	There are several versions of the synthetic population as each new generation produces a unique result due to the stochastic nature of some generation methods used during the generation process.
Purpose of the data	In a practical scenario, this dataset is used in the AGRICORE project to initialise the agent-based modelling scenario. The dataset provides the simulation orchestrator with the true size and accurate representation of the population to be generated, thus creating a realistic simulation environment that will consequently yield realistic results.
Type of data	Numeric and categorical data.
Form of the data	Dataset is organised in tabular format. Each row contains the information for a single farm and each column is an attribute of the agent.
Format of the data	Synthetic populations are exported as CSV files.
Origin of the data	Synthetic population generation module from AGRICORE tool.
Dataset stability	Once a synthetic population has been created, no further modifications are performed.
Size of data	179 MB

Table 14 Crop representativeness dataset

Crop representativeness dataset	
Identifier	AGRICORE_CropRepresentativeness_IDE_5, AGRICORE_CropRepresentativeness_UNIPR_6, AGRICORE_CropRepresentativeness_IAPAS_7, AGRICORE_CropRepresentativeness_AUTH_8
Dataset description	The crop representativeness dataset contains a detailed analysis of the crops present in a specific use case. The information source of this dataset is the original microdata sample available for each use case coming from FADN, RICA, or RECAN. For each individual crop code represented in the FADN guide, an analysis of specific associated variables is performed. The objective is to orient this analysis towards the main crop indicators to understand the representativeness of each crop in a certain region based on agricultural, productive, structural, crop distribution or economic factors. For all the farms included in the original sample, a weighted extrapolation is performed, thus obtaining the total cultivated area, the total production, the total value of sales associated, the number of holdings that utilise such crops, and the mean number of crops combined with the crop analysed. This information is useful to establish a ranking and compare the crops with higher importance in the use case to those with low relevance. The analysis is performed for both conventional and organic production methods, and values are presented in standardised units: €, Ton, Ha. One single dataset is created for each use case, with a similar format, building rules, and structure.
Purpose of the data	The function that this dataset fulfils within the project is to support the process of product grouping. Due to the large number of variables related to crops available in microdata, it is necessary to simplify the model and alleviate the computational cost of running simulations by reducing the number of variables through crop grouping and creating product groups. These product groups are actually new artificial crops comprising a set of individual crops that share some peculiarity: similarity of cultivation, low representativeness, etc. In turn, for each original variable associated with each individual crop, a homologous

	variable has been created for the product group. These variables contain the sum of the homologous variables for all the crops composing the product group. The process of composing product groups must be sustained on objective criteria. In some cases, using crops of interest or crops associated with specific subsidies is enough, but when treating crops with no special importance or when analysing marginal crops, a deeper understanding of crop distribution should be obtained. This crop representativeness is used in these cases to support crop grouping according to economic, structural or productive factors.
Type of data	Numeric data for crop indicators; text data for crop names and crop descriptions.
Form of the data	Dataset is organised in tabular format. Each row is dedicated to a single crop. Each column is dedicated to one crop indicator.
Format of the data	Crop representativeness analyses are exported as CSV files.
Origin of the data	Microdata from FADN, RICA and RECAN.
Dataset stability	These reports are the outcome of a data analysis. The purpose of the information contained is to be interpretable, and no further data transformations have been required.
Size of data	27KB Andalusia; 20KB Italy; 23KB Greece; 22KB Poland;

Table 15 Product groups dataset

Product groups dataset	
Identifier	AGRICORE_ProductGroups_CAAND_9, AGRICORE_ProductGroups_UNIPR_10, AGRICORE_ProductGroups_UTP_11, AGRICORE_ProductGroups_AUTH_12
Dataset description	This dataset contains information relative to the product groups generated. For each product group generated in the use case encompassing both crops and livestock, the product groups file contains information about the features of the synthetic crop. It is possible to characterise a product group according to the following characteristics: Arable, Cereal, Perennial, Livestock food, Meadows and pastures, Milk, Other and Fixing nitrogen. Given the features of the original individual crops, product groups are tagged with the above-mentioned categories in alignment with these characteristics. Additionally, there is another information field indicating whether the product group is in a conventional or organic production regime. There is a unique product groups dataset for each use case, and their content depends on the product groups created, which in turn depends on the presence and representativeness of individual crops in the use case microdata.
Purpose of the data	This dataset is used by the simulation engine to get information about the characterisation of the product groups defined and to evaluate the cultivation possibilities of each product group during the simulation runtime.
Type of data	Text for all variables included in the dataset.
Form of the data	Dataset is organised in tabular format. Each row is dedicated to a single product group. Each column is dedicated to one variable.
Format of the data	These datasets are exported as CSV files.
Origin of the data	Product groups are defined for each use case.

Dataset stability	These datasets are immutable once they are defined. Simulation engine loads the information contained, but no data modifications are required.
Size of data	736 B Andalusia; 810 B Italy; 642 B Greece; 608 B Poland;

Table 16 Product mapping dataset

Product Mapping dataset	
Identifier	AGRICORE_ProductMapping_IDE_13, AGRICORE_ProductMapping_IDE_14, AGRICORE_ProductMapping_IDE_15, AGRICORE_ProductMapping_IDE_16
Dataset description	This dataset contains the aggregation rules for each product group. The information contained links to individual FADN crops codes with the specific product group associated with each use case. The information presented includes the FADN crop code, its description, the product group associated, and the description of the product group. The aggregation rules are defined both for conventional and organic crops. There is one Product Mapping file for each use case.
Purpose of the data	It is a configuration file. The information contained in this dataset is used to perform the aggregation task during the microdata preprocessing. The synthetic population generator module aggregates variables to form product groups based on the established links between individual FADN crops and their respective product groups.
Type of data	Text for all variables included in the dataset.
Form of the data	Dataset is organised in tabular format. Each row is dedicated to a single FADN crop. Each column is dedicated to one variable.
Format of the data	These datasets are exported as CSV files.
Origin of the data	Product groups are defined for each use case.
Dataset stability	These datasets are immutable once they are defined. Simulation engine loads the information contained, but no data modifications are required.
Size of data	16 KB Andalusia; 30 KB Italy; 18 KB Greece; 18 KB Poland;

Table 17 Subsidies dataset

Subsidies dataset	
Identifier	AGRICORE_Subsidies_CAAND_17, AGRICORE_Subsidies_UNIPR_18, AGRICORE_Subsidies_AUTH_19, AGRICORE_Subsidies_UTP_20
Dataset description	This dataset contains, for each use case, all the information about subsidies required by the simulation engine. The dataset contains several variables, including subsidy Description, subsidy FADN code, if the subsidy is Coupled, the Associated product group (in case it is coupled), the Economic compensation of the subsidy in euros per hectare, the First year in which the subsidy figures in data and the Last year the subsidy was in force. For the case of decoupled subsidies, the Economic compensation was obtained through a data analysis. The value set is the average value registered in the dataset. In the case of coupled subsidies, links between coupled subsidies and product groups have been found. Firstly, identifying the FADN crop codes coupled with the subsidy. Secondly, computing the economic compensation due to this adherence to such subsidy. Finally, computing the weighted Economic compensation according to the crop representativeness of each product group.

	There is one subsidies dataset for each use case. Both organic and conventional product groups have been considered.
Purpose of the data	The subsidies dataset is used to describe each subsidy and set the related configuration information.
Type of data	Numerical and text information
Form of the data	Dataset is organised in tabular format. Each row is dedicated to a single subsidy-product group coupled pair. Each column is dedicated to one variable.
Format of the data	These datasets are exported as CSV files.
Origin of the data	The origin of the data is a data analysis of product groups defined and microdata for each use case.
Dataset stability	These datasets are immutable once they are defined. The simulation engine loads the information contained within, but no data modifications are required.
Size of data	2,2 KB Andalusia; 1,5 KB Italy; 1,5 KB Greece; 832 B Poland;

Table 18 Crops cost dataset

Crops cost dataset	
Identifier	AGRICORE_CropsCost_CAAND_21, AGRICORE_CropsCost_AUTH_22, AGRICORE_CropsCost_UTP_23
Dataset description	This dataset contains the production costs associated with each product group in euros per hectare. For Andalusia, Greece and Poland use cases, this has been generated from microdata samples. Average production costs we obtained by solving an optimisation problem in which total production costs and total production by crop were introduced. The data obtained was computed after product group obtention.
Purpose of the data	Crops cost data was used to fill some agent attributes not available in the use cases microdata files.
Type of data	Numerical and text information
Form of the data	Dataset is organised in tabular format. Each row is dedicated to a single product group. Each column is dedicated to one variable.
Format of the data	These datasets are exported as CSV files.
Origin of the data	The origin of the data is a data analysis of product groups defined and microdata for each use case.
Dataset stability	These datasets are immutable once they are defined. The synthetic population generator module loads crop cost information to fill crop cost parameters on the agent attributes.
Size of data	332 B Andalusia; 328 B Greece; 328 B Poland;

5 FAIR Data

This section aims to present how the AGRICORE consortium should save the information in order to make the data Findable, Accessible, Interoperable and Reusable (FAIR).

5.1 File formats

The file format to be employed is a primary factor to generate FAIR data. The file format should be accessible in the future so that the selected formats should be non-proprietary, open, with documented standards, in common usage by the research community, using standard character encodings and uncompressed.

The following table includes the recommended and acceptable formats that the AGRICORE consortium adopts following the UK Data Service guidance on recommended formats.

Table 19 Acceptable file format

Type of data	Recommended formats	Acceptable formats
Tabular data with extensive metadata (Variable labels, code labels, and defined missing values)	• SPSS portable format (.por) • Delimited test and command ('setup') file (SPSS, Stata, SAS, etc.) • Structured text or mark-up file of metadata information, e.g. DDI XML file	• Proprietary formats of statistical packages: SPSS (.sav), Stata (.dta), MS Access (.mdb/accdb) • Matlab formatted datafiles (.MAT), • R data files (.rdata or .rda; R workspace) • GAMS datafiles GAMS Data eXchange (GDX), • GIS datafiles Esri Shapefile .SHP, .DBF, .SHX, JSON-sat • NLOGIT/LIMDEP data files .lpj .sav • SAS datafiles (.sas) • ODBC datasource • GRETL datafiles • GAUSS datafiles • Eviews datafiles.
Tabular data with minimal metadata (column headings, variable names)	• Comma-separated values (.CSV) • Tab-delimited file (.tab) • Delimited text with SQL data definition statements	• Delimited text (.txt) with characters not present in data used as delimiters • Widely-used formats: MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf), OpenDocument Spreadsheet (.ods)
Textual data	• Rich Text Format (.rtf) • Plain text, ASCII (.txt) • eXtensible Mark-up Language (.xml) text according to an appropriate Document Type Definition (DTD) or schema	• Hypertext Mark-up Language (.html) • Widely-used formats: MS Word (.doc/.docx) • Some software-specific formats: NUD*IST, NVivo and ATLAS.ti

Documentation and scripts	• Rich Text Format (.rtf) • PDF/UA, PDF/A or PDF (.pdf) • XHTML or HTML (.xhtml, .htm) • OpenDocument Text (.odt)	• Plain text (.txt) • Widely-used formats: MS Word (.doc/.docx), MS Excel (.xls/.xlsx) • XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHMTL 1.0
Audio data	• Free Lossless Audio Codec (FLAC) (.flac)	• MPEG-1 Audio Layer 3 (.mp3) if original created in this format • Audio Interchange File Format (.aif) • Waveform Audio Format (.wav)
Image data	• TIFF 6.0 uncompressed (.tif)	• TIFF other versions (.tif, .tiff) • RAW image format (.raw) • Photoshop files (.psd) • BMP (.bmp) • PNG (.png) • Adobe Portable Document Format (PDF/A, PDF) (.pdf) • JPEG (.jpeg, .jpg, .jp2) if original created in this format • GIF (.gif)
Video data	• MPEG-4 (.mp4) • OGG video (.ogv, .ogg) • Motion JPEG 2000 (.mj2)	• AVCHD video (.avchd)

Tabular data and Textual data have been commonly handled in implementing the activities planned in the AGRICORE project. Hence, the next recommendations have been followed by the consortium be make datasets easier to understand and export:

- Do not put more than one table on a worksheet.
- Create charts on new sheets. Do not embed them in the worksheet with the data.
- Include a header row with a clear title for each column.
- Upon manipulating Excel worksheets, be careful when deleting the content of row or column cells. The simple deletion of the content of the cells may yield missing values when importing the data in other software packages (i.e., Stata) using selected procedures. Please take care of deleting the whole rows and/or columns.
- Be mindful of the possibility that data are not written following the international notation employing the comma (,) as the separator of thousands and the dot (.) as the separator of decimals.
- Relying on a data conversion programme like Stat/Transfer (https://stattransfer.com/) may facilitate data management.

5.2 Metadata: Data document

Clear and detailed documentation is essential for data to be understood, interpreted and used. The data document describes the content, formats and internal relationships of the data.

The following template describes the documentation required within the project to document the datasets used and generated. This information will be indexed in the Confluence platform. If, during the elaboration of the data formats, any field is not determined yet, the label TBD should be included. In case any field is provisional, the * should be included at the beginning of the text.

At this stage of the project, it can be confirmed that most of the information was gathered from public sources for their characterisation, which are generally available online. However FADN metadata were essential to generate synthetic populations, which are non-public datasets. Those public datasets were characterised and fed the ARDIT tool under the framework of Task 1.8. To this end, a "Master Table" was provided in the first version of the DMP (D9.2) and then was updated to include more relevant data for the characterisation (see D1.3, D1.4, D1.5 and D1.6). Even if the consortium is aware of the heterogeneity of the existing data provided by different sources, the consortium has made its best effort to gather the following information to ensure these domains are covered

- Dataset information
- Content description
- Technical description
- Access

Dataset information	
Name	Name of the dataset
Data contained	Describing the data collected and the agroupation of data
Dataset structure	Detailing tables structures and main indications to help understanding the reading of the dataset
Dataset format	How can this data be available: format to be used
Data source access	How can this data can be accessed: Web, download, individual access...
Source	Source of the dataset
Generation process	Process followed to gather/produce the dataset
Author	Author of the data
Maintainer	Institution/company in charge of maintaining
Last actualization	Date
Update frequency	Period of time
Periods covered	Time when the dataset started to gather information
Release date (past and expected future ones)	Figure detailed in days
Additional information	Relevant information to be included for a better understanding of the data source
Rights	Any known IPR, licences or restrictions on the use of the data (if applicable)
Access information	Where and how to access the data (if applicable). Publicly available/Access request required
Embargo	Description of the embargo that will be held on the research data (i.e., previous publication on paid-for journals and later published openly) and the period that this embargo will entail (If applicable).

Publication information	Sharing policies among the following options: Green Open Access; Gold Open Access; Non-shared; Published in paid-for journals; Other (Describe)

In addition, the next template will also be filled in by each partner characterising an existing or produced dataset for the purpose of promoting FAIR principles:

Dataset identifier
What naming conventions do you follow?
Will keywords for searching the dataset identifier be provided to optimize the opportunity for re-use?
What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how. (If applicable)
Will this dataset generated and/or used in the project be made openly available as the default? If positive, How?
What methods or software tools are needed to access the data?
If there are restrictions on use, how will access be provided?
What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?
How will the data be licensed to allow for the widest re-use possible?
When will the data be made available for reuse?
Are the data produced and/or used in the project useable by third parties, in particular after the end of the project?
How long is it intended that the data remains reusable?

5.3 Data sharing: Publication

The AGRICORE consortium fully embraces the H2020 requirement for Open Access publishing, following the guidelines presented by the EC. The project has ensured '**green**' and '**gold**' publishing, i.e., self-archiving in one or more repositories. The associated costs of publications will be claimed as part of the H2020 grant.

The project will make its datasets of results publicly available through the following repositories:

- The project website: https://agricore-project.eu/
- The European data sources Index Module developed in Task 1.8
- The project public repository in Zenodo
- Public software repositories such as GitLab and GitHub
- The dissemination of the software tools is not limited to the distribution of the source code but also includes an extensive plan for promoting the adoption of the proposed technologies by other researchers and developers. This has been done in the frame of the planned dissemination and communication activities and especially during the clustering and workshop activities.

In any case, the partners have not published their results when the publication of the research data violates any confidential restriction or IPR of any of the partners. For instance, those results derived form the FADN microdata.

6 Allocation of resources

6.1 Responsibility and roles

IDE is the project coordinator and is in charge of the management of the Confluence site.

Ayesa Advanced Technologies (AAT) is responsible for maintaining and updating the DMP. Therefore, AAT has been in charge of continuously updating the DMP to produce a final version by the end of the project by providing clear guidelines on the management of personal data and compliance with the General Data Protection Regulation (GDPR). The DPO of the project provided by AAT has an email channel to efficiently manage everything related to data for the duration of the project. The email to which all concerns about data management can be addressed is dpo_agricore@ayesa.com

All partners are responsible for updating and contributing to the Confluence site of the project. Partners have been also requested to deliver periodically:

- Pre-print manuscripts of any accepted publication
- Slides and posters shown at conferences
- Raw data supporting the figures is papers and deliverables
- PhD dissertations generated in the frame of the project

In case new personnel are assigned to a relevant role, responsibilities with respect to the DMP are also taken over. For details on the management roles and structure of AGRICORE see D10.1. In case the contact person for a certain dataset is leaving the project, the institution of the original contact person will take over the responsibility and will assign a new contact person.

7 Ethical aspects

In this section, the ethical aspects and legal compliances about DMP are presented. None of the data generated during the execution of the AGRICORE project contain information on individuals or companies, so every data was anonymised. The technical data does not represent or use any human being or corporate entity. Moreover, the AGRICORE consortium has done its best to adapt data management to the European GDPR.

With respect to the IPR, the whole project is using open-access data and ensuring data generated will be released as open access.

Only when the achievement of the main objective of a partner is jeopardised by making specific parts of the research data openly accessible, the consortium would have evaluate the reasons for not giving access in order to protect the partner's commercial activity. However, this case has not occurred.

7.1 Data, samples and equipment exchange between EU and non-EU parties

During the project, the consortium has kept an account of the scientific samples and data exchanged with non-EU parties, which would be stored on the Confluence platform on the page "Data exchanged between EU and non-EU members". This would be available to the EC services at any time. The potential exchange between non-EU parties would be closely evaluated before proceeding, requesting permission from the concerned entities. Nonetheless, private data (not already publicly available) would not be shared with non-EU partners unless strictly necessary and always after receiving the required authorisations. This has not been occurred because any data has exchanged with a non-EU party.

All the ethics issues will be developed in WP11 and addressed and reported in D11.1, D11.2 and D11.3.

8 Other issues for data management

At the end of the project, the AGRICORE project does not make use of other information requiring a special treatment different than the one already explained within this document, and no other issues have been found. If any other concerns are raised, the DPO of the project will provide the best recommendation to address the situation.

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

Deliverable Number	Deliverable Title	Lead beneficiary	Type	Dissemination Level	Due date
D9.2	Data management Plan	AAT	Report	Public	M6
D10.1	Project management handbook	IDE	Report	Confidential	M1