D1.7 Systematic approach for the Identification and filling of information gaps through participatory research actions

agricore

Deliverable Number	D1.7
Lead Beneficiary	AXIA
Authors	AUTH, IDE, IAPAS, UTP, CAAND
Work package	WP1
Delivery Date	M35
Dissemination Level	Public

www.agricore-project.eu

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

Document Information

Project title	Agent-based support tool for the development of agriculture policies
Project acronym	AGRICORE
Project call	H2020-RUR-04-2018-2019
Grant number	816078
Project duration	1.09.2019-31.8.2023 (48 months)
Deliverable Authors	IDENER Team, AXIA Team, CAAND Team
Deliverable Reviewers	IDENER Team, UNIPR Team, AUTH Team

Version History

Version	Description	Organization	Date
0.1	Deliverable template proposal	AXIA, CAAND, IDE	12/04/2022
0.2	Final ToC and deliverable structure	AXIA, CAAND, IDE	15/04/2022
0.3	First Complete draft version	AXIA, CAAND, IDE	07/06/2022
0.4	Revision and comments	UNIPR, AUTH	02/07/2022
0.5	Implementation of changes and corrections	AXIA, CAAND, IDE	18/07/2022
0.6	Language review	IDE	22/07/2022
0.7	Final exportation and formatting	IDE	23/07/2022
1.0	Deliverable completed	AXIA, CAAND, IDE	30/07/2022

Disclaimer

All the contributors to this deliverable declare that they:

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

Executive Summary

AGRICORE is a research project funded by the European Commission under the RUR-04-2018 call (part of the H2020 programme). The project proposes an innovative approach to apply agent-based modelling to increase the capacity of policymakers to evaluate the impact of agricultural measures within and outside the framework of the Common Agricultural Policy (CAP). The AGRICORE suite stands out for being highly modular and customisable. Thanks to its open-source nature AGRICORE can be applied to a multitude of use cases and easily upgraded as future needs arise.

The assessment of agricultural policy impact is made through an agent-based simulation, which replicates the characteristics, distribution, and interactions of the real populations of interest. To define the simulation environment, it is necessary to generate a synthetic population of agents which interact between each other and with external modules. This requires the use case designer to provide a set of data inputs, such as, the agent's attributes of interest, the aggregated distributions of the real population and other inputs for the external modules. These inputs are obtained through the data sources characterised in ARDIT (a central storage point for locating useful agricultural datasets). However, when designing future use cases, data sources in ARDIT that fit the use case framework might not exist for some input(s), and subsequently will need to be considered as information gaps.

Such information gaps can be filled using information collected via different activities, including depth search for information/data sources, contact with stakeholders and ad-hoc participatory research activities. The design and execution of these activities strictly depends on the particularity of each use case. This deliverable describes the methods and resources employed in developing the AGRICORE pilot use cases. Future users will therefore have all the relevant information to guide them through the process.

Finally, in addition to providing design considerations, some recommendations regarding scheduling and execution of the ad-hoc participatory research activities are also included. The recommendations are based on feedback received from various stakeholders, the experience gained during the design and execution of participatory research in the pilot use cases and the first participatory research results.

Abbreviations

Abbreviation	Full name
ABS	Agent-based simulation
ANCs	Areas facing natural or specific constraints explained
AoI	Attribute of interest
ARDIT	Agricultural Research Data Index Tool
ARMA	The Agency for Restructuring and Modernisation of Agriculture (in Poland)
CAP	Common Agricultural Policy
DEM	Big Data Extraction
DFM	Big Data Fusion
DWH	Data warehouse
EEEA	Farm Structure Survey
ESCYRE	Spanish Survey on crop surfaces and yields
FEADER	European Agricultural Fund for Rural Development
FEAGA	European Agricultural Guarantee Fund
GDP	Gross Domestic Product
GHG	A greenhouse gas
GUI	Graphical User Interface
IAM	Impact assessment module
INE	National Statistical Institute
LFA	Less Favoured Areas
LSU	The livestock units
NUTS	Nomenclature of Territorial Units for Statistics
NVZs	Nitrate Vulnerable Zones
PR	Participatory Research
RAMSAR	Convention on Wetlands of International Importance, especially as Waterfowl Habitants
RDP	Rural Development Program
SIGPAC	Geographical Information System for Agricultural Plots
SIPEA	Information System on Organic Production in Andalusia
SPG	Synthetic population generation
SW	Software

List of Figures

List of Tables

Table of Contents

1 Introduction

The AGRICORE project will provide a tool for modelling and simulating how public policies affect the agricultural sector at regional, national and EU levels. It will also consider the wide diversity that exists between farms located in different geographical areas and/or dedicated to growing different crops. This will be achieved by implementing an agent-based model and simulation environment populated by synthetic populations (SP) of agricultural holdings replicating the characteristics, distribution, and interactions of the real populations of interest. Within the SP, each agent represents an individual farm (Agricultural Holding) as an autonomous decision-making entity that individually assesses its own context and makes decisions based on its current situation and expectations.

The AGRICORE tool stands out for two main reasons. Firstly, to better replicate the target population the most cutting-edge ICT techniques and methods will be applied to optimise the generally long and tedious parametrisation and calibration phase required by current agent-based tools. Secondly, the AGRICORE tool will be created as a highly modular and customisable suite, and it will be released as an open-source project so future users can transparently update and improve the tool as required.

The efficient parametrisation and calibration process of the model is achieved by making use of multiple information sources, such as, EU agricultural statistics, geo-referenced datasets, other regional/national databases, results from previous projects and stakeholder knowledge. Unfortunately, not all of these data sources are publicly available or do not have the data structure or level of detail needed to parametrise and calibrate the models. For this reason, the design and execution of information gap filling (IGF) activities, including participatory research (PR) actions, are usually required to collect the necessary data for simulating the proposed use cases.

Deliverable 1.8 explained the detection of the information gaps affecting the three Use Cases investigated in the AGRICORE project, and the design of the participatory research activities proposed to populate them. Moreover, the design of these activities requires a previous analysis of the target population (also presented in D1.8) in order to know the most efficient way to collect the required information. Among the proposed activities, conducting survey campaigns and additional contact with stakeholders could be undertaken. Furthermore, the planning of those activities was presented in D7.1 which included planning and scheduling the use cases with special emphasis on the survey campaigns, the method for identifying and contacting main stakeholders and a monitoring plan. There is a close relationship between WP1 and WP7, thereby, resulting in a fluid exchange of information, coordination and joint work between the partners involved in both work packages.

The current deliverable presents a systematic approach for the identification and filling of information gaps through participatory research actions, which is generically applicable to any additional use case. It has been elaborated based on the feedback from several stakeholders, the experience gained during the design and execution of participatory research in the agricultural sector and the PR results obtained from the first AGRICORE Use Cases. The purpose of this deliverable is to provide any person or institution with the necessary tools and knowledge to create a new use case with the AGRICORE suite. Thanks to this systematic approach and the highly modular and customisable IT structure, future users will be able to exploit the full potential of the tool by upgrading and adapting it to future use cases.

2 Construction of an AGRICORE use case

As explained in the introduction, the AGRICORE suite is primarily based on an agent-based model that replicates the interactions between farmers. In this section, the AGRICORE suite is described in more detail and its potential applications are also presented.

Figure 1 provides the functional diagram of the AGRICORE tool, where the most relevant modules and their interconnections are depicted. Firstly, the Synthetic Population Generator (SPG) module is responsible for generating the synthetic population for the agent-based simulation (ABS), which replicates the distribution and characteristics of the population of interest. As such, multiple datasets are used to parametrise and calibrate the model. Each agent is defined by a set of attributes and composed of two dimensions:

- The structural dimension defines the agent as an agricultural enterprise, and it is where the long-term decisions are made.
- The agro-economic dimension defines the farm as an agricultural production unit, and it is where short-term decisions are made based on the economic health of the farm.

Both dimensions are influenced by farmer behavioural aspects including risk aversion and willingness to innovate.

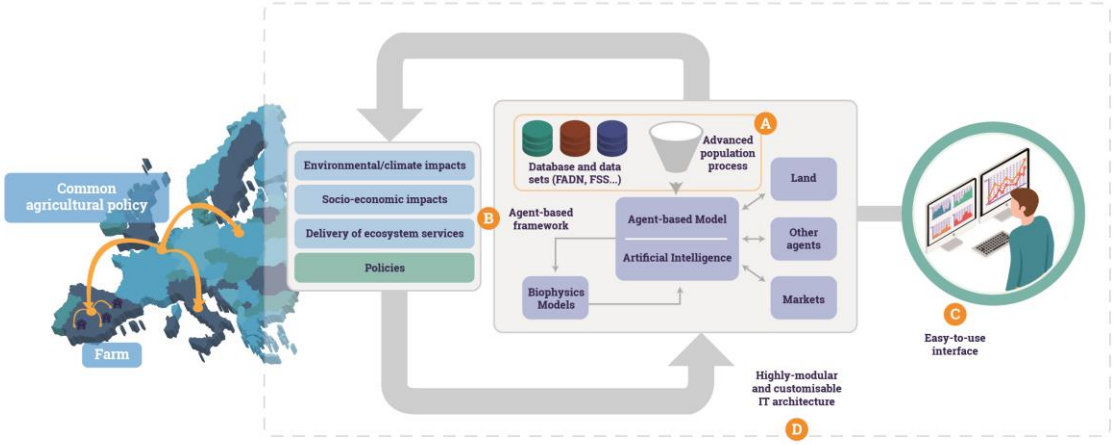

Figure 1 Diagram of the AGRICORE tool.

To provide context for the agents and their interactions there are several modules that determine the environmental simulation conditions. These modules are listed below:

- Agricultural policy module: it translates the agricultural policy(ies) to be analysed and provides the simulation with a legal context, such as, the requirements for an agent to apply for a measure.
- Land market module: this module simulates the market in which agents exchange land.
- Product market module: it simulates the market in which the prices for agricultural inputs and outputs are determined.

- Biophysical model: this model reproduces the evolution of crops and livestock due to the meteorological conditions and the agro-management actions associated with the technology chosen by the farmer.
- Meteorological module: this module replicates the meteorological conditions of a region.

Another important AGRICORE element is the Impact Assessment Module (IAM), which encompasses an environmental IAM, a socio-economic IAM and an ecosystem services IAM. These modules provide a set of KPIs to measure the impact of the simulation against several areas of interest e, such as, environment, ecosystem services and socio-economic factors. These KPIs are the main outputs that the tool will present to the user through a set of visualisation libraries. This will be achieved through an intuitive and easy-to-use graphical user interface (GUI). Therefore, to build a new use case, future users will only need to provide the necessary data sources to build the synthetic population, initialise the different modules, run the simulation and observe the resulting KPIs to evaluate the aspects they are interested in.

The GUI design feature makes the AGRICORE tool accessible to a wide group of users and entities who are not necessarily technically minded, such as, politicians, regulatory bodies, and agri-food cooperatives. Furthermore, thanks to the fact that it is a modular and open-source tool, it is highly adaptable and can be used for multiple purposes. The main purpose being the evaluation by policymakers and researchers of the impact of agricultural measures at regional, national, and European levels. The use of the AGRICORE tool should allow these stakeholders to predict the outcomes of the policy instruments being formulated and to customise these instruments until the simulated outcomes meet the set policy objectives.

The results of agricultural policy simulations could also be used by other policymakers to design public policies in sectors closely related to agriculture, such as, food consumption, transport, industry, and the environment. In addition, the use of the tool could also be extended to other private institutions (e.g., financial institutions) that are interested in designing support or assistance packages for the agricultural sector and who wish to maximise the expected outreach and effectiveness of such initiatives.

A further use of the AGRICORE tool is to predict the diffusion of certain types of farming, crops, and livestock. As a result, agri-food cooperatives and consultants might be interested in using this tool to adapt their services and facilities. Furthermore, the scientific community is another potential user of this tool. Among its research uses are: i) studying the impact on the landscape, ecosystem, and certain species; ii) predicting plagues and diseases; and iii) even testing new models that improve the predictions or better replicate real conditions. Finally, it should be noted that throughout the project, stakeholders from all the aforementioned profiles, have been contacted to receive their suggestions so their knowledge could contribute to the development of the AGRICORE suite.

3 ARDIT & Data warehouse

In addition to the previously explained agent-based model the Agricultural Research Data Index Tool (ARDIT) and the Data Warehouse /DWH) are two essential software developments in the AGRICORE suite due to their close relationship with the ABS setting and execution. Both software elements are also strictly linked and functional to the SPG module.

ARDIT is a platform that catalogues agricultural data sources, providing users easy access to data sources of interest. Through this tool, the user can access the ARDIT catalogue of datasets to check their characteristics such as spatial and temporal scope, resolution, aggregation level, last available update, etc. The tool distinguishes between socio-economic and geo-referenced databases and includes datasets at the European, national, and regional levels. In addition, the user can access the characterised datasets (if public) to retrieve the data and he/she can also characterise new data sources in ARDIT. For this purpose, it has a user guide that covers both the characterisation process and the definition of each of the characteristics associated with the data source. ARDIT is composed of two versions: the Global Indexer, which is focused on creating a fully public and accessible portal of datasets related to agriculture, where users will be able to perform searches and filter datasets based on multiple attributes and properties; and the Local Indexer, which provides the tools necessary for storing new datasets in the user's local DWH and allows the user to manage a personal catalogue of datasets. In relation to the AGRICORE project, approximately 300 datasets have been characterised in ARDIT.

The Data Warehouse is the tool responsible for the storage, extraction, transformation and loading of data within the AGRICORE project. The relationship of this module with the SPG module comes from the fact that the DWH is where the data necessary to generate the synthetic population is stored. There are two modules that perform this connection: the Data Extraction Module (DEM), which is responsible for searching datasets and extracting the necessary data, and the Data Fusion Module (DFM), which searches for correlations between the data to obtain probability distributions required to assign values to synthetic agents.

The importance of ARDIT and DWH lies in their connection to the SPG procedures, as the latter identifies the required data using ARDIT, and once identified extracts it from the DWH to start the synthetic population generation. Thanks to the ARDIT API, parallel data source searches can be performed based on the definition of the spatial and temporal scope of the use case, which will be configured by the user through the AGRICORE GUI.

Once the user chooses the datasets to be used, the DEM is able to search the dataset and extract the data and the DFM will extract the probability functions to pass them to the SPG module, which generates the synthetic population needed for a specific use case.

4 Initial audit of information required for the execution of an AGRICORE use case

The most difficult task for the user is the initialisation and parameterisation of the tool with a population of agents. Therefore, this section and the following section show the user the steps required to build an AGRICORE use case. Section 4 presents the information required to create an AGRICORE use case, while Section 5 explains how to fill information gaps a priori when specific required information is not available.

All the inputs required by the tool are established a priori and all must be initialised, even if they are not essential for the analysis to be simulated by the use case. Initialisation can even be performed with default values such as estimates or dummy values, but in this case, the performance of the model could be compromised, leading to less accurate results. However, due to the possibility of having no data for certain module inputs, Section 5 describes the methods used to resolve information gaps in the AGRICORE pilot use cases. This will provide future users with a set of guidelines and recommendations regarding the treatment of information gaps.

As mentioned previously, thanks to the modularity and the open-source nature of the tool, it is possible to change its inputs and outputs for adaptation to future use cases. Focusing on the main purpose of the tool, which is to analyse the impact of agricultural policies, Figure 2 provides an activity UML diagram which illustrates the process to design a use case and to understand the necessary information to carry out a simulation.

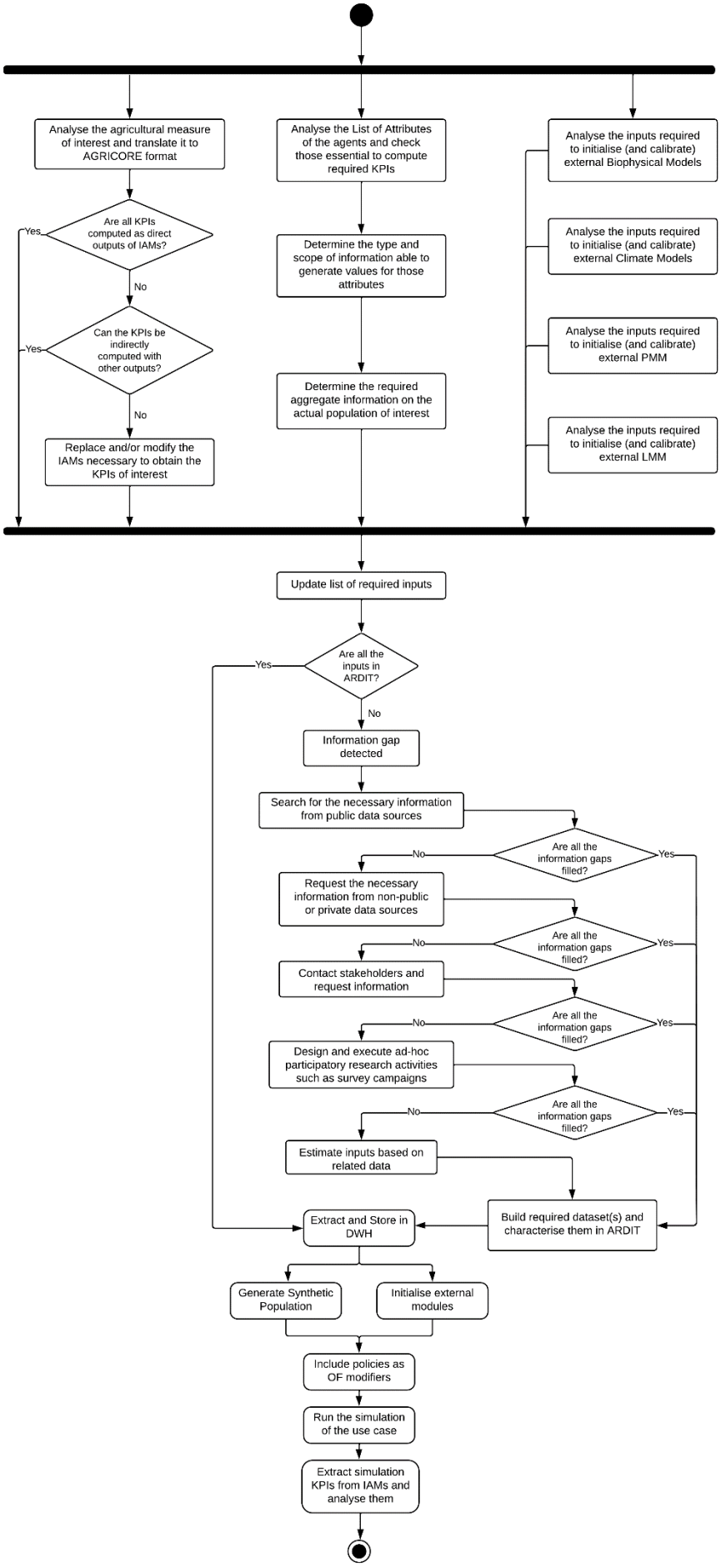

Figure 2 General process to determine the required information to create an AGRICORE use case.

The AGRICORE tool must be provided with three main types of data inputs. They require a set of prior analyses to determine what information is needed, which is represented by the three vertical branches between the horizontal bars in the UML diagram (Figure 2). These types of inputs are:

1. Agent Attributes of interest (AoI): they characterise the elements (SW objects) that make up the agent and the context that affects its actions. Those defined for the AGRICORE project are gathered in Table 1 (over-page). As observed, they are classified according to the element of the Agricultural Holding (AH) that they are related to, such as, farm holding structure, farm owner, farm manager, parcel, crop, livestock, products, economic-financial module and ecosystem. Moreover, inside each of these elements (SW objects), the attributes are divided into:
    a. Parameters: these attributes are fixed or change over the long term.
    b. States: they change according to the environment and decisions of the farmer during the simulation.
    c. Agro-management decisions: these attributes are directly determined by the farmer's actions.
    d. Disturbances: these attributes depend on external conditions that the agent cannot control.
    e. Outputs: they are the results from the simulation of an agent's actions subject to its states and external disturbances.
2. Aggregate data: this data corresponds to the total number of exploitations, area and production grouped by NUTS 2 or NUTS 3 regions. This data must be provided for the crop(s) or livestock(s) of interest according to the policy measure, but also for the most representative crops and livestock of the region(s) to be analysed.
3. Inputs from external modules: this is referred to all the necessary data to initialise the external AGRICORE tool modules.

Table 1 Attributes of the objects which define each agent (Source: own work).

Objects	Agricultural Holding Structure	Agricultural Holding Owner(s)	Agricultural Holding Manager	Crop enterprises	Livestock enterprises	Output Products	Economic financial module	Ecosystem
Parameters	Number of owners, probability of generational renewal, geographic location (coordinates, NUTS3, further granularity).	Gender, Grade of innovativeness, Risk aversion level.	Gender, Grade of innovativeness, Risk aversion level, education level.	Type, regional cultivation standards, average regional yield.	Type, regional breeding standards.	Type	(re)investment propensity, size synergies, rate of interest, tax rates, WD_{min}	Soil properties: number of layers, layer thickness, max. bulk density, clay, sand, silt, organic carbon Soil types. Aquifers' quality
States	Economic size (FADN), Type of Farming (FADN), land structure (total area, parcels), livestock units, machinery capacity, regular workforce.	Age, probability of generational renewal	-	Status of permanent crops (age and health).	Livestock herd status (species and ages).	Stock levels	Assets, liabilities, equity. Solvency, Liquidity and Profitability indicators.	Current classification of ecosystems that make the Agricultural Holdings located in them potential recipients of AES. (e.g. nitrates-polluted areas)
Agro management decisions	Land ownership management (buy/sell), available capital. allocation of resources to enterprises, land management (rent/lease), contracted machinery,	-	-	Production Technology (if more than one has been considered in the model)	Production Technology (if more than one has been considered in the model)	Production utilization (sales, farm use, farm consumption, changes in stock)	Investment, loans, withdrawals	-
Disturbances	Land prices, production factors prices, output product prices, public policies.	-	-	Plagues, meteorological conditions	Unexpected deaths, meteorological conditions	Outputs from external agents (and imports)	Taxes, accountancy regulations.	Deviations of abiotic Factors (temperature, rainfall, etc), plagues, patogens, others.
Outputs	Socio-economic impact (labour, rent), environmental impact (land use, emissions, water intake, pollution)	-	-	Actual Yield	Actual Yield	Product Revenue	Cash flow, profit/loss, balance sheet.	Ecosystem services impact

5 Detection and analysis of information gaps

The project began by firstly preparing a list of the required inputs to simulate the use case. A manual search for available information sources was then conducted to help initialise the attributes of the agents, generate the synthetic population and initialise the different modules. Furthermore, this process was undertaken in two search phases where the useful data sources found were characterised using the ARDIT tool. This characterisation task must be completed to use data sources that are not included in ARDIT. Consequently, future users who encounter an information gap will benefit from the work done by previous users who have already filled their information gaps. To characterise a data source, in addition to the dataset of interest, additional metadata which is detailed in the ARDIT User Manual must be specified. Metadata of interest includes

- **Title**: the actual name of the dataset.
- **Description**: a brief description of the dataset, its purpose or the text-based description of data collected.
- **Issued**: date of formal issuance or publication of the specific dataset.
- **Last update**: most recent date on which the dataset was changed, updated or modified.
- **Dataset type**: the dataset can be of 2 types: GEO-REFERENCED, when data are directly related to a geographical point, and SOCIOECONOMIC, for statistical or economic datasets.
- **Producer**: the institution which publishes or maintains the dataset.
- **Link**: the hyperlink of the referenced dataset.
- **Language**: the original language in which the data and metadata are available. In the case of a national dataset in its local language, the characterisation will be performed in English, and the original name of the table will be reported between brackets, next to the original title.
- **Periodicity of publication**: the frequency with which each data is updated. It can be chosen between different periods (i.e., annual, monthly, biennial, etc.), but it also includes the item "irregular", which is used when there is no regularity in the publication of data.
- **Catalogue**: this field depicts the catalogue to which the dataset belongs, such as EUROSTAT, LUCAS, FADN, ETC.
- **Spatial resolution** (only for geo-referenced datasets): minimum spatial separation resolvable in a dataset, measured in metres.
- **Temporal resolution**: it is intended to provide a summary indication of the temporal resolution of the data distribution as a single value.
- **Resource type**: this field depicts the nature or genre of the resource, such as dataset, collection, image, text, etc.
- **Temporal extent**: it indicates the period to which the specific data refers.
- **Geographical coverage**: here, the geographical scope of the data is indicated by listing the applicable spatial regions from continents to NUTS 3 regions. Only one level of spatial coverage is needed. For example, if the user indicates 'Europe' as continental coverage, there is no need to report each European country; the only important thing is to make sure that the dataset contains data for all the countries in that continent (as per regions of each country).
- **Access rights**: a declaration of the rights that concerns how the distribution is accessed: publicly available or access request required.
- **Format**: the file format of the distribution. xls, csv, html, pc axis, spss, tsv, (pdf if it is the only format available).

Thanks to ARDIT the detection of information gaps is greatly simplified as it improves the search for multi-origin information sources. When the user starts to define the use case, they must specify their geographic and temporal scope, amongst other things. Based on the parameters initially defined, a query to ARDIT is launched which then provides the relevant datasets. If no compatible dataset is found for any of the model inputs, the user will be notified via the interface. In such instances, the user might consider these inputs as information gaps and should start searching for new data sources to populate them. The filling methods and resources employed in the development of the AGRICORE project, including commonly detected issues and mitigation strategies, are described in Sections 6 to 8. These sections correspond to the main three IGF activities to be carried out and are presented in increasing order based on their resource requirements. The implementation of these activities needs to be suitably customised according to the particularities of the use case.

Another type of gap corresponds to the different module outputs when they do not completely cover the KPIs required to assess the impact of the analysed agricultural policy. This gap is not due to a lack of information but to the specificity of the policy instruments, which is outside the scope of the AGRICORE model and modules. This issue should be detected in the model set-up step which lists the required information to build the use case. Thanks to the modularity and open-source nature of the tool, the user could improve the implementation of any module or directly substitute it with another external module which is able to compute the missing KPIs.

6 Further search for information/data sources

This section addresses the activity of autonomous data search through publicly available data sources. This activity includes different search types of varying complexity, which are explained in conjunction with the resources and methods used to construct the AGRICORE pilot use cases.

When an information gap arises, the first step involves checking the available information stored in ARDIT. Although it may seem contradictory, ARDIT has useful information that can search for required external data sources. For this reason, a quick search in the datasets catalogue should be performed to find any related datasets. The ARDIT catalogue may include datasets that only partially apply to the use case, with some characteristics that are not relevant to its scope. For example, there will be instances of suitable datasets that do not coincide with the temporal extent of the use case scope. The same situation could be encountered with the crop, livestock or geographical coverage. Therefore, the characterisation of those related datasets, mainly their links and their producers, could be further investigated to look for the missing data. By following this approach, reliable and verified data sources are used, and the characterisation process is simplified as existing characterisations can be reused.

If this search method is unsuccessful, another option involves contacting the institution in charge of generating and/or maintaining the data. It might be that the information exists but has not been made public. This point surrounds contacting stakeholders and is addressed in the following section.

When the required data cannot be found using common data sources, the next step would be to search from scratch. and the following points must be considered:

- Data should come from official sources. These usually have a publicly available methodology that allows for understanding the tabulation of the data and its limitations.
- Precision required. Before starting the search, the required level of data resolution that can be accepted should be considered, as official data can be found based on the farm census or based on surveys. Typically, when dealing with information gaps on aggregated data or data that are not relevant to the measure to be analysed, survey-based data are sufficient.
- The format of the data. Data should preferably be tabulated in CSV or Excel format to facilitate uploading into the model from the DWH.
- The importance of the crop/livestock in the population of interest. If the search is for data referenced to a type of crop or livestock, its role in the study region should be considered. In other words, if it is a marginal crop or a recently introduced crop (less than 5 years old), it is likely that no specific data will be available.
- Methodology. The methodology with which the data has been collected is fundamental. It may be that, for different dataset versions, the same producer has followed different data collection methods, has collected different data, or has changed the way they present the data over time. This also includes other considerations, such as, when data collection has been paused and resumed, or if there has been a considerable delay between data collection and data publication (e.g., the Covid-19 pandemic situation).
- Geographical coverage. Not all data sources have the same scope in terms of geographical coverage. Therefore, it should be noted that data with NUTS 3 resolution are often collected and published by institutions at the corresponding regional level. It is common that national institutions at a higher level also contain such data, while it is less common for supranational institutions even though there are some cases. For example, European institutions such as EUROSTAT usually produce data at the European or country level but not at the NUTS 3 regional level.

Apart from public administration institutions, there are many entities that produce data which can be assigned, such as, universities, research institutions, NGOs, private institutions, and associations. To help guide future AGRICORE tool users fill information gaps, a few institutions which are usually responsible for producing and publishing data are listed below. It should be noted that these institutions are named generically, and their nomenclature in each country or region may vary.

- (Regional) Ministry of Agriculture, Livestock, Fishing and Rural Development
- (Regional) Ministry of Energy
- (Regional) Ministry of Industry
- National/Regional Institute of Statistics
- Geographical Information System (GIS) in agriculture
- Agricultural Guarantee Fund
- Annual Agricultural Statistical Survey
- National Aeronautics and Space Administration (NASA)
- European bodies: BISE, CORINE, CRU TS, ECA&D, Agri4cast, LUCAS, SoilGrids, OECD, EUROSTAT, FaoStat
- Results of European Research Projects

Finally, public reports and studies can also be used as a resource in the search for data sources under this IGF activity. However, these documents usually do not present a tabular data structure, and if they do, they are usually in plain-text formats making it difficult to incorporate them into the DWH. Moreover, this search is often more time-consuming as it involves reviewing documents. The data presented are often survey-based or extracted from data sources already analysed so they may not be overly useful. From experience, the most useful reports are those produced by public bodies, although smaller organisations such as associations and cooperatives also often publish reports. It should be noted though that their data can be biased as it is extracted directly from their associates or cooperative members. In conclusion, it is rare to find a report that has been prepared with exactly the same requirements and constraints of a particular use case.

7 Liaison with stakeholders

Stakeholder liaison is an essential activity in the development of a use case and should start right from the beginning. This section explains how AGRICORE addressed this activity and which points must be considered by future users to achieve a mutually beneficial collaboration. The first step in planning this activity is to analyse the targeted stakeholders to decide the type of collaboration to be requested and what results can be expected from each collaboration. In AGRICORE the stakeholders have been classified into the following 6 groups:

- **Policymakers**: members of the regional or national government, especially from ministries or departments related to agriculture, who oversee making new policies.
- **Farmers**: this type groups individual farmers, agricultural associations and cooperatives and associations of farmers.
- **Scientific community** (Universities, research institutions, etc.).
- **Consultancy and advisory agencies**: they are companies that offer technical advice in relation to agriculture, such as farm design, changing the type of production, improvements to increase production, etc.
- **NGOs**: they are non-profit organisations interested in the diffusion and consciousness-raising of promoting sustainable and eco-friendly agricultural practices.
- **Clusters**: they are partnerships of projects funded under the same call where it is possible to exchange information.

This IGF activity requires prior knowledge regarding potential stakeholder backgrounds, as its successful development depends more on the quality of the contacts rather than the number of contacts. Considering the speed of establishing contact and forming a relationship nor the usefulness of the contact can be known a priori, it is important to initiate the contact as soon as the Use Case development starts. Additionally, a series of tips to consider during the development of this activity are listed below to help boost and create leverage with the stakeholders identified:

- Organise the stakeholders. It is important to monitor the contact process and automatise it as much as possible. There are two measures which make this process easier. Firstly, create a contact list that can be updated to record contact information and the results obtained. Evidently, the user can organise this information using the format s/he prefers and include additional information if it is considered worthwhile. A table with the following columns was used in AGRICORE:
    - Organisation: name of the institution to which the contact belongs.
    - Type of stakeholder: one of the 6 groups into which the stakeholders have been grouped.
    - Contact: name of the contact, his/her position and the contact information (email, telephone number, address, etc.).
    - Contact status and approach: dates when contact was initiated and how (by email, by phone, in-person, etc.) and the current status of the relationship (already made, in process or still to be made).
    - Collaboration: describe the reason for making contact and the results of the collaboration.
    - Expected impact for the stakeholder: description of aspects of the use case that could be of interest to the stakeholder.

Secondly, the user can prepare a generic yet customisable slide presentation for the use case which also describes the purpose of your contact. The presentation should contain the scope, motivation and expected results of the use case. It can be customised by emphasising the most relevant aspects that may be of interest to each stakeholder. Therefore, the contact process is improved by focusing on gaining the attention and interest of the recipient.

- Schedule the contact process. The contact process may extend for weeks or even months, so the early anticipation of the contact initialization process is a good practice to save time. As mentioned in the previous section, if the required information was not found following the first search for data sources (resources of producers with similar datasets) then it is recommended to contact those producers directly. In parallel, the search process should continue by exploring other data sources in case the contacted producers are unable to supply any relevant information.
- Define the request in detail. The contacts made during AGRICORE have demonstrated that better results are achieved when the request for information is more detailed. Data can be assessed from many different points of view, so it is important to specify what you exactly need and use unambiguous terms. To support your request and ensure that the recipient clearly understands your requirements, a good idea is to reference similar data, however, indicating the different aspects you are interested in (similar data but for different types of farming, different geographical scope or different temporal scope).
- Do not only ask for data. Even if the contact was unsuccessful regarding data, important information can possibly still be obtained. For this reason, after receiving an answer to a data enquiry, the reply might ask for other contacts either within or outside their institution. Stakeholders usually have considerable knowledge about the agricultural sector and could help the UC designers source external organization/s that might have relevant information. Furthermore, they can possibly inform you if they know if such data even exists (this will potentially save significant search time as on occasions such data is not collected in the same format or with the specifications that the UC designer requires).
- Contact with future prospects. Whether the contact was successful or not it may be possible for the stakeholder to collaborate in later phases of the use case. For example, executing ad-hoc PR activities or validating the simulation results. As such, if they have shown interest in the use case then you should consider staying in contact for future collaborations.
- Monitor the contacts. It is important to stay in regular contact with long-term contacts or with those that will collaborate in future phases of the use case. It is particularly worthwhile to track the circumstances of the contact person (new contact information, planned leave, change of position in the entity s/he works at, etc.) and to anticipate possible problems with any future collaboration.
- Prioritise your contacts. Those contacts who are most likely to possess the required information or who are already known to the user should be the first contacted. There is a greater likelihood of obtaining the information via this approach, and often in a more agile manner. Similarly, if you get a secondary contact notify them that you have been referred by your primary contact. Additional sources for identifying contacts are listed below:
    - Websites on public bodies ((regional) ministries, local offices, etc.). On these websites, the user can find the organigram of the organisation or a consultation mailbox to which address her/his queries.
    - Websites for other organisations. In these cases, there is usually a general contact form that can be followed.
    - Scientific papers. If the user finds a paper related to the use case, it is possible to contact the authors.

- CORDIS and Participant Portal. On these European Commission web pages, the user can filter the research projects by the topic and contact their authors or partners.

All of these recommendations are intended to guide future users in obtaining the information required with a certain guarantee of success. Furthermore, since policymakers are the main beneficiaries of the AGRICORE tool, the section below describes further helpful guidance on how best to approach them and obtain their commitment to collaborate. In addition, the following section also focuses on how to achieve a wider diffusion for the conducted use case.

In summary, contact with stakeholders is a very broad activity that due to its uncertainty requires considerable scheduling and monitoring. For this reason, forward planning and conducting tasks in parallel is vital in sourcing good quality contacts. As with the autonomous search for data sources, if these first contact efforts are not fruitful and information also cannot be found via more exhaustive information searches, it is suggested to consider alternative participatory research activities (refer to Section 8.1).

7.1 Improved contact with policymakers

The above section highlights that consistent and clear communication with different stakeholders is of high importance within the AGRICORE project. On the one hand, our stakeholders are one of the main sources for data acquisition, and on the other hand, the different stakeholders, and especially policymakers, will be the final end-users of the AGRICORE tool. From the beginning of the project, it was clear that the success of an AGRICORE use case greatly depends on the knowledge, feedback and cooperation with policymakers. This stakeholder group act as external experts and can state the requirements and standards that need to be fulfilled by the AGRICORE tool. Moreover, the creation and implementation procedure of the CAP measures must be considered to understand the roles of the policymakers taking part. The Directorate-General for Agriculture and Rural Development (DG AGRI) in the European Commission is the authority at the European level responsible for defining CAP measures and supervising their implementation in the European Union territory. Its different departments oversee defining the guidelines for the CAP measures and the indicators for their evaluation. This information is then sent to the different member states that will design and implement these policies under those guidelines and their administrative structures.

The roles and responsibilities of the parties involved throughout the policy design and implementation process have been simplified into a table of policymaker profiles based on the experience gained during the AGRICORE project. Table 2 presents the policymaker profiles that future users might need to contact to request information for future use cases or to promote the AGRICORE tool. These are the most relevant generic profiles for the development of an AGRICORE use case and their main functions and competencies are also included. The nomenclatures and functions of the profiles may change depending on the use case, especially their geo-administrative scope. However, it has been verified that organization charts for the public institutions responsible for designing, applying, and monitoring CAP measures are similar across EU, national and regional levels. In addition, to aid the understanding of the organisation chart and the contact with these policymaker profiles in a real use case, the following information has been added:

- Firstly, the table includes a classification of the profiles according to their position in a generic organisation chart, with 1 being the highest-level profiles and 4 being the lowest level profiles. The high-level organisational units (levels 1 and 2) are headed by politically appointed technicians, who are usually relieved of their position when the political party in government changes. In contrast, the lower levels (3 and 4) are units usually comprised of non-political technicians who tend to remain in their job position despite changes in

government. Therefore, it is typically possible to maintain a more stable and lasting relationship with them over time.

- Secondly, Figure 3 reflects how the institutions in charge of managing CAP measures in EU countries are organised. As such, future users will know which institutional level to address in each use case.

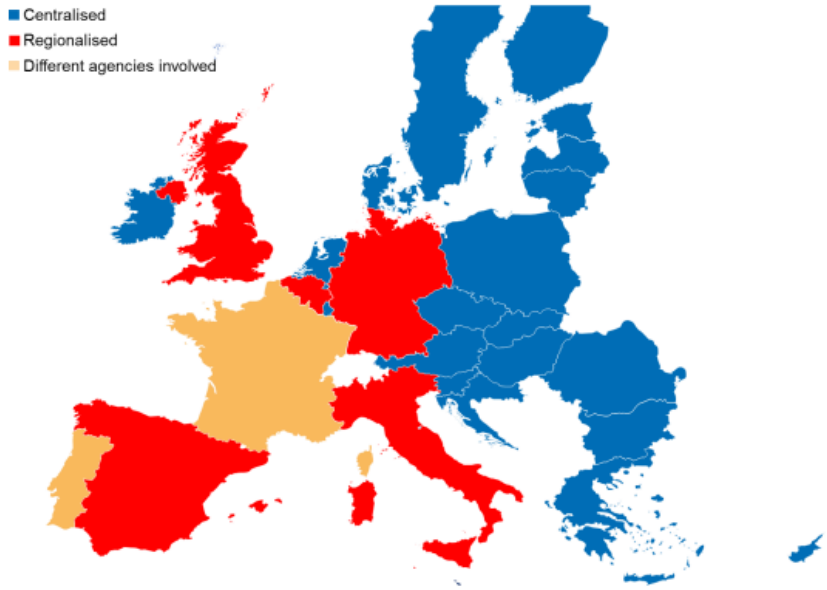

Figure 3 Administrative structure by Member State (Source: Analysis of administrative burden arising from the CAP).

Table 2 Profiles of policymakers (Level in a generic organisation chart: 1. High political position, 2. General management position (technician appointed politically), 3. General coordination position, 4. Service Area position).

Position	Level	Roles

General Secretariat for Agriculture and Food	1	• Responsible for CAP:
		  ○ The development and coordination of horizontal CAP issues, in particular those concerning the Common Market Organisation for agricultural products and CAP direct payments and rural development policy support measures.
		  ○ Monitoring of the CAP as well as the programmes established by the European Union in the agricultural field.
		• Design, coordination, and strategic planning, as well as the preparation or commissioning of operational studies of sectoral policies and strategic sectors.
		• Evaluation of the quality, efficiency, and performance of public policies in strategic agricultural sectors (evaluable and results-oriented public action).
		• Foresight and technological monitoring of sectoral markets that are considered a priority in agricultural matters.

General Secretariat for European Rural Development Funds	2	• Management of Common Agricultural Policy (CAP) aid, guaranteeing that aid reaches beneficiaries within the deadlines. • Ensuring that the requirements of CAP measures are met.
Directorate-General for Direct Aid and Markets	2	• Establishment of guidelines for the development and management of the Geographical Information System for the identification of agricultural parcels. • Monitoring and analysis of agricultural markets.

		• Developing, coordinating, assessing and promoting the use of more environmentally friendly techniques.
		• Developing the department's responsibilities for:
		  ○ Supporting the economic profitability and competitiveness of agricultural holdings and the management of agri-food markets, in particular through the design and implementation of the CAP instruments.
		  ○ Regulating agricultural production means (soil use, agricultural machinery, and fertilisers).
		  ○ Regulating the production, importation, exportation, certification, and commercialization.

Direct Payments Department	3	• Management of CAP payments, cross-compliance requirements, conditionality, and intervention under the CAP Strategic Plan. • Analyse the climate change impact on the agricultural sector. • Analysing, monitoring, and preparing proposals for changes in regulations
CAP Department	3	• Reforms of the Common Agricultural Policy, rural development and supervision.
Plant breeding and protection Department	3	• Plant breeding and seed production, fertilization, and assessment of its impact on the environment. • Registration of plant protection products, phytosanitary supervision, and protection plants, as well as the registration of fertilisers and plant conditioners.
Department of Organic Agriculture and Food Quality	3	• Management of organic farming and organic production. • Registration and protection of geographical indications.
Department of Agricultural Markets and Deputy	3	Regulation of agricultural markets under the Common Agricultural Policy and the market and selected products and organized forms of cooperation, as well as agricultural statistics and agricultural market research
Technical support Department	3	Support under the RDP 2014-2020 financial instruments and the CAP Strategic Plan and technical assistance RDP 2014–2020.
Department of analysis and statistics	3	Analysis and management of the data under the competence of each directorate general.
General technical secretariat	4	• Management of communication with citizens. • Direction and management of documentation and information.

In regard to identifying and filling in information gaps necessary for the proper impact assessment of agricultural policies, the main idea is to assess and collate their needs and requirements in a systematic way to guide and further enrich the AGRICORE design. To ensure that the initial contact with any policymaker follows the same structure, a cover letter for the project is a good approach (see Appendix 1). After establishing the first contact, further communications with policymakers should focus on specifying common targets and needs. Furthermore, if deemed necessary, a questionnaire can also be attached to the presentation letter to collect initial feedback and understand the policymaker's needs.

8 Participatory Research activities

This section considers several ad-hoc PR activities, which could be used as a last resort for the user, should none of the above activities be successful. These activities are often more resource-intensive and largely uncertain, plus the resulting data may even be inaccurate. Their implementation is usually constrained by the resources available and conditioned by the representativeness of the data obtained through the different activities. However, in many use cases these activities may be unavoidable due to their particularities or the inputs that are required by the AGRICORE tool.

The ad-hoc PR activities are explained in general terms, although some illustrative examples from the pilot use cases are included. In regard to implementation in future use cases, these activities would need to be properly customised.

8.1 Design of ad-hoc Participatory Research activities

The objective of these activities is for the user to build the necessary data set for the use case. Therefore, it is highly recommended that the user has the necessary expertise in the fields involved to ensure a correct design and execution of the activities. In addition, considerable knowledge is required to ensure the reliability of the data obtained. The user can base the development of the activity on previous research/studies and through the collaboration of stakeholders. Furthermore, before undertaking such an activity, the user should accurately assess the real need for such data in achieving the expected results. The effort may be disproportionate with regards to the resource consumption required and/or the accuracy of the data obtained. This could be further amplified if the inputs will not have a significant effect on the simulation results.

The first of these activities is related to the estimation or extrapolation of data. This is the least resource-intensive task, but the accuracy of the data obtained can vary greatly depending on the input data and methods used. In many cases, this is a viable approach as correlated data are available to help estimate the missing data. For example, the total production and acreage of a crop are proportionally correlated and knowing one can help estimate the other. This type of estimation is very common when one has temporal data from different years to the particular year in question. In these cases, a regression algorithm is typically used to treat the data.

In addition to estimating continuous values, discrete values can also be estimated, especially for classification. Below is an example of the population analysis for one of the project pilot cases. The classification of farm types according to slope, production (known data) and crop density (missing data) was estimated from the average production of the farm types. Finally, it should be noted that more complex estimates can be made if one has a thorough understanding of the data and the agricultural sector in question. More complex methods include using artificial intelligence techniques such as clustering methods and deep learning.

Another ad-hoc activity investigated during the development of the project pilot use cases has been the implementation of survey/interview campaigns. This type of activity is the most demanding of all due to the extensive deployment of resources required. This section explains the necessary design and execution steps. The final analysis of the data collected has been left to the users' discretion, since this will depend to a large extent on the survey design and whether the expected data has been obtained. This, in turn, depends on the particularities of the use case and the magnitude and typology of the information gaps detected. The steps required to prepare the survey campaign are explained below, assuming that the user has thoroughly analysed the policy to be evaluated (target population, requirements, geographical and time scopes, benefits for applicants, the process of adherence to the measure, etc.).

8.1.1 Analysis of the target population

In addition to studying the policy measure(s) to be evaluated, it is fundamental to know their evolution over recent years and the intrinsic features of the agricultural sector as a whole. This background helps to characterise the target population of the use case, which then determines the selection of the sample population and to a large extent the scope of the survey campaign. Types of information of interest are listed below:

- The number of potential beneficiaries in recent years.
- The number of actual beneficiaries in recent years.
- Geographical distribution of the (potential) beneficiaries and non-beneficiaries: those regions with a higher concentration of beneficiaries will probably have exploitations with similar features and the same applies for those with fewer beneficiaries.
- Typical data classification of the involved sectors. It is important to know if the exploitations are classified according to certain features as they can be reused during the classification in the PR activity design. This classification, together with the territorial division (municipalities, agricultural regions, NUTS 3 regions, etc.), is the most common way to organise available data in the sector.
- Gender and age balance for the involved sectors.

While this information may be available from the previous data collection process, it is also often found in reports on the evolution of previous measures or on the sector(s) to which the measure to be assessed applies. Moreover, this analysis also includes research on how farmers in the sector tend to manage their farms. This information does not need to be detailed and usually includes the organisation/s which they interact with most, the scheduling of farm work and farm costs, amongst other factors. If the user of the tool does not know this information about the sector, it is best to contact some of the parties involved to obtain a general overview. This information will be very useful when planning questionnaire distribution, since the objective is to adapt it as much as possible to the farmer. For example, by knowing the agricultural calendar, it will be possible to launch the survey campaign during the farming season with the least workload and thus maximise the number of survey responses.

8.1.2 Selecting the sample to be investigated

Once the size and main features of the target population are known, the portion of that population to be surveyed must be selected. The first step is determining the minimum sample size population for it to be sufficiently representative of the real population. In AGRICORE, the pilot use cases established 10% of the target population as the number of farmers to be investigated. Furthermore, the sample can be composed of farmers from different groups according to the particularities of the use case. In the AGRICORE project, two strategies were followed:

1. **Extract the sample from the leading farmers in the sector**. This strategy assumes that these farmers are a reference and that the remaining try to imitate them. Thus, if their characteristics are studied, one can get an overall understanding of the sector. This approach was applied in a nationwide use case where the target population was mostly small farmers who were difficult to contact.
2. **Extract the sample from the group of beneficiary and non-beneficiary farmers for the measure to be analysed**. This was used in two other pilot use cases and in each case the sample was composed of different proportions of both groups. Following this approach makes it possible to compare both farmer profiles, allowing the understanding of the motivations for adopters and the barriers for non-adopters.

It is important to state that the selection of the sample must be as unbiased as possible. To this end, the most common and effective way is to select the respondents randomly. Obviously, the distribution of this sample can be defined in order to conduct the survey campaign in regions or sectors of interest or that maximise the number of survey responses. For AGRICORE, this distribution has been made based on the different parameters listed below:

- **Location**: the territory to which the measure to be analysed applies has been divided into NUTS 3 regions, agricultural regions or municipalities, and a target number of surveys in each region has been defined.
- **Type of exploitation**: as the target population is usually classified into several groups according to some productive features, it is interesting to distribute the sample proportionally to the weight of those groups in the target population.
- **Gender balance**: since the agricultural sector is generally highly male-dominated, it is important to maintain a representation of women in the sample to understand also their specificities. Therefore, both genders should be surveyed proportionally according to the total number of male and female workers.

These parameters will also serve to monitor the evolution of the survey campaign. If the desired distribution of interviews is not being met, the survey campaign can be intensified for certain sample population strata. Finally, the selection of the sample also partly determines how the survey will be executed. If the user has sufficient resources or the sample is concentrated in certain areas, it will be possible to conduct a physical or telephone survey campaign (generally requires more time). However, if resources are scarce or the sample is spread over a large territory, a mass distributed electronic survey campaign may be more feasible.

8.1.3 Design of the questionnaire

The design of the questionnaire must be based on the information gaps that need to be filled and the characteristics of the sample population. This step is crucial because a poorly formulated or confusing questionnaire is unlikely to provide the required data. Although the design depends on each unique use case, some recommendations are included below.

1. **Reuse previous work**. During the literature review for the use case, it is possible that the case builder might find other researchers who have also conducted a similar survey campaign. It is highly recommended to spend some time looking for research similar to the use case to be simulated. Stakeholders who may have already been involved in tasks like obtaining survey results should also be contacted. This is particularly interesting because basing the survey design on a previously conducted study helps to justify and support the final design, especially if the research in question yielded satisfactory results. Nevertheless, the author should be advised of the survey request purpose and the case builder should bear in mind that the information received is only a reference and that it should be modified to adapt as closely as possible to the particularities of the use case.
2. **Consider the distribution of the survey**. The survey design should vary depending on whether the survey is in-person, over-the-phone or online. Firstly, the length of the questionnaire should be considered, as longer questionnaires can be inconvenient for the respondent, especially when conducted by telephone or online. The second point to consider is the clarity of the questions: any doubts on the part of the respondent can be clarified in face-to-face and telephone surveys, but not during online surveys. Finally, the design format of the questions can also affect the results: online and face-to-face surveys are visually supported, and questions can be formulated using tables and accompanied by explanatory pictures. However, this is not possible for telephone surveys so the questions in this format must be direct, clear, and easily comprehensible.

3. **The profile of the respondents**. As was mentioned previously, the pilot use cases have different sample populations, and their features must be considered when the questionnaire is designed. This influences the type of information which can be asked for. For instance, farmers with large exploitations usually conduct more precise financial management practices and can be asked for more specific data on this aspect. In small exploitations the financial management is more traditional and typically data scarce. The above though only partially determines the distribution of the questionnaire. For example, an online questionnaire will be relatively straightforward for a young sample population that is more familiar with new technologies, whereas it may pose some problems for older respondents.
4. **Length and structure of the questionnaire**. A questionnaire should be of sufficient length to cover the information to be collected, no more and no less. It is sometimes preferable to obtain data to fill in information gaps by other means allowing the survey campaign to only address the most important questions. In addition, substantial data can be obtained if the questions are formulated correctly, since it is possible to deduce information from the answers to several questions or to condense a lot of information with table formats. As for the structure, it is essential to correctly order the questions and avoid confusion for the respondent. A good practice is to divide the questions into specific sections according to the topic and then order them from the most general to the most specific. In addition, it is recommended to add a brief description at the beginning of each section, so the respondent has some context regarding what they are going to be asked.
5. **Clarity of the questions**. The clarity in the wording of the questions is essential in obtaining the required data. Therefore, questions should be kept as short as possible and ask directly for a specific piece of information. In addition, they should be formulated using simple language and vocabulary that the respondents understand. Consistency should also be maintained throughout the questionnaire. For example, when asking for a magnitude always use the same unit of measurement. Making many cross-references between questions should also be avoided. A cross-reference might be avoided by putting the linked questions following each other. The purpose being to avoid ambiguities and confusion on the part of the respondent, and potentially invalidating the answers.
6. **Types of questions**. It is vital that the right kind of question is chosen for each piece of required information (open vs closed questions, multi-answer vs single-answer questions, etc.). A suitable type of question makes it easier for the respondent to give his/her answer. In general, it is suggested to avoid open-ended questions, as they provide room for subjective interpretation by the respondent and require additional time for further analysis. Regarding closed questions, care should be taken when preparing the possible answer options, as a wrong choice could bias the answers.

Finally, it is considered beneficial to share a draft version of the questionnaire with the stakeholders involved in the use case. While this is not a specific tip for the questionnaire design, it allows the chance to incorporate stakeholder feedback and improve the survey, including its distribution and approach surrounding the sample population.

8.1.4 Pilot survey and final design

The next step is to test the designed questionnaire with a small percentage of the sample population to check if the expected data are obtained and if there are any design errors. In the pilot survey campaign developed for the AGRICORE use cases, the following points were checked:

1. Whether the length of the questionnaire was adequate to maintain the attention and interest of the interviewee.
2. Whether the distribution channel was appropriate according to the selected design for questions.

3. Whether the farmers interviewed actually knew the information they were being asked about. Also, to evaluate if they were reluctant to answer any of the questions.
4. Whether any question was poorly formulated or open to misinterpretation.

This pilot survey methodology provides the user with the best chance of receiving the required data. To this end, the necessary modifications must be implemented based on the findings of the pilot survey. In the AGRICORE use cases it was necessary to modify some questions and reduce the length of the questionnaire, including changing the distribution channel of the surveys. The latter is definitively fixed at this stage.

8.2 Planning and scheduling of Participatory Research

The duration of the participatory research is variable because it depends on the success of the initial activities to be conducted. For this reason, the user should periodically review the planning of the participatory research, updating the next steps and re-assessing the objectives. In the pilot use cases, the first action was to determine which information gaps could be more difficult to fill and then focusing more effort on the IGF activities related to those gaps. Less resource-consuming activities should always be investigated first. However, it might be that the missing data needs to be finally obtained via ad-hoc PR activities which are more expensive and time-consuming.

Furthermore, the execution of the IGF activities must also be planned according to their characteristics. For instance, if the user cannot obtain certain data after an initial quick search on the internet, then planning becomes critical. If s/he knows that an official public institution generates the required data, then they should contact them well in advance because long response times are normal. At the same time, if the user knows that the execution of a survey campaign is essential in obtaining some data, s/he should start to design the survey as soon as possible due to the extensive planning required. For this reason, it is worthwhile to begin with simple IGF activities and progress to more resource-intensive tasks if results are not initially achieved. Therefore, the quickest means to source data are tested first, and longer and more complex activities can be started well in advance.

In addition to the above time management practise, it is recommended that the case builder establishes participatory research checkpoints. Therefore, during the monitoring phase the current status of the activities can be compared against the information that was planned to have been already achieved. This can help to detect possible delays or problems as addressed via a risk assessment, which is detailed in Section 8.3.

Finally, the execution of a survey campaign might involve a planning stage with the personnel who will be responsible for performing the interviews (i.e. face-to-face or telephone surveys). The first step would be to explain the questions to them, clarify any doubts that may arise and make them clear on the overall objective of the survey. Moreover, it is advantageous if the interviewer has a background in the agricultural sector to be surveyed as they can better resolve respondent doubts. Once the survey has been explained, the second step would be to jointly set deadlines and objectives in order to establish the checkpoints mentioned above.

8.3 Execution and monitorisation of the Participatory Research activities

Once the implementation of participatory research activities begins, it is essential to monitor them and assess which information gaps have been filled. To aid this task, periodical meetings are held to review the set checkpoints, evaluate possible risks and problems that may have arisen, and implement possible alternative actions. In addition, during these meetings, the planning of

new activities is decided based on the results obtained from the previous activities for each information gap.

The potential risks of such participatory research, together with possible mitigation and prevention actions are presented below in Table 3. In this table, there are two columns, "Probability" (Prob.) and "Impact" (Imp.), which measure the probability of occurrence for the risk and the impact it could have on the development of the use case. This is measured at three levels: high (H), medium (M) and low (L). Furthermore, it should be noted that this table is dynamic and could be modified in monitoring meetings according to the status of the project.

Table 3 Risk assessment and mitigation actions

Risk #	Risk	Prob.	Imp.	Mitigation action
1	Delays in the execution of the tasks.	M	M	Adapt the planning (dates and procedures) of the tasks and their development to the current status and foreseeable issues.
2	Lack of data to initialise the ABM simulations.	L	H	· Checking the availability of the necessary data to initialise the ABM inputs after collecting the available data source · Checking the availability of the necessary data to initialise the ABM inputs after designing Participatory Research activities to fill in the detected information gaps.
3	Difficulties in managing face-to-face interactions with relevant stakeholders.	M	M	Preparing and planning these interactions by telematic channels to carry them out when it was possible.
4	Not considering the particularities of the use cases in the ABM implementation.	L	H	· Compiling the requirements (features of the beneficiaries, KPIs...) obtained from analysed Measures. · Contacting relevant stakeholders, especially policymakers, to track possible updates in requirements. · Monitoring the inclusion of the provided requirements in the different modules.
5	Not obtaining the expected data from the Participatory Research actions.	M	M	· Defining an alternative Participatory Research action to the one already proposed that allows the collection of the desired data or, failing that, a representative sample of those data. · Monitoring the development of the planned Participatory Research activities. · Proposing alternative ways to obtain this information, such as estimations.
6	Difficulties in reaching the target number of answered questionnaires.	M	H	· Conducting the surveys by directly contacting farmers that belong to the target population and facilitating their responses to the questionnaires (time, place, personal interviews…) · Adapting the questionnaires to the issues encountered. · Looking for additional respondents.
7	Not considering the dependency between the different tasks of the use case in terms of time and results.	M	H	Elaborating detailed planning and schedule of the use case development, including monitoring the tasks in progress.
8	Unavailability of resources (means of contact to conduct the survey, stakeholder's collaboration) that were considered in the planning of the execution of Participatory Research.	L	H	Not designing the Participatory Research activities based on the same resources (diversification of resources).
9	Not finding stakeholders (academic institutions, technical services from the	L	M	Design some standard simulation scenarios and carry out the impact assessment according to the existing mechanisms.

	Commission…) willing to participate in the testing of the platform.			
10	Not having the necessary stakeholders' collaboration for the tasks	M	M	· Agreeing in advance with them on the collaboration(s) they will carry out in the use case. · Preparing alternatives to the expected collaborations if some stakeholders fail to meet what was agreed on (diversification).
11	Obtaining incorrectly answered or incomplete questionnaires.	L	M	· Continuous review of the more recent answered questionnaires to classify them as valid or not. · Collect extra questionnaires to ensure a representative sample is obtained.
12	Obtaining contradictory or confusing information after the analysis of the questionnaire responses.	L	H	Contacting stakeholders (policymakers, agricultural associations, and technicians) that could help to filter outliers and discard those results considered unrealistic.
13	Difficulties in contacting agencies and farmers to conduct questionnaires due to Covid-19 restrictions.	M	M	An intensive campaign to encourage respondents to participate in the online questionnaire.
14	Data obtained from participatory research is not representative of the whole territory of the use case.	L	M	Conducting extra survey campaign to ensure a representative sample is obtained.
15	Data received from the questionnaire is not easily interpretable.	L	L	Look for ways to improve data quality and interpretation methods.

9 Conclusions

This deliverable presents a systematic approach for identifying and filling information gaps through participatory research actions. It especially focuses on the work necessary to set up any use case using the AGRICORE tool. Although all this work greatly depends on the particularities of the use case to be addressed and the information available, this deliverable brings together a set of generic guidelines intended to orientate future users of the tool. These guidelines are based on the experience acquired throughout the development of the three project pilot use cases. As such, the recommendations are provided to avoid and mitigate potential problems, and to gain maximum benefit from the actions to be conducted. Furthermore, the deliverable is addressed to a general audience, so it is not necessary to have extensive knowledge on the agricultural sector.

Since the AGRICORE tool is highly modular and open source, the deliverable starts by briefly describing its component modules. In summary, the modules can be modified by users or replaced by third-party models if they do not fit the needs of the use case. To achieve this, the user must first assess whether the outputs obtained from the simulation are able to analyse the impact of the agricultural measure to be studied. Once the tool modules are chosen, the information inputs needed for the simulation will consist of the attributes of interest of the agent, aggregated data of the target population and inputs from the modules. These inputs should be initialised with data from the data sources characterised in ARDIT, which can be found within the AGRICORE tool during the simulation setup. Therefore, if no sources are available for a given input, it will then be considered as an information gap.

Following the identification of information gaps, the information gap filling activities that were carried out in AGRICORE were presented, such as further searches of public and non-public data sources, contact with stakeholders and ad-hoc activities (survey campaigns and estimation of the inputs). All these activities are described in order from least to most resource-consuming, together with recommendations for their design and implementation with the aim of obtaining the desired data in the shortest possible time. Therefore, emphasis is placed on efficient activity planning. Simple tasks are to be addressed first and if the results are not obtained then more extensive tasks follow. The process then returns to continuing with simpler tasks for the next activity in order of resource consumption. This procedure optimises the planning and execution of the activities. In addition, the development of these activities, especially the survey campaign, requires regular monitoring to mitigate deviations against the objectives set and the unnecessary waste of resources.

Finally, considering that the potential tool users will be policymakers, this deliverable includes a specific section which lists the profiles that are likely to work more closely with the tool and its application. Liaison with policymakers will be conducted following the same process used for other stakeholders, however, it is also important to understand the organization chart for the respective administrative institutions. Depending on the purpose of the communication the most relevant contact person can therefore be directly approached. Communication will be mainly made to request information and to promote the tool and its application in the use case to be studied. Regular contact is necessary (before, during and after the development of the use case) to receive their feedback and adapt the use case in accordance with their needs. Therefore, a standardised approach is recommended which includes using a cover letter and short opinion questionnaires.

10 References

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

Deliverable Number	Deliverable Title	Lead beneficiary	Type	Dissemination Level	Due date
D1.8	Use case participatory research actions	CAAND	Report	Public	M18
D7.1	Use case planning and set of involved stakeholders	AAT	Report	Public	M25

Specifically, D1.8 describes the information gaps detected for the three AGRICORE pilot use cases, and the ad-hoc participatory research actions designed to fill those gaps.

D7.1 describes the aforementioned three pilot use cases and summarises the established contact with the respective relevant stakeholders for each UC.

11 Appendix A - Cover letter

AGRICORE Project

Agent-based tool for the development of agriculture policies

info@agricore-project.eu

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under grant agreement number 816078.

Dear Sir/Madam:

I am contacting you to introduce you to AGRICORE, a European H2020 project that aims to help design better agricultural policies. The AGRICORE project is developing a tool for modelling and simulating different instruments and measures associated with the Common Agricultural Policy (CAP), both at regional, national and EU levels. This is an agents-oriented approach that allows taking into account the wide diversity that exists between farms located in different geographical areas and/or dedicated to different crops.

The objective of the AGRICORE tool is to be able to test different alternative CAP instruments a priori, i.e., before their approval and implementation, in order to analyse the impact that each of these alternatives could have on the farm economy, land and agricultural prices, the environment and the social development of rural areas in Europe. For this purpose, an agent-based model has been implemented (each agent represents an individual farm as an autonomous decision-making entity) in which agents are created constituting synthetic populations that mimic the distribution, characteristics and interactions of the population of real farms of interest. This synthetic population is based on the data extracted from multiple agricultural EU statistics, geo-referenced datasets, other regional/national databases, results from previous projects and stakeholders' knowledge.

The AGRICORE tool aims to involve all agents in the agricultural sector in both its development and its subsequent use. For this reason, the tool has an intuitive and user-friendly interface that does not require a background in computer science to be used. Furthermore, the tool is developed with a view to future improvements, being open-source and based on a highly modular and customisable IT architecture. This allows for implementing improvements in the modules and replacing them with more specific ones adapted to future use cases.

Given the interest that the AGRICORE tool can have in the socio-economic development of the agrifood sector, as well as the modelling of policy strategies in the field, the AGRICORE project cordially invites you to participate in its development and continuous improvement process, so that once completed you can use this tool to know the real impact of current and future measures, and at the same time optimise the use of existing economic resources, adjusting it to the needs of the sector.