MEF4CAP

Monitoring and Evaluation Frameworks
for the Common Agricultural Policy

Deliverable D4.3

Description of design and results of demonstration cases

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101000662.

Document control page

Document control page	
Deliverable	D4.3 Description of design and results of demonstration cases
Document title	Description of design and results of demonstration cases
Version	V1.0
Responsible author	Ifigeneia-Maria Tsioutsia, AgroApps
Contributors	Electra Athanatsiki, AgroApps
Date of delivery	28.02.2023
Type	Text
Language	English (UK)
Rights	Copyright "MEF4CAP – Monitoring and Evaluation Frameworks for the Common Agricultural Policy"
Status	( ) In progress (X) For review ( ) Approved

Version history			
Version	**Implemented by**	**Date**	**Description**
V1.0	AGROAPPS	24.02.2023	First draft

Contents

Executive Summary

This document aims at describing the design and initial results of demonstration cases. Specifically, the purpose of this deliverable is to provide detailed descriptions per demonstration case with the findings and the extend to which project's objectives have been met and how the pathways can contribute to future monitoring and evaluation.

Since the demonstration cases are still in the implementation phase, this document considers to be the first version of the respective deliverable and an updated version will be submitted at the end of the project.

The sections of this deliverable are:

Section 1 – Introduction

Section 2 – Description of Demonstration Cases

This section includes a detailed description of each demonstration case along with the objectives and the expected outcomes as well as the main activities that will or have been performed during the implementation phase.

1. Introduction

The new Common Agricultural Policy (CAP) involves continuous monitoring and data analysis to identify potential risks and deviations from expected outcomes, instead of relying solely on periodic checks or inspections. This evolution can be integrated with various technologies such as remote sensing, IoT sensors, machine learning algorithms, and integration with various datasets to enhance the accuracy and efficiency of monitoring.

In the context of CAP and future policies, more focus is placed on assessing and addressing specific indicators related to environmental, social and economic sustainability. For instance, remote sensing and satellite imagery could be used to monitor land use changes, crop yields, and soil health indicators, while IoT sensors could provide real-time data on water and energy consumption, animal welfare, and greenhouse gas emissions. Machine learning algorithms could be applied to these datesets to identify patterns and predict potential risks and deviations from expected outcomes. In addition, the combination of technologies could be used to support participatory monitoring and evaluation, where farmers and other stakeholders are involved in the monitoring process. For example, mobile applications could be developed to allow farmers to report on their farming practices and input data on crop yields, water and energy use, and other indicators.

Overall, the combination of technologies and the collection of data from farmers could provide a more accurate and efficient way of monitoring and evaluating the effectiveness of CAP and future policies. This is the aim of demonstration cases, to identify potential risks and opportunities for improvement to future monitoring and evaluation framework.

2. Description of Demonstration Cases

2.1. Demonstration Case #2: Ireland

The Irish demonstration case aims to expand the capacity of accessing farmer's data directly from the Farm Accountancy Data Network (FADN) data collection. Specifically, if focuses on developing linkages with dairy processors to directly access data such as milk sales and input purchases for participating dairy farms. To do so, a group of 5 conventional dairy farms have been engaged that were willing to allow access to their data flow between processor and the FADN system in order to collect various data and make this process more efficient and economical.

The data that have been collected throughout the implementation of the demonstration case are FADN and additional farm-level variables. There are two aspects to the "technologies" considered in this demonstration case:

- Streamlining of the data collection process by collecting relevant data directly from e.g. the dairy processor for Irish dairy farms in the case of data (financial, technical and environmental) currently collected and inputted by the FADN data collector. This template could be followed for other input suppliers.
- Collation of relevant data collected through the Teagasc National Farm Survey (Irish FADN and additional data) into an interactive data dashboard for stakeholders (including farmers).

This demonstration case has been delivered in conjunction with a number of relevant stakeholders/ end-users in Ireland, including farmers, dairy processors and research and advisory staff from Teagasc (the Irish Agriculture and Food Development Authority).

Objectives and expected outcomes

The main objectives of the Irish demonstration case are the collection of data to calculate and report economic and environmental Key Performance Indicators (KPIs) for Irish dairy farms and the development of an attractive and informative sustainability dashboard for farmers and policy makers, designed in conjunction with relevant stakeholders.

By fulfilling the aforementioned objectives the Irish demonstration case would achieve the following outcomes:

- An electronic data collection process would make it easier to collect farm-level data for data administrators, researchers and policy makers.
- The proposed process would also reduce the burden on the farmer of supplying processor held data manually to data collectors e.g. invoices for relevant inputs and outputs.
- This framework would streamline the process of combining this data with data from other sources, including data that only the farmer can supply. This would be helpful in

the design and calculation of relevant sustainability indicators and would also reduce the likelihood of any errors occurring in data collection and assembly.

- The collation of relevant data and improvements with regard to data visualisation would be very useful for farmers and other stakeholders including those involved in the design and implementation of relevant policy.
- This could be a means to develop a data collection process that is scalable, making it feasible and more affordable to track sustainability across a larger pool of farms.
- The development of a user-friendly data interface would become a tool to help track, analyse, and display data (KPIs) and allow farmers and advisors to gain deeper insights into overall farm sustainability. This would help to identify actions to drive sustainability improvements.

Main actions performed

The main actions that have been performed during this period for the Polish demonstration case are the following:

- Data visualization workshop: An online workshop to demonstrate farm data visualization with respect to key farm performance indicators and to present the benefits of streamlining the data collection process. The aim of this workshop was to discuss the feasibility and possibility of streamlining the (FADN) data collection process (in the context of the Irish Teagasc National Farm Survey); to collate a list of relevant economic and environmental KPIs for stakeholder use (e.g. farmers, processors, researchers, advisors and policy makers); to discuss stakeholders' needs in terms of an interactive data dashboard; to present prototype dashboard developed using Microsoft Power BI and to agree on a strategy to allow for the collection and collation of relevant data more efficiently.
- Demonstration case national workshop preparatory meeting: The dashboard has been developed and the workshop has been planned. The aim of this national workshop is to have an in-depth discussion of the Dutch dashboard and the possibility of developing a similar template for the Irish case.
- Data dashboard development: Discussions have been held on the objective of developing an attractive and informative sustainability dashboard for dairy farmers and a presentation by Soops and WUR colleagues has been performed demonstrating the data dashboard that is being developed in the Netherlands. Furthermore, preparatory meeting was held to explore stakeholder requirements for the development of the data dashboard.

2.2. Demonstration Case #1: Poland

The Polish demonstration case aims to support farm-level management and policy makers regarding the management of applications at the farm. Specifically, it focuses on the optimal fertiliser management on farm-level based on the information of input and output of nutrients collected on plot level. The proposed way to achieve this, is to combine data derived from the

Farm Accountancy Data Network (FADN) imported by the Paying Agencies (PAs) and additional information collected by the farmers.

The data that have been collected throughout the implementation of the demonstration case are:

- Farm fields and crops registered in PA: Data collected by PA regarding crop production linked with farm fields. This data assists in monitoring crop production rotation and is declared by farmers applying for direct payments.
- Input of artificial and organic fertilisers per field and crop: Data on amount of natural and artificial fertilisers and soil improvement application per field and crop. This information is collected additionally to standard FADN book on farm level with assistance of auditors.

Following one of the general objectives of the CAP related to sustainable management of natural resources and climate action a list of impact indicators relating to biodiversity, water and soil is part of the current Common Monitoring and Evaluation Framework (CMEF). Focus area-related evaluation questions, such as *"to what extent have Rural Development Programme (RDP) interventions supported the improvement of water management, including fertiliser and pesticide management"* and *"to what extent have RDP interventions contributed to reducing GHG and ammonia emissions from agriculture"* directly points to necessity of monitoring farming practices related to fertiliser management. As it was depicted in Delivery 3.1 European Court of Auditors' Report on Greening Scheme advised that farmers should only have access to CAP payments if they meet a set of basic environmental norms and future indicators should be better targeted. This indicates that fertiliser management should be monitored on farm level to provide accurate information on CAP and measures impact on natural resource management. Up-taking specific fertiliser management practices are individually decided by farmers and require information on input needs based on input – output nutrient balances on crop/ plot level. In case of FADN, provision of such information requires collection of more detailed information on fertiliser use and therefore, creates additional burdens. However, this additional information is essential to optimize fertiliser management on farm level and farm incomes. In the demonstration case connection of additional information included in FADN with administrative register of plots and crops collected by PA are to be a test of plausibility of such system in provision of precise information on agricultural land management. Based on this exercise, obstacles and benefits of expanding proposed solution to all FADN farms are to be identified. In overall, demonstration case conforms to use of more accurate data to access the contribution of CAP interventions to policy objectives achievement.

The involved parties in the Polish demonstration case are the Polish PA (IT experts) that provides technical information regarding transfer of data to FADN, and agricultural experts that provide indicators needed for the calculation of nutrient input – output depending on specific crop production and yield, including flow of nutrient with organic matter.

Beneficiaries (end-users) of the new information flow are the farmers, agricultural advisory and policy evaluators.

Objectives and expected outcomes

The main objective of the Polish demonstration case is to present the possibilities to use administrative data to avoid increase of burdens related to data collection needed for future monitoring and evaluation of the CAP. It is expected that the demonstration case will allow farmers to recognise benefits from direct transfer of administrative data to FADN. Namely, transfer of data collected by PA regarding farm plots and crops combined with additional data regarding fertiliser application would allow for calculation of primary nutrients balances. Finally, information on nutrient balances would support management of fertiliser application.

This new information would directly reveal possible overuse of fertiliser that has negative impact on farm incomes and on environment. From other side, insufficient application of fertiliser would result in soil degradation. Thus, collection of series of these new indicators would support optimal crop rotation and cultivation of fertiliser catch crops.

Main actions performed

The main actions that have been performed during this period for the Polish demonstration case are the following:

- Presentation of the demonstration case #1 idea: The demonstration case #1 was presented to the team in the light of project goal.
- Analysis of potential administrative data sources to enrich FADN: Analysis of possible administrative data resources of PA to be connected with FADN.
- Analysis of indicators that can be added to FADN with help of administrative data: Analysis of preliminary outcomes of WP1 and WP2 to list new indicators and possibility to obtain additional data.
- Proposal of additional indicators to be tested on farm-level: Indicators that are of high concern were chosen and analysed with external expert regarding additional information.
- Linking administrative data with FADN – concept: The concept of gathering and connecting administrative data with FADN database was proposed and tested.
- Selection of farm for demonstration case #1: Auditors were instructed what farm should be chosen for demonstration case #1 and contracted with farmers.
- Adjustment of administrative data transfer to FADN: The concept of transferring administrative data directly from PA was modified, including farmers as intermediary.
- Meeting with auditors to prepare broad list of farms with exemplary accounts and experience of cooperation: List of farms/ farmers to potentially be included in demonstration case was prepared.
- Methodological consultation regarding indicators and their utilisation: The concept of inclusion of new indicators in annual report for farmers was elaborated.
- Meeting with auditors to establish list of farms/ farmers and provide clarification on additional tasks: Final list of farms/ farmers participating in demonstration case and information on indicators.

- Preparation of GDPR agreement to be signed by participating farmers: GDPR agreement prepared and distributed for acceptance.

2.3. Demonstration Case #1: the Netherlands

The Dutch demonstration case focuses on the sectors of organic dairy and arable farmers, by enriching, combining and crossing data from existing sources, i.e. FADN, with alternative sources of information (i.e. economic/ environmental/ sustainability data, fertiliser/ antibiotic use, etc.). Information requirements has been delivered, i.e. the specific data standards and requirements that must either met or developed in order to produce and provide to the farmers the required information.

Specifically, this demonstration case comes to provide an answer to the following questions:

- Can sensor data be integrated with economic data?
- Is it feasible to integrate the economic data automatically into SITRA system?
- Can this data be visualized in a web presentation (dashboard)?

The Netherlands has the most advanced software and data organization currently available (ARTIS in the Dutch FADN) and will extend that with new types of data, especially sensor data, tested with 5 organic dairy farmers in the FADN.

The data that have been collected/ used throughout the implementation of the demonstration case are:

- Farm, farmer profiles and parcel information: Full record of the (legal) status of the farm, tax status (VAT obliged or not, etc.), its owners including data about the farmer (name, national id, age, address, fiscal income, etc.) and data on static parcels' characteristics (cadastre data, use title (owned, rented, etc.)). In some cases, the legal structure can be quite complex (one farm, more farmers or one farmer owning several farms; farmers and farm households are not necessarily equivalent: a farm can have two farmers that live or live not in the same household, e.g. husband and wife, father and son).
- Inventory and building information: Includes the monitoring and management of all production materials, equipment, buildings and other infrastructure. Inventories of animals, chemicals, fertilisers, seeding and planting materials. Data are typically used to calculate the use of inputs during a period (accounting year, often the calendar year, but sometimes a harvest year like May-May or July-July) by adjusting the purchases and sales with the beginning and end-inventory.
- Purchase and sales data (UBL/ paper invoices, self-created bills, dispatch notes): Provides information on transactions with other parties: suppliers of inputs and services (including payments to landowners, personnel, taxes, etc.) and cooperatives, food processors or other farms that buy products or services (like contract work) from the farm. Some farms create invoices for customers themselves. Invoices contain data on amounts and prices paid for individual type of products and services as well as their volumes. This includes fertilisers, pesticides, anti-biotics, water, etc. Used to calculate

costs and sales (revenue), prices, quantities, use/ consumption, and environmental impacts. Supports reconciliation with, and validation of, farm management information and financial accounts. The data set may also include the invoices itself (hard copy or digital) and other evidences escorting the items of the inventory. Specifications and particularities of production materials, chemicals, fertilisers, etc., that support the interpretation of e.g. accountancy data through mapping and matching.

- Bank data (and cash book data): Data on individual bank payments (including credit cards, etc.) that specify which trading partners (suppliers of inputs and services, buyers of products and services) have been paid or made payments to the farm. In principle, digitally available to farmers (and with authorisations to their software and accountants) under PSD2. Some farms (especially those in short supply chains or regions where cash is still king) do cash transactions that are noted down on paper or in a cash register.

- Production and operating data: Provides information on inputs and outputs (use, consumption, production, environmental impact, etc.) of e.g. agricultural, horticultural, dairy or veal enterprises (that are branches/ activities within a farm). Often a lot of the input use (and sales) can be linked directly to certain crops and types of animals, but in some cases (e.g. fertilisers) a farmer has to note the allocation if interested in gross margins and costs of production of these branches or at field level. Typically, includes pesticides and fertilisers use, on large farms sometimes also hours worked of personnel and machinery (fuel consumption).

- Accountancy data – balance sheet: Provides information on e.g. farm assets and debts (valuations) including livestock, crop and other product inventories, loan balances, etc. The balance sheet gives the farm manager a "snapshot" of the net worth on a specific data. The net worth is the value of all assets on the farm less the amount of money owed against those assets.

- Accountancy data - cash flow: Cash flow statements provide information on the farm or farmers household on liquidity flows, e.g. farm sales and expenses, non-farm income and expenses and debt payments, but also changes in loans and money spend on investments. The projected monthly cash flow statement can be used to look ahead to the next year of operations. By projecting a cash flow for the next year, potential cash shortfalls can be noted and appropriate changes in the farm operation can be analysed.

- Accountancy data – sustainability report (including material balances): Provides information on the environmental, social and governance (ESG) indicators of the farm. Indicators are calculated on the basis of invoices and other accountancy data. Key performance indicators typically include use of antibiotics, pesticides, fertilisers (N, P, K) including manure, water, etc., where needed, material balances (obliged for organic farms) that provide information on material flow through the farm. In combination with data on buildings and other technologies used, proxies can be estimated for the emissions of CO_2, NH_3, fine particles, etc.

- Accountancy data – profit and loss account and income statements: Provides information on e.g. farm income and expenses, interest payments, livestock and grain inventories, costs of depreciation, accounts payable and receivable. Year-to-year profits are calculated on the income statement (i.e. profit/loss statement). The income statement is used to calculate net family (farm) income.

- Accountancy data – income taxes: Farm income and expenses, non-farm income and expenses, interest payments, depreciation schedules, resulting in the savings of the farm family (or families).

The services developed in this demonstration case help FADN managers responsible for monitoring the CAP and other policies to upgrade the FADN to the FSDN (Farm Sustainability Data Network) as proposed in the EU Farm to Fork Strategy. For this monitoring it is important that the FSDN has access to farm level sustainability data and that the farmer is happy to provide these. That means that data should be easy to handle (with robotic accounting) and that it is useful in farm management (dashboard).

At a later stage the services could also be used by all European farmers (that have a bank account and could be in the VAT system, that is roughly farms with a turn-over that is higher than € 25,000) to show their compliance in certification systems for sustainability and the CAP eco-schemes or CAP 2nd Pillar subsidies.

The blending of accounting data and sensor (and satellite) data in a dashboard for farmers can also be upgraded to regional projects in emission management, to commercial farm information systems that are used by farmers for compliance in certification (e.g. organic farms, commercial schemes like GlobalGap, Animal Welfare schemes etc).

The services are tested on a limited number of farms and in a limited number of countries, which means that it has to be upgraded to all farms in the FADN and can be upgraded to other countries. As well as to farms outside the FADN in commercial software (farm management information systems).

The following implementing partners contribute to this demonstration case:

TEAGASC: the Agriculture and Food Development Authority – is the national body of Ireland providing integrated research, advisory and training services to the agriculture and food industry and rural communities.

IAFE-NRI: Institute of Agricultural and Food Economics, an independent scientific research centre, supporting the Polish Ministry of Agriculture and Rural Development in designing and monitoring national agricultural and rural policies.

WEcR: Wageningen Economic Research, a socio-economic research institute that offers governments and companies (socio)economic insights and integral advice for sound policies and better decision-making in an innovative way.

Soops: Science Object Oriented Products & Systems, a software development company and expert in object technology, supporting organisations in translating their needs into suitable information and communication technology.

The following end users benefit from this demonstration case:

FADN systems: The managers of the Farm Accountancy Data Networks have to upgrade their FADN to a Farm Sustainability Data Network, according to the Farm-to-Fork proposal of the

European Commission. The demonstration case show how indicators can be collected in a cost efficient and auditable way (robotic accounting). It also shows that new types of data, like sensor data, can be combined into a dashboard. As they rely on the voluntary collaboration of farmers it is important to provide farmers also with useful information for their management from the dashboard prototype developed by the demonstration case.

Farmers benefit from a dashboard that provides insights to support their farm management and, based on a data space concept (ref. https://gaia-x.eu/what-is-gaia-x/core-elements/data-spaces), puts them in control of the data of their farm by managing data flows to and from the farm. Some farms (especially in the Netherlands where ammonia emissions are an important policy topic) are very interested in the actual (measured) emissions and might benefit from validating emissions with the sensor data.

Paying agencies benefit from the demonstration case in situations where future policies on e.g. CAP eco-schemes have to be monitored with farm level data, as the results of the demo-case can be used in audit or certification processes and results can be uploaded to the IACS system.

Certification control bodies (like in organic farming) benefit as the demonstration case makes clear how the farm accounts could be used to inspect farms with auditable data (e.g. with mass balances). This helps to reduce administrative costs. In the longer term it helps to better prepare a farm visit for inspection with an administrative mass balance check in advance, to have digital inspections (e.g. in situations of Covid or Bird flu) and apply artificial intelligence in risk based control programs.

National authorities that have to implement a FAST tool in farming for nutrient management, as obliged from 1.1.2024, might benefit from the demonstration case example of integrating such environmental accounting in standard farm software.

Software companies that provide Farm Management Information Systems or Farm Accounting software benefit as the results of the demonstration case could help them to upgrade their software.

Objectives and expected outcomes

The main objective of the Dutch demonstration case is to deliver a digital information flow that would provide better data for monitoring and policy evaluation and assist the policy maker to have better regulation in reducing use of inputs like antibiotics, pesticides, nutrients that cannot be observed from the outside by satellite.

Main actions performed

The main actions that have been performed during this period for the Polish demonstration case are the following:

- Investigate whether sensor data can be integrated with economic data:
    - Technical development of SITRA, based on ARTIS.
    - Order and install sensors with 5 farmers
    - Access by Soops to FADN data of 5 farmers formalised.

- Show the usefulness of an accounting approach in monitoring agriculture
  - Describe the administrative world in agriculture
  - Show the potential of robotic accounting in combining financial and environmental accounting
  - Investigate options to combine financial accounting and farm management information systems
  - Show that relevant Farm2Fork indicators and (in NL) mass balances (organic farming) can be calculated.
  - 4 weeks moving average of mineral balance N.
- Development of dashboard
- Receive feedback for the dashboard from 5 farmers
- Environmental devices and sensors acquired calibrated and deployed at FADN farms
- Environmental devices and sensors calibrated
- Measuring schema, installation protocol and data management in SITRA
- Environmental devices and sensors deployed at FADN farms
- Prototype dashboard development, combining mass balances, accounting and sensor data for experimenting with farmers
- Establish FADN data access (Netherlands)
- Mass balances integration for 5 (organic) farms participating in FADN
- Workshop and follow-up with farmers

2.4. Demonstration Case #2: Greece

The Greek demonstration case will focus on the use of digital tools in support of future CAP Monitoring and Evaluation. An agro-environmental data management platform will be demonstrated, capable to collect, share and visualise farm level cultivation related data (e.g. farm calendar data including parcel geometry, crop type, harvest, use of pesticide/irrigation/fertilizers, geo-tagged photos) along with data from relevant external sources referring to CAP indicators (e.g. soil quality, biodiversity index, water bodies, GHGs emissions). The collected farm level data will allow the extraction of farm/regional aggregates/statistics relevant with future CAP performance monitoring. Selected outcomes will be shared with individual farmers acting as rewards/incentives for further data sharing of farm level data.

The Greek demonstration case has complementarity with the Spanish Demonstration Case 2. Similar approaches have been designed for calculating similar indexes. However, there are differences in cultivation types and sources of information. The information services of the two versions of the demonstration case are designed, implemented and utilised in parallel and there are no dependencies. Exchange of ideas and conceptual approaches are exchanged from an early stage.

The data that have been collected/ used throughout the implementation of the demonstration case are:

- Fertilisation application: Date-time, type of chemical, dose
- Pesticides application: Date-time, type of chemical, dose
- Irrigation: Date-time, dose
- Land management: Ploughing, fallow land, etc.
- Harvests: Date-time, quantity, quality
- Other cultivation activities: Pruning, grazing
- Crop type: Olive groves, peach, kiwi
- Satellite based Earth Observation: Calculated indexes (e.g. NDVI)
- Photographs: Photos taken by mobile devices, annotated and associated with specific location, growth stage, cultivation type.

The services developed in this demonstration case aim to:

- Support subsidies control and compliance checks based on “traffic light” scheme, optimising and reducing physical checks for PAs.
- Improve efficiency and performance for monitoring CAP indicators.
- Provide capacity on combining and processing heterogeneous information.
- Reduce administrative burden.
- Accelerate payments, achieve right income allocation and reduce bureaucracy.
- Extract farm/ regional statistics on the applied practices, e.g. use of pesticides, fertilisers, irrigation
- Provide rewards as incentives to farmers/ advisors through access to regional statistics and other useful info (e.g. early warnings on pest infestations).

This approach is tested with a selection of approximately 60 Greek farms, members of Neuropublic’s network, that are located in the region of Northern Greece allowing to experiment with the extraction of regional aggregated outcomes from their farm logs. The selected crop types are olive groves and fruit orchards (peach and kiwi). The data collection, processing, sharing and visualisation tools will be developed on top of Neuropublic’s smart farming portfolio of services.

The end-users that benefit from this demonstration case are farmers, farmer organisations and cooperatives, Paying Agencies and Policy Makers, Integrated Administration and Control.

Objectives and expected outcomes

The main objective of this demonstration case is the evaluation of innovative agricultural data sharing approaches. That is, to deploy information flow mechanisms among systems operating at farm level (e.g. FMIS, including EO data) with the respective systems at regional level (e.g. IACS) and with CAP monitoring and evaluation frameworks (e.g. Integrated Farm Statistics).

There are two main expected outcomes from this demonstration case. The first one is the evaluation of innovative agricultural data sharing approaches, focusing on the improved efficiency and performance for CAP indicators monitoring; the capacity on combining and processing heterogeneous information; the reduction of administrative burden; the efficient and effective controls of aid applications and the acceptance of data sharing. The other outcome is the several potential benefits for farmers, such as faster payments, provision of regional statistics and farm-advisory services that will lead to reduced bureaucracy and right income allocation; and for Paying Agencies such as reduction of administrative burden, compliance with CAP targets/indicators, achieved performance.

Main actions performed

The main actions that have been performed during this period for the Greek demonstration case are the following:

- Selection of targeted areas and crop types: Approximately 60 Greek farms selected from Northern Greece. The selected crop types are olive groves and fruit orchards (peach and kiwi).
- Identification of data products of interest: Identify which data products may act as incentives for farmers to consent sharing their data.
- Identification of relevant (open) data repositories: Biodiversity indexes – Observations of selected species (Farmland Bird Index – FBI) for years 2018-2021 from 4 observatories in the area, Natura 2000 areas, Water Bodies, Wildlife sanctuaries.
- Adaptation/ Integration of data repositories in a user-friendly dashboard.
- Integration of datasets – Improvement of methods for aggregates extraction.
- Identification of which environmental indexes are more relevant for policy monitoring.
- Identification of CAP indicators
- Selection of indicators: Based on the activities of WP1, WP2, and WP3 a subset of CAP indicators has been selected.
- Identification of data sources: Identify appropriate set of farms/ farmers/ cultivation types that comply with a set of requirements, e.g. farmers are using a farm management information system which contains the necessary information items.
- Specification of data products: Farm calendar data processing methods have been specified that will allow to extract the required outcomes (indicators).

- Specification and implementation of Information System: This information system supports data collection, processing, calculation and visualisation of CAP indicators.
- Evaluation and end-user feedback on the use of the developed Information System: The information system will be demonstrated and utilised be relevant stakeholders. Their feedback will be collected through a structure procedure, e.g. with the use of questionnaires and/ or interviews.

2.5. Demonstration Case #2: Spain

The Spanish demonstration case has developed a digital farmbook that allows cooperative members (farmers) and their advisors to follow-up their performance with regards to the use of fertilisers, pesticides and water consumption and ensure their alignment with policy demands.

A secure workflow to exchange information between farmers and cooperatives was needed for this purpose. Therefore, legal data exchange contracts have been developed to ensure that the farmers are well informed about the management of their data, ownership and that the cooperative gets explicit consent to use them. In these contracts, each cooperative is also informed about its duty to comply with the EU GDPR.

So far, only historical information coming from the collaborating farmers’ old-farmbooks has been used to feed the GIS. These farmbooks contain mainly phytosanitary data as this was mandatory from 2012 onwards but, in some cases, information on fertilisers is also available. As soon as the digital farmbook is ready, these farmers will start to fill the information regarding water, phytosanitaray and fertilisers consumption, feeding the GIS in real time.

In Spain, from July 2023, every action a farmer takes on his farms will have to be recorded in the digital logbook within 30 days. This information, mainly concerning the use of pesticides, fertilisers and water, will be aggregated into thematic GIS dashboards so that the cooperative's advisors can see at a glance what is happening on their members' plots in real time. This will allow more preventive rather than corrective actions to be taken, thus improving the efficiency of farm management.

There are mainly two type of data sources:

- In-farm data:
  - crop and variety (non-personal data from old farmbooks),
  - area (LPIS – hectares and geometry),
  - yield (amount),
  - water use (application dates and volumes)
  - fertiliser use (application dates, amounts and product type)
  - phytosanitary use (application dates, amounts and active substances)
  - General farm data (owner, machinery, workers, etc.
  - pesticides consumption.

- Out-farm data: earth observation (Sentinel-2, NDVI), meteorological (AEMET & SiAR) and soil information (LUCAS and regional databases when available).

Spanish Co-ops (Agri-food Cooperatives of Spain) is the organisation that represents at national and European level the social and economic interest of the Spanish Agri-food cooperatives, it has developed, in collaboration with five farmers from the Cuatro Rayas cooperative (wine cellar) and other IT companies, the digital farmbook (Hispatec Analytics) and the GIS platform (ESRI).

The main end-users that benefit from the services developed in the demonstration cases and cooperative's farmers and their advisors.

Objectives and expected outcomes

The main objectives of this demonstration case can be categorised as general and country specific.

General objectives:

- To showcase the potential of further simplifying and reducing the IACS burden.
- To implement futuristic procedure involving regular and systematic observation for tracking and assessment, over a period of time, in compliance with eligibility criteria, commitments and other obligations, that allows conclusions to be reached on the eligibility for the aid or support/ payment requested (e-declaration).
- To integrate data flows derived from various sources (e.g. earth observation services, in-situ sensing networks and farm level data), to overcome limitation of the current approach.
- To deploy and evaluate information flow mechanisms among systems operating at farm level (e.g. FMIS) with the respective systems at regional level (e.g. Integrated Administration and Control System) and with CAP monitoring and evaluation frameworks (e.g. Integrated Farm Statistics).

Country case objectives:

- To develop an easy-to-use digital farmbook which integrates in a GIS both in-farm and out-farm data for a better decision making (farmers and cooperatives advisors)
    - In-farm data: crop, area (LPIS), yield, water / fertilizer / phytosanitary / pesticides consumption, treatment / application dates.
    - Out-farm data: earth observation (Sentinel 2), meteorological (AEMET & SiAR) and soil information (LUCAS and regional databases when available).
- Farmbook API development to communicate the required information to the administration. The adequate aggregation of farmers' data will result in indicators for CAP monitoring and evaluation purposes.
- A friendly system for tracking data at farm level in cooperative frameworks and available to be replicated beyond grapevines sector.

Main actions performed

During this period, there has been a lot of work in the development of a beta version for the digital farmbook and it is constantly updated into improved versions. Once the final version of the digital farmbook is developed (affected by several regulations that are to be yet published) it will be facilitated to Cuatro Rayas' farmers so they can provide feedback on usability, usefulness and level of satisfaction. Several meetings have already been held in order to retrieve their expectations and to constantly update them on the demonstration case progress (beta version of the farmbook, GIS, etc.). The cooperative advisors are also present in these meetings given that the GIS platform will allow them to improve their assessment and recommendations.

Finally, the GIS system was implemented and tested with previous mandatory and/or voluntary data from the Cuatro Rayas cooperative (pesticides and fertilisers). The dashboards are still in process of fine-tuning. Throughout the implementation phase several surveys have been launched to farmers to understand their needs and collect useful information to estimate the GIS data storage needs.

The main actions that have been performed during this period for the Spanish demonstration case are the following:

- Online meetings with Cuatro Rayas cooperative farmers and their advisory services: Several meetings with five farmers (several of them also work as advisors) form the Cuatro Rayas cooperative to present them the project's concept and plan. Updates on regulation requirements and status.
- Development of a digital farmbook beta version: Preliminary version of the digital farmbook.
- Meeting at the cooperative facilities with advisory services and farmers: Meeting in Cuatro Rayas cooperative facilities to present the farmbook beta version and collect their feedback.
- Training sessions in ArcGIS online: Training sessions to show them the GIS platform performance.
- Surveys: Launching of questionnaires asking farmers about their approaches.
- Development of data exchange contracts: Based in EU Regulation 2018/1807 and COPA-COGECA code of conduct.
- Gathering Cuatro Rayas farmers' historical data: Non personal data collection from 2020/ 2021/ 2022.
- GIS platform feed and fine-tuning: Fine tune of GIS platform dashboards and enrichment.

2.6. Demonstration Case #3: the Netherlands

Privacy preserving infrastructures can play an important role in addressing the issue of data sharing in the agricultural and food system. The Federated Learning (FL) method, often used as a metaphor of a data train, aims to facilitate such a privacy preserving infrastructure and support data sharing for specific data spaces that consists of interactions. These interactions are hybrid and could concern humans and machines such as data owners (data stations), data gateways, algorithmic models, (trains itself) and data users, such as policy analysts (passengers).

Nonetheless, there are no issues with sharing metadata with a Findable, Accessible, Interoperable and Reusable (FAIR) format, since only relevant data will be made available for various specific data users in order to support a certain decision-making step, such as monitoring and evaluation. The formulation of the decision-making step is crucial since this would be the train itself, that will cover an indicator for monitoring and evaluation of certain indicators. Next steps could concern models that serve for a prediction or forecast of an indicator.

Data sharing in the context of FADN - Incentives for farmers to participate on FADN

The farm accountancy data network (FADN) monitors farm income and business activities and is an important informative source for understanding the impact of the measures taken under the Common Agricultural Policy (CAP).

Aggregated FADN statistics are available in the public domain[1]. However, national Liaison Agencies responsible for conducting FADN surveys in FADN Liaison Agency collect many more relevant indicators that are very valuable for among others policy evaluations, but are not accessible via the EU FADN, since the EU FADN operates on the basis of the lowest common denominator in terms of overall data provision. National level data could be made available to, for example, a policy analyst at EU level by linking micro-labs of national FADNs that provide access to individual data.

However, for measuring the performances at farm level, the lack of qualitative and descriptive micro-data at farm level is an issue. For example, EU FADN does not provide sufficient information on disposable income of agricultural households because of political resistance and fear of farmer refusal[2]. Therefore, there is a need for a unified system, that could collect and combine socio-economic and environmental indicators at national level using them to monitor, assess and evaluate the performances of farms. Within the framework of FADN the data owners are the farmers (i.e., FADN members). The data controllers (i.e., FADN Liaison Agencies) determine the intended purpose of use of these data collected among FADN members. On the other hand, data processors (i.e. policy analysts) are able to use these data based on the policy set by the data controllers. The incentive for farmers to participate are many from financial rewards to gaining overall knowledge generated surrounding their own farms to improve their business (e.g. benchmarking to further optimize processes). Summarized, FADN is an instrument of the European Commission to evaluate the income of famers and understand the policy impact on their businesses.

Data Sharing and GDPR

[1] Commission, European, 2022, FADN Public Database (SO): European Commission
[2] O' Donoghue, Cathal, et.al, 2016, Farm economic sustainability in the European Union: A pilot study. Studies in Agricultural Economics 118(3):163-171.

Data sharing and integration that is compliant with the General Data Protection Regulation (GDPR) is one of the key challenges within the FADN network. Generally, the issues for farmers concern benefits of data sharing without compromising on privacy abiding by GDPR regulations[3]. Given the advantages of data sharing, issues still remain that deal with ethical (am not allowed to), political (I don't want to), administrative (I don't have resources) and technical (I can't) (Sullivan, et al. 2011). These issues could be addressed well if instead of sharing data by farmers (and FADN Liaison Agencies), enabling shared algorithms that use these data by merely accessing in a privacy preserving way. To enable such a setting it is important that the data from farmers are interoperable with other FADN Liaison Agencies to be understood as well as by algorithms as humans. Therefore, to share these data in such a framework, metadata based on ontologies is essential. However, until date, limited ontologies available in the public domain hampering the introduction of seamless data sharing and integration of technologies.

It is expected that these ontologies could have added value by facilitating interoperable and flexible systems while preserving privacy. Nevertheless, the implementation of interoperable and flexible systems might also shift power relations within the existing FADN. Therefore, to overcome this hurdle, extensive negotiations between the data owners and data users are needed to settle data governance agreements, such as data ownership, confidentiality and sovereignty. Reaching an agreement is a challenge especially when many actors with various and sometimes conflicting interests are involved, for example, evaluation agency versus farmers.

Current issues with data sharing

Data sharing and integration is one of the key challenges within the FADN network. There exists sensitive data that has different political consequences (For example farm income and pesticide use are sensitive in nature also due to environmental changes). The individual data is sensitive due to privacy and confidentiality reasons (for example GDPR requirements). The latter is addressed with data train concept that is demonstrated in this study. The first one is addressed with an acceptable expert question that is designed in co-creation with the FADN members who are involved in this study. This study focuses on additional data besides what is required for EU FADN to derive variables for analysis.

The issue for policy makers and government bodies is mainly to gather data from farmers without compromising on privacy and abiding by the GDPR regulation. Data has become an asset for farmers and at the same time it has also become governance issue for policy makers. This means that it is personal for farmers to share their data and at the same time it also becomes challenging for authorities to make policies to tackle environment changes and global warming[4].

Apart from this, for the additional variables, there are no common standards that are used by all farmers or data providers alike that make the integration of data possible for interpretation of results generated from algorithms that run on them. For this purpose, domain specific ontologies and controlled vocabularies play a key role in harmonising these data at diverse

[3] Atik, Can Martens, Bertin, 2021, Competition Problems and Governance of Non-Personal Agricultural Machine Data: Comparing Voluntary Initiatives in the US and EU. J. Intelll. Prop. Info. Tech. & Elec. Com. L. 12:370
[44] Rotz, Sarah, et al., 2019, The Politics of Digital Agricultural Technologies: A Preliminary Review. Sociologia Ruralis 59(2):203-229.

locations. However, as of now there are no standardised ontologies that exists within the FADN network that is commonly accepted by every FADN Liaison Agency in Europe. Since data is often sensitive and highly confidential it makes it challenging for implementing Europe wide guidelines. To overcome this hurdle, extensive negotiations are necessary between FADN Liaison Agency and data owners. Reaching such an agreement is often challenging especially when many FADN Liaison Agency with various and sometimes conflicting interests are involved. By the application and demonstration of FL, this case seeks to understand how data on national level could be accessed to analyse one or more socio-economic sustainability indicators. Therefore, this study attempts to demonstrate the implementation of FL with FAIR data that should support the development of tools for policy analysis for the CAP. For this we start from three data stations located in different countries, namely: the Netherlands, Poland and Ireland. This will allow the FADN network to judge the potential of FL for the agricultural system.

This study evaluates the potential pathway for the privacy preserving FL method, also known as the data train, that analyses education of a farmer and its influences on off-farm income for three EU FADN Liaison Agency. The evaluation should provide understanding to FADN analysts in the context of the CAP. The privacy preserving method that is investigated in this study is an architectures that promises to allow semantic interoperability between information of different FADN Liaison Agencies, while simultaneously being able to perform analysis with higher quality attributes such as accuracy. Therefore, this study will show the organisation of a FL set-up that should support data sharing within the FADN.

How data train differs from a traditional approach

The different FADN Liaison Agencies will be able to use relevant policy data across the value chain supported by FAIR. The FAIR guiding principles allows humans and machines to automatically find and reuse data[5]. Originally these principles were meant for scientific data management and stewardship activities, since then it has been widely used for any digital object for reusability purpose[6]. To enable these principles data train concept was initially introduced in the form of a personal health train infrastructure[7, 8, 9, 10].

Since data is an important asset, it is highly desirable to have access to a limited set of parameters derived from these data within the data stations owned by data owners restricted in a formalized setting. Such requests are generally brokered by smart contracts hosted in a centralised server. Traditional client/server models describe how a server provides data and data-services to one or more of its clients. Models on the client side use these data to provide meaningful aggregated statistics¸ which is derived knowledge represented on websites. Examples of such a setup includes web servers, mail servers, etc. In general requests are fired from clients to servers to be processed and results are sent back to clients.

---

[5] Wilkinson, M.D., et al., 2016 The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3:160018.
[6] Wittenburg, Peter, et. al., 2020 State of FAIRness in ESFRI Projects. Data Intelligence 2(1-2):230-237.
[7] Beyan, Oya, et al., 2020 Distributed Analytics on Sensitive Medical Data: The Personal Health Train. Data Intelligence 2(1-2):96-107.
[8] Deist, T. M., et al., 2020 Distributed learning on 20 000+ lung cancer patients - The Personal Health Train. Radiother Oncol 144:189-200.
[9] Shi, Z., et al., 2019 Distributed radiomics as a signature validation study using the Personal Health Train infrastructure. Sci Data 6(1):218.
[10] van Soest, J., et al., 2018 Using the Personal Health Train for Automated and Privacy-Preserving Analytics on Vertically Partitioned Data. Stud Health Technol Inform 247:581-585.

The key difference between the traditional approach and data train approach is that within this framework data remains on its physical location and never leaves, instead, algorithms move over the internet to gather outcomes from individual data stations for informed decision making for the authorised users. These algorithms therefore require explicit permissions from the data owners go grant access to them to be run on their data. Often these requests are integrated in the form of smart contracts to be automatically brokered on data owner's behalf.

Within this study, we aim to define a minimal set of technical requirements and implement an example use case within the domain of accounting data (i.e., FADN). Here we will learn about the influence of education on the off-farm income from three European FADN Liaison Agency (Ireland, Poland and The Netherlands). To show the working principle, a data train framework is implemented using supplementary FADN data available on national level.

We present the concept of data train as an analogy to FL. For this study, we reuse the infrastructure that is provided by Vantage6, nonetheless there are many other concepts that are similarly called Swarm Learning (SL). It was found that this concept is provided by tech organisations like Alibaba and Hewlett-Packard. SL is a decentralised Machine Learning (ML) framework that enables organisations to use distributed data to build ML models by leveraging blockchain technology that facilitates the sharing of insights captured from the data rather than the raw data itself. SL is a biologically inspired Artificial Intelligence (AI) approach based on the behaviour of social insects like ants and bees[11]. FL also works on a similar rule while the term was first introduced in Google AI's blog[12]. FL, however, requires a central server that coordinates the participant nodes and receives model updates. Moreover, Swarm Learning should address problems that are related to security and fault tolerance of FL. Because of the presence of a central custodian in the architecture, the FL infrastructure might be vulnerable to malicious attacks or software failures[13]. Another initiative that is building on top of FL is the Federated Scope platform. This platform, based on an event-driven architecture, should provide easy-to use interfaces and flexible customization of tasks for supporting learning safely and effectively[14].

Summarising, the FADN is an important source for understanding the impact of policy measures under the CAP. The collaboration allows to monitor farm income and business activities. Data sharing and integration that is compliant with GDPR regulations is one of the key challenges within the FADN network. Generally, the issues concern benefits of data sharing without compromising on privacy abiding by GDPR regulations. Privacy preserving infrastructures can play an important role in addressing the issue of data sharing in the food and agricultural system. The FL method, aims to facilitate such an infrastructure and support data sharing for specific data spaces that consists of interactions. A key precondition, and also a promise of FL, is the presence of semantic interoperability. The incorporation of the FAIR principles, such as a FAIR Data Point and harmonised data models, such as ontologies, could facilitate semantic interoperability to some extent and support flexible systems. One possible point of attention

[11] Warnat-Herresthal, Stefanie, et al., 2021 Swarm learning for decentralized and confidential clinical machine learning. Nature 594(7862):265-270.
[12] McMahan, Brendan; Ramage, Daniel, 2017, Federated Learning: Collaborative Machine Learning without Centralized Training Data. Google, ed, Vol. 2022: Google.
[13] Han, Jialiang, et al., 2022, Demystifying Swarm Learning: A New Paradigm of Blockchain-based Decentralized Federated Learning. arXiv preprint arXiv:2201.05286.
[14] Xie, Yuexiang, et al., 2022 Federatedscope: A flexible federated learning platform for heterogeneity. arXiv preprint arXiv:2204.05011.

is the shift in power relations, which could give room for fraud causes and require extensive negotiations on data governance aspects, such as data ownership and data sovereignty.

Approach

There are three main components within the adopted data train concept, namely data stations and data trains and semantic rails. In this study, the Vantage6[15] was reused as a privacy preserving FL infrastructure for secure information exchange. This infrastructure was hosted at proximity of STATION-1, STATION-2 and STATION-3. Privacy policy and collaborations for data sharing were defined on the server as a starting point to build this demonstration case. In Figure 1 an architectural view represents the set-up and its main actors, steps and flows.

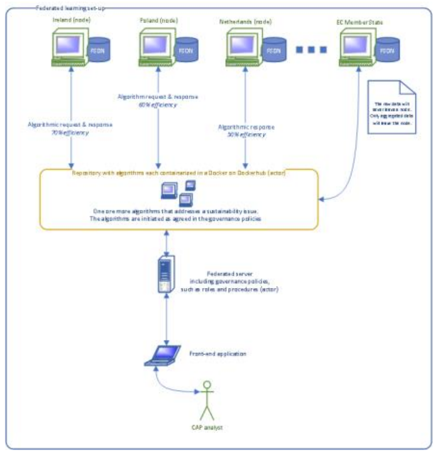

Figure 1: Architectural view of the FL set-up adopted for the CAP analysis context.

Data Stations

The data stations are represented by three national research institutes, acting as Liaison Agencies, hosted at different physical locations, namely in Poland (The Institute of Agricultural and Food Economics), Ireland (Teagasc – the Agriculture and Food Development Authority) and The Netherlands (Wageningen Research). The data stations provide access to supplementary FADN data and facilitate running of algorithms by respective model users (i.e., researchers) that would run on these stations. All three data stations have the statuary task to collect FADN data.

A summary statistics analysis of the three data stations should provide insights on the distribution for the variables concerning education, successor of the farmer, size of the farm and whether there is off-farm income. These variables are further elaborated in this report as part of the semantic rails.

Data Trains

For this demonstration case, we see users of data train as policy makers or government bodies who are responsible to get estimates on each data stations for policy making and are referred as model users. Model developers are researchers who are hosted at research institutes

[15] https://Vantage6.ai/

specialized in building predictive models and publish it on open access servers to be used by others. The models are seen as the data trains.

The model that is developed in this study to demonstrate the concept, is an algorithm that generate simple summary statistics, a linear regression and a correlation matrix given the considered variables. The algorithm is developed with iterations using test data and real data. This algorithm is encapsulated is a docker image that is pushed to a docker repository which is a publicly available registry such as Dockerhub[16]. All source code implementing this frame can be reused and can be found at the repository maintained. This implementation framework can be copied and used to implement other algorithms of choice. Other templates using python framework can also be found at this repository.

In order to conduct meaningful analysis and potentially create impact, we need to understand which variables to use and which correlations are relevant to look for. The steps to align and discuss these factors were undertaken in co-participatory way with the FADN Liaison Agencies that represent the three data stations. First, the feasibility of the case was discussed while taking existing implementations and documentation into account. Next, the approach was agreed upon to design and develop an infrastructure for demonstration purposes. For the context of FADN, alternative domain modelling approaches were explored for capturing semantics and building a resilient semantic rail. Finally, the prototype for this case was developed and conclusions were drawn. These conclusions are based on the analysis of potential contribution of FL that supports the monitoring and devaluation of the CAP with interoperable sharing of sustainability and economic data.

Semantic Rails

To highlight the aspect of our data-intensive set-up and emphasize FL as a methodology, it is important to note the understandability of data. The semantic rails should address the question of how the user would discover and meaningfully integrate data[17]. As was pointed out in the introduction of this paper, it is key to understand the success factors of harmonizing data on a national level for the three data stations. This requires a methodology to adopt ontologies to include relevant information for an interoperable federated set-up with structured format. Interestingly, the harmonisation effort resulted in that we had to resort to lowest denominator for education as well as off-farm income.

All data stations were hosting data in a Comma Separated Values (CSV) format, while the description was modelled in Unified Modeling Language (UML). In order to have the data infrastructure semantically interoperable, the use of a FAIR Data Point was considered for the metadata from all data stations[18]. The metadata were made available on a locally hosted FAIR Data Point by reusing existing ontologies, such as the Dublin Core.

[16] https://hub.docker.com/
[17] Janowicz, Krzysztof, et al., 2015 Why the data train needs semantic rails. AI Magazine 36(1):5-14.
[18] da Silva Santos, Luiz Olavo Bonino, et al., 2022 FAIR Data Point: A FAIR-Oriented Approach for Metadata Publication. Data Intelligence:1-21.

FADN and Indicators

Some policy related expert questions that are considered to query on the FL infrastructure are listed below:

- Which variables influence the farm income in the different countries and what are the differences? Why are these differences?
- What are possible ways to measure and predict the off-farm income, for job growth and rural poverty with a FL set-up and semantically interoperable data? Off-farm income is income obtained from activity unrelated to the farm business at different reporting levels (farm level, regional, national).
- What are the most effective information management strategies to analyse and predict off-farm income obtained from activity unrelated to the farm business at different reporting levels (farm level, regional, national) for job growth and rural poverty with a FL set-up and semantically interoperable data?
- How do the different data stations compare in the prediction of off-farm income at different reporting levels (farm level, regional, national) with a FL set-up?
- How do long-term decision makers, such as policy makers, perceive predicted values for the off-farm income for the results derived from different data stations?
- What are the factors to predict off-farm income at different reporting levels (farm level, regional, national) for the 3 countries, the Netherlands, Poland, Ireland?
- Which variables influence the benefits regarding education in the different farms/regions/countries and what are the differences? Why are these differences?

Results

Semantic Rails facilitate initial steps for interoperability

As mentioned earlier, the variables of interest are selected in a co-creation way, meaning that each FADN Liaison Agency is asked for input. The different variables in the dataset contain information on an anonymized unique identifier, plot surface, farm size, off-farm income, education and successor.

For plot surface, as a unit of measure, hectares (ha) was used. For farm size the standard yield (SY) was used, which is a standardized measure for the economic size of agricultural companies based on the yield that is achieved on average on an annual basis per crop or animal produce. The farm size is expressed in tonnes (t) Euros. The off-farm income variable has a Boolean data-type, meaning that it indicates whether there is off-farm income (1) or not (0). Information on education is provided with 2 variables, namely whether there is education benefitted in the relevant year and the highest education benefitted. The variable on education for the relevant year has a Boolean data-type, meaning there is education benefitted (1) or no education is benefitted (0). For the variable representing the highest education, there is an enumeration from 0 to 3 that contains a hierarchy of levels of education, or missing data (0). The variable successor is also represented with an enumeration that contains different categories from 1 to 4. Further details on the variables are shown in Figure 2.

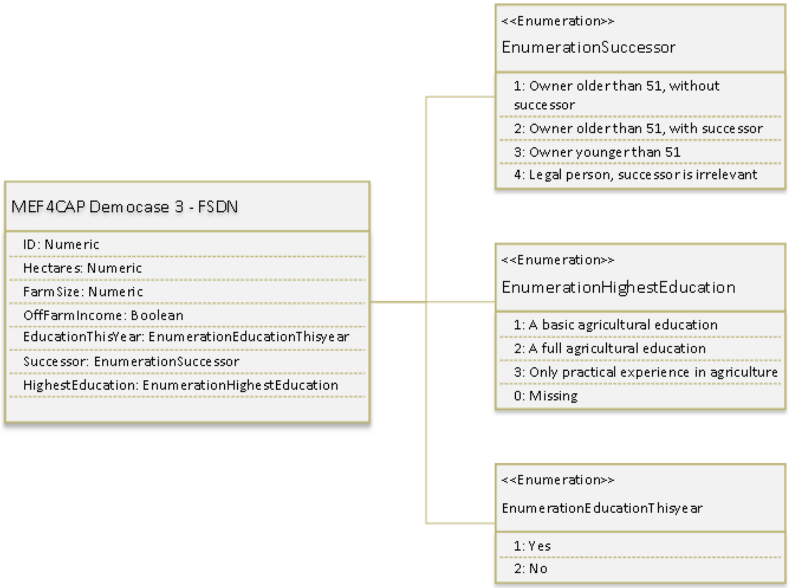

Figure 2: An UML class diagram that represents the variables of interest

As mentioned, as a starting point to facilitate semantic interoperability, a FAIR Data Point was used on the dataset level to describe the dataset with metadata. A snippet of the description is shown in Table 1 and indicate the reuse of an existing list of terms derived from the Dublin Core (dcterms), that is focussed metadata standardisation. The Dublin Core is a well-known example of metadata standards, which started already in 1995 by providing commons via the web on a variety of topics such as languages, structure, metadata mapping, application profiles[19]. Each data station within this setup is hosting a section of this metadata.

dcterms:description "This dataset is created with the intention to analyze FADN data on national level to support CAP policy making. In this case the data is about off-farm income and some other variables that could influence the off-farm income. The data is used for the purpose of the MEF4CAP project.";

dcterms:hasVersion "0.1";

dcterms:language <http://id.loc.gov/vocabulary/iso639-1/en>;

Table 1: Snippet of metadata to ensure semantic interoperability on the dataset level.

Data Stations facilitate aggregated FADN statistics

Statistically Explain the picture and explain each variable, what it means and how the distribution looks like from a FADN perspective.

[19] Arakaki, Felipe Augusto, Plácida Leopoldina Ventura Amorim da Costa, and Rachel Cristina Vesu Alves, 2015 Evolution of Dublin core metadata standard: an analysis of the literature from 1995-2013. International Conference on Dublin Core and Metadata Applications, 2015, pp. 220-222.


```
       ID              Hectares          Farm_Size       Off.farm_income Successor Highest_education
 Min.   : 25848   Min.   :  0.00   Min.   :    0.0   0:772           1:358     0:285
 1st Qu.: 73336   1st Qu.:  1.00   1st Qu.:   91.0   1:714           2:398     1:276
 Median : 91595   Median :  9.00   Median :  268.5                   3:348     2:330
 Mean   :128458   Mean   : 24.18   Mean   :  535.8                   4:382     3:306
 3rd Qu.:106069   3rd Qu.: 30.75   3rd Qu.:  619.5                             4:289
 Max.   :517879   Max.   :720.00   Max.   :13669.0
 Education_this_year       Station
 1:743               STATION-1:498
 2:743               STATION-2:489
                     STATION-3:499
```

Figure 3: Summary statistics of the three data stations and selected variables

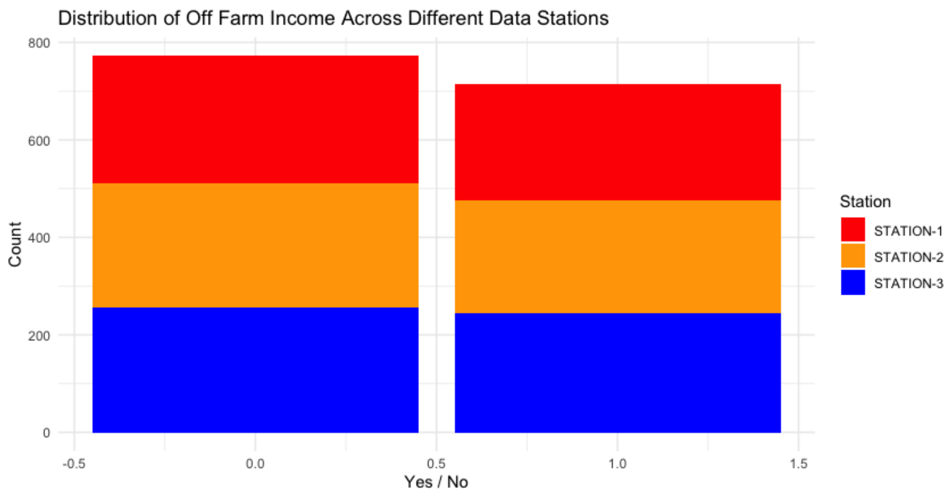

Figure 4: Distribution of off-farm income across different data stations

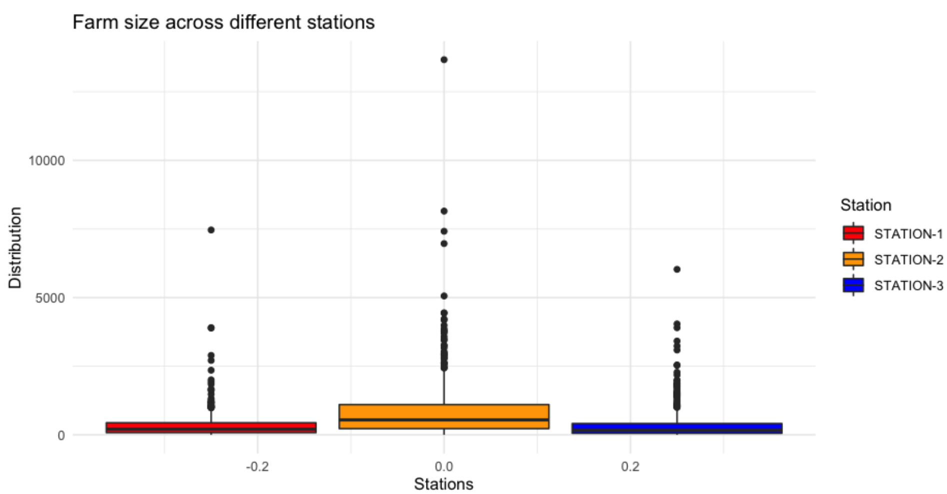

Figure 5: Boxplot with farm size expressed in tonnes (t) Euros across different data stations

Data Train facilitate ML with a Generalised Linear Model

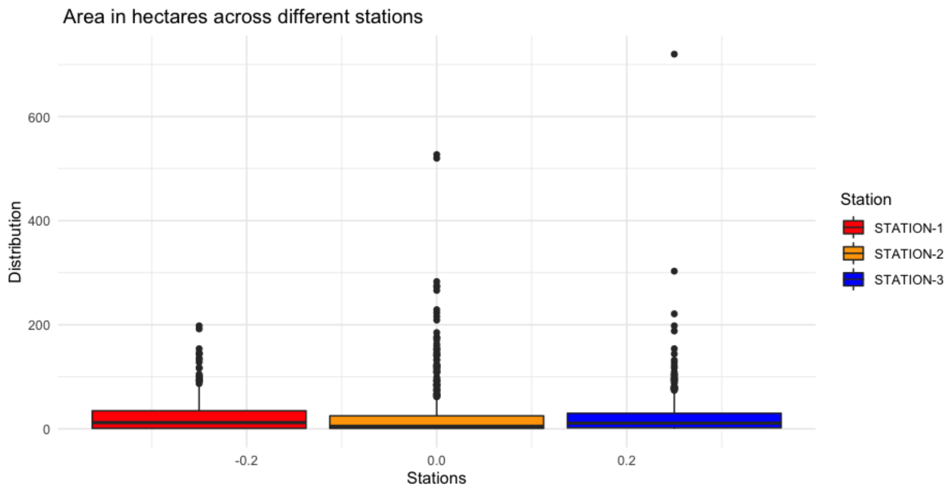

Figure 6: Boxplot with plot surface expressed in hectares (h) across different stations

Conclusions

The issue of data sharing in the agricultural and food system can be addressed with privacy preserving infrastructures. It is expected that the Federated Learning (FL) method could facilitate such an infrastructure with possibilities for semantic interoperability. For the first time, the effects of economic trends are studied with a FL set-up to exploit Farm Accountancy Data Network (FADN) at national level to improve EU policy making under the Common Agricultural Policy (CAP). The set-up consists of nodes interacting between each other, i.e. the Liaison Agencies who are responsible for conducting FADN surveys in FADN Member States. These interactions could concern data owners, data gateways, algorithmic models and data users, such as policy analysts. In this study, Vantage6, as a privacy preserving FL infrastructure, is used for secure information exchange. This infrastructure was hosted by all the partners at proximity of three stations, respectively, STATION-1 the Netherlands, STATION-2 Poland, and STATION-3 Ireland. Simple statistical analysis was used to demonstrate the federated set-up within the agricultural accounting domain. Main results of this case study contribute to pathways for privacy-by-design infrastructures for FADN monitoring and evaluation. In specific, the indicators off-farm income and level of education of a farmer are analysed. Additionally, the results contribute for semantically interoperable data sharing within the agri-food system. These insights are to some extent generalizable for other farm business activities to understand the impact of the measures taken under the CAP.

There are currently three main issues that are observed in a FL setup, namely data democratization, limited AI models and efficiency of the tools.

Firstly, in most FL systems, there is not much emphasis on democratization of data, which seems to be of paramount importance for data harmonization. Other authors, see Han, et al. (2022), Warnat-Herresthal, et al. (2021), Dänschel (2020) suggest that current constraints in FL require that data should be in a structured format (e.g., CSV) and that the order of variables in that data should be maintained. As mentioned in the results section, this resulted in persisting to lowest common denominator for off-farm income as well as education. Aside from this, it is important to consider the data type as it poses a problem when ML models use this data (for example, a ML model might expect a number, but the value is in characters). To address some of these issues, modern data formats have been proposed, such as linked data formats like RDF, where data can be represented as triples. The advantage of this approach is that the data for ML models does not need to be in a predefined tabular structure, but can be expanded as it is. However, in order to be able to use these data, an additional SPARQL layer must be integrated into the machine learning models, which first identifies the required variables and determines whether they are available at all from the data owners. This does not preclude still taking into account pre-processing steps such as missing values and other data formatting issues. When different data owners collaborate in a federated setting using same data in different contexts additional layers can be added on top of data stations where automated deciphering of data based on ontologies can be carried out using modern data standards. Mentionable examples within the medical and health domain are the Fast Healthcare Interoperability Resources (FHIR) specification and the Observational Medical Outcomes Partnership (OMOP) common data model. This is currently missing in the FL setup for the food- and agricultural domain.

Secondly, so far, there are only a limited number of AI models implemented in the FL environment. Some of them are commercial in nature that should provide solutions for a range of organizations. Another open-source model that have just been made available in a federated environment is the Generalized Linear Model (GLM)[20]. For large datasets, simple models (summary statistics) are often sufficient for demonstration purposes because complex statistical model algorithms do not scale for large datasets. Apart from that, most complex statistical models are designed to run on single devices in a centralized setting. Modern algorithms like deep learning models need to be redesigned to leverage the power of a FL setup.

Thirdly, the efficiency of programming languages and tools play an important role in operationalizing a scalable infrastructure. Most ML models are created using languages such as Python or R. These languages allow a researcher to quickly create models for research purposes. However, these models are difficult to operationalize because they have issues with data structures, as most of them work with tabular data and are incompatible with web data formats such as JSON. While there are some packages in these languages that take care of some of the problems, they are not inherently efficient. Most of the models created in a FL environment are dockerized. Docker provides an environment that allows a ML model to be reusable and reproducible by taking into account all the dependencies that a ML model requires. However, since both Python and R are interpreted languages, Docker images created with these languages are very large, resource intensive, and require good network bandwidth and CPU resources. Recently, efforts are being made to create models in modern compiled languages such as Golang and Rust. Since the ML models created using these tools are in binary format, all dependencies are included in them, making these models more efficient both CPU and in terms of network bandwidth. In future it is important that more ML models be build using compiled languages.

Note that in the Farm to Fork strategy, the Commission announced its intention to convert the FADN into a Farm Sustainability Data Network (FSDN). The FSDN initiative will expand the scope of the current FADN network to also collect farm level data on environmental and social farming practices, as well as provide farmers with reporting on their farm performance. The data collection methodology will be in line with the existing one for the current FADN.

[20]Cellamare, Matteo, et al., 2022 A Federated Generalized Linear Model for Privacy-Preserving Analysis. Algorithms 15(7).

Although increased attention to expand data collection efforts some national FADNs will capture more data than required for FSDN. This will allow the FADN network to judge the potential of FL whether it is proven or further demonstration is needed to exploit the potential for the agricultural system.

There is a potential contribution of data stations in the concept of data train that supports interoperability and sharing of sustainability and economic data to support monitoring and evaluation for European policy analysts. However, despite these promising results, questions remain that should be taken into account before expanding the limited set of variables. Main side conditions for the success of the data train concept is the availability of relevant stakeholders, availability of aggregated data on national level as well as the farm level.

From an organisational perspective, it is important to prevent possible fraud cases (intentionally and unintentionally) that could become easier to happen or already happen in the current system. Especially, in case of extending the type of participants in the ecosystem, such as authorities and industries. For example, within the Netherlands there have been fraud cases on manure that were brought forward as swindles with figures to get more subsidy. Although the news articles have been criticised on the image damage for the sector, it still had a negative impact[21]. Possible mitigations to address this could be taken with a solid data governance framework, with a special attention on data quality and data validity.

Another limitation for this study could be the scope of data-sharing, which could be broadened from FAIR to other frameworks. Several national, international and European initiatives are in seek of ecosystems that ensure alignment between the interest of the individual and the rest of the ecosystem for data-sharing. Key requirements are trust between participants, data interoperability and data sovereignty (Braud, et al. 2021; Tardieu 2022). However, for the European initiative GAIA-X, special attention is needed on the strategy[22]. Another relevant framework is the iSHARE Scheme which is, mainly in the Dutch logistics sector, a collaborative effort to improve data-sharing with guidelines on identification, authentication and authorisation of inter-organisational data attributes. The iShare Scheme is a combination of agreements on functional, technical, operational and legal agreements to which participants should adhere[23].

To develop a full picture of data-sharing concepts, additional studies will be needed that assess comparatively the differences between approaches, its key benefits and disbenefits. For example, the Californian approach for the protection of consumers privacy seems to have different requirements and could complement the GPDR. From an architectural, more attention could be given in future studies on the different approaches like centralized, distributed and decentralized, besides federated[24].On the other hand, the Solid project aims to "empower an equitable, informed and interconnected society". The protocol, which is published, realises a space where individuals can maintain their autonomy, control their data, privacy, applications and services to accomplish their needs.

[21] Havermans, Onno, 2021 Onderzoek: boeren sjoemelen op grote schaal met mest en dat wordt nauwelijks gecontroleerd. Trouw.

[22] Autolitano, Simona, and Agnieszka Pawlowska, 2021 Europe's quest for digital sovereignty: GAIA-X as a case study. IAI Papers 21(14):1-22.

[23] Jansen, Vincent; Westenbert, Remco; Nijs, Pieter; Hoppenbrouwer, Denise, 2019 iShare Trust Framework. In Goals and scope of the iSHARE Scheme. i.

[24]Gao, Y, et al., 1993 Comparison and analysis of centralized, decentralized, and federated filters. Navigation 40(1):69-86.

The FAIR Data Point could be complemented with approaches such as the Minimal Interoperability Mechanism (MIM's) to achieve semantic interoperability on the data level for decision making. Especially for cross domain interoperability, such as the Internet of Things and Smart Cities, it is expected to have significant benefits for the society[25].

2.7. Demonstration Case #4: Spain

In order to test the monitoring of Spanish Strategic Plan eco-scheme (CAP) "Low Carbon Agriculture – (P1) Increasing the carbon sink capacity of pastures by promoting extensive grazing eco-scheme" payments, the Spanish demonstration case supports sheep breeders in cooperatives to easily demonstrate to payment agencies that their herds comply with the requirements to receive this aid, through the development of an innovative methodology.

For the extensive grazing eco-scheme, on eligible hectares of permanent or temporary pasture on which grazing is declared, the following must be carried out:

- The areas must be grazed on a daily basis with the holding's own animals for a minimum period of 120 days per year, either continuously or discontinuously. For this purpose, the grazing time is calculated at farm level, so that all the hectares of pastureland used for grazing must be grazed, but not necessarily each one of them must be grazed every day of the year.
- Respect the following intervals for the agro-ecological regions:
    - In wet pastures, between a minimum stocking density of 0.4 livestock units (LU) per hectare and a maximum stocking density of 2 LU/hectare.
    - On Mediterranean pastures between a minimum stocking density of 0.2 LU/hectare and a maximum stocking density of 1.2 LU/hectare.
- If the farmer does not intend to graze all the animals on his holding, he must state at the time of application the number of animals on his holding, by species, with which he intends to graze.
- Beneficiaries must include in the single application form a list of the agricultural parcels used for grazing by the animals on each livestock holding.

To carry out this demonstration case, three farms of sheep breeders' members of the EA Group cooperative have been engaged. The holdings are located in the Extremadura Region, at the Alburquerque, Villanueva de la Serena and Torrejón el Rubio municipalities. Some relevant features of the participating livestock farmers:

- Flocks of around 1.000 sheep.
- One or two batches (250-300 sheep each) will be monitored in each flock using GPS trackers.

[25] Kalatzis, Nikos, et al., 2019 Semantic interoperability for iot platforms in support of decision making: an experiment on early wildfire detection. Sensors 19(3):528.

- The complete cycle will be monitored (postpartum recovery, mating, gestation and lambing).

All of them fell into the agro-ecological region of Mediterranean pastures, as it can be appreciated on the map below.

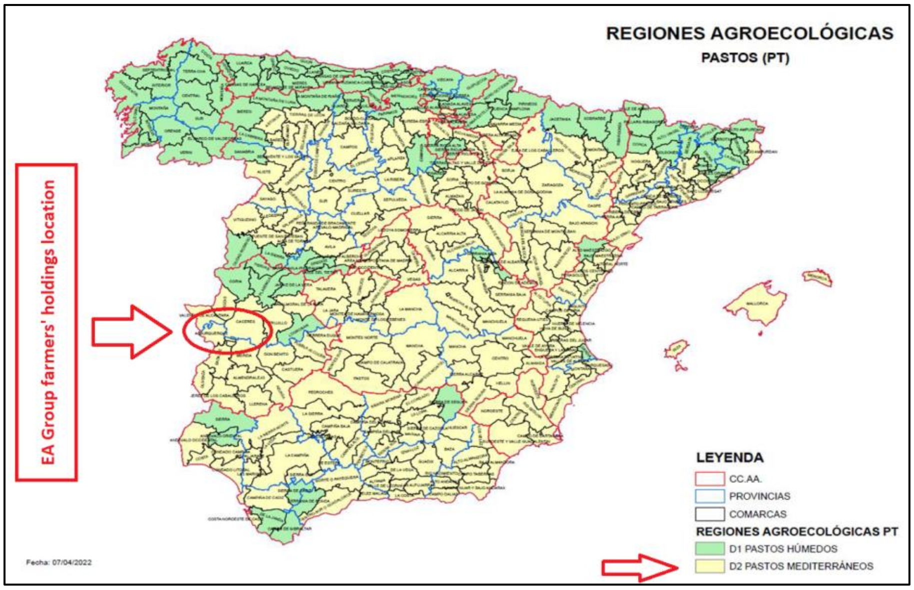

Figure 7: Farmers' holding location

In the practice, this fact means that the intervals of minimum and maximum stocking of these holdings should stay between 2 and 12 sheeps per hectare.

Thus, to prove that the farmers comply with all the abovementioned requirements of the extensive grazing eco-scheme, the Spanish Demonstration case developed a system that integrates the near-real-time herds position (GPS trackers) in the field and the areas with the most intensive grazing activity for sheep breeders to improve the understanding of the movements of the herds.

Specifically, the Demonstration case defined the workflow to translate the herd position collected from GPS devices attached to some sheep into geo-referenced information ready to be incorporated in a Geographical Information System (GIS). Once the position was processed and ingested within the GIS environment, it brought together herds position, Satellite data and farmers' LPIS information. Indexes derived from satellite imagery support the identification of those most-intensive grazing areas within the LPIS parcel and the information derived from the positioning system will show the actual movements of the herd in the field. This also may become a probe of grazing on that parcel.

The data that have been used are in- and out-farm data, including: LPIS, Herd characteristics (type, size, ages), feeding and growing average indicators by herds, Sentinel-1, Sentinel-2, GPS data tracks, meteorological information (AEMET network), etc.

At least 25 GPS trackers/necklaces have been displayed over these farms to keep track of different batches from the flocks during the whole production cycle of the animals. The GPS trackers use different technologies such as SIGFOX (some of them provided with a SD memory card for those areas with lower connectivity) or GSM, providing not only the position, date and time of the batches in near-real time, but also animal temperature and accelerometer parameters.

In addition, farmers play a relevant role for the final assessment by providing their first-hand opinion on the economics and the level of administrative burden derived from the proposed system.

The planification of the demo was started in early 2022 and its development is foreseen till the end of the project life (2023).

Spanish Co-ops (Agri-food Cooperatives of Spain is the organisation that represents at national and European level the social and economic interest of the Spanish Agri-food cooperatives) develops, in collaboration with three livestock farmers (end-users) from the EA Group cooperative (sheep breeding) and other IT companies, a monitoring methodology based on a GIS platform (ESRI).

The demonstration case shares in common with demonstration case #2 a GIS system where the information is gathered and analysed in real time. Moreover, similarly than in demonstration case #2, for the Spanish Strategic Plan eco-scheme (CAP) "Low Carbon Agriculture - (P1) Increasing the carbon sink capacity of pastures by promoting extensive grazing eco-scheme", there is a specific requirement to provide information to the administration through the mandatory digital farmbook, in order to get the payments.

In particular, the dates of entry and exit of the animals to and from each agricultural grazing parcel shall be recorded in the digital farm logbook (no later than 15 days after the end of the grazing period). Thus, the development of the digital farmbook will be not only useful for the purposes of demonstration case #2 but also for those of demonstration case #4.

Objectives and expected outcomes

The main objectives of this demonstration case are to develop a replicable methodology for herd monitoring in extensive livestock production cooperatives, integrating into a GIS, EO data (Copernicus) and near-real time GNSS positioning services; to easily demonstrate to the paying agencies that the herds are grazing outside at least 120 days per year with a cattle load between 0,2-1,2 LSU/ha; to provide proxy-information about the livestock carbon sequestration in the soil taking the average amount of manure/ slurry produced by the sheep to put in value what is returned to the soil in the grazing process (extraction vs fixation); to provide information on the degree of intensification or abandonment of grazing areas (land use); and to extract useful conclusions on the performance and suitability of the different technologies.

Main actions performed

The main actions that have been performed during this period for the Spanish demonstration case are the following:

- Set the time plan and milestones
- Meetings for definition and agreement on terms of the collaboration and participants

- Set the livestock collars and correct issues
- Set the GIS and design its performance while feeding it with the info coming from the GPS trackers
- Meetings with the policy makers and RDs follow up

3. Conclusions

All four demonstration cases have different approaches and test various technologies but the goal is one; to explore and determine whether the adaptation of those technologies could be facilitated through these cases.

Despite the fact that the demonstration cases are still in progress, initial feedback is that new technologies and data collection processes could improve the monitoring and evaluation frameworks in several ways:

- Remote sensing: Remote sensing technologies, such as satellite imagery and drones, could be used to gather data on land use, crop health, and environmental conditions. This could help to identify patterns and trends that could inform policy-making.
- Machine learning: Machine learning algorithms along with other external data could be applied to monitoring and evaluation framework to identify patterns and predict future trends. This could help to improve accuracy of forecasting and policy-making.
- Participatory data collection/ In-situ data: Participatory data collection methods, such as mobile applications, crowdsourcing or even FADN, could be used to collect data directly from farmers. This could help to reduce the burden of data reporting and improve the accuracy of data.

Overall, the integration of new technologies and data collection processes could help to improve the accuracy, timeliness, and transparency of the data management and processing. This could provide a more reliable and comprehensive source of information for agricultural policy-making, ultimately leading to more effective and sustainable policies.

