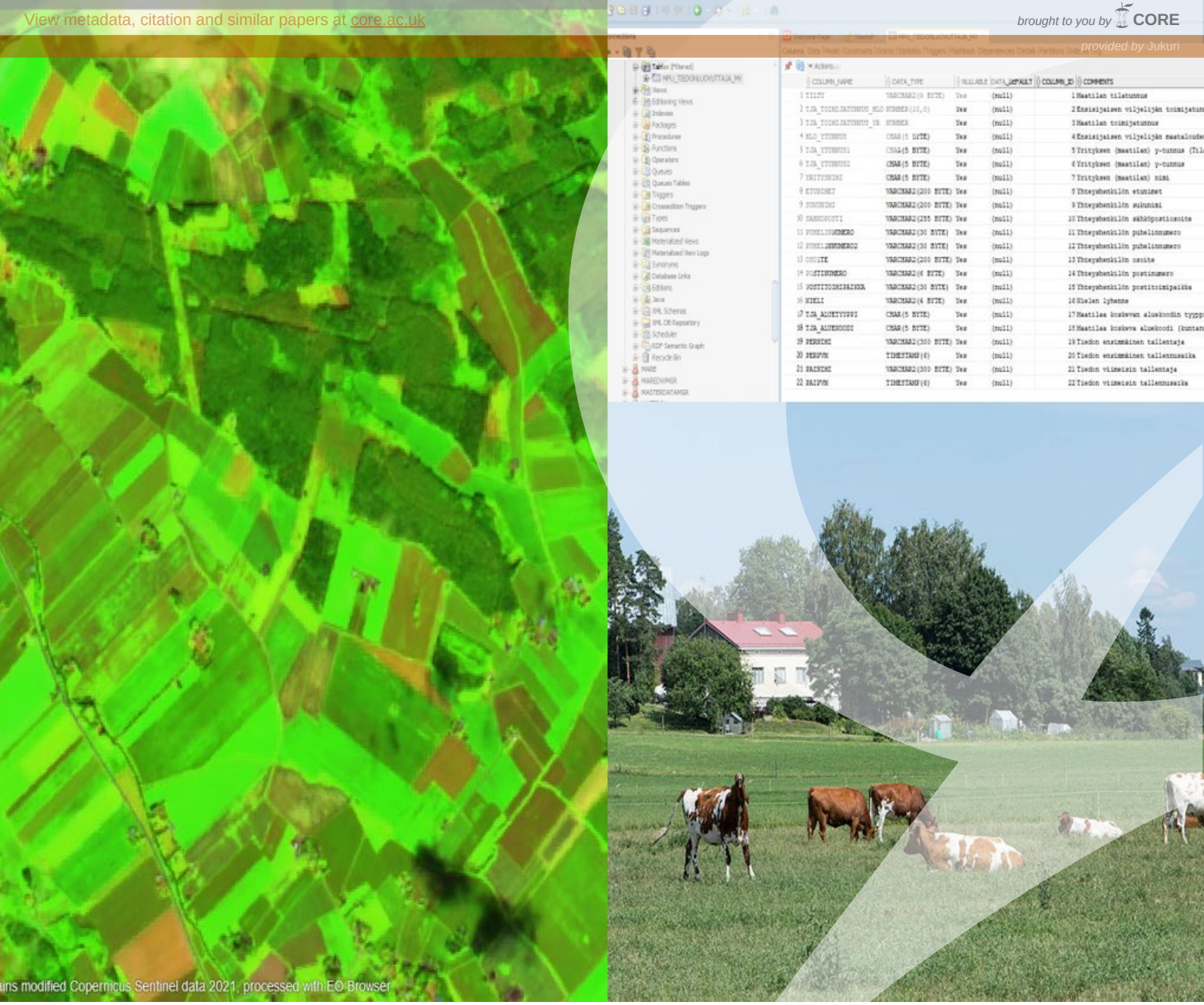


Natural resources and bioeconomy studies 28/2021

# Modernisation of agricultural statistics

## Final report

Anna-Kaisa Jaakkonen, Jaana Kyyrä, Pasi Mattila, Esa Katajamäki,
Maria Yli-Heikkilä, Hanna Kastikainen and Mika Sulkava

Luke
NATURAL RESOURCES
INSTITUTE FINLAND

Natural resources and bioeconomy studies 28/2021

# Modernisation of agricultural statistics

## Final report

Anna-Kaisa Jaakkonen, Jaana Kyyrä, Pasi Mattila, Esa Katajamäki, Maria Yli-Heikkilä, Hanna Kastikainen and Mika Sulkava

Natural Resources Institute Finland, Helsinki 2021

**Recommended citation:**

Jaakkonen, A-K., Kyyrä, J., Mattila, P., Katajamäki, E., Yli-Heikkilä, M. Kastikainen, H. & Sulkava, M. 2021. Modernisation of agricultural statistics : Final report. Natural resources and bioeconomy studies 28/2021. Natural Resources Institute Finland. Helsinki. 35 p.

Modernisation of agricultural statistics
Finland Grant agreement 2018.0212
Final report 20.8.2020
Co-funded by Eurostat

Anna-Kaisa Jaakkonen, ORCID ID, https://orcid.org/0000-0002-4024-8815


ISBN: 978-952-380-196-7 (Print)
ISBN: 978-952-380-197-4 (Online)
ISSN 2342-7647 (Print)
ISSN 2342-7639 (Online)
URN: http://urn.fi/URN:ISBN:978-952-380-197-4

Authors: Anna-Kaisa Jaakkonen, Jaana Kyyrä, Pasi Mattila, Esa Katajamäki, Maria Yli-Heikkilä, Hanna Kastikainen and Mika Sulkava
Publisher: Natural Resources Institute Finland (Luke), Helsinki 2021
Year of publication: 2021
Cover photo: Sentinel map, Erkki Oksanen (cows), print screen from database
Printing house and: publishing sales: Juvenes Print, http://luke.juvenesprint.fi

# Summary

Anna-Kaisa Jaakkonen, Jaana Kyyrä, Pasi Mattila, Esa Katajamäki, Maria Yli-Heikkilä, Hanna Kastikainen and Mika Sulkava.

Natural Resources Institute Finland (Luke), Helsinki

The "Modernisation of agricultural statistics" project had three main goals: (1) modernisation of the statistical register of agricultural and horticultural enterprises (Mapu), (2) identification of new data sources that use agricultural statistics, and (3) the better utilisation of farm related geospatial data to define the location of farms. In addition, an international seminar was held in Poland during the project in cooperation with other countries carrying out a similar project and with representatives of Eurostat.

The project was divided into work packages in accordance with the main goals. The first work package included the modernisation of Mapu. This work package was further divided into two main tasks based on data sources: administrative data and direct data collection. Before the concrete modernisation of systems and databases, we identified how well the definition of a farm used in Mapu corresponds with the definition of a farm used in EU statistics (Article 2 of regulation (EU) 2018/1019 on integrated farm statistics). In practice, the definition of a farm entity used in Mapu in accordance with the Integrated Administration and Control System (IACS) corresponds well with the definition used in EU statistics. The most significant difference is that Mapu mainly does not include fur farming, reindeer husbandry and beekeeping. Apart from fur farming, the significance of these sectors is low, and this project does not recommend that enterprises engaged in these sectors be added to Mapu. Fur farms are included in the Business Register of Statistics Finland, which can be used to produce statistics of fur farms. If statistical obligations related to fur farming expand later, for example, as a result of the SAIO regulation, adding fur farms to Mapu should be reassessed.

Mapu's databases and related data flows were modernised during the project. The removal of direct database rights to datasets of agricultural administration for reasons related to information security was a significant change regarding administrative data. The Hansolo loading platform was built as a new solution. The administrative data required for agricultural statistics is copied regularly in Hansolo, from which data is available for Mapu and agricultural statistics. In addition to databases and data flows, the data content was modernised. Adding the contact details and business ID of farm parties, such as farmers, to the data content was the most significant addition. The data system of the administrative body of rural industries, belonging to IACS, is the primary data source. Data about animals is obtained from animal registers. In other words, units (farms and horticultural enterprises) in Mapu are the same as units in IACS.

Some data in Mapu can be obtained through direct data collection processes in conjunction with the survey for horticultural statistics. During this project, the data collection application, databases and data flows of this survey were modernised regarding all process phases, ranging from data collection to publication. As a new feature in data collection for horticultural statistics, farmers were able to log in to the online application through strong electronic identification using their banking codes, for example. As a result, it was not necessary to send usernames and passwords separately to respondents. In 2020, the data collection application was developed further by introducing an authorisation feature. In practice, this means that farmers can authorise another person to respond to the survey on their behalf.

In the second work package, we identified the broader use of data about farm coordinates to define the location of farms in agricultural statistics. The aim is to present data on maps better than before, while considering the data privacy of individual farms.

In the third work package, we examined what new data sources could be used more effectively than at present in agricultural statistics. According to the examination, data in the parcel databank about the use of crop protection products apparently represented the actual use of products better than data collected through the current data collection process. The use of parcel specific data obtained from farms for various purposes is developing and, as a result, parcel data may also become available as source material for statistics by new electronic means.

In addition, we developed a method to calculate the field-specific monoculture area. Compared with the previous method version, the use of the QuantumGIS program was a new factor. It helped to accurately define the area used for a single crop in consecutive years, also when the crop was only grown in part of the parcel in question during the three years under examination. The method for calculating the monoculture area developed in this project will be used in the 2023 IFS data collection process to produce data about crop rotation.

Considering the future of agricultural statistics, it will be interesting to see how new production sectors, such as insect farming, will possibly be included in the statistics. This also applies to questions related to production methods: in the future, food may be produced by enterprises other than conventional farms, such as by enterprises engaged in vertical farming. Environmental issues have steered agricultural activities in the past and will continue to do so during the next few years. For example, sequestering carbon in farmland may expand beyond perennial grasslands and fallows. Data production related to products sold, combined with statistics of environmental impact arising from agriculture, presents a challenge in agricultural statistics.

Growing and new statistical needs, together with fewer resources, will steer statistical production during the next few years towards the use of new data sources and the identification of data production that is "sufficiently good" in terms of quality and finances.

Modernisation of agricultural statistics co-funded by the Eurostat (Finland Grant agreement 2018.0212). This publication is the final report of the project (20.8.2020).

**Keywords:** agriculture, statistics, farm register, farms, Eurostat, Grants, geolocation, geospatial data, statistical database, administrative data, data collection

# Contents

# 1. Statistical register of agricultural and horticultural enterprises (Mapu)

The Finnish statistical farm register is entitled the register of agricultural and horticultural enterprises (often called "Mapu"). The Mapu register has a long tradition: it and its predecessor, the statistical farm register, were established in 1972 based on farm data from the previous year.

Currently, Mapu's data sources include administrative registers and an annual statistical survey sent to horticultural enterprises. When this project started, Mapu's database structures were outdated, and their manual updating was time consuming. What is more, outdated hardware was used to run the data collection application for the statistical survey sent to horticultural enterprises and the database structures were impractical.

Merging Mapu's data with datasets outside the scope of agricultural administration was difficult due to nonharmonised identification data. In addition, it is known that not all sectors that, according to the regulation on integrated farm statistics (IFS), are, for example, within the scope of the sampling frame for agricultural accounting are covered by Mapu.

One of the goals of this project was to fix the aforementioned deficiencies as comprehensively as possible. A second goal was to modernise Mapu-related databases and data flows both in terms of data obtained from administrative registers and data collected through the statistical survey sent to horticultural enterprises.

Another goal was to improve the opportunities to merge data, for example, with data of Statistics Finland and to identify how Mapu could possibly be expanded to cover new sectors that it has not included: fur farming, reindeer husbandry and beekeeping.

## 1.1. Identification of new data sources

Collecting data directly from farmers is expensive and increases the response burden. Other optional data sources should therefore be identified for statistical purposes, and their usability and quality should be examined at the same time. The aim in this project is find and evaluate new data sources.

## 1.2. Use of geospatial data

Farm data includes volumes of geospatial data, such as the location of farms (farm buildings), livestock buildings and field parcels, including field crops. The aim is this project, for example, to identity how well the location of a farm building represents the location of production and how crop rotation can be calculated using parcel data. The use of geospatial data also involves data protection issues and the blurring of geospatial data.

# 2. WP1 a Modernisation of the register of agricultural and horticultural enterprises

The purpose of this work package was to modernise the data content, databases and data flows of the Mapu register and to define its target groups, i.e. farms. The majority of data about farms and horticultural enterprises is updated based on data in administrative registers. In addition, data about cultivated areas and crops is obtained annually by collecting data directly from horticultural enterprises, and this data collection process is used as a data source for Mapu. In this project, the Mapu register was modernised regarding both data updated based on administrative data and data collected directly from horticultural enterprises in line with this project's goals. The first phase of the work package was to examine and clarify the definition of a farm in the Mapu register.

## 2.1. The definition of a farm

In Finnish agricultural statistics, the definition of a farm is based on a number of different definitions of a farm, the most significant of which include the definitions of Eurostat's agricultural statistics, the FAO's World Programme for the Census of Agriculture and the administration of agricultural subsidies. In the Mapu register, the definition of a farm has, in practice, been in line with the definition used by the administration of agricultural subsidies. Next, these definitions of a farm will be presented, after which we will examine how well the definition used in the Mapu register fulfils the needs of the EU's agricultural statistics.

### a. Definition of a farm in the EU regulation on integrated farm statistics

*According to Article 2 of regulation (EU) 2018/1091 on integrated farm statistics, 'agricultural holding' ('farm') means a single unit, both technically and economically, which has a single management and which undertakes economic activities in agriculture in accordance with Regulation (EC) No 1893/2006 belonging to groups A.01.1, A.01.2, A.01.3, A.01.4, A.01.5 or to the "maintenance of agricultural land in good agricultural and environmental condition" of group A.01.6 within the economic territory of the Union, either as its primary or secondary activity. Regarding activities of class A.01.49, only the activities "Raising and breeding of semidomesticated or other live animals" (with the exception of raising of insects) and "Beekeeping and production of honey and beeswax" are included.*

### b. Definition of a farm by the FAO

Another definition of a farm significant for agricultural statistics is given in the FAO's World Programme for the Census of Agriculture 2020, Volume 1, Programme, concepts and definitions
(http://www.fao.org/3/a-i4913e.pdf):

*"An agricultural holding is an economic unit of agricultural production under single management comprising all livestock kept and all land used wholly or partly for agricultural production purposes, without regard to title, legal form or size. Single management may be exercised by an individual or household, jointly by two or more individuals or households, by a clan or tribe, or by a juridical person such as a corporation, cooperative or government agency. The holding's land may consist of one or more parcels, located in one or more separate areas or in one or more territorial or administrative divisions, providing the parcels share the same production means, such as labour, farm buildings, machinery or draught animals."*

*There are two types of agricultural holdings: (i) holdings in the household sector – that is, those operated by household members; and (ii) holdings in the nonhousehold sector, such as corporations and government institutions. In most developing countries, the majority of agricultural production is in the household sector. The concept of "agricultural holding" is therefore closely related to the concept of "household".*

c. **Definition of a farm by the administration of agricultural subsidies** (source: https://www.ruokavirasto.fi/globalassets/tietoa-meista/asiointi/oppaat-ja-lomakkeet/viljelijat/tuet-ja-rahoitus/oppaat-ja-esitteet/viljelijatukien-hakuopas-2020_saavutettava.pdf)

In direct subsidies and programme based farmer subsidies, a farm covers all units managed by a single farmer and used for agricultural activities, located in a single member state. In other words, all production units (agricultural holdings) of a single farmer or individual farmers located in Finland are regarded as a single farm.

If a farmer operates several units used in agricultural activities in different parts of Finland, and they constitute a single administrative entity, i.e. they are managed by the farmer, data about these units only needs to be provided using a single application.

d. **Definition of a farm in the register of agricultural and horticultural enterprises**

In the Mapu register, the definition of a farm is in line with the definition of the administration of agricultural subsidies (IACS). In practice, the definition is as follows:

**Registered unit (farm):**

A farm covers all units managed by a single farmer and used for agricultural activities, located in a single member state. In other words, all production units (agricultural holdings) of a single farmer or individual farmers located in Finland are regarded as a single farm (in line with the IACS definition).

**Restriction based on industrial sectors:**

Mapu includes the following sectors in accordance with the standard industrial classification (TOL/NACE):

- 011 Growing of non-perennial crops
- 012 Growing of perennial crops
- 013 Plant propagation
- 014 Animal production (apart from 0149 (Raising of other animals))
- 015 Mixed farming

**Lower limit of a farm:**
Mapu covers all farms engaged in agricultural production. No lower limit has been defined for a farm in Mapu. Mapu acts as a sampling frame for agricultural statistics and, in this way, a lower limit can be defined separately for each statistics. In statistics that represent the structure of farms, the lower limit of a farm has been defined in accordance with the standard output (SO). The lower limit is an SO of EUR 2,000.

The aforementioned definition of a farm in Mapu fulfils the characteristics of the EU's IFS regulation and the FAO definition. These definitions are compared in the following table.

| Definition of a farm | Unit | Industrial sectors | Notes |
|---|---|---|---|
| EU IFS regulation | A single unit, both technically and economically, which has a single management and which undertakes economic activities in agriculture in accordance with Regulation (EC) No 1893/2006 belonging to groups | A.01.1, A.01.2, A.01.3, A.01.4, A.01.5 or to the "maintenance of agricultural land in good agricultural and environmental condition" of group A.01.6 within the economic territory of the Union, either as its primary or secondary activity. Some activities belonging to group A.01.49. | Agricultural sectors have been described in more detail here than in other definitions. |
| FAO: | An agricultural holding is an economic unit of agricultural production under single management | all livestock kept and all land used wholly or partly for agricultural production purposes, without regard to title, legal form or size. | |
| Mapu/IACS | A farm covers units managed by a single farmer and used for agricultural activities, located in a single member state. | Agricultural activities | |

In practice, the definition of a unit in the EU's IFS regulation ('a single unit, both technically and economically, which has a single management') is identical to the definition in IACS/Mapu ('units managed by a single farmer and used for agricultural activities, located in a single member state').

Instead, the definitions differ from one another with regard to the profitability of sectors related to agricultural production or agriculture. In particular, Mapu does not include farms or enterprises belonging to sector 01.4 (Animal production) and its subcategory 01.49 (Raising of other animals). The most significant Finnish sectors missing from Mapu are fur farming, reindeer husbandry and beekeeping. Then again, enterprises engaged in these activities are partly included in Mapu through field crop production, for example. During this project, no changes were made in the definition of a farm in Mapu. However, any needs for later changes, considering the profitability of different sectors, have been assessed in Section 2.4.

## 2.2. Data content and publication schedule of the register of agricultural and horticultural enterprises

**The modernised Mapu register acts as a basic statistical register for farm statistics.** The main tasks of a basic statistical register are (a) to define the population/sampling frame, and (b) to ensure that basic registers include the identification data required to merge units included in the basic statistical register with other registers.

To carry out these main tasks more effectively, it was decided that the data content of Mapu be expanded during this project. Another key reason for expanding the data content was the removal of Luke's direct database rights to administrative registers. For example, contact details of farmers were added to Mapu's data content. Previously, the contact details required were retrieved directly from administrative databases. In addition, certain common variables, such as data about organic farms, needed in different agricultural statistics were added to the data content. Needs arising from agricultural accounting, in particular, were considered during the modernisation of the data content.

Challenges in the compilation of Mapu result from the updating cycles of data source systems, administrative registers mainly. During the project, we discovered that there is a need for **a preliminary Mapu register**, which can be used to compile agricultural statistics for the current year. Currently, **the final Mapu register** has only been completed in April of the following year (n+4) when poultry data has been obtained.

In the future, the purpose is to complete the preliminary Mapu register for the current year in August when crop areas and cattle and pig data are available. Sheep and goat data is preliminary, and poultry figures can be estimated based on the previous year's data. Using this data, the preliminary SO and production sector of farms can be calculated. These are needed in sampling for agricultural statistics, for example.

The final Mapu register will continue to be completed in April of the following year.

**Data content of the modernised Mapu register**

**a) Data about farms**

- Business ID from farming administration
- Business ID from data merged by Statistics Finland
- Municipality
- Part of municipality
- Coordinates
- Legal form
- Type of ownership
- Production sector (national)
- Organic or not?
- Standard output (SO) of the farm in accordance with the national definition
- Total area of forest land on the farm
- Total area of roads and unproductive land on the farm
- Total plot area on the farm
- Total area of other land use on the farm

Derived calculated variables:

- Rounded location coordinates
- SO of the farm in the previous year
- SO of the farm in line with the production sector
- Production sector
- Livestock units on the farm
- Total area of basic parcels on the farm
- Total area of land parcels on the farm
- Total area of rented fields on the farm

**Field parcel data**

- Municipality
- Total area of basic parcels
- Ownership of basic parcels
- Cultivated areas per crop and production method

**Animal data**

- Number of animals per animal category and production method (organic production/conventional production)

### b) Contact details of farmers

- First names and last name of the principal farmer
- Email address
- Telephone numbers
- Postal address
- Language
- Business ID
- Operator ID

**In addition, identification data about the farmer's farm has been merged with farmer data:**

- Farm ID
- Name of enterprise/farm

Contact details of farmers are a new feature in the data content of Mapu. Mapu does not include personal identity codes of farmers. Instead, the operator ID acts as a unique identifier, which is an individual party specific surrogate key used in the field of administration.

### c) Data about operators/person involved in farm

- Farm ID (using this data farm parties can be connected to farms, to which they are parties)
- Operator ID (an individual party specific identifier used in the field of administration)
- Role on the farm
- Year of birth
- Gender
- Language
- Legal form of the party

In addition to the farmer, persons involved in farm includes the farm owner and the farmer's spouse. Often, the farm owner is the same as the farmer, but the farm can also be owned by a company.

## 2.3. Data structures and flows of the register of agricultural and horticultural enterprises

Mapu's databases and related data flows were modernised almost completely during the project. In addition to modernising the register technically, the goal was to improve the ability to merge farm data. As already stated in this report, the main data sources of Mapu are registers of agricultural administration and the statistical survey sent to horticultural enterprises. Next, we will describe how databases and data flows regarding the Finnish Food Authority's registers were modernised in this project. Changes regarding the statistical survey sent to horticultural enterprises have been presented in Section 3.

### 2.3.1. A new database solution for administrative data

In Luke's statistics, a limited number of employees have had rights to directly access databases managed by the Finnish Food Authority. These databases consist of data about agricultural subsidies paid to farms and animal registers, for example. For reasons related to organisational changes (transfer of databases from the National Land Survey of Finland to the Finnish Food Authority, which also acts as a supervisory authority) and data protection, these rights to directly access the databases were removed.

A new solution, **the Hansolo loading platform**, was therefore developed for using administrative data. Hansolo provides Luke with a window to the most important data sources, i.e. administrative databases. Technically speaking, Hansolo is a separate database. Structures similar to administrative database structures have been built in Hansolo regarding the data required in statistical production. Data is updated in the structures through a daily extract, transform, load (ETL) procedure. In agricultural statistics, administrative data is used protected from Hansolo, and the supervisory authority cannot view what data the producers of agricultural statistics use.

The data content of Hansolo cannot be expanded if any new statistical data is added to administrative registers. In the future, Hansolo can also be used for data transfers between other authorities, such as between Statistics Finland and Luke's agricultural statistics. However, this is still under development.

### 2.3.2. Linking the Business Register of Statistics Finland with the register of agricultural and horticultural enterprises

One of the key goals of the project was to make it easier to merge Mapu with other datasets. In practice, this was built so that identification data (business ID) used in the Business Register of Statistics Finland was added to Mapu's data content. During the project, business IDs from Statistics Finland were not added to Mapu's databases, but a plan to update business IDs and the database structures were completed. Business IDs will be updated in the database during this year.

Data sources used in the statistics on the finances of agricultural and forestry enterprises of Statistics Finland include registers of the tax administration and the Ministry of Agriculture and

Forestry of Finland and datasets of Luke's agricultural statistics. Available registers within the scope of the Ministry of Agriculture and Forestry of Finland include the statistical register of agricultural and horticultural enterprises maintained by Luke and data included in the Integrated Administration and Control System (IACS) system maintained mainly for the administration of agricultural subsidies, including customer data and farm specific crop production, domestic animal and subsidy data. This structure of data systems is also called the rural industry register. IACS includes business IDs for farms, the party/owner of which is an enterprise, i.e. the legal form of which is a limited liability company, limited partnership or general partnership. (Figure 1). However, these farms comprise a very small part of all farms.

In merging registered data from agricultural and tax administration, the largest challenge has been the merging of data based on two different registered units. Agricultural administration mainly uses the farm ID as the registered unit, whereas the tax administration uses the personal identity code or business ID.

"Farm" mainly represents a physical unit, which consists of an entity comprising specific hectares, buildings and domestic animals, while "person" is not an unambiguous concept. "Enterprise" cannot always be defined unambiguously: although an enterprise has a business ID, it may also, in certain rare cases, comprise an entity of several legal units (business IDs). In terms of businesses, a farm can be an enterprise's local unit or operational unit. In the Business Register of Statistics Finland, "local unit" means an enterprise operating in a specific geographic location or its part. "Operational unit" has the same meaning as "local unit", while it is also defined in accordance with the industrial sector. The Business Register includes identification data for local units and operational units. The farm ID does not usually change when, for example, the farm ownership passes over to the next generation, while the business ID does.

Agricultural subsidies are applied for and paid on the basis of farm data, but they are always paid to a person or legal unit. This is why, in certain (rare) cases, the amount of subsidies can be difficult to allocate to a specific farm if the owner applies for subsidies for several farms. As such, "farm" is not a legal unit, and it therefore cannot have any legal form, although the farm register includes such data. If the owner changes or there are several owners, the legal form is not always unambiguous.

The customer register of agricultural administration includes data about all "roles" acting on a farm separately, i.e. their personal identity code and role type. Primary role types include the responsible farmer, the farm owner and the spouse.

Currently, tasks are distributed in merging data in the financial statistics of agricultural and forestry enterprises of Statistics Finland so that Statistics Finland merges data with a farm ID received from the field of administration of the Ministry of Agriculture and Forestry of Finland and other data with a business ID or personal identity code. This data is primarily merged using the subsidy applicant's business ID or personal identity code. One challenge is that farm roles can be parties to several farms and enterprises, or a single farm can submit several tax returns based on the act on income tax on agriculture or the business income tax act. Therefore, it is not always easy to identify the farmer in charge of a farm's finances or the enterprise in question. The most difficult cases are examined one by one.

Expanding agricultural statistics to financial statistics is a great leap, in which merging data from agricultural statistics (Luke) with financial statistics and tax data offers extensive high-quality data to compile statistics about agriculture from a broader societal perspective. This is why a key, the business ID, is needed to merge data. The accuracy, method and usability of merging will be iterated over time.

The business ID(s) can only be added to Mapu reliably if agricultural administration starts to use business IDs (one or more) alongside farm IDs in their activities.

Differences and challenges related to the business ID or farm ID of farms have also been studied in Sweden (Swedish Board of Agriculture). Their conclusions are fairly similar to ours: producing good registers calls for cooperation between decision makers and researchers. Good registers in turn help to reduce the data collection burden and costs in agricultural statistics.

The project's conclusions/recommendations for the use of business IDs in the future:

- Business IDs that are connected to farms owned by enterprises will primarily be imported into the new Mapu register from the Finnish Food Authority's databases.
- The business ID obtained through data merging carried out by Statistics Finland will be used as the secondary business ID.
- Data merging carried out by Statistics Finland will be used, where possible.
- Merged data will be imported into Mapu through the Hansolo platform.
- Crosschecks will be carried out and the usability of business IDs in data merging will be improved.
- The expanded use of business IDs will be investigated, meaning that the principal farmer is linked to their business ID, the owner is linked to their business ID, etc.
- Business IDs can be used to identify the financial activities and the full-time/part-time operations of farms in more detail.
- Dialogue over the use of business IDs in agricultural administration as well will continue between Statistics Finland and the Finnish Food Authority.

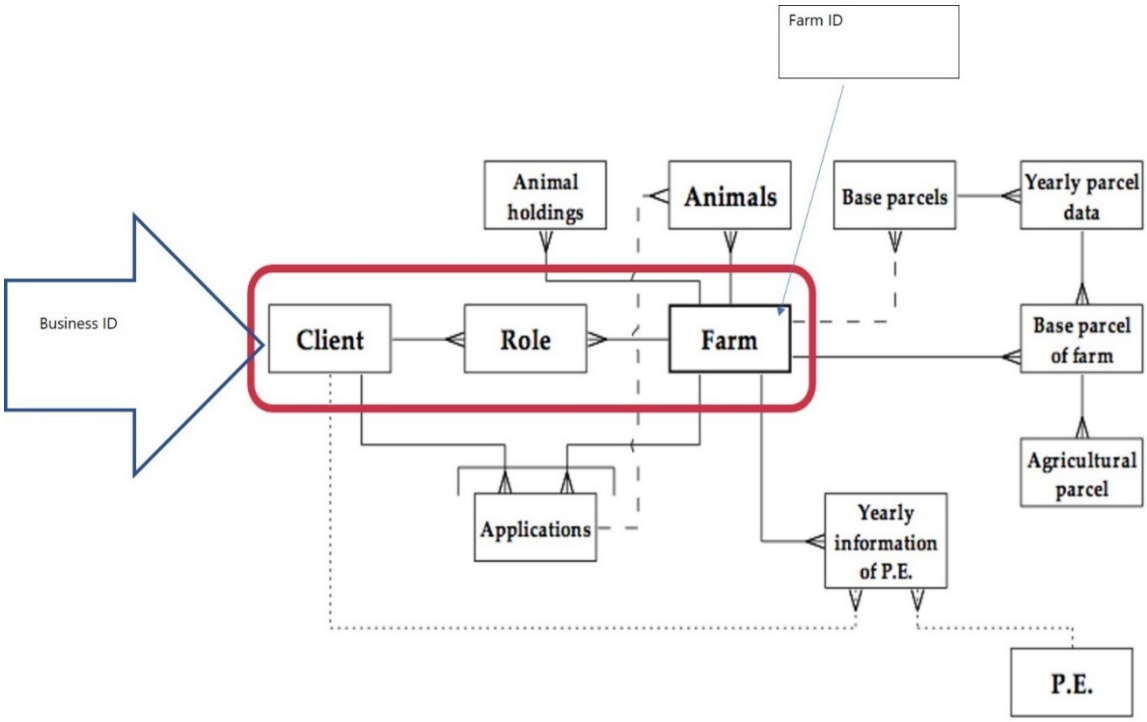


**Figure 1.** Business ID connected to a person (client) in agricultural administration systems.

### 2.3.3. Database structure of the register of agricultural and horticultural enterprises

The database structures of the modernised Mapu register have been built in TilastoP, a centralised database for the processing and storage of data. (Figure 2)

The content of the modernised Mapu register is located in the MAPUMGR schema. Its satellites include the KMMMRG schema, which consists of data about crops and fields, and the ELÄINTIETOMGR schema, which consists of data about farm animals. Data is retrieved into these schemas from copies of administrative databases located in Hansolo. In addition, Mapu's data is supplemented using business IDs retrieved from Statistics Finland through separate runs.

Data obtained through data collection related to horticultural statistics and agricultural accounting is saved in tables in PUREMGR and MLMGR schemas. Currently, Mapu is only updated to a small extent through statistical surveys. Data about greenhouse areas is mainly updated using data collected through horticultural statistics. If the availability of administrative data decreases for some reason, data can be collected and supplemented for Mapu through statistical surveys.

The JULKAISUMGR schema includes database structures related to the publication of statistics.

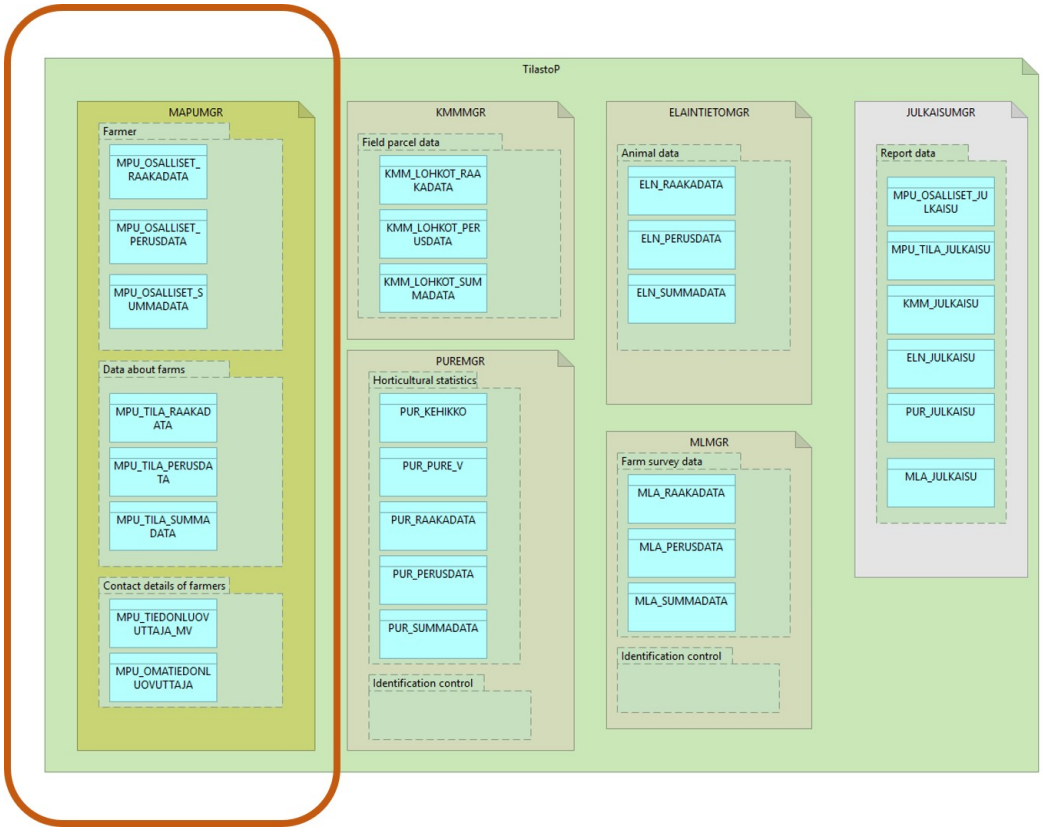


**Figure 2.** Structure of the modernised the register of agricultural and horticultural enterprises

The MAPUMGR schema includes separate table groups for data about farm parties and the farm. Data obtained directly from data sources is saved in raw data tables. Processed unit level data is located in basic data tables and summed data in sum data tables. Data ready for publishing, encrypted and summed at specific levels is imported into JULKAISUMGR tables, from which statistical data is published.

Contact details of farmers are used in the materialised MPU_TIEDONLUOVUTTAJA_MV view located in MAPUMGR. Behind this materialised view, a data flow is located, through which contact details of enterprises and farmers are imported into Hansolo from the shared customer data system (Asti) located in the Finnish Food Authority's environment and, with regard to contact details of farmers, from data located in the agricultural subsidy app. The Asti system uses data located in the Population Information System and the Business Information System, Finland's official registers. ETL procedures are used to import data masses into Hansolo overnight. Changed data is updated every hour during office hours. From data imported into Hansolo, contact details of farmers are filtered into the MPU_TIEDONLUOVUTTAJA_MV view.

Not all farmers apply for agricultural subsidies, which is why their contact details cannot be obtained through data loaded in MPU_TIEDONLUOVUTTAJA_MV. Contact details of these cases are saved in the MPU_OMATIEDONLUOVUTTAJA table.

## 2.4. Expanding the coverage of the register of agricultural and horticultural enterprises to new sectors

During this project, we did not change the definition of a farm in the Mapu register. Instead, we examined the coverage of the definition of a farm compared to the definition used in the EU statistics regulations. During this examination, we discovered, as was already known, that the sectors of fur farming, reindeer husbandry and beekeeping for honey and beeswax production, for example, were mainly missing from Mapu. The significance of the sectors currently missing from Mapu in Finland and the possibility to add these to Mapu will be discussed below. Fur farms, reindeer farms and beekeepers were last included in agricultural accounting in 1990. In addition, we assessed whether insect farming, a sector launched in Finland in recent years, has expanded so significantly that it should be added to Mapu as a new sector.

### 2.4.1. Fur farms

According to statistics of the Finnish Fur Breeders' Association, there were 967 fur farms in Finland in 2019. Finland is the third largest fur producer in Europe. As a producer of certified fox fur, Finland is the largest in Europe and the second largest in the world after China.

Some fur farms operate in conjunction with agricultural farms. According to the results of the 2016 Farm Structure Survey, 240 farms were engaged in fur farming in addition to agricultural production. In the Farm Structure Survey, fur farming is listed as a secondary occupation in agriculture. According to the definition of a farm in the IFS regulation (2018/1091), fur farms are classified as agricultural farms. In other EU laws related to agricultural statistics, such as the regulation on economic accounts for agriculture (EAA) and the regulation on statistics on agricultural inputs and outputs (SAIO, under preparation), the definition of a farm refers to the IFS regulation. Therefore, the EU's agricultural statistics use a harmonised definition of a farm.

According to the SAIO regulation under preparation, reporting data about the number of fur animals to Eurostat as part of domestic animal statistics will become mandatory. In addition, working hours accumulated on fur farms, for example, are included in agricultural working hours (ALI statistics) that are reported to Eurostat in conjunction with economic accounts for agriculture (EAA).

In practice, including fur farms in Mapu would mean the launch of a new data source or new data sources in updating the register. One challenge is that fur farms and agricultural farms

use different identification data. Fur farms are already included in the Business Register of Statistics Finland, in which the business ID is used as an identifier. The Business Register included (in 2018) 893 enterprises engaged in fur farming. As a statistical authority, Luke can have access to this data located in the Business Register and use it, for example, as a sampling frame during data collection. In this project, Mapu was developed so that it will be easier to merge data between Mapu and the Business Register using business IDs.

We recommend that the Business Register of Statistics Finland will still be used as a sampling frame when collecting data related to fur farms. If the SAIO regulation or other statistical obligations increase statistical obligations related to fur farms, for example, regarding the number of fur animals, Luke will obtain the contact details required for data collection from the Business Register.

### 2.4.2. Reindeer farms

According to the Reindeer Herders' Association, there were 4,354 reindeer owners in Finland and some 1,000 households received a significant livelihood from reindeer husbandry between 1 June 2018 and 31 May 2019. Citizens of a state belonging to the European Economic Area (EEA) who are permanently living in the reindeer husbandry area and reindeer herders' cooperatives can own reindeer in Finland. Reindeer owners from different areas, i.e. parties, form cooperatives that are reindeer husbandry units of different sizes in terms of their area and the number of reindeer. These cooperatives are responsible for reindeer husbandry in their respective areas.

The Ministry of Agriculture and Forestry of Finland sets the maximum permitted number of living reindeer for ten years at a time, considering, for example, the sustainable production capacity of winter pastures. Between 2010 and 2020, the permitted number was 203,700 living reindeer. According to the Reindeer Herders' Association, the number of living reindeer has been lower than 200,000 in recent years.

In the Farm Structure Survey, reindeer husbandry has so far been included in other business activities. According to the 2016 Farm Structure Survey, 300 farms were engaged in reindeer husbandry in addition to agricultural activities.

According to the definition of a farm in the IFS regulation (2018/1091), reindeer farms are classified as agricultural farms. The regulation's definition of a farm states: Regarding activities of class A.01.49, only the activities "Raising and breeding of semidomesticated or other live animals" (with the exception of raising of insects) and "Beekeeping and production of honey and beeswax" are included.

Considering EU statistics, reindeer farms do not, at present, need to be added to the Mapu register. The data content of the Farm Structure Survey does not include questions related to reindeer farms, and no variables related to reindeer have been planned in the SAIO regulation under preparation. In addition to Finland, reindeer husbandry is mainly carried out in Sweden in the EU and in Norway in the EEA. Measured by volumes, reindeer husbandry is not any significant activity in the EU, and reindeer farms do not need to be added to the Mapu register due to any EU obligations.

### 2.4.3. Beekeeping

It is estimated that there are some 3,400 beekeepers in Finland. This estimate is based on the number of members (2,750) in the Finnish Beekeepers' Association and on their estimate that

some 80 per cent of Finland's beekeepers are members of the association. There are around one hundred professional beekeepers and approximately 70,000 hives in Finland.

According to the definition of a farm in the IFS regulation, apiaries are classified as farms. In Finland, there are only few professional apiaries, and they are engaged in small-scale activities. This is why data about apiaries is regarded as a non-significant (NS) variable in Finland, and data about the location of apiaries on farms has not been reported to Eurostat. The SAIO regulation will not include any provisions on beekeeping.

Beekeeping is a fairly small-scale activity in Finland, are there are roughly 100 professional beekeepers. For this reason, we recommend that apiaries should not be added to the Mapu register at present. If the activity expands or there are new national or international data needs, adding apiaries to Mapu should be reassessed.

### 2.4.4. Insect farming

Finland approved insects as a food product in the autumn of 2017. Initially, it was believed that insect based food products will expand rapidly in the markets. Only the use of farmed whole insects for human consumption is permitted in Finland. As insects are classified as novel foods, a novel food approval must be applied for and obtained within a specific time for their use as food.

According to the Finnish Food Authority (2019), there are roughly nine inspect species approved at present as a food product in Finland, including house crickets (Acheta domesticus), bees, Western honeybee larvae (Apis mellifera) and mealworms (Tenebrio molitor).

Currently, insect farmers can register as **food business operators**, whose activities are within the scope of the food legislation and supervision. Insect products produced in this way can be marketed as food products. Operators who breed insects for use as food, i.e. primary producers, must **report their primary production location** to their municipality's food supervisor before any activities can be started. If a producer is already engaged in primary production in any production sector, such as pig farming or greenhouse production, they must make an addition to their primary production location report, indicating that they intend to start insect farming for food production purposes. This addition can be made by sending a free-format notification to the municipal food supervisor. In addition, primary producers must register in the Finnish Food Authority as **an operator in primary production in the feed sector**, even if they produce insects for use as food. **In the EU**, bred insects are classified as **agricultural animals**. However, they are not within the scope of agricultural statistics at present.

New insect companies and products entered rapidly in the markets, and there have been more than 50 registered insect operators in Finland. However, many companies have already been discontinued, and insect farming is only a small-scale activity at present, covered by roughly 20 companies. If insect farming was included in the Mapu register, data about relevant companies would be obtained from the Finnish Food Authority's registers through the farm ID or customer ID. Operators have a business ID or a personal identity code.

The significance of insect farming may increase on a European scale. At least, it is expected that the European Food Safety Authority approves insects for human consumption. At this point, Finland will not add insect farming to Mapu due to low production volumes, but it is ready to reconsider if the level of activity expands significantly.

# 3. WP1 b Modernisation of data collection for horticultural statistics

Data collection for annual crop statistics has been divided into two parts in Finland. Data collection for crop production statistics covers cereals, as well as grass, oil and protein crops, among others, while data collection for horticultural statistics comprises outdoor and greenhouse production of horticultural crops. Agricultural area statistics are mainly based on the IACS register (Table 1.)

| The name of statistics | Home page | Eurostat eDamis Web-Forms |
|---|---|---|
| Crop Production Statistics | https://stat.luke.fi/en/crop-production-statistics | Table1 – Crops from arable land |
| Agricultural area | https://stat.luke.fi/en/utilised-agricultural-area | Table4 –Utilised agricultural area |
| Horticultural Statistics | https://stat.luke.fi/en/horticultural-statistics | Table2 – Vegetables<br>Table3 – Permanent crops for human consumption |

**Table 1.** Annual crop statistics (ACS) in Finland

Horticultural production is carried out on 3,000 farms in Finland, fewer than 900 of which are greenhouse enterprises. Data collection for horticultural statistics is carried out in Finland by means of a census, as it would be very difficult to obtain a reliable sample from a highly non-harmonised population. The census has remained fairly unchanged since 1984, and the first web census was conducted in 2008. The new data collection application built during this project is the third ever developed.

Following telephone interviews, the total response rate to horticultural statistics has always been high, roughly 98 per cent. Approximately half of all farms respond online. The response rate has been high as horticultural farmers consider reliable statistics to be important in their sector. Cooperation with horticultural associations has been important: they use statistics in their work and encourage their members to respond to census.

The need for modernising the online data collection form arose from the previous form becoming technologically outdated. In addition, building the application required external employees and was expensive. One of the aims of the project was to build a new data collection application, which could be maintained and modified independently in statistical services. The data collection application (Figure 3) was built using the **Lomo form engine**, provided by the Finnish Food Authority, and the modernised census was built to feature two languages (Finnish and Swedish). At the same time, databases and data flows connected to this data collection process were modernised (Figure 3, Figure 4).

Like its predecessor, the new application includes a number of **checks**, the purpose of which is to produce high quality data, immediately during the data collection phase. IACS areas reported by farmers were pre completed in the online form which made responding quicker, as often it was only required that farmers checked their surface areas and entered their harvest volumes. Most checks were fairly simple: harvest volumes relative to the surface area. However, checks were particularly challenging with regard to greenhouse production, as a single

greenhouse can be used to grow several crops during a single year. Nevertheless, the checks were highly successful. The better data can be obtained directly from respondents, the less it is necessary to correct and validate data afterwards.

The usability of the new application was also tested by a small group of horticultural farmers, and the application was fixed based on feedback received. A professional tester was used to observe farmers who used the saving application. During testing, it was discovered that "excess checks", i.e. too strict and complicated cross-checks, made responding difficult or even impossible.

What was new in data collection for horticultural statistics was that farmers were able to log in to the online application using strong electronic identification based on official personal data connected to the principal farmer. The Suomi.fi identification service used is run by the Digital and Population Data Services Agency. Another advantage of strong electronic identification is that it was not necessary to send usernames and passwords separately to farmers.

Respondents were mainly able to use the identification service, as a similar process is used in many other current systems in Finland. The process was more challenging for older respondents who are not yet familiar with strong electronic identification. Another problem was that only the person who received the survey was able to respond. This problem will be solved in the autumn 2020 survey, in which an authorisation feature will be built. This means that farmers can authorise another person, such as the farm manager, to respond to the survey on their behalf.

The application was built so that telephone interviewers can also use it. The data collection process in November and telephone interviews in December proceeded without any problems.

A paper questionnaire was excluded for the first time from the cover letter sent to farmers. Before this, it was always sent as reference, even though responses were given online. This did not present any problems, as farmers were already familiar with the data content of the survey. Two reminder text messages were sent. Later, reminders will also be sent via email. Then again, text messages reach respondents better than emails.


**Figure 3.** Horticultural statistics: the new application (available in Finnish and Swedish)

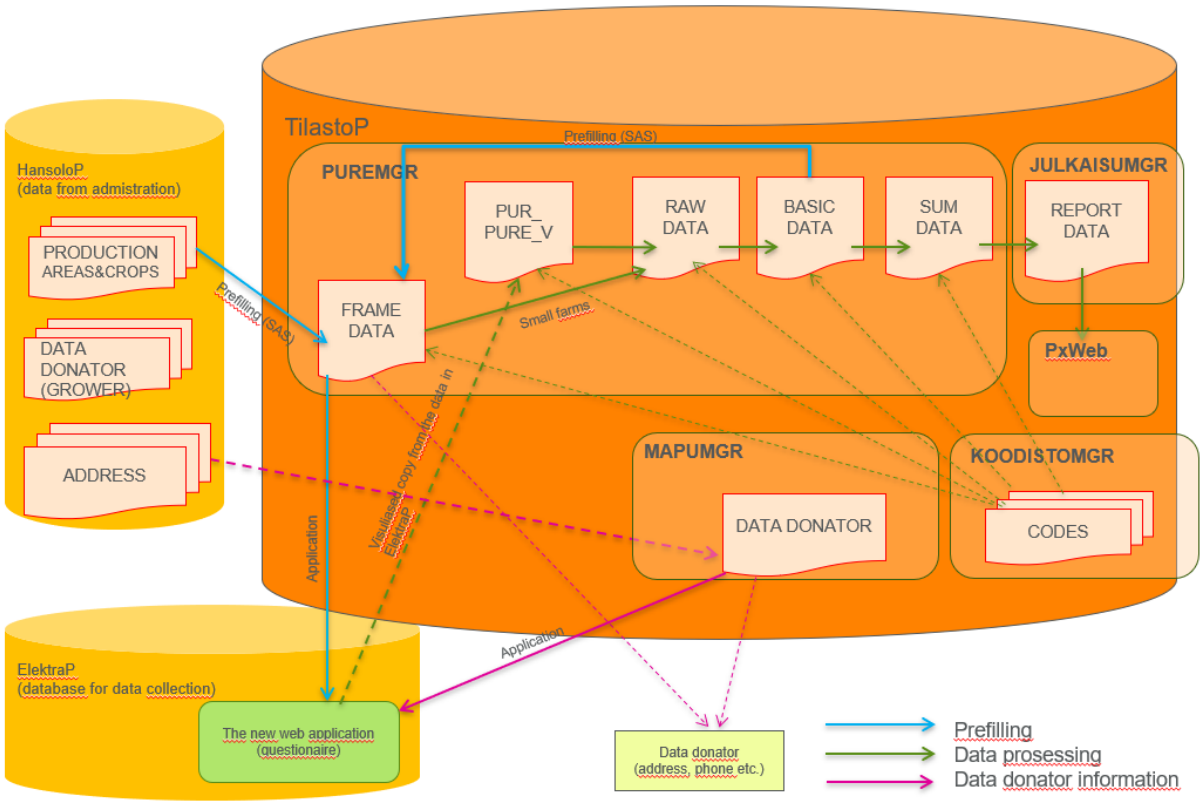


**Figure 4.** Statistical production process / data flows in horticultural statistics


**Figure 5.** Database structure of horticultural statistics (report data is coming)

## Conclusions

- The modernisation of the data collection application calls for hard work and diverse skills, especially if databases and the processing of data, ranging from data collection to publication, are to be modernised at the same time.
- The basic principle of the modernisation is to speed up future data collection processes, and its benefits can already be seen: for example, the 2020 data collection application was completed ahead of schedule.
- The new login method is secure and user-friendly.
- The introduction of an authorisation feature in the autumn of 2020 allows principal farmers, i.e. respondents who originally received the survey, to use Suomi.fi authorisation to authorise another person or enterprise to respond to Luke's data collection surveys on behalf of their enterprise or farm. The authorisation is an electronic power of attorney, whose data is saved in an authorisation register. When a person logs in to the form, the service checks whether the identified person is authorised to respond on behalf of the enterprise in question.
- There are significant benefits and savings: a single app can be used to give responses and conduct interviews.
- There were also other savings, as no paper questionnaire was no longer sent to respondents. This did not present any problems among respondents.
- The checks included in the data collection application improve the quality of raw data and save time during data checks.

# 4. WP2 a Geolocation of farms

Areal unit assigned to an agricultural holding is a key element in agricultural statistics where data is aggregated into administrative boundaries. In this task we studied how the region code is determined in IACS in relation to the farm, field parcels and/or livestock building locations. In an ideal case, the areal unit should correspond to the largest volume of farm's agricultural production or the statistical unit in question. Assigning agricultural production to its real geolocation would reduce statistical bias when drawing inferences from statistics based on aggregated data. Here we have developed methods to acknowledge the problem of wide distances of agricultural production of agricultural holdings when designing sampling survey. However, based on the number of agricultural farms operating on wide distances, this problem is marginal at the moment. In overall, the true value of this study was its contribution to planning sample surveys which is an elemental task in GSBPM. Efficient survey planning in agricultural surveys requires accurate information on true locations of production holdings. Complementing existing registry data, our analysis on alternative farm geolocations increases the accuracy of geospatial registry data, which can be utilized in sample surveys of agricultural statistics.

## 4.1. Data analysis with IACS data

In Finland, the national Integrated Administration and Control System (IACS) operated by Finnish Food Agency holds various geospatial information on agricultural holdings. IACS provides data on location of agricultural holdings and livestock buildings in addition to annual agricultural land use (Land Parcel Identification System, LPIS). Despite wide coverage, the IACS register contains uncertainties related to locations. We were informed by the IACS administration that the location of an agricultural holding is primarily set to the postal address of the farmer. In some cases, the location information was originally not available to the administration and therefore, it was set to the centroid of the corresponding municipality (LAU-2), or the largest field of the holding.

We analysed LPIS data from the year 2017. Here are a few findings:

- In 2017, there were 70 240 farms in the LPIS
- Average number of parcels per farm: 14
- Average size of a parcel: 2.5ha
- Average distance of parcels from the farm: 3.7km
- Average distance between parcels of a farm: 6km
- Max. Distance of parcels of a farm: 757km
- 85% of farms have all the parcels within 10km distance
- 95% of farms have all the parcels within 20km distance
- 2.7% of farms have some parcels farther than 30km
- 21% of farms have parcels in two or more municipalities
- 2% of farms have parcels in two or more locations (at least 30km distance in between locations)

To summarize, 95% of the farms have all the parcels in the vicinity. Only a fraction of farms are running multiple farm entities with considerable geographical distance.

Algorithm to a parcel based geolocation of a farm:

1. Calculate Euclidean distance of each parcel from the farm location
2. Order distances. Put the first parcels to Cluster 1. This is the nearest cluster to the farm. If there is a gap of 30km or more, set those parcels to Cluster 2. And so forth.
3. Calculate the area of each cluster (sum of parcel areas).
4. Choose the biggest cluster (by area). That is the most important cluster of parcels for the farm.
5. Choose the biggest parcel within the biggest cluster. The centroid of that parcel will be the new alternative geolocation of a farm.

For livestock buildings, we analyzed data from the year 2019. Here are a few findings:

- In 2019, total number of livestock buildings: 25346
- Mean distance between livestock building and farm: 6km
- Median distance between livestock building and farm: 0km
- 80% of distances between livestock building and farm is < 4.5km
- Max. distance between livestock building and farm: 652km
- 12 605 farms have livestock buildings
- 3776 farms have more than one livestock building
- 1 664 (13%) of farms have more than two livestock buildings
- 9% have more than three, 6% have more than four
- 6% of farms have livestock buildings in municipality(ies)
- 0.8% of farms have livestock buildings only in municipality(ies)
- The median distance of between farm and livestock building: 6km

To summarize, most of the farms and livestock buildings are located near to each other. In some cases, the livestock buildings fall to the neighbouring municipality. There was also some evidence of strategic placement of production. For example, one farm had a poultry house close to a cooperative egg production factory in a region known for its poultry farming within a considerable distance from the farm, whereas the fields were located near the farm. In some rare cases, one farm entity has multiple livestock buildings on great distances. As an outcome of these analysis with IACS data, we have Python codes to help when designing sample surveys.

## 4.2. INSPIRE grid

In accordance with draft IFS-regulation, a method to place a farm by its geographical location to an INSPIRE 1 x 1 km grid has been developed. Thus, the geolocation of a farm can be recorded as an INSPIRE cell code. We rounded the coordinates of the location of the holding to a two-dimensional equal area grid based on the ETRS89 Lambert Azimuthal Equal Area (LAEA) projected coordinate reference system (EPSG:3035). We chose LAEA projection because at the European level it has gained acceptance. For example, European Environment Agency provides national grids in Shapefiles with LAEA projection. Our Python3 programme truncates coordinates to the nearest kilometer and join coordinates to a cell codes as follows: The cell code "1kmE4730N5131" identifies the 1km grid cell with coordinates of the lower left corner: X=4730000m, Y=5131000m. We tested the accuracy of our algorithm by overlaying the original coordinates onto a grid from EEA's Shapefile. The validation results confirmed that resulting cell codes were equal to our hand-crafted cell codes. The algorithm developed was computationally efficient. For 52,600 farms it took 1min 23s to calculate the cell codes on 2,9 GHz Intel

Core i9 with 32G RAM without parallelisation. The reference place of the holding in the administrative records is the holder's residence.

## 4.3. Assigning privacy-preserving location

For dissemination purposes, we also studied privacy-preserving methods to attribute location. Before the new IFS-regulation the location of the agricultural holding for transmitting purposes was defined in the European Parliament and Council Regulation (EC) No 1166/2008 as the latitude and longitude coordinates within an arc of 5 nautical minutes (around 9km) to avoid the direct identification of an individual holding. In addition, if a latitude and longitude location contained only one agricultural holding, then this holding was to be attributed to a neighbouring location containing at least one other agricultural holding. The location was projected to ETRS89 SRID 4258. In the FADN framework similar disclosure procedure is used. For data dissemination purposes, we developed a Python program for rounded nesting of farms. The coordinates of each farm are first rounded to certain nautical minutes (e.g. 5 or 10 minutes), and then it is checked whether it has an adequate number of neighbours (e.g. 3, 5 or 10) in the same location. If the farm needs to be relocated to a location with more neighbours, the new location will remain within its original area (e.g. municipality or NUTS3 area). It remains an open question, which parameters and appropriate aggregation (area) to use in spatial analysis of information. In future work embracing the protection of confidentiality and quality aspects for data dissemination, we need to tackle the challenge of highly uneven spatial distribution of agricultural holdings.

# 5. WP2 b Geospatial data and crop rotation

Land parcel data obtained from the field parcel register of agricultural administration helps to identify, on a farm-specific basis, the field area in which a single crop has been grown continuously over several years. Crop production is regarded as monoculture if a single crop has been grown for at least three years in succession. The area used in crop rotation can be calculated by deducting the monoculture area from the total area. Only crops, the monoculture of which is considered detrimental, are included. Crops included in the monoculture area are cereals, potato, sugar beet, legumes, outdoor vegetables and strawberry. For example, perennial grass crops are not included in the monoculture area.

## 5.1. Use of geospatial data to define the crop rotation area

Previously, the monoculture area was calculated on the basis of land parcel areas without using any geospatial data to indicate the location of land parcels. This method is described in report "Reduction of response burden by replacing survey questions with register data: Cases of crop rotation, non-regular non-family labour, and number of pigs, sheep and goats. Grant agreement n. 08414.2015.004-2015.531". This method was further developed by adding the use of geospatial data linked to land parcels to it.

Currently, nearly all farmers in Finland use the electronic subsidy application system of agricultural administration, in which the location of land parcels for crops is reported in electronic format. Data about the shape and location of land parcels is saved in the database of agricultural administration in the form of geospatial data. This data can be used to identify the area in which a single crop has been grown in successive years.

## 5.2. Defining the monoculture area from land parcel data

When developing the method, 2016, 2017 and 2018 were used as reference, as geospatial data was available from these years. The area of crops included in the monoculture area was examined on a farm-specific basis. For example, wheat sections from 2018 and the parcels in which these sections were located were selected from data about each farm under examination. Autumn wheat and spring wheat were regarded as a single crop. As a result, the parcels that also included wheat sections in 2016 and 2017 were selected. It was possible that a single parcel selected was used by different farms in 2016, 2017 and 2018.

At the next stage, data from 2016, 2017 and 2018 about all wheat sections located in the selected parcel was compared to identify the area that was common for the three wheat sections of the three successive years. Compared with the previous method version, the use of the QuantumGIS program was a new factor. It helped to accurately define the area used for a single crop, also when the crop was only grown in part of the parcel in question during two of the three years under examination. The wheat monoculture area calculated in this way was allocated to the farm which used the parcel in question in 2018. The monoculture areas of all crops under examination were defined in the same way. Finally, the total amount of monoculture areas of different crops was calculated for each farm.

The method for calculating the monoculture area developed in this project will be used in the 2023 IFS data collection process to produce data about crop rotation.

# 6. WP3 New data sources – data from farms and other sources

## 6.1. Using data sources of farms

The content of data about crop production is steered by the needs of crop production and requirements set out in laws and subsidy conditions. Crop planning software usually provides the following parcel-specific crop production data:

- Crop or other use
- Area
- Tillage, irrigation and other similar activities (method, date and other similar data)
- Planting (volume, date, seed treatment)
- Use of artificial fertilisers, manure and other fertilisers (fertiliser type and nutrient content, application method and rate, date, nutrient balance)
- Crop protection (product, rate, treated area, date)
- Harvesting (yield, date)
- Working hours (to be added as a new feature)

Farm and crop production data has various applications, for example, in administration, research and media in addition to statistical production. It would therefore be cost-effective to collect data for all these purposes via a single channel. Several proposals for the collection of data have been presented for different purposes, while the lack of funding has often been an obstacle.

ProAgria, an advisory organisation, has collected crop production data from volunteers into a parcel databank, which farmers can use as reference in developing their farm activities (https://proagria.fi/sisalto/lohkotietopankki-1230 ). Data included in the parcel databank has also been used in research. The use of parcel-specific crop production data as source material for statistics on the basis of crop protection data in the parcel databank will be assessed later.

The Shared Waters research project (https://www.samassavedessa.fi/en-US) aims to build a digital data system, from which farm-specific data significant in terms of nutrient leaching would be available. Parcel-specific data about cultivation methods obtained from farms would act as initial information. In addition to technical implementation, the project examines the legality and acceptability of the collection and use of data.

One action during the new EU agricultural policy period is the provision of a digital nutrient tool (Farm Sustainability Tool for Nutrients, FaST) for farmers (https://ec.europa.eu/info/news/new-tool-increase-sustainable-use-nutrients-across-eu-2019-feb-19_en). The tool aims to offer help in the planning and management of the nutrient economy to improve the efficiency of the use of nutrients and reduce adverse environmental impact. Each member state will build the tool using, for example, existing digital crop planning tools. The coordinated implementation of the tool across the EU may also enable the collection of data processed using the tool for use as source material in statistical production and other purposes.

**New data sources on farms?**

The volume of data generated on farms has increased rapidly through technological development. Machinery and equipment increasingly produce electronic data about their operations.

What is more, various measuring instruments and sensors are available to monitor activities and conditions, such as the weather, growth and soil, or animals and indoor conditions of livestock buildings. This data is increasingly used in business activities related to agricultural production, and it helps to improve the efficiency of food production and boost financial profitability (https://www.forbes.com/sites/timsparapani/2017/03/23/how-big-data-and-tech-will-improve-agriculture-from-farm-to-table/#2bc594d45989). So far, data generated on farms has mainly been accessible to manufacturers of equipment used to collect data. For diverse uses of data, it would be good if as much data as possible was available for various purposes, including the production of statistics. This would benefit from establishing broadly used standards for the storage and transfer of data and ensuring information security, without individuals who transfer data needing to worry about their data being used without proper authorisation (https://www.sciencedirect.com/science/article/pii/S0308521X16303754). The aim is to create well-functioning practices for the distribution of open data, for example, using MyData practices (https://mydata.org).

## 6.2. Datasets in the public sector

Public datasets are available online in more diverse ways than before. For example, Statistics Finland releases its statistics online and also provides their source material for use in research and statistical production (tilastokeskus.fi/tup/tutkimusaineistot). Agricultural structural statistics have already used data from the register of completed education and degrees to identify the level of education of farmers, and the 2020 agricultural accounting will use data from the Incomes Register to define the working hours of salaried farm employees.

The farmers' holiday and stand-in scheme allows farmers to take annual leave and to call on help from farm relief workers, for example, during illnesses. Data about working hours completed by farm relief workers is available from the Farmers' Social Insurance Institution (Mela), and it has been used as a data source for agricultural labour force statistics. Mela also takes care of matters related to occupational accidents in agriculture, meaning that it can therefore also act as a source for this type of data. Like many other organisations, Mela produces openly accessible statistics related to its activities (https://tilastot.mela.fi).

Under laws on agricultural statistics, Mela is obligated to disclose any data required for statistics. However, not all data sources have been mentioned beforehand, and the party maintaining a data source may be reluctant to disclose data if it is confidential or otherwise sensitive. It is therefore important that parties disclosing data know for what purposes their data is used and that it is certain that their data is disclosed for safe processing and retention.

For parties maintaining statutory registers, it is also important that laws enable the disclosure of data for statistical production.

Satellite data is an example of openly accessible data that has various applications (https://sentinel.esa.int). Luke has also launched various projects to investigate the use of satellite data in research and statistical production, for example, to measure harvest volumes and the surface properties of fields.

## 6.3. Assessing the usability of parcel data in crop planning software

### 6.3.1. Crop planning software for farms

Currently, data related to operations of agricultural and horticultural enterprises is often processed and saved in electronic format using software designed for planning and monitoring farm activities. In crop production, parcel specific data about cultivation activities, such as the use of fertilisers and crop protection products, is recorded. If this electronic data could be made available for further processing directly from planning software, farmers would not need to provide it separately, for example, by responding to statistical surveys.

In Finland, agricultural planning software and services are often provided by software companies and ProAgria advisory organisation, whose Wisu crop planning program is available to farmers as a web service. Some Wisu users have sent data about their farm's parcel-specific crop production activities to ProAgria's parcel databank. mentioned above in section 6.1. In its web service, farmers can use summarised data to assess the impact of different cultivation methods and production inputs on crop production results.

### 6.3.2. Sources of data about the use of crop protection products

Data about the use of crop protection products in Finnish agriculture has been collected from farmers in conjunction with 2013 and 2018 crop and horticultural surveys. Statistics compiled on the basis of this data include data about the volume of crop protection products used and the area treated with them in producing the most important crops of Finland's agriculture and horticulture production. The use of parcel notes as a data source for statistics could be an alternative to collecting data about crop protection products and other crop production activities. To assess this alternative, ProAgria disclosed data about the use of crop protection products in 2018, when similar data was also collected by sending a survey to farmers. Data obtained from ProAgria's parcel databank was compared with the data obtained from the survey. As a result, it was possible to assess the usability of parcel-specific data in statistical production and the properties of data obtained through the surveys.

#### Number of crops and crop protection products in source material

Dataset of the parcel databank included data about the use of crop protection products from 258 farms. Of these, 67 farms had participated in the 2018 crop survey or horticultural survey. Of these, 60 farms had data in the parcel databank about the use of crop protection products regarding those crops and crop protection products, data about which was collected through the crop survey or horticultural survey. These farms had a total of 3,867 combinations of a farm, crop, parcel and crop protection products. After parcel-specific data was summed up on farm level, there were 668 combinations of a farm, crop and crop protection product.

As many as 40 farms had data about the use of crop protection products in datasets of both the parcel databank and the crop protection product survey. Of these, 15 farms had more crops treated with crop protection products in the parcel databank than in the survey. Most of these farms had only one crop more in the parcel databank, while the difference was greater on three farms. The number of reported crop protection products was higher in the parcel databank than in the survey. Data was reported about 13 crops, of which 11 involved more crop protection products in the parcel databank than in the survey. One crop had as many products and

one crop had more products in the survey than in the databank. The parcel databank included 338 and the survey data 263 combinations of a crop and crop protection product.

The more data to be registered there is, the more probable it is that not everything is registered. Details that only apply to a small part of the cultivated area can easily remain unregistered. Rarely used products are easier to forget. Furthermore, the registration of data also includes the risk of misspelling. This is why it is better that data is registered as few times as possible. In particular, larger farms have a large number of parcels, and it would be important that software produced the necessary crop-specific summaries regarding the use of crop protection products, for example. If farmers need to make summaries on their own, some data may easily be left out.

When the datasets were compared, we discovered a case, in which the parcel databank indicated that seven different crop protection products had been used on different parcels for a single crop, three of which had been applied to these parcels twice. The majority of the total area of the crop in question had only been treated once using a specific product, whereas the survey data indicated that the entire crop area had been treated with this product and no other products had been used.

If farmers need to separately enter data about each parcel in the system, not all data may be registered regarding farms with a large number of parcels. This risk can be reduced by transferring data online between planning software and farm machinery. If the volume of fertilisers in a fertiliser applicator or crop protection products sprayer is adjusted in accordance with data entered in planning software, the risk of planned actions registered in parcel data not being updated in accordance with the actual situation is lower. Overall, the use of data in situations where data needs to be correct improves the quality of data. As a result, any errors can be detected, avoided and fixed. Electronic systems can also include checks that show a message if there are any errors or deviations.

Farms are increasingly ordering crop protection product spraying and other farming activities from external contractors. This means that farmers are not always aware of all the details, such as the products and volumes used. In these situations, it would be good if data about each activity completed could easily be sent from the contractor's machinery to the farmer.

### Volume of crop protection products used

A total of 37 farms had identical combinations of a crop and crop protection product in the parcel databank and the survey data. There were 226 of these combinations of a farm, crop and crop protection product in both datasets. In 116 cases, i.e. in roughly half of all cases, volumes differed between the parcel databank and survey data by at most ten per cent. In 39 cases, the volume was more than ten per cent lower in the survey data than in the parcel databank, and in 71 cases it was more than ten per cent higher.

In the parcel databank, the applied amount of crop protection products is indicated per hectare, while in the survey farmers indicate the total amount of each product applied to a crop. The hectare-specific volume is suitable for the parcel databank, as its data is saved per each parcel and application. Usually, the hectare-specific volume remains unchanged every time an product is sprayed on a single parcel. In the survey, farmers report the total volume of each product per crop over the entire growing season, which means that the volume reported includes all applications of the product on all parcels, on which the crop is grown. The hectare-specific volume may have varied per parcel and application, which is why it may be easier for respondents to report the total volume than the hectare-specific average. The total volume

may be easy to assess on the basis of the total consumption of the product in question, and the calculation of the hectare-specific volume would require additional work.

The parcel databank included 1,379 cases of plant protection product application, in which the application rate was below the lower limit indicated in the application instructions of the product in question, and 749 cases, in which the volume was above the upper limit. Regarding application rates within the permitted range provided in application instructions, the rate was closer to the lower limit in 1,787 cases and closer to the upper limit in 442 cases. Therefore, it appears that the application rate is more often lower than higher compared with than the average rate provided in application instructions.

Considering the entire survey dataset, the applied volume of a plant protection product reported by a respondent corresponded to ab application rate that was below the lower limit provided in application instructions in 6,520 cases and above the upper limit in 1,339 cases. A total of 14,222 cases were within the permitted range provided in instructions, of which 10,688 cases were closer to the lower limit and 3,150 cases were closer to the upper limit. The higher number of cases below the lower limit may be explained by respondents having reported the hectare-specific application rate, although the purpose was to report the total volume applied. As the treated area is usually more than one hectare and often many hectares, any incorrectly reported volume per hectare is clearly lower than the correct total volume. This was taken into account when correcting the survey data so that, if the total volume reported by a farmer corresponded to an ab application rate that was clearly outside the application rate provided in application instructions and this volume was within the hectare-specific range provided in the instructions, the volume reported was assumed to be the hectare-specific rate, and the total volume was calculated on the basis of this assumption. Any volumes missing from the survey data were corrected in accordance with the average application rate provided in application instructions, and this may have resulted in volumes being higher than the actual volume used in many cases.

### Possible uses of parcel-specific data

The quality of the data from the parcel databank used in this project was higher than that of average data provided by crop planning software. If parcel data provided by software is to be used as source data for statistics, farmers must be aware of this, and they must record parcel specific data following the same level of accuracy as they do when responding to statistical surveys. Considering information security, it is also important that farmers know for what purposes data about their farm is used. It would motivate farmers to provide parcel data if they did not have to respond to a separate statistical survey.

A crop-specific summary of the use of crop protection products could also provide interesting information for the farmer. As a result, software companies should add a functionality to compile the necessary summary of crop protection products to their software. As farmers have access to summaries prepared for statistics, they can also check and confirm them.

The parcel databank could possibly be used as a source for certain more detailed data that is used to supplement data collected from farmers. This data could include the volume of each product applied to each crop. If data about volumes was obtained from the parcel databank, data about crop-specific products used should only be collected from farmers. It would be important to identify in more detail how well the farms included in the parcel databank represent all farms and how well the cultivated area of the farms included in the databank represents the entire cultivated area.

Farm software already includes interfaces, using which data is transferred automatically, for example, to certain subsidy application forms of agricultural administration or to agricultural economic statistics. Similar interfaces could also be built for the transfer of data about the use of crop protection products and fertilisers and data about other farming activities. The growing significance of environmental aspects in agriculture has increased the need for data about environmental impact. Therefore, this data could be used in research and administration in addition to statistics. It could be useful to build a single shared data transfer channel to collect this data for different purposes. The extensive use of data would be an additional criterion for building interfaces. If farmers experience that they benefit from such interfaces as there would be fewer surveys, software companies could represent these interfaces as an asset to promote their software and possibly be ready to build these interfaces to their software at their own expense. Furthermore, the diverse use of data would provide farmers with more motivation to keep software data correct and updated. Farmers could also be rewarded for providing data about their farms by disclosing other data that is useful for farm activities and created using different data sources. An easy-to-use data collection system would enable data to be collected more extensively and frequently than at present, which would increase the volume and quality of available data.

The development of data systems and services opens up new opportunities to use parcel-specific data. The use of data is often associated with the examination of farm-level activities, for example, regarding finances or environmental impact. However, datasets produced during these processes can also be used for other purposes, such as statistics. Providing the FaST tool, built as part of the EU's new agricultural policy period, for all recipients of agricultural subsidies may offer a new opportunity to collect farm-level data in electronic format.

## Conclusions

- Data in the parcel databank about the use of crop protection products apparently represented the actual use of products better than data collected through the crop protection product survey.
- In many cases, data in the parcel databank included products, the use of which was not reported through the crop protection product survey. In some cases, farmers reported that the entire cultivated area was treated using a single product applied to the majority of the area, while part of the total area was actually treated using other products.
- In the parcel databank and the survey data, the application rate of crop protection products used is often lower than the average rate provided in application instructions. Replacing the applied amount missing from the crop protection product survey data by an applied amount based on the average application rate provided in application instructions may, in many cases, have produced a volume higher than the actual volume.
- The use of parcel-specific data obtained from farms for various purposes is developing and, as a result, parcel data may also become available as source material for statistics by new electronic means.

# 7. WP4 Seminar of modernisation of agricultural statistics

Poland, Latvia and Finland jointly organised the "Modernisation of Agricultural Statistics" workshop in Olsztyn, Poland, on 8 and 9 October 2019. Statistics Poland, and especially its Statistical Office in Olsztyn, was responsible for primary arrangements. In addition to representatives of Eurostat, producers of agricultural statistics from 16 countries participated in the workshop, representing Austria, Bulgaria, Finland, Germany, Greece, Hungary, Ireland, Latvia, Lithuania, Malta, the Netherlands, Norway, Poland, Romania, Slovakia and Slovenia.

The purpose of the workshop was to present good practices related to the production of agricultural statistics in different countries. The use of geospatial and satellite data is a good example.

Luke had six presentations:

- "Modernisation of agricultural statistics " – Anna-Kaisa Jaakkonen, Senior Statistician and Arja Anttila, Senior Specialist, Luke Finland
- Data collection in Agricultural Statistics, Senior Specialist Arja Anttila
- Farm planning software and internet services – data source for agricultural statistics, Pasi Mattila
- Use of geospatial field parcel data of IACS in the investigation of crop rotation, Pasi Mattila
- Statistical Farm Register system – Finland, Pasi Mattila
- Merging farm register data and administrative data, Anna-Kaisa Jaakkonen, Esa Katajamäki

In addition, Pasi Mattila acted as a moderator in the "Modern methods of data collection" session.

All workshop presentations can be downloaded from https://olsztyn.stat.gov.pl/en/seminars-and-conferences/workshop-modernisation-of-agricultural-statistics-olsztyn-poland-89102019/.

The workshop agenda can be found from Appendix 1.

## Appendix 1

## Agenda Workshop "Modernisation of agricultural statistics

Workshop "Modernisation of agricultural statistics"
EU Grant number 2018.0218

Date: 8-9 October 2019
Venue: Omega Hotel, Olsztyn, Poland
Organizer: Statistics Poland

7 October

Arrival of participants to Warsaw and Gdańsk
Transfers to Olsztyn

8 October

**7:00 – 9:00** Registration

**9:00 – 13:00 Opening session**

**9:00 – 10:30**

- Introduction/welcome - Artur Łączyński, Director of Agriculture Department, Statistics Poland Marek Morze, Statistics Poland, Director of Statistical Office in Olsztyn
- "The role of Eurostat in modernising agricultural statistics" - Marjo Kasanko, Deputy Head of Unit Agriculture and Fisheries, Eurostat
- "Agricultural census 2020 in Greece" – Lemonia Dionysopoulou, Head of the Agriculture, Livestock, Fishery and Environmental Statistics Division, ELSTAT
- Discussion

**10:30 – 11:00** Coffee break
**11:00 – 13:00**

- Introduction – moderator Artur Łączyński, Director of Agriculture Department, Statistics Poland
- "Modernisation of agricultural statistics " – Anna-Kaisa Jaakkonen, Senior Statistician and Arja Anttila, Senior Specialist, Luke Finland
- "Modernisation of agricultural statistics" – Kaspars Misans, Vice-president, Central Statistical Bureau of Latvia
- "Modernisation of agricultural statistics" – Artur Łączyński, Director of the Agriculture Department, Statistics Poland
- Discussion
- Summary/conclusions – moderator and participants

**13:00 – 14:00** Lunch break

**14:00 - 16:30 Session 1. Modern methods of data collection**

**14:00 – 15:00**

- Introduction – moderator Pasi Mattila, Senior Scientist, Luke Finland
- "Reduction the burden of respondents and improvement of statistics quality by effective use of administrative data sources" – Anita Raubena, Deputy Director of Agricultural and Environmental Statistics Department, Central Statistical Bureau of Latvia

- ”Farm planning software and internet services as a potential source of field-parcel-specific data for agricultural statistics” – Pasi Mattila, Senior Scientist, Luke Finland

**15:00 – 15:30** Coffee break

**15:30 – 16:30**

- Discussion
- Summary/conclusions – moderator and participants

**9 October**

**9:00 – 10:30 Session 2. Geospatial and satellite data**

- Introduction – moderator Artur Łączyński, Director of Agriculture Department, Statistics Poland
- „Use of geospatial field parcel data of IACS in the investigation of crop rotation” – Pasi Mattila, Senior Scientist, Luke Finland
- „Possibilities of using satellite data for agricultural statistics' needs” – Tomasz Milewski, Expert, Statistics Poland
- Project on „Satellite use for harvest estimation in Germany'' – Ute Walsemann, Germany
- Discussion
- Summary/conclusions – moderator and participants

**10:30 – 11:00** Coffee break

**11:00 – 12:30 Session 3. Methodological issues**

- Introduction – moderator Barbara Domaszewicz, Deputy Director of Agriculture Department, Statistics Poland
- “The concept of data validation mechanism developed/within the grant project: Implementation of business architecture for ESS Validation” - Dariusz Miziołek, Expert, Statistics Poland and Marek Panfiłow, Expert, Statistics Poland (Statistical Office in Olsztyn)
- “Merging farm register data and administrative data” – Anna-Kaisa Jaakkonen, Senior Statistician, Luke Finland
- Discussion
- Summary/conclusions – moderator and participants

**12:30 – 13:30** Lunch break

**13:30 – 14:30 Session 4. Modernisation of the statistical farm register**

- Introduction – moderator Kaspars Misans, Vice-president, Central Statistical Bureau of Latvia
- “Statistical farm register system”– Anneli Partala, Senior Statistician, Luke Finland
- Discussion
- Summary/conclusions – moderator and participants

**14:30 – Closing session – summary**

**10 October**

Departure of participants


Natural Resources Institute Finland
Viikinkaari 4
FI-00790 Helsinki, Finland
tel. +358 29 532 6000