

Modernisation of agricultural statistics in Hungary

Final report

Eurostat grants for 2017

Objective: 08.4 – Provide quality agriculture, fisheries and forestry statistics

Module (DTM): 08.4.11 Agricultural statistics

Agreement number: 2018.0214

Authors:

Márk STOLL

Ágnes PATAY

Content

Introduction.....	3
Evaluation of the methodology before 2020	4
Scope of the project	7
Task 1	7
Task 2	7
1.1. Updating farm register system	9
Development of a new Agricultural Farm Register	9
Using the linked database as a frame for Agricultural Census 2020	19
AC2020 Pilot survey, February 2019	19
1.2. Improving data processing and validation systems.....	22
Data validation during data collection phase	22
Data validation during data processing phase	24
1.3. New thresholds.....	25
2.1. Improved sample design.....	28
Sample Design	28
New stratification	30
Household activity	32
2.2. Exploitation of new estimation methods	32
2.3. New survey districts, geo-coding, geolocation of farms	33
2.4. Integration of existing administrative data sources	34
Summary	36

Introduction

The methodology of the Hungarian agricultural data collection system has been almost unchanged for decades. Therefore there was a clear need that methodology needs to be adapted to the changed structure of agriculture, and the availability of several good quality administrative databases.

Hungarian Central Statistical Office (HCSO) launched its project on the preparation for Agricultural Census 2020 in 2018 February, with the aim to establish a new system for the Hungarian agricultural statistics for the Agricultural Census 2020 and the period after that.

In order to accomplish that HCSO planned to achieve the following objectives within the framework of the project:

Task 1:

- updating farm register system
- improve data processing and validation systems

Task 2:

- improve sample design
- exploitation of new estimation methods
- geocoding/geolocation of farms
- integration of existing administrative data sources

The main purpose of the activities above was to have a modern, cost effective data collection methodology, not only for the Census, but also beyond 2020.

Evaluation of the methodology before 2020

In Hungary farm threshold had been almost the same since 1972, though the importance of the households' agricultural activity has decreased significantly. There were almost two million farms in 1972 in Hungary, and the number decreased to about 570 thousand in 2010 and the trend continues. It was estimated that there were about 400 hundred thousand farms in 2016 – based on data of the last Farm Structure Survey (FSS), 2016.

Simultaneously a concentration has also happened, which means the number of bigger farms has been quite stable, while the number of small farms has fallen significantly.

In the last Agricultural Census in 2010 more than 2.2 million addresses were visited in order to explore the potential agricultural activity. According to the results:

- data of about 570 thousand farms were recorded,
- there were more than a million addresses with only household agricultural activity (below the threshold),
- and another more than half a million addresses which did not have any agricultural activity

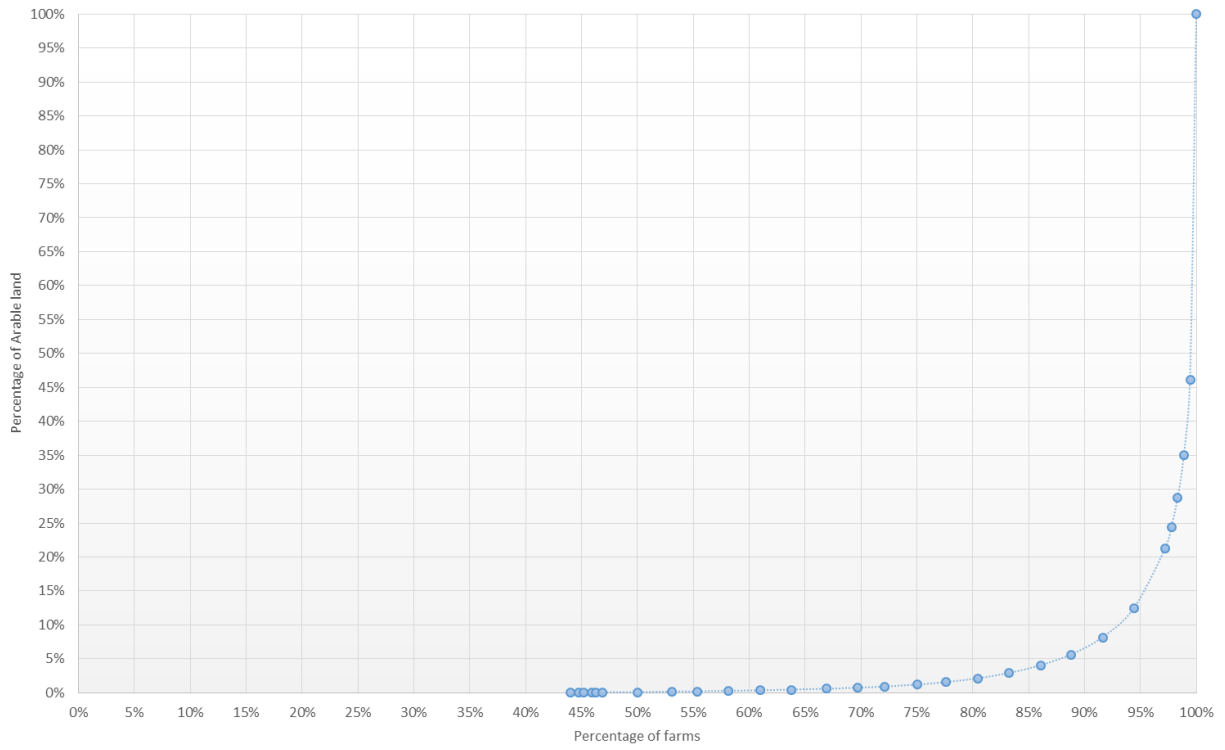
It meant that during the last Agricultural Census in 2010 more than one and a half million addresses were visited without any agricultural activity or agricultural activity only below the threshold.

In addition, due to the low thresholds 72% of all farms used less than 1% of arable land in 2010, and other main variables had a similar distribution.

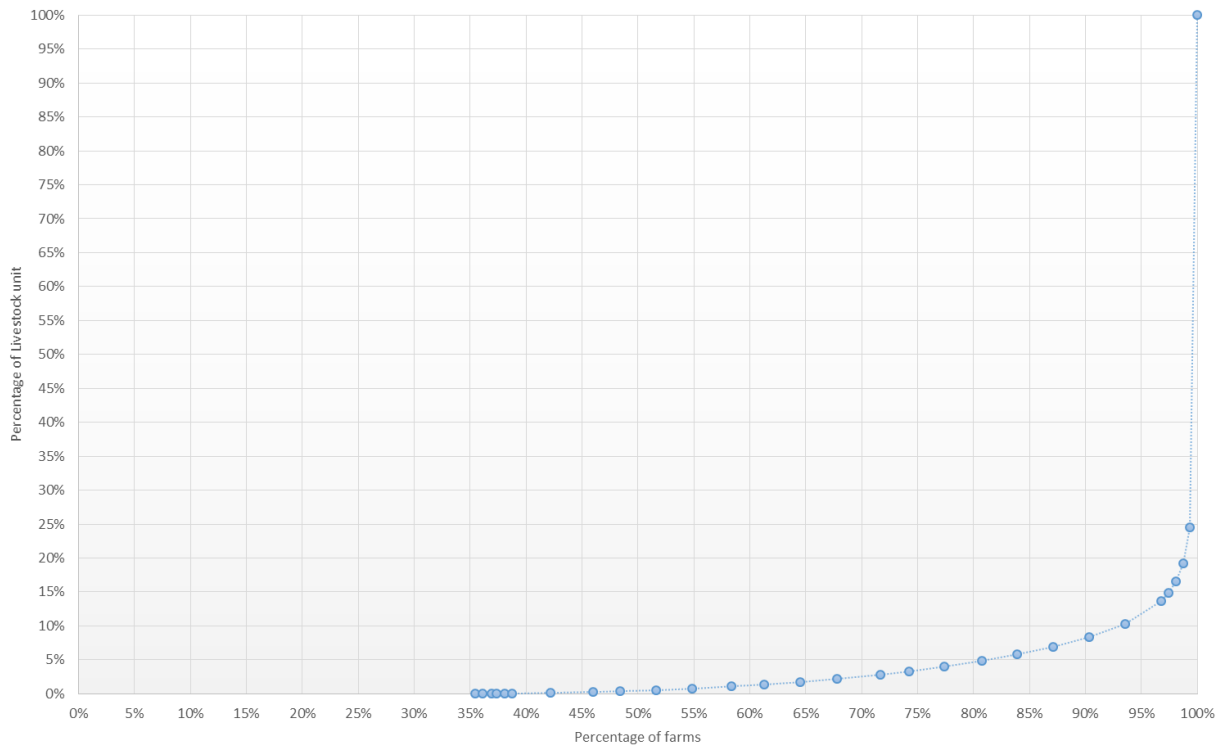
All of the above meant that the data collection system was not very resource efficient. A lot of resources were spent on contacting the high number of addresses that were not a very important part of the agricultural production universe.

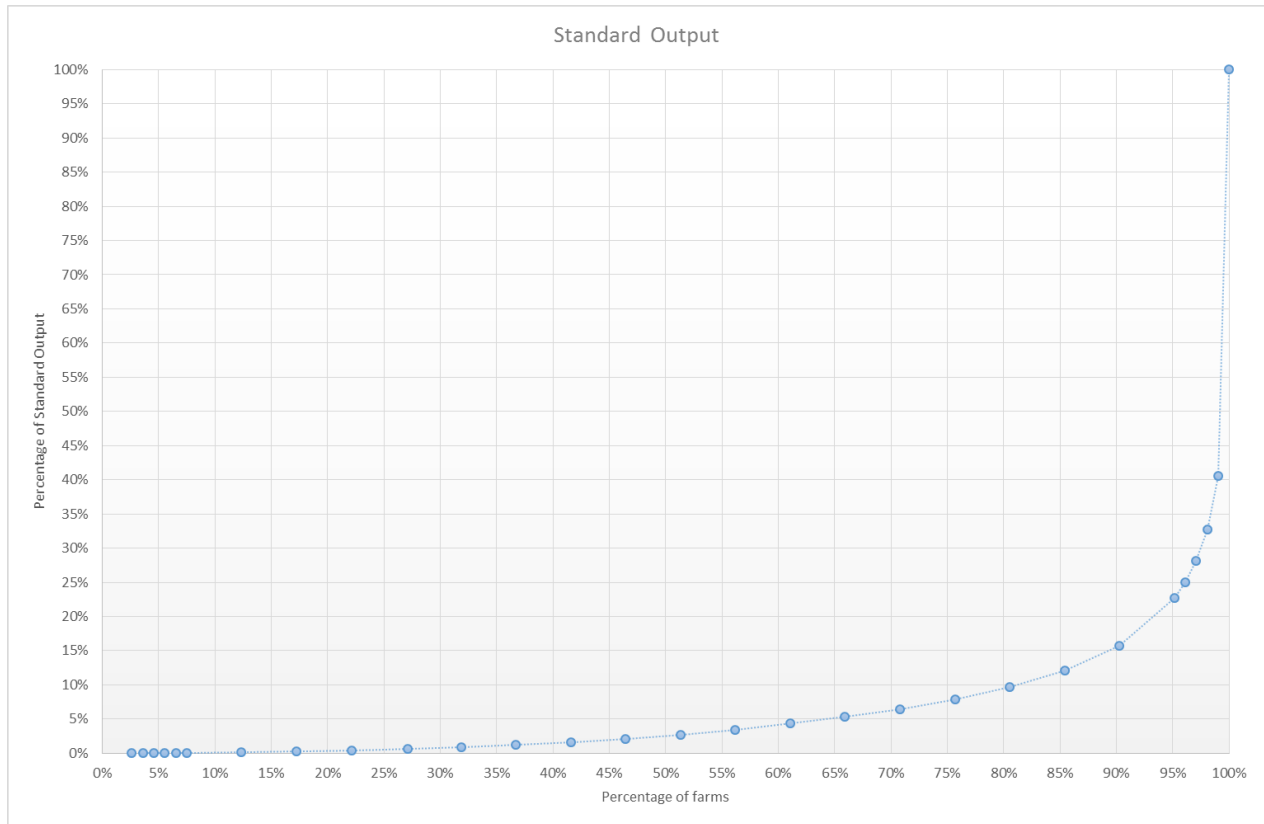
On the graphs below it can be seen that arable land, livestock and standard output are concentrated in a small percentage of bigger farms.

Arable Land



Livestock Unit





Until now there was not enough information on farming activity, there were not enough acceptable address lists of small farms in Hungary. Because of the absence of a regularly updating Farm Register, HCSO carried out agricultural sample surveys as a cluster sample, which meant that all households were visited in the selected enumeration districts in rural areas.

The result was that many addresses were visited unnecessarily. During the last decade administrative registers, databases have been created, and their quality has evolved, so now it is possible to use them as an address list or information on the population frame.

Scope of the project

The main goal of the project was to establish a new system of Hungarian agricultural statistics for the Agricultural Census 2020 and the period after that.

The project consisted of the following tasks and sub-tasks:

Task 1

1. Updating farm register system:

Improve Farm Register so that it updates more frequently. The goal was to increase response rate and quality of data to make data collection more cost effective (mainly existing farms shall be contacted). It was the core of the whole methodological improvement, as it was considered as a cornerstone of the agricultural statistics.

2. Improving data processing and validation systems:

This process was strongly related to the previous objective, because as a result of administrative database linkages, administrative data will be used for validation and data processing purposes more effectively. Within the framework of the project, it was evaluated which of the currently existing databases can be used for this purpose.

In order to treat non-response better the use of different imputation methods was also evaluated.

3. New farm threshold:

New farm threshold, for the agricultural statistics was established. Hungarian thresholds have been unchanged for decades and it was one of the lowest in the EU currently, therefore the revision of the current practice was necessary in order to focus on the relevant agricultural activity.

Task 2

1. Improved sample design:

Within the framework of the project, a new sampling methodology for agricultural statistics beyond 2020 was developed.

2. Exploitation of new estimation methods:

It was evaluated how reweighting methods can be used in the future according to handling unit level missing data.

3. Geocoding/geolocation of farms:

It was evaluated how Farm Register can be connected to the main HCSO Address Register. Currently the agricultural register uses its own address database. Connection with the Address Register was considered beneficial as it stores a geocode for each address.

4. Integration of existing administrative data sources:

In connection with the development of the farm register, other administrative data sources can be used more effectively, as linkage between the different databases can be strengthened. The integration of administrative data sources was quite difficult before because of the lack of available identification information. At the beginning of the project it was expected that if each farm data within the register can be updated by administrative data, then the sampling frame quality will be better.

In the following chapters, we will describe the progress achieved within the framework of the project.

1.1. Updating farm register system

Development of a new Agricultural Farm Register

The main limitation to develop the methodology of agricultural statistics was that before the project there were no updated information on the population of farms in Hungary. This made the change of the sampling methodology impossible. So it was decided not only to improve the current register, but to develop a completely new application.

Previous Farm Register:

- was based on the last census (AC2010) information
- stored only basic register information (name, address, legal status, etc.)
- was not linked to other databases
- one unit was stored in one record
- maintenance was done per record (inside the program)
- did not update regularly and automatically by any database (survey information, other registers, administrative databases)
- the application was outdated, it was developed 20 years ago

The status of this register could support only the data collection methodology when whole surveyor districts were selected, where all addresses were visited two times per year, irrespectively whether there were agricultural activity on them or not. As it was a big burden both for the respondents and HCSO, the main task was to develop a regularly and automatically updating Farm Register, in which all units that might carry out agricultural activity are registered, and therefore can be easily contacted. It was supposed that this way visiting other addresses becomes unnecessary.

HCSO started to elaborate the specification of a new Farm Register application. A few team members visited the Austrian Statistical office in order to gain useful practical experiences in connection with the Austrian Farm Register. This visit helped to formulate the requirements toward a well-functioning farm register. After the visit on the first internal meetings it was agreed that the main objectives in connection with the newly developed register are the following:

- main required functions:
 - o regular update of information on the population of farms (using data from administrative databases and data collections)

- performing queries to list farms with certain characteristics
- producing address lists for data collections
- the software should communicate via interfaces with other HCSO internal applications
- historical data storage is necessary in order to have an overall view on the evolution of certain records over time
- it shall be developed in Java

The detailed specification of the register was completed by September 2018. It contained the following main information:

- list and data structure of the administrative databases that shall be used by the Farm Register
- required data transformation of the administrative databases
- types of interfaces that should be developed, in order to ensure communications between other internal HCSO applications
- planned environmental diagram of the different connected internal applications
- needed functions in connection with the different user interfaces

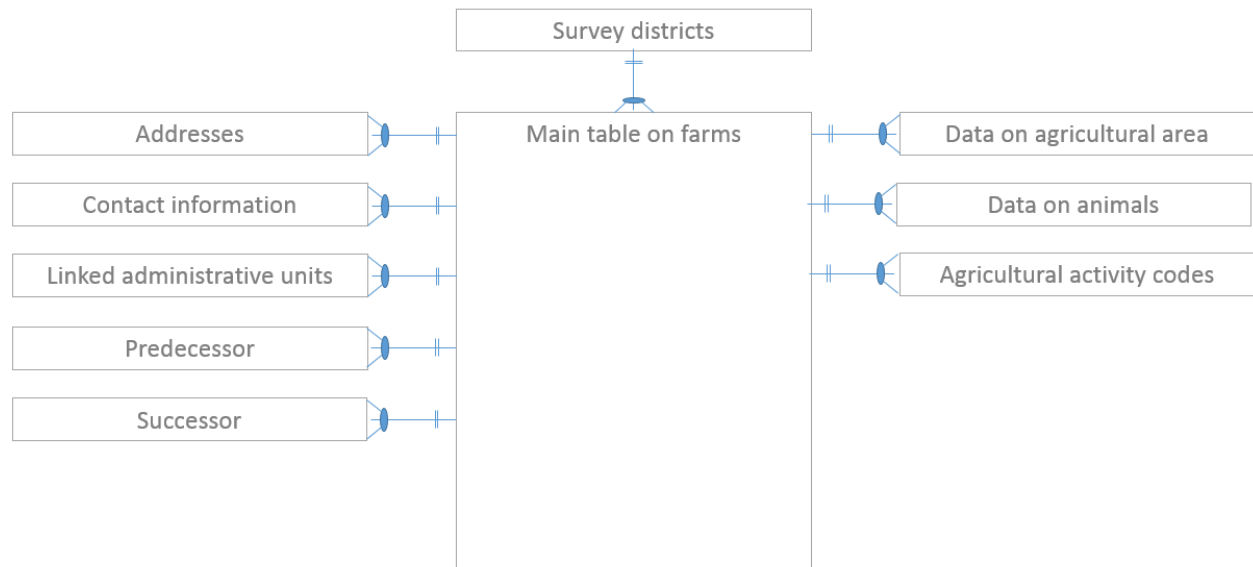
The development of the application was outsourced and the public procurement was successful. The software development started in July 2019 and was planned to be finished by the end of 2019.

New Farm Register:

- it is based on the records of all available, good quality administrative registers and the previous Farm Register
- it is stored as a relational database
- it updates regularly and automatically by database linkages – between the register and the external registers or administrative data sources – within the application
- all maintenance and update processes are carried out within the application
- beside basic register information it stores also data on agricultural production (coming from administrative sources and data collections)

It is assessed that the new Farm Register will be a good sampling frame for sample surveys beyond 2020 and will also be a good basic database that will be used for validation and imputation.

Data model of the Farm Register is planned to be the following:



As there were delays in the software development, the modified date is the end of July 2020.

Data sources of the new Farm Register

Within the framework of the project, the inventory of databases that could be used in the new Farm Register was completed.

As first step the possible available administrative databases were detected, all of that can contain relevant and good quality information for the regular update of the Farm Register.

The identification of these databases was carried out in strong cooperation with the Hungarian Agricultural Ministry, as databases are owned by many institutions and organizations that are governed by the Ministry.

The work started with expert level meetings and several possible databases were identified. Among them, some were selected as a source of register information in the future, based on the criteria that they:

- update frequently,
- contain good quality information on names and addresses
- contain ID numbers, that can be used for record linkage

According to the Hungarian Law on official statistics (2016. CLV.) administrative databases can be used for the purpose of official statistics without any constraints. Members of the official statistical Service can get any of these administrative databases with their unique identifiers free in electronic form. In case of

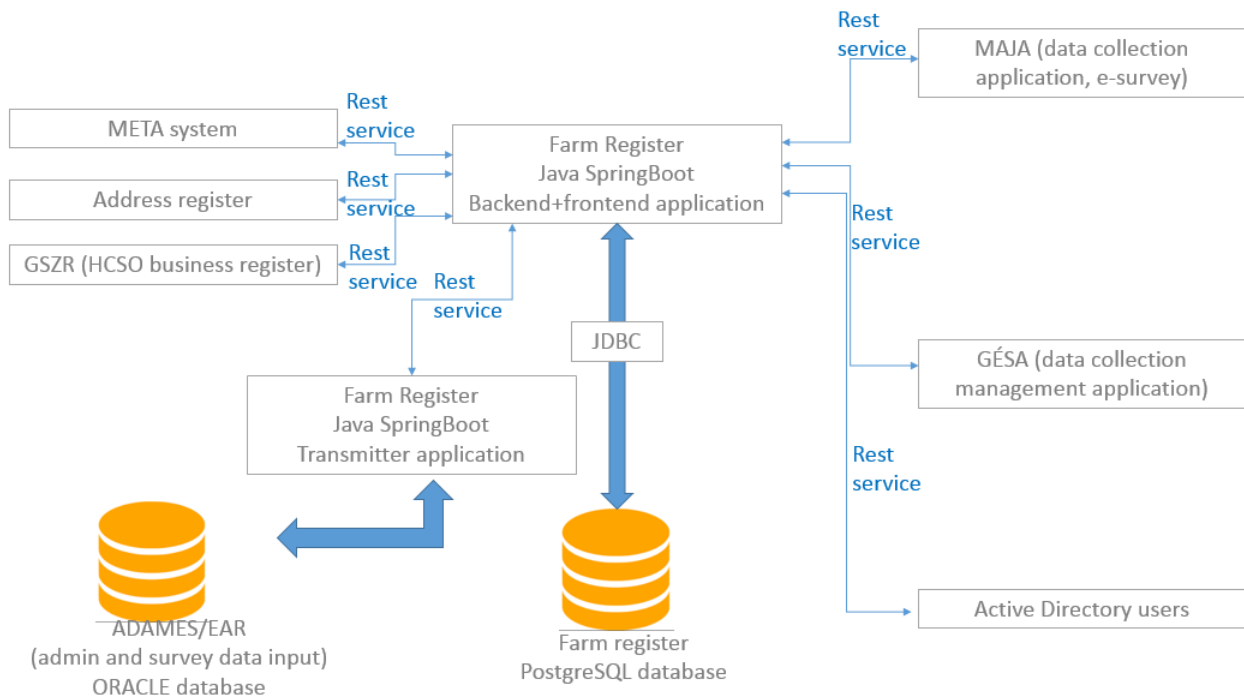
transmission of administrative data the detailed data structure, frequency, contact persons, etc. are laid down in cooperation agreements.

The following step was to conclude cooperation agreements separately with the different organizations, and HCSO concluded cooperation agreements with the respective organizations by the end of 2018, in order to have a frequent data exchange in the future.

HCSO has its own application for secure database transfer. The administrative data files (csv, xlsx) enter into HCSO database via the application named KARÁT, and data transformations are done in another application called ADAMES (that has similar logic than Pentaho software of Hitachi).

Both the incoming and the transformed data are saved into HCSO ORACLE database. The Farm Register updates its information by uploading the database table into its own table structure under its own schema.

The Farm Register will have the following system architecture:



The following databases will be transferred to HCSO regularly (annually, semi-annually or quarterly) to update the Farm Register:

- Hungarian Chamber of Agriculture
 - o List of members: according to a national legal act, it is compulsory to be a member of the Hungarian Chamber of Agriculture if one carries out agricultural activity. Therefore it is

considered to be quite an exhaustive list. The problem is that it contains mainly persons with agricultural activity, not farms. About 340 thousand clients in each period.

- Hungarian State Treasury
 - o SAPS (Single Area Payment Scheme): it contains annual information on persons and legal entities who applied for subsidies. The database also contains the area of crops, for which they applied for subsidies. It is considered to have a very good coverage for agricultural area and producers, but has the same problem as the previous database, namely it contains mainly persons with agricultural activity, not farms. About 160 thousand clients in each period.
 - o Other income supports scheme: it contains annual information on persons and legal entities applied for subsidies other than SAPS. It is a good additional information on farmers, even those who have only animals. Quantitative data can be used for data validation as well, though it is not a complete database at all for agricultural area and livestock. Only a part of the totals is present in the database. It contains mainly persons with agricultural activity, not farms. About 40 thousand clients in each period.
- National Food Chain Safety Office
 - o Animal registers: it is the most complete in case of bovine, the number of other animals (pig, poultry, sheep) is only in part in the database, as in Hungary it is not compulsory to register all animals. It contains mainly persons with agricultural activity, though more persons can have the same licence, not farms. It contains mainly persons or legal entities having livestock, and there is also information about where animals are kept. About 120 thousand clients in each period.
 - o Licensed traditional small-scale producers: after registration, members can have a tax reduction if they only produce annually below a certain value. It contains mainly persons with agricultural activity, though more persons can have the same licence, not farms. About 240 thousand clients in each period.
 - o Family farms: family members farming together, who can have special financial support and credit options. About 30 thousand clients in each period.
- Organic farming registers: compulsory for an organic producer if they want to have certification for their products. It contains mainly persons with agricultural activity, not farms. About 2 thousand clients in each period.
- Vineyard register: register of wine producers. About 30 thousand clients in each period.

In every period the first step is to evaluate what are the changed records.

Before data transformation, those records that have not changed since the previous period will be excluded from the data process. The other records can be grouped into three main categories:

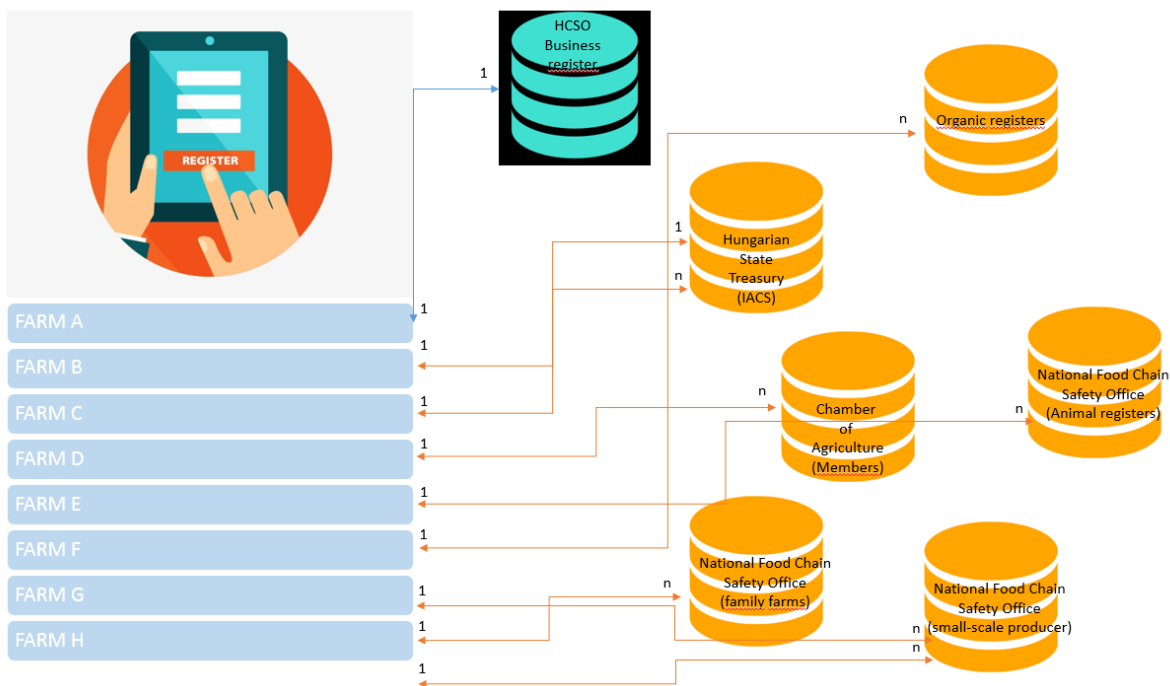
- new units (have to be linked to an existing farm in the register)
- modified units (name and address information have to be updated)
- ceased units (the administrative unit is not existing in the period, but remains linked to a certain farm in the register)

Beside the administrative databases the new Farm Register will use HCSO's own registers, namely:

- Business register
- Address register

Linkage of administrative data sources to the previous Farm Register

By the first quarter of 2019 all records of each administrative database mentioned previously had been linked to the previous Farm Register (the list of farms based on data of AC2010). The ones that had no possible connections to the register had been added to that as new farms.



Naturally all administrative databases have their own structure, and they keep the relevant information according to their own requirements. Therefore in the very first moment the entire process was designed – reception of data, data transformation, validation and data linkage – for each database separately, one

by one. It was standardized how typical (possible) problems – e.g. mistyped information in addresses – would be handled in the future.

The biggest challenge of data linkage was the fact that there was no unique identification number of Hungarian farms that was recorded and existed in all administrative databases. Each database had a different identification key, which sometimes occurred in other databases, sometimes did not. Fortunately there were IDs which were more or less general in many databases (e.g. IACS identification number), but there was no guarantee that any ID, which was not the database's own ID, was filled out completely in the given database.

On the other hand, the previous Farm Register did not contain any information on the different administrative ID numbers, therefore these identification numbers could not be used for the linkages for the first time. That made the linkage process very difficult in the first round. The new register will contain all required identification numbers that were linked to the farms, therefore it will not be a problem.

Because of the previous reasons, for agricultural enterprises (legal entities) linkage could be done based on identification numbers, for instance tax number. For private holdings linkage was done mainly by using name and address information, which made the process more difficult.

Considering the private holdings the most difficult issue was that the units of the Farm Register are agricultural farms, but all administrative registers contain information on persons or enterprises. It was hard to decide whether persons belonging to the same family were farming separately or together. It had to be handled also when the administrative sources contained outdated information (e.g. deceased person, wrong addresses). It also occurred that an administrative database had information on persons who are not involved in any agricultural activity, but – from various reasons (e.g. tax) – were recorded themselves in the databases.

[Linkage rules used in case of private holdings with names and addresses](#)

Names and addresses were always brought to a common format before the record linkage.

This mainly meant the separation between family name and first name (for instance female names in Hungary can be changed in four different ways after marriage) and title and prefix. This also meant converting them into capital letters, and removing extra spaces (at the beginning or end, or more than one between names), etc.

In order to link records by names effectively, names of farmers and administrative units were split up into the following parts:

Title
Prefix
Family name 1
Family name 2
First name 1
First name 2
Original name format

Addresses were converted into HCSO's own address format that is used in HCSO's address register. If the certain part of the address could be found in the Address register it was given that identification number. The administrative unit was compared to the farms of the Farm Register and searched for the same or the most similar name and address.

Results were categorized into the six groups, depending the strength of the similarity: from 1 when they were exactly the same to 6 where the house number differed by 3.

Similarity of names were grouped into nine categories and several subcategories depending on similarity. It was decided which combination of the address and name similarity meant a strong relationship (automatic linkage), what categories should be checked manually, and what is considered no link.

Linkage with ID numbers

ID numbers were compared to all ID numbers connected to certain Farm Register units. If an administrative record could be linked to several other units in the register the linkage was checked manually.

The following IDs are present in the different databases that can be used for linkage:

Database	Type of ID						
IACS	IACS client ID	Tax number	Personal tax number				
Chamber of Agriculture	IACS client ID	Tax number	Personal tax number	Small scale producer ID	Family farm ID	Vineyard register ID	Food Chain Safety ID
Small scale producers	IACS client ID	Tax number	Personal tax number	Food Chain Safety ID	Small scale producer ID		
Family farms	Family farm ID	Tax number (farm)	Personal tax number (farm)	Tax number (family member)			
Animal register	IACS client ID	Tax number (Farmer)					
Vineyard register	IACS client ID	Tax number	Personal tax number	Vineyard register			
Organic register	Tax number						
Business register	Tax number						

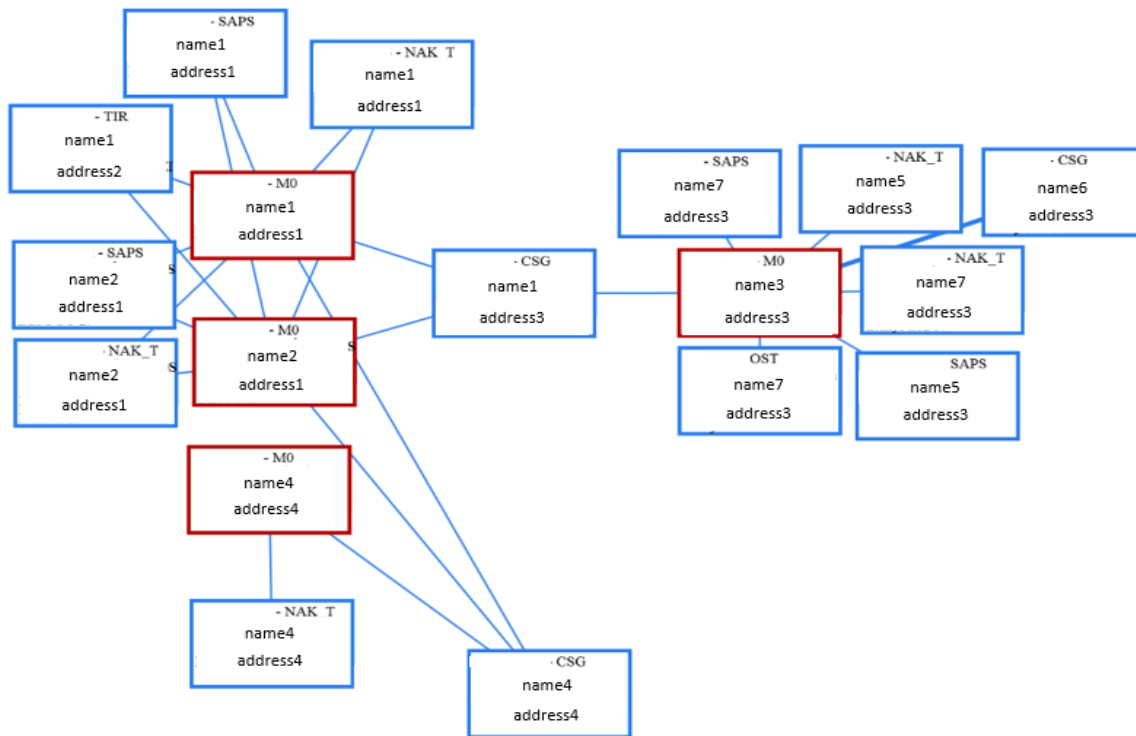
Efficiency of the linkage

After the first round of linkages the result was the following:

Database	Total number of records	Automatically	Manual	No result	Manually rejected	Added as new farm
IACS-SAPS	162 188	100 042	17 261	43 242	1 643	44 885
IACS-Other	35 715	18 499	3 503	13 306	407	13 713
Chamber of Agriculture -	340 222	278 478	23 118	37 436	1 191	38 627
Chamber of Agriculture -	133 710	59 076	9 051	63 830	1 753	65 583
Small scale producers	233 866	165 356	23 439	43 514	1 560	45 074
Family farms	22 824	14 742	7 787	242	43	285
Animal register	121 599	94 013	14 836	12 077	673	12 750
Vineyard register	30 704	23 989	2 057	4 529	129	4 658
Organic register	3 176	3 152	10	13	1	14

Splitting up of network of administrative units

At the end of the linkage of all administrative databases, in some cases the linked units resulted in a complicated graph. The units within a graph were connected because they had the same address and had for example same name or family name. And in other case the name was the same but it existed on different addresses in different databases.



The problem with such a big group was that it was difficult to decide whether they were really farming together and where to find them among their several addresses. Therefore it was decided that the biggest graphs should be split up, and they will be contacted separately during the Agricultural Census.

After this split the question was how to decide which names and address to use when contacting the farm during AC2020.

Deciding which address to choose from the network for contacting farms

As more administrative units were linked to the same farm, it could happen that names and/or addresses were different. Contact names and addresses had to be determined, where the farm would be contacted during Agricultural Census 2020. The contact names and addresses were determined according to the following priority:

1. Name/address in IACS database (as it was considered the most relevant)
2. Most frequent address
3. Priority of the other databases
4. Lower ID number as it was supposed to be older

Main features of the new Farm Register application

In the future the new register application will be used to manage and update the list of farms, and their administrative units.

Records of a farm, and linked administrative units will appear in the following way:

Paradicsom Béla Tomato Tom

Törzsszám	Farm ID number	M2764731	Megjegyzés	Comments
Név	Name	Paradicsom Béla	Típus	Type of farm
Üzemeltető törzsszám	Operator in case of hatchery		Állapot	Status
Rétegg kód	Stratum in case of private farms	kiemelt egyéni	Körzet	District in case of private farms
Árbevétel	Turnover in case of enterprises (from Business register)		Létszám kategória	Number of employees in case of enterprises (from Business register)
Admin TEÁOR	Types of activity in case of enterprises (from Business register) - administrative		Stat TEÁOR	Types of activity in case of enterprises (from Business register) - statistical
Gazdálkodási forma	Legal type		Adatszolgáltatási készség	Respondent status

[Szerkesztés](#)
[Adatátadás GÉSA-nak](#)

Address	Contact information	Admin units	Activity codes	Animals	Crops	Predecessor	Successor	Log of modifications
Címek	Elerhetőségek	Kapcsolatok	M-kódok	Állatfajok	Földterületek	Jogelődők	Jogutódok	Adatváltozási napló

[+ Új cím létrehozása](#)

Típus	Megye	Irányítószám	Település	Teljes cím
Székhely cím	Heves megye	3335	Bükkszék	Piros utca 88.

Showing 1 - 1 of 1 items.

[← Vissza](#)

Three groups of users will be defined with different types of rights:

- viewer: can search in the list of farms, download data
- manager: import and process administrative databases, can modify information on farms, and make the linkages inside the program
- administrator: manages the system, grants user right

The main tasks that will be carried out inside the program:

- managing data on existing farms: updating their contact information and status
- searching for information, performing queries: for data validation or for carrying out sample surveys
- maintaining survey district: by having coordinates for addresses all new farms will be grouped to the closest survey district
- process new administrative information: updating the list of farms and their information, linking new administrative units to existing farms, creating new farms
- process information coming from data collection: updating the list of farms and their information, updating linked administrative units
- maintaining the activity of farms: to assign enterprises for different data collection

[Using the linked database as a frame for Agricultural Census 2020](#)

After the database linkages there were around 750 thousand units in the new register database, but this was supposed to be only a possible list of addresses to ensure the full coverage of agriculture. Most of them will not reach the threshold. According to the estimations and modelling there will be around 270 thousand farms surveyed within the framework of AC2020, who will reach the new threshold with their agricultural activity. This means that AC2020, besides its fundamental purposes, will also serve as a good opportunity to evaluate the quality of the accessible administrative database. By using the new Farm Register in the future, less addresses need to be visited.

[AC2020 Pilot survey, February 2019](#)

In February 2019 a pilot survey was carried out in order to evaluate the quality of the database linkages, and activity data stored in the administrative databases. The purpose of the survey was to detect the possible problems that might occur at the time of the AC2020.

The number of farms in the sample was 5 099.

According to the evaluation of the results the main problems that have to be solved by AC2020 were the following:

- explain well and clearly the definition of farm to the respondents: sometimes it was difficult to respondents to report data on all administrative units linked to the farms, as the administrative databases contains natural persons not farms in case of private farms,
- merging and splitting up of farms: it might happen that certain linked administrative units are not farming together and want to report their data separately. The same problem might also occur, namely that administrative units were not linked to a certain farm, so at the time of AC2020 it is solved on the e-survey to add additional units or split them into more farms.
- wrong addresses, foreign addresses: it is difficult to find farmers if they do not have updated address or they are living abroad. In case the surveyors cannot collect the data these units have to be imputed.
- rented area: in many databases the owner of the land is presented. If this person cannot give proper information on the farmer, who cultivates the land, the data of these units have to be imputed as well.

During the pilot survey farmers had four options to classify the linked administrative units, as:

- belongs to the farm
- belongs to the farm, but expired
- belongs to another known farm
- belongs to other unknown farm

But respondents classified records as expired in a very high share, even if according to the respective database they were still existing. Therefore during Agricultural Census 2020 this category will not be used.

Based on the experience of the pilot survey, respondents of AC2020 have to evaluate linked administrative units in the first part of the questionnaire. Units of the three main administrative registers are prefilled (Hungarian Chamber of Agriculture, Hungarian State Treasury, Licensed traditional small-scale producers). On the questionnaire it has to be reported, whether the linked unit (1) belongs to the farm, (2) belongs to another known farm, or (3) belongs to other, unknown farm.

Farms also have an option to add additional units, not linked to the farm by HCSO, and all these added units will be evaluated within the register after the data collection phase.

4. Persons related to the holding in the records of the Hungarian State Treasury

if there is administrative data pre-filled:

Please indicate in case of the following client codes of the Hungarian State Treasury that they belong to the holding or not. If not all codes listed, please add the missing ones.

Name	Client code (10 digits)	*Acceptance	Name of the other holding
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

*1. Belong to the holding (the holder own client code or client code of other persons with whom farming together) 2. Belong to other farm known (farming separately) 3. Belong to other holding not known

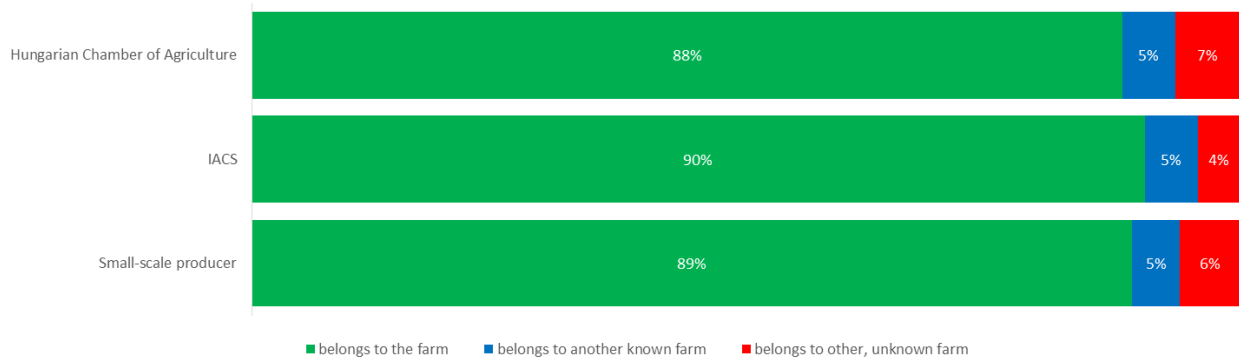
Does the holding have any other client codes related?

if there is no administrative data pre-filled:

Please list here the name and client codes related to the holding for which subsidies were requested from the Hungarian State Treasury.

Name	Client code (10 digits)
<input type="text"/>	<input type="text"/>

At the time of writing this report (26.06.2020) results show that farmers accept the linked units in a very high share. Only in case of bigger private farms the linked units are reported as belonging to another known farm in higher share than the average (~20%). This mainly means they report that family members are farming separately.



For most of the farmers living outside the country, the opportunity to fill the questionnaire online was provided by email. Problems with rented area will be treated in the data processing phase.

1.2. Improving data processing and validation systems

Data validation during data collection phase

Fully electronic data collection reduces time of data flow, so during AC2020 data will be transferred to HCSO database continuously. It means that experts are able to check them parallel to the data collection. As several data consistency checks are running at the time of data input, data entered by respondents have a very good quality.

As a result of the new Farm Register the administrative information, including information on production, will be used for data validation parallel to the data collection, even for some main variables the verification is done inside the data collection program at once.

The data collection program notifies the respondents, in case they forget to report certain areas or animals about which HCSO has information from any of the administrative databases. The error message warns the respondents that according to HCSO's information she/he might have. It ensures that data is reported exhaustively by all farms, and reduces the need of imputations.

A használt földterület művelési ágak szerint (2020. június 1.)

Melyik művelési ágba tartoznak/tartoztak a gazdaság által használt földterületek az alábbiak közül? ?

Művelési ág	Egyik sem
Szántó és szántóként használt kert	<input type="checkbox"/> Igen <input checked="" type="radio"/> Nem
Üvegház, fóliaház, fóliasátor (konyhakerti üvegház nélkül)	<input type="checkbox"/> Igen <input checked="" type="radio"/> Nem
Gyep	<input type="checkbox"/> Igen <input checked="" type="radio"/> Nem
Gyümölcsös	<input checked="" type="checkbox"/> Igen <input type="radio"/> Nem A KSH rendelkezésére álló adatforrások alapján Ön rendelkezhet ilyen területtel: gyümölcsös. Biztosan nem használ ilyen területet?
Szőlő	<input type="checkbox"/> Igen <input checked="" type="radio"/> Nem
Konyhakert (konyhakerti üvegház is)	<input type="checkbox"/> Igen <input checked="" type="radio"/> Nem
Erdő	<input type="checkbox"/> Igen <input checked="" type="radio"/> Nem
Nádas	<input type="checkbox"/> Igen <input checked="" type="radio"/> Nem
Halastó	<input type="checkbox"/> Igen <input checked="" type="radio"/> Nem
Gombatermesztésre használt alapterület	<input type="checkbox"/> Igen <input checked="" type="radio"/> Nem

Incoming data are continuously monitored and compared individually to other administrative databases. In case of agricultural enterprises and big private farms, inconsistencies are resolved in cooperation with the respondents, by email or phone. In the future, based on the experiences of Agricultural Census 2020, HCSO will evaluate the completeness and quality of different administrative databases, and use them instead of data collection in order to reduce response burden and data collection costs.

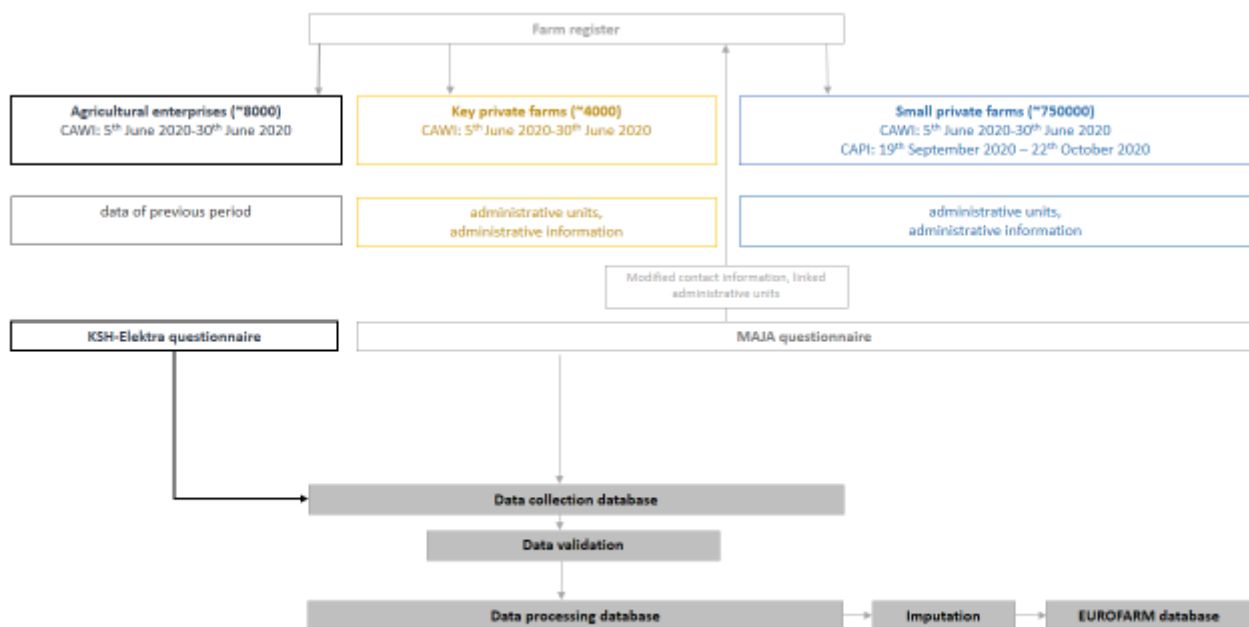
Data is collected via two different data collection applications. Agricultural enterprises report their data in KSH-Elektra system, while private farms report that in KSH-MAJA application (a newly developed data collection application). Data of both groups of respondents are coming into HCSO database, where the original reported data is stored before data validation. Therefore editing rate can be calculated, and the most problematic questions on the questionnaire can be detected.

In case of agricultural enterprises the following were prefilled in the online questionnaire:

- data of previous period: as the regular annual data collection is full scale in case of enterprises, all of them had previous data to be compared with during AC2020

In case of private farms the following were prefilled in the online questionnaire:

- administrative information: whether they have data in other databases
- linked administrative units: as in case of private farms it can be 1:n, so farmers are asked to validate them on the questionnaire



Data validation during data processing phase

In order to treat non-response and missing data, possible imputation and reweighting methods were evaluated. Previously, during FSS 2016, a donor imputation method was used, but only for elements about which former information was available. Previously, during the annual surveys, missing data were not treated in a harmonized way, different methods were used for agricultural area and livestock variables.

Due to the lack of reliable auxiliary variables and considering the unknown reliability of all administrative production data, it was decided not to use any reweighting methods in the short term (between 2020 and 2023). During this period the current imputation method will not change, but will be extended for all missing units and the whole imputation process will be carried out in HCSO's own data processing software. As for annual surveys, handling of missing data will be done uniformly, expert estimation for agricultural area variables will not be used in the future.

All imputation methods that are built into HCSO's own data processing software were evaluated for the purpose of agricultural statistics by methodological literature.

Taking into consideration the available information, the following unit level imputation methods will be used in the future for agricultural statistics in Hungary:

1. Census and structural surveys: nearest neighbour imputation will be used, because:
 - non-response rate considered to be low (around 1%)
 - there are variables that strongly correlate
 - it can be easily implemented within HCSO's data processing software
 - donor selection can be based on administrative information (production information or geographic information)
2. Annual surveys: imputation with data of previous period with trend will be used, because:
 - non-response rate considered to be low (1-2%)
 - times series are available
 - in case there is no data from previous period nearest neighbour imputation will be used (donor selection also will be based on administrative information)

1.3. New thresholds

Based on the last census information 72% of the farms used less than 1% of arable land in 2010. Therefore the increase of the threshold helps to avoid to visit a large number of small farms that are not a significant part of the universe.

Regulation (EU) 2018/1091 contains the list of the physical thresholds and the required coverage for the census.

In case of Hungary these thresholds would not ensure the 98% coverage criteria, therefore frame extension was needed.

Calculated coverage with EU thresholds would be the following:

	Number of farms	Agricultural area	Livestock in LSU
2010	203 967	97%	92%
2013	184 195	98%	94%
2016	176 372	98%	95%

According to calculations the current low thresholds would provide more than 99% coverage for utilised agricultural area and livestock unit also. But the modelling of the coverage by using different thresholds and variables resulted in that the current low threshold can be increased.

During the determination of the new thresholds the last census information was used, and the observed trend since then was also taken into account: the number of small farms has fallen continuously.

The following requirements were evaluated:

- The coverage exceeds the 98%,
- Select threshold variables that will be regularly updated and can be provided by good quality administrative databases. Which means categorization will be possible for all records in the Farm Register before the future sample data collection.

After investigating 46 different options, the following new thresholds were determined:

Hungarian threshold	
Productive area (arable land, kitchen garden, orchards, vineyards, grassland, forest, ponds, reed total)	1 ha
Orchard	0.25 ha
Vineyard	0.1 ha
Aromatic and culinary plants, herbs, flowers, ornamental plants, seeds, nurseries	0.2 ha
Potatoes	0.25 ha
Fresh vegetables and strawberries	0.25 ha
Greenhouses	100 m ²
Cultivated mushrooms	100 m ²
Cattle	1
Sheeps	10
Goats	10
Pigs under 20 kg	30
Pigs above 20 kg	2
Ostriches	3
Hens	50
Turkeys	30
Ducks	100
Geese	50
Rabbits	50
Agricultural service in the previous 12 months	

The new thresholds were consulted by the main stake-holders (ministry, government offices, chamber of agriculture), who considered this change very beneficial, as data collection can focus on the real agricultural activity in the future. Based on preliminary estimation there will be around 270 thousand farms in AC2020, 170 thousand above the IFS threshold and 100 thousand in the frame extension.

In order to have comparable time series, data of previous IFS were recalculated. In the tables it can be seen that main variables will have a good coverage even with the higher threshold though with significantly less farms to contact.

Recalculated data with new threshold compared to original data, %

	Number of farms	Standard Output	Agricultural area	Livestock unit	Total AWU
2010	61%	97%	99%	98%	78%
2013	60%	97%	99%	98%	79%
2016	64%	98%	99%	99%	82%

According to livestock unit it also can be seen that almost all main variables will have nearly 100% coverage with the higher threshold, only goats and bee hives variables will have lower coverage.

Recalculated livestock data with new threshold compared to original data, %

	Livestock unit	Cattle	Pig	Sheep	Goat	Hen	Other poultry	Bee hives
2010	98%	100%	98%	98%	75%	95%	97%	43%
2013	98%	100%	99%	98%	72%	95%	98%	42%
2016	99%	100%	97%	98%	77%	95%	97%	36%

The new threshold means the number of family labour force will decrease similarly, as will the number of farms, but the non-family labour force will hardly change.

Recalculated labour force data with new threshold compared to original data, %

	Total AWU	Family labour force, person	Permanent labour force, person	Temporary labour force, person	Family labour force, AWU	Permanent labour force, AWU	Temporary labour force, AWU
2010	78%	63%	100%	99%	71%	100%	99%
2013	79%	62%	99%	98%	71%	99%	98%
2016	82%	65%	99%	99%	74%	99%	98%

2.1. Improved sample design

Sample design

Before 2020, a cluster sampling method was used in case of Hungarian agricultural statistics. According to this method surveyors' districts were selected and all addresses within selected districts were visited to detect agricultural activity above the threshold. The reason of this method was that there was no regularly updated Farm Register.

On the other hand the share of CAWI questionnaires is quite low, therefore agricultural data collection relies on not only online but also personal interviews. Therefore travelling costs needed to be also taken into account when new sampling methodology was elaborated.

In the table below it can be seen that share of CAWI is higher only in case of special and structural surveys, when much more resource is spent on communication.

Data collection	Share of CAWI
FSS 2013	2%
FSS 2016	15%
Orchard survey 2017	51%
June survey 2018 - annual	8%
December survey 2018 - annual	11%
June survey 2019 - annual	11%
December survey 2019 - annual	12%

According to our assumptions, the new Farm Register will be a good list of addresses on the population of farms. Therefore it will not be necessary to visit every single address in a surveyor district in order to detect relevant agricultural activity. Based on the administrative information it will be possible to identify the addresses where agricultural activity may occur, and only farms having any agricultural activity will be visited and surveyed.

The Agricultural Census will also serve to verify the reliability of address information of administrative databases. According to the result of the pilot survey in 2019, the new register will be an appropriate although slightly over-covered sampling frame, as used databases contain some outdated existence information that cannot be filtered out before the data collection.

At first stage, redesigned sampling methodology will cover agricultural data collection only in the period 2020-2023. In order to decide between cluster sampling and stratified random sampling methods several options were evaluated, with different share of personal interviews and with different size-stratification.

For this investigation, last census database information (2010) was used to evaluate variables which have to meet the given precision requirements of annual surveys and FSS2023. In case of enumeration districts, also data of 2010 were used, to estimate average travel and questionnaire costs per personal interview. Estimated total cost of the following methods were compared:

- stratified random sample,
- one-stage cluster sample,
- two stage cluster sample (both equal and different sampling rates for different size categories inside the districts)

During the analysis it was assumed that agricultural enterprises and the key private farms (which are the biggest private farms) would be fully surveyed.

Naturally the stratified random sample gave the smallest sample size both for annual surveys and FSS2023. Due to the fact that the travelling costs are higher than questionnaire costs, the total cost difference between one-stage and two stage cluster sample was almost negligible, and the different sampling ratio for different size categories also did not cause considerable cost reduction. Therefore considering the estimated cost besides the actually low share of CAWI questionnaires it was decided to keep the present one stage cluster sample until 2023. Though as a result of the address list coming from the new register it will need less resources compared to the former practice when all address were visited inside the enumeration districts.

The number of smaller farms is decreasing continuously, so the last census database was slightly outdated for this investigation. Therefore after the Agricultural Census 2020 this investigation will be rerun based on the new data and the new district structure.

The preliminary results show clearly that the change in sampling methodology to stratified random sampling is only possible if share of CAWI questionnaires increases significantly.

During the Agricultural Census, almost 400 thousand farms or households have the possibility to complete online questionnaires. The HCSO self-developed new data collection program is planned to ensure that completion of the questionnaire is more convenient. An extensive communication campaign and incentive tools serve the purpose of raising the share of online questionnaires. Previous experiences show that those who completed online questionnaires during the FSS, will do that in annual survey also (if they are in the sample).

The long-term plan, after 2023, is to use stratified random sample where stratification will mean geographical (county level) and size stratification, and the methods of the data collections mainly will be CAWI or phone assisted CAWI questionnaires.

The above mentioned would require a good and regularly updated sampling frame, with exact information on the size of agricultural activity and also a sufficiently high share of CAWI questionnaires. As it is not the case right now in Hungary, the plan was rejected in short term.

New stratification

Within the framework of the current methodology the four following size (and legal) categories are distinguished:

- Agricultural enterprises (regardless their sizes),
- Key private farms
(around 3 000 farms with significant agricultural activity – using key-private farms thresholds),
- Non-key private farms (farms above the EU thresholds)
- Households with agricultural activity (between the national and EU threshold)

Agricultural enterprises and key private farms are fully surveyed in case of all agricultural surveys which are relevant to them. Sample surveys are carried out only in case of non-key private farms. According to current plans this will be the practice in the future at least until 2023.

In parallel with the determination of new farm threshold the former stratification methodology have been revised also.

During this revision the main focus was on non-key private farms, a category which was a very mixed set. It caused high CV values for some relevant variables and consequently caused unnecessary high sample sizes both for FSS and annual surveys. The basic preliminary idea was to split this layer into at least two categories. Similarly to the new threshold determination it also had to be considered to select variables that would be updated regularly and existed in good quality administrative databases. It is required for the categorization of all records in the Farm Register before the future sample data collection.

As first step the roof of this non-key private farm category was investigated. (For this investigation SamplingStrata and Stratification packages of R were used.)

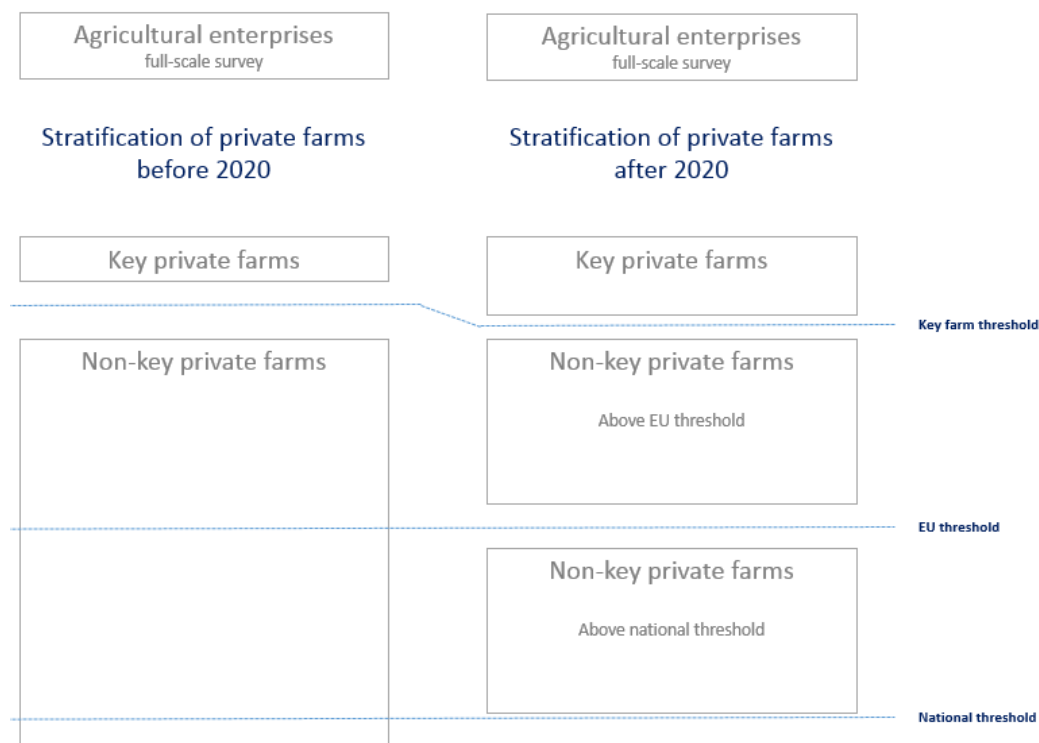
Based on the last census database, simulations were run and it was evaluated, if the key private farm threshold is changed a bit, how it affects the sample size for an FSS. (For this simulation stratified random

sample selection was used). According to the results of the modelling, if the key private farm category is extended from 3 000 farms to about 4 000 (by decreasing the thresholds for some variables), sample size of the non-key private farms would decrease significantly. Based on the investigation of different options, new key-private farm thresholds were finalized.

As next step it was investigated, how the non-key farm layer can be split into at least two size categories. Due to the fact that frame extension category had to be defined for agricultural census to provide the 98 percentage coverage for UAA and livestock unit, the non-key farm layer was split between the threshold of IFS in the legal act and our newly defined farm threshold. Therefore the EU threshold can be used as a layer definition.

As third step, similar simulations were run to see whether there is any gain if the new two categories were split into more subcategories. The frame extension category is very narrow between the two thresholds, so not surprisingly a further breakdown did not cause any noticeable effects. In case of category between the key private farm thresholds and IFS thresholds it was seen that further breakdown can have little effect. Considering that cluster sample will be used until 2023, it was decided not to use more size-categories until then.

Final cluster of farms for agricultural data collections for 2020-2023:



Agricultural enterprises are fully surveyed electronically in all agricultural surveys which are relevant to them. During this project we concentrated on the new stratification and sampling methods for private farms so the renewal of stratification of enterprises has been postponed until after the Agricultural Census. However, it is a future aim to further reduce the burden on respondents, and have a stratification of enterprises similar as in case of private holdings. It would reduce the sample size for smaller enterprises as well.

Household activity

Agricultural activity under the national threshold might be still significant. Because of the data needs of Agricultural and National Accounts, information needs to be collected on households, who have small agricultural activity under the thresholds within the framework of AC2020.

The questions were thoroughly investigated and it was decided to collect information on this kind of household activity parallel to IFS. Between two IFSs data will be estimated.

For this purpose, basic information on UAA and livestock number will be collected from all units, even if they do not reach the threshold. In addition a small sample (1%) were selected, therefore in 2020 it will be possible to check if there are any significant household activity in the addresses, which are not listed on any administrative databases.

2.2. Exploitation of new estimation methods

In order to treat non-response units and missing data, it was evaluated how re-weighting methods can be used in the future instead of the imputation methods or expert estimates we used earlier.

Right now the quality of record linkages is not completely clear. The completeness and timeliness of production information which is present in the administrative sources is also unknown. The Agricultural Census will serve to verify the reliability of quantitative information of administrative databases. Therefore, in the short term an imputation method is planned to be used for handling missing units, which considering the low non-response rate is not only acceptable but more practical. After AC2020 it will be possible to evaluate which auxiliary information from which sources could be used for any re-weighting method and also to decide whether it will be necessary or not. Preliminary tests shows the SAPS database could be a high quality auxiliary information for all crop variables, but apart from the cattle register there is no other animal register that is considered complete and reliable enough.

2.3. New survey districts, geo-coding, geolocation of farms

Current survey districts were manually created for Agricultural Census 2010.

When creating the current survey districts the main boundary condition was that each district contained more or less a similar number of addresses. Address means all available addresses in the district, regardless whether there is any agricultural activity or not. This resulted in that there can be big differences between the numbers of farms among districts, because the structure of agricultural activity and density of farms differs region by region, or even village by village.

For Agricultural Census 2020 new districts were created using ArcGIS software and each district will contain approximately a similar number of records of our new register database.

For this purpose, 750 thousand records on the address list were geo-coded, using HCSO's address register which contains coordinates for each address.

The algorithm that made the automatic district generation worked according to the following constraints:

- number of records was evaluated in each settlement. If the number was higher than 60 then addresses were grouped into more than one district. In case the number was lower, then only one district could have been created.
- number of addresses in each district should be 40 +/- 10%
- addresses grouped in the same district shall be accessed easily, therefore main address routes were also taken into account

After that the algorithm used Oracle PL/SQL Spatial technology and SDO_NN function. The result was that it defined survey districts based on the distance of points. After algorithmic creation of new districts finished, some manual work was also needed to modify districts if it was necessary. For the purpose a special spatial application was developed to allow regrouping of certain addresses.

The final districts were established by April 2020.

Farms were geo-coded based on their main, administrative address. In case of new farms an algorithm within the Farm Register will calculate the distance between the addresses, and cluster farms in the district of the closest address.

After the Agricultural Census a new district structure will be created which will be based on the number of farms instead of the number of records of our new register. This new structure will be used for sample design and sample selection for annual surveys until 2023.

For FSS2023 we are planning to use stratified random sample and if the survey cost will be manageable we would like to use this method for annual surveys also from 2024. In this case the maintenance of survey districts will not be necessary from then.

Geocode farms in case of EUROFARM database will be done also with the use of the Address Register.

2.4. Integration of existing administrative data sources

The administrative sources that will be used for the new Farm Register contain also data on agricultural production. It is planned that these data will be stored in HCSO database and aggregated at farm level, therefore data coming from data collections can be more easily validated. It is also expected that more and more data collections can be substituted by administrative data in the long term.

Within the framework of Agricultural Census, HCSO will use the following administrative databases to replace data collection in case of the following variables:

- organic area,
- organic livestock,
- rural development measures,
- participation in other environmental schemes,
- farmer is beneficiary as young or new farmer,
- permanent grassland eligible for financial support,
- grapes for wines (PDO/PGI)

Out of the 311 EUROFARM variables, 61 are planned to be collected from administrative sources, which means 20% of the total. 17 variables are non-existing or not significant.

Data validation and estimation methods will also be improved. Fully electronic data collection reduces time of data flow, so data will be transferred to HCSO database continuously. As a result of the new Farm Register the administrative information, including information on production, will be used for data validation parallel to the data collection, even for some main variables the verification is done inside the data collection program at once.

The following administrative sources will be used for validation:

- Hungarian State Treasury:
 - SAPS (Single Area Payment Scheme),
 - Other income supports scheme,
- National Food Chain Safety Office,
- Animal registers.

Summary

The project has resulted in the complete redesign of Hungarian agricultural statistics. In the future it will mean less respondent burden, better quality and more timely data, and will require less resources.

Direct data collection costs (paid to enumerators, coordinators, etc.) are estimated to decrease by 25% in nominal and 44% in real term.

Register

The linkage of databases provided a good address list to contact all farms in the population for AC2020 and beyond. The address list of possible farms gave an opportunity to contact less respondents, by visiting 70% less addresses than in the time of the last census ten years ago. The register application, and incoming future update of the different administrative databases will ensure that all new units will be included, and all information regarding Hungarian farms will be updated. In the future it will save costs and human resources, and also reduce respondent burden, as with a properly updated population frame stratification, sampling will be more effective. This was not possible until 2020.

New farm threshold

It makes possible to comply with the required coverage, with a significant reduction of respondents. The number of farms will decrease about 36%, while the required 98% coverage of utilized agricultural area and livestock will be kept. This results in a big reduction in respondent burden within the framework of AC2020 and beyond.

New stratification and sample design

It is expected that having four strata, having more farms in the key farm strata, and dividing stratum below the key farm level into two (above the EU threshold and between the EU and national threshold) will result in a smaller, better designed sample and better estimation of agricultural data.

Thanks to the regularly updating Farm Register, in the future it will not be necessary to visit all addresses in a survey district, but only those ones that exist in the register and have agricultural activity.

After 2023 it is also planned to switch to stratified random sample instead of cluster sample, as it would reduce further the required sample size. But it requires higher CAWI share, or data collection by CATI, as travelling costs can be high in case of stratified random sample if data is collected by CAPI.

Use of administrative data

HCSO has been using administrative data for agricultural statistics before (namely organic data, and data coming from Vineyard Register). But the linkage of databases makes possible to use administrative information more exhaustively. In the future the SAPS data, and Bovine Register data is planned to be used instead of data collection. This will significantly reduce respondent burden from June 2021, as data on sown area, and cattle will be not collected via questionnaire.

Besides substituting data collection by administrative data, validation can be also improved. Therefore those databases that do not provide full scale information on certain variables (like the number of poultry or pigs in the Animal registers) can be used also for validation purposes, not to miss data at item level on the questionnaire.

All of the above means the respondent burden and data collection costs are reduced already for 2020, but there will be further reduction in the future. HCSO is committed to use more and more available administrative data in the future.