FINAL TECHNICAL AND METHODOLOGICAL REPORT ON IMPLEMENTATION OF THE ACTION CARRIED OUT WITHIN THE GRANT AGREEMENT

AGREEMENT NO:	**2018.0218**
DATE OF SIGNING OF THE AGREEMENT:	**18.07.2018**
PARTIES OF THE AGREEMENT:	• **STATISTICS POLAND** • **EUROPEAN COMMISSION - EUROSTAT**
TITLE OF THE ACTION:	**MODERNIZATION OF AGRICULTURAL STATISTICS**
ELIGIBLE PERIOD:	**14.12.2018 – 13.12.2020**
REPORTING PERIOD:	**14.12.2018 – 13.12.2020**

Project leader:
(name and surname)

Leading unit: **DEPARTMENT OF AGRICULTURE**

Director of leading unit:
(name and surname)

Warsaw, 12.03.2021
(place and date)

„The action described in this report has been implemented with the financial assistance of the European Union. The sole responsibility lies with the author and the European Commission is not responsible for any use that may be made of the information contained therein.

Table of contents

I. General information

On 20 June 2018, an agreement number 2018.0218 was signed between the European Commission, represented by Mr. Marcel Jortay, Director of Directorate E - Regional and Sectoral Statistics, and the Central Statistical Office, represented by President of the CSO, for the action entitled "Modernisation of agricultural statistics". The eligible period for the implementation of activities covered by this agreement was from 14.12.2018. until 13.12.2020. The overall objective of the action was to improve the statistical system for the activities related to the Integrated Farm Structure Survey (IFS) and ESS requirements by updating and modernizing the **Agricultural Survey Sampling Frame (Pol. OdBR)**, making greater use of administrative registers and non-administrative sources, improving existing methods of collecting information and exchanging good practices related to agricultural statistics.

The following specific objectives were pursued:

1. Modernisation of OdBR by extending the sampling frame with spatial data and its integration with the sampling frame for social research, in order to obtain a wider range of information about the farm users and the households with the farm user.
2. Updating of OdBR with data from administrative registers and non-administrative sources for livestock data.
3. Creating agricultural statistics on the basis on administrative livestock units (cattle, sheep, goats).
4. Exchange of good practices in the field of conducting farm structure surveys, including agricultural censuses.
5. Consolidation of the data collection process in IFS and representative agricultural surveys.

II. Description of the completed works in the project and their results.

1. Modernization of OdBR by extending the sampling frame with spatial data and its integration with the sampling frame for social research, in order to obtain a wider range of information about the farm users and the households with the farm user.

These works were divided into two separate activities, implemented independently:

a) Modernization of OdBR by extending the sampling frame with spatial data.
b) Integration of OdBR with the sampling frame for social survey to obtain a wider range of information about the farm user and the households with the farm user.

Ad 1a)

Methodological part.

The main goal of this task was to increase the quality of the agricultural survey frame (OdBR) in terms of agricultural holdings and holder’s residence spatial location.

Statistics Poland has at its disposal a wide range of spatial data with location of addresses. Since the 2010-2011 census round Statistics Poland maintains Spatial Address Databases (Pol: PBA), which hold locations for buildings with dwellings. PBA is the spatial representation of the NOBC subsystem of the TERYT register. NOBC holds information on addresses, buildings and dwellings. The databases are updated quarterly based on information received from county (former LAU1) authorities on changes in the state of buildings (notifications about new buildings or buildings that cease to exist). PBA is the main in-house source of address reference for all survey frames.

Information on dwelling location is not sufficient for all survey frames. The agricultural survey frame cannot be geocoded using PBA data alone, since not all addresses in the frame are connected to dwellings. Therefore Statistics Poland had to seek out external address data sources. The most mature nationwide address point dataset is disseminated by the national mapping agency – the Head Office of Geodesy and Cartography. Since 2012 municipality (former LAU2) authorities are obliged to maintain the Register of Localities, Streets and Addresses (EMUiA) with locations of address points on their territory. Datasets prepared by municipality authorities are aggregated on national level by the mapping agency and distributed via the State register of boundaries and areas of territorial division units of the country (Pol: PRG). This data source covers the whole country and the data comes from the lowest territorial division unit (municipality). Address data is maintained locally at the minicipality level, therefore it is presumed, that the location accuracy of the source is reliable.

Statistics Poland acquires the PRG address point dataset on a regular basis for various purposes, including georeferencing survey frames. Regular analysis of the dataset proved that its quality is increasing over time and while it is still not perfect, it is the most complete external source of address points.

Addresses in statistical survey frames are described by identifiers from the TERYT register, which is maintained by official statistics. The identifiers describe: administrative units, localities and streets (if applicable). To facilitate record matching between the survey frame addresses and external address point sources, these sources need to contain the same identifiers.

Address points in PRG are distributed in GML (Geography Markup Language) format. Datasets consist of the following feature classes: address points, localities and streets. The data contains TERYT identifiers but they are not stored in a way, which allows unequivocal assignment of identifiers to addresses: TERYT identifier of the locality (SIMC) is available only in the locality feature class, TERYT identifier of the street (ULIC) is available only in the street feature classes. There are no uniform identifiers to connect addresses to localities and streets and the address point feature class references localities and streets only by name. Therefore the address point, locality and street tables can only be matched using names (character strings) in order to transfer locality and street TERYT identifiers to the address point feature class. Table matching by character strings is the least effective way of matching, mainly due to typing errors. Address characteristics in the GML files are provided in a non-explicit way. Datasets are disseminated at voivodship (province) level, but localities are not described with the municipality identifier, and streets do not have neither the municipality nor the locality identifier (or at least name). Same locality and street names appear multiple times within one voivodship (province). This problem persists since the data source has been published by the national mapping agency, and while the

quality of data increases in terms of completeness and correct names, the absence of identifiers in the key address point feature class still hampers the usability of the data source.

The agricultural survey frame (OdBR) acquires address data from various sources, including the Agency for Restructuring and Modernisation of Agriculture (ARMA). Therefore the quality of addresses is not always satisfactory. Statistical Office in Olsztyn regularly performs activities aiming at increasing quality of address data in the survey frame. These include record matching by dwellings with the NOBC dwelling subsystem of the TERYT register to assign locality and street TERYT identifiers. When OdBR addresses are matched with NOBC, they also receive the NOBC dwelling logical identifier to facilitate instant matching with PBA address points. Other activities aiming at increasing quality of OdBR address data include matching with the population register (PESEL) by dwellings.

Description of the task implementation

1) Point locations for agricultural holdings and holder's residence,

In order to assign georeference to addresses in OdBR a series of matches have been performed.

First, the addresses were matched with PBA address points by the NOBC logical dwelling identifier, which yielded a result of over 86% geocoded addresses. Second, the addresses were matched with address points from the Register of Localities, Streets and Addresses (EMUiA) by municipality, locality and street identifiers and the address number. This increased the outcome to almost 94% matched addresses. Over 6% of OdBR addresses remained unmatched, which implied unsatisfactory quality of OdBR address data or insufficient completeness of address point data sources. The unmatched addresses were further analysed: over 60% seem to have correct address information, while almost 40% have various information deficiencies, including: lack of the SIMC identifier for localities with a locality name present, lack of the ULIC street identifier with a street name present, lack of any address information apart from the municipality identifier, lack of or wrong address number.

The OdBR addresses have undergone further processing in Statistical Office in Olsztyn to increase their quality. Afterwards another series of matches have been performed – this time synchronizing the reference dates for the datasets (previously matches were performed with several versions of address point datasets). The reference date has been set to 31.12.2018. Again, the OdBR addresses were matched with PBA address points, this time yielding a result of 89% geocoded addresses (3 p.p. increase). Afterwards they were matched with EMUiA address points, which increased the outcome of geocoded addresses to 94,5% (0,5 p.p. increase). The unmatched addresses were analysed: over 73% seem to have correct address information, while almost 27% still have information deficiencies mentioned in the previous paragraph.

The second matching operation proved that the quality of OdBR has increased in two aspects: the overall quality of addresses including their identifiers (less addresses with information deficiencies) and the integration of addresses with the NOBC dwelling database. However, the overall matching result is still not satisfactory. The second matching yielded only 0,5 p.p. more geocoded addresses than the first match. The addresses that previously did not match with the NOBC dwelling register have then matched partially with EMUiA address points. The second matching resulted in more

addresses matched with NOBC but then less unmatched addresses were geocoded with EMUiA data. This leads to the conclusion that further works should be focused on increasing the quality of OdBR addresses and/or seeking out other address point data sources, because it appears that the set of addresses that cannot be geocoded is roughly the same in both matching attempts (0,5 p.p. difference).

2.) Agricultural holding locations from non-address data sources.

Both address matching operations resulted in approximately 5% of unmatched addresses for holding locations and holder's residence locations. The share of unmatched holding locations is approx. 5,5%, while the share of unmatched holder's residence locations is approx. 4,5%. There is no possibility to improve the latter without enhancing the address quality in the agricultural survey frame (OdBR), but there are other data sources, which can deliver agricultural holding locations.

The Agency for Restructuring and Modernisation of Agriculture (ARMA) maintains a database of agricultural producers, which provides connections between agricultural holdings and cadastral parcels. Cadastral parcel geometries are available in the Land Parcel Identification System (LPIS) also maintained and made available by ARMA.
There were 5,5% agricultural holdings for which there were no locations determined based on matching addresses in the survey frame with address points from various sources. These holdings were further processed to acquire spatial location from LPIS via the database of agricultural producers.

For 62,3% of these holdings, a centroid of the largest cadastral parcel listed in the agricultural producers database was assigned as holding location. The remaining holdings, which constitute 2% of the whole survey frame (OdBR) had no reference to any cadastral parcel.

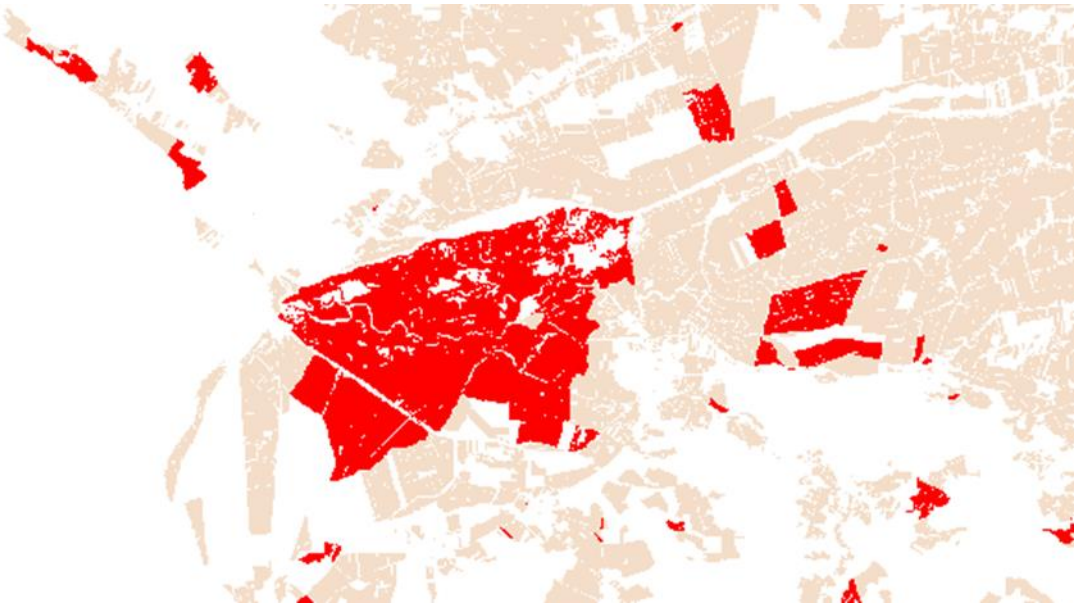
Matching operations performed on addresses and parcels resulted in assigning location to almost 98% of agricultural holdings (either based on address points or parcels). 2% of agricultural holdings remain unidentified on the map. These holdings have no reference to cadastral parcels in the agricultural producers database and holding address data is either missing, incomplete or incorrect. Existence of these (almost 33000) holdings should be verified. Since there is no possibility to locate them or provide sufficient address or parcel data, they should probably be removed from the survey frame.

3.) Polygons for agricultural holding extents.

Another aspect of georeferencing agricultural holdings was to use cadastral parcel data to determine holding extents using parcels. The initial plan was to use parcel data from the Land Parcel Identification System (LPIS) maintained by the Agency for Restructuring and Modernisation of Agriculture (ARMA). LPIS holds boundaries of whole parcels, regardless of the actual parcel area used for agriculture. Aside from LPIS, ARMA made another spatial dataset available to official statistics with extents of arable areas, grasslands and afforested agricultural areas. Since this dataset gave a better picture of which areas were actually used for agricultural production, it has been decided to use it in place of LPIS.

To connect spatial data for agricultural parcels to holdings, a database of agricultural producers was used. This database is also maintained by the Agency for Restructuring and Modernisation of Agriculture (ARMA). During a preliminary analysis of the database it was discovered that there are a lot of cases, where a single agricultural parcel belongs to more than one holding. It was decided that all holding <-> parcel relations have to be maintained in the process of determining holding extents, therefore a many-to-many relation table was exported from the database with several million records.

To develop the methodology for determining agricultural holdings' extents, a relation table for one voivodship (province) (opolskie) was extracted. A pilot aggregation of parcel geometries to holdings confirmed the fact that some parcels belong to more than one (sometimes several holdings). The relation table contained within its attributes information on area of a given parcel, which is assigned to an agricultural holding. In theory it was possible to divide parcels belonging to more than one holding proportionally to the share of area belonging to each holding. However there was no possibility to determine how to correctly divide the areas due to two problems. First, the area shares for holdings did not always add up to the total area of the parcel, so there was no possibility to ensure a correct proportional division. Second, there was no possibility to determine how to divide the parcels spatially. The majority of holdings which owned a share of a parcel, had other parcels in distant areas not neighboring the parcel to be divided. Therefore even if it was possible to perform a proportionally correct division, there was no way to know how to draw boundaries between parts and which parts should belong to which holdings.



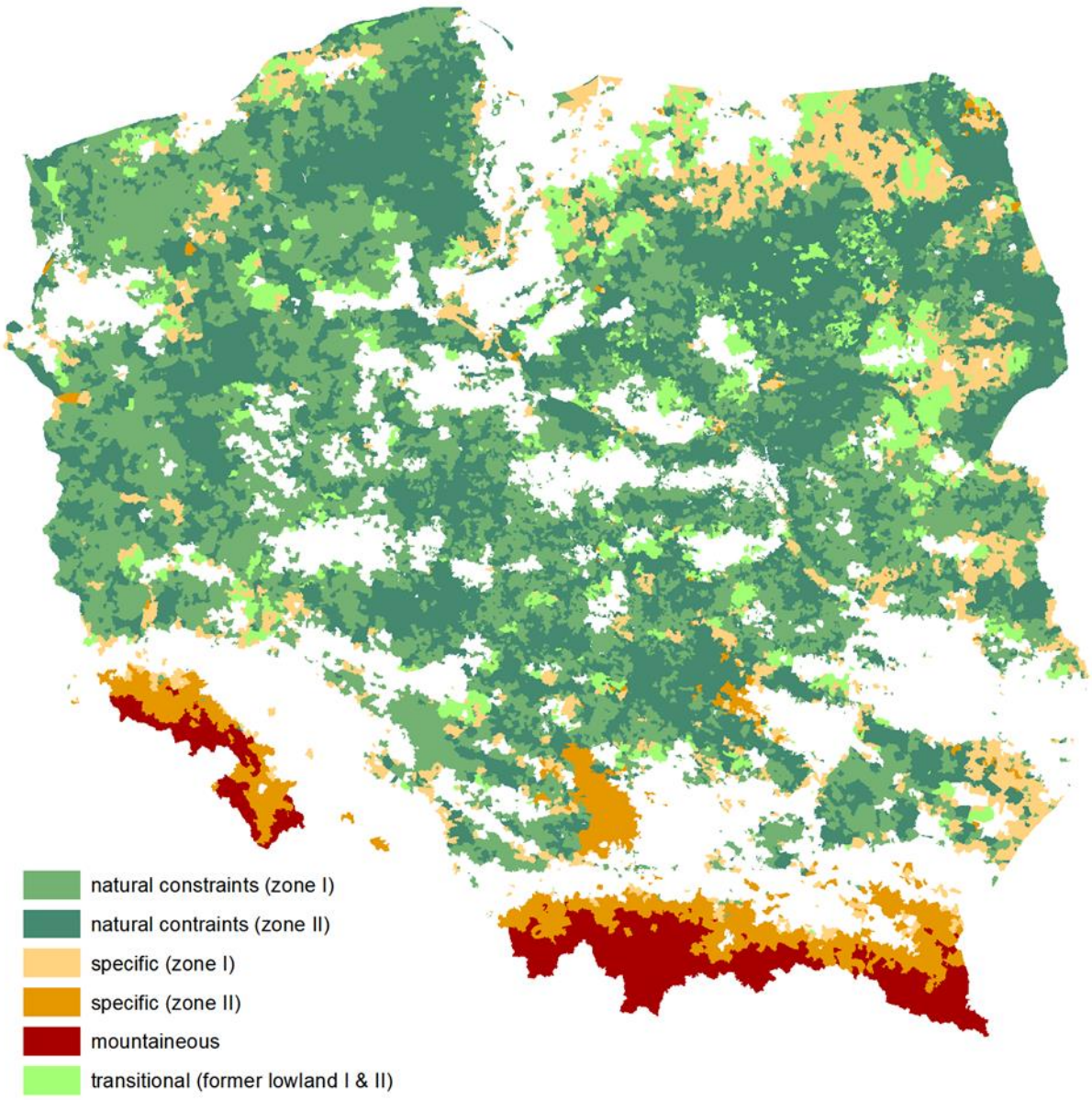
An example of large agricultural parcels belonging to more than one holding (marked in red), Warta river valley near the city of Kostrzyn, Poland

Taking the pilot aggregation into account, it was decided that in cases of shared parcel areas, the whole parcel area will be assigned to each holding that has a relation to it. The result will not be correct in terms of the actual holding area but it is the best possible way to show (and not lose information about) where agricultural production for each holding takes place.

Extents were aggregated for over 1,3 million agricultural holdings. Holding to parcel relation data was available for both 2018 and 2019, so the aggregation was performed for both years. Data for 2018 yielded a 99,7% share of holdings for which it was possible to determine the extent. A slightly worse share (99,41%) was acquired for 2019 data.

4) Less Favoured Areas' boundaries.

Another product of the project is the boundaries of Less Favoured Areas (LFA). Those areas are defined as groups of cadastral units. Boundaries of cadastral units are disseminated by the national mapping agency as part of the State register of boundaries and areas of territorial division units of the country (PRG), therefore the LFA dataset is a product of polygon aggregation using cadastral unit boundaries. The LFA delineation was updated in 2019, therefore a new dataset had to be produced. LFA boundaries will be used to assign the LFA attribute to holdings in the Agricultural Census 2020 and subsequent Farm Structure Surveys.



Less Favoured Area map

Results of work and recommendations.

As a result of works in the project, matching operations performed on addresses and parcels resulted in assigning location to almost 98% of agricultural holdings.). 2% of agricultural holdings (almost 33000) remain unidentified on the map. These holdings have no reference to cadastral parcels in the agricultural producers database and holding address data is either missing, incomplete or incorrect. Existence of these holdings should be verified in Agricultural Census 2020.

In terms of determining the size of a farm on the basis of cadastral plots, a pilot aggregation of plot geometry for farms was performed. The result of the work allows us to conclude that this method of determining the size of a farm should not be recommended for the future.

The result of the work in the project is a new LFA dataset, which will be used to assign the LFA attribute to farms in the Agricultural Census 2020.

Ad 1b)

1. Methodology for integration of the Agricultural Survey Frame (OdBR) with Integrated Population Database.

Abbreviations used

Abbreviation	*Full name*
PESEL	Personal identity number (Social Security number)
NOBC	Address identification system for streets, real estate, buildings and apartments
OdBR	Agricultural Survey Frame
IPD	Integrated Population Database
SEZ	Variable records system
SIMC	System of identifiers and place names
ULIC	Central street catalogue
KRUS	Agricultural Social Insurance Fund
IACS	Integrated management and control system

The goal of integration of IPD with OdBR is to obtain a wider range of information about the holder and about persons remaining in households and families with the user by place of residence. Additionally, the integration process allows for subject verification of the OdBR in relation to the holder.

For the implementation of the above goals development of new methodological solutions was assumed, including:

a) methods of integrating two frames by working out an integration key/keys,
b) methods for determining the number of holders of a farm and members of its household by residence.

The result of the activities will be the development of methodology for updating the OdBR based on a fixed number of holders and members of its household.

Therefore, in this process, the most important activity is to check the fact of the presence of the holder in the integrated population database, compliance of the holder's residence address established in the IPD and the address in the OdBR, and in the next stage identification of the user's household composition that will allow determining the residence address of the holder and his/her family.

1.1 Methods of combining (integrating) OdBR and IPD data and connecting keys

Data sources have different structure, IT solutions, objective and subjective scope, function depending on the purpose for which they were created. The key action is therefore to properly compile the information they contain, which is possible in the data integration process.

The integration consists in combining data from many heterogeneous sources (concerning the same or similar entities) so as to obtain a uniform, coherent image of the collected data and increase their scope.

Statistical data integration is justified when:

1. the reference periods of both integrated sets are the same,
2. collections refer to two identical or different but partly overlapping populations.

Data integration is carried out at the micro level, i.e. data on a single object is combined. This is the most effective way of combining data, which, however, requires the availability of a unique key for combining data or a set of common features that allow for unambiguous identification of the examined object. In the second case, these are the features that occur in both sets, are characterized by the same (or very similar) definition and agreement in terms of the value of the features. A set of common features does not always fully identify units, but on the basis of specially defined criteria, e.g. similarity of record pairs, it is possible to indicate the same units or units very similar to each other. Combining by identifiers is considered the basic method. By design, unique identifiers directly point to specific entities. E.g. PESEL will be the basic identifier for persons, and REGON for national economy entities. The condition for combining data sources by identifiers is the existence of a unique identifier that does not contain empty values in the reference database/register. In this way, a synthetic set is created containing full information for the examined units, containing all specific features from different sets.

In the case of IPD and OdBR integration, both of the above assumptions are met, i.e. the integrated sets have the same reference period and contain a partially overlapping population for holders. In addition, both integrated databases use the NOBC statistical database (address identification system for streets, real estate, buildings and apartments) to determine the correctness of address data.

For the purposes of IPD and OdBR integration, a data connection was assumed in the scope of persons who are holders using first of all the PESEL number appearing in both integrated databases. In the case of unpaired records of holders occurring in the OdBR, an alternative method of combining records was used, which consisted in comparing the set of common features found in both sets, in the form of address of residence, sex and date of birth.

Identified holder records have been appropriately marked in the integrated population database by adding a field containing the relevant information.

As a result of pairing carried out in the OdBR set, people who were not found in the IPD were identified, including those without a PESEL number. In addition, records were identified that in IPD had markers eliminating such units from the target population. They were e.g. persons who went abroad, receiving benefits outside Poland, or persons who appeared only in one register used to build the integrated population database. Such records were subjected to detailed analysis and then excluded from further integration process.

The results of the work carried out in the field of the initial integration of OdBR and IPD are described in point 3. Appropriate pilot study for this task will be carried out in the second year of project implementation.

1.2 Method for determining information on the holder and his/her household

To implement the action, which is to determine the method of developing information on the holder and members of his/her household, decision was made on using the so-called “Family”. Commencing work on introducing such a solution, i.e. creating a database containing information on family structure at the lowest level of the territorial division of

the country based on data from administrative registers, was caused by information needs and cyclical provision of data on families.

It should be added that determining the full population of families and their demographic and social structure at the lowest possible level of territorial division of the country can be obtained only on the basis of the results of population and housing censuses, i.e. at intervals of about 10 years. In the inter-census periods, basic data on families can be compiled on the basis of information from current sample surveys (e.g. European Survey on Living Conditions - EU-SILC, Household Budget Survey. However, these surveys are conducted on too small samples, preventing presentation of more detailed data, especially at lower levels of territorial aggregation.

Works on the construction of the "Family" domain were carried out as follows:

a) selection of data sources and preparation of auxiliary variables - the process of separating families was carried out using the integrated population database (derived from administrative collections), PESEL register and the result set from the last General Census 2011 containing information on marital status, spouses and parents;

b) work with the PESEL register - development of the input set included e.g. conversion of the legal marital status variable, and supplementation, verification and deduplication of data on variables related to spouses, in particular the spouse's PESEL number;

c) development of assumptions for the so-called parentification process, i.e. combining child records with parent records. In the process of parentisation, as a result of the matching examination of parents' surnames and first names carried out within the same address, taking into account the age difference between potential children and their parents, the parents were assigned to the children;

d) extracting biological families - the stage of assigning people to a specific family by giving them a unique family identifier and determining the position of particular persons in the family, including:

- determining which generation a person belongs to,
- separating families such as marriage or partnership,
- adding children to families or creating new families of single mothers or fathers with children,
- territorial assignment of separate families.

The final effect of the work was the creation of the FAMILY table. The table contains basic data about each of the identified families, i.e. information about the commune of residence, type of biological family and number of children.

1.3 Assumptions for pairing data from registers

For the purposes of conducting the process of determining the composition of the household/family of the holder, at the first stage of work the main assumptions and management steps were defined, including:

a) identification in the integrated population database - based on the identifier (PESEL number) and alternative connection keys - holder,
b) designation of integrated people records,
c) selection of additional variables used to extract members of the holder's family,
d) adding the family identifier from the integrated population database to the collection,
e) generating an integrated set,

f) analysis of unpaired cases.

The process of extracting members of the household/family of the household user is based on the analysis of variables in terms of the person and the established family composition based on the family identifier. In addition, the home address of the household user and assigned family members is checked. As part of this process, the records of people living at the same address are identified, followed by family members. This approach is dictated by the fact that in the "Family" table created by the process of parentalization other people living at the same address are not identified and properly marked if they do not form a family. They can be, for example, persons from outside the biological family forming separate households.

Particular emphasis was placed in this process on checking the consistency of the address features of both integrated sets due to the fact that this is the main variable that allows people to be assigned to the same household and family. Works in this area included:

- coded address - an address that has the NOBC apartment identifier, or in the absence of this identifier at least the SIMC city code,
- matching address - the coded address for a single record, which appears on both sides (OdBR and the integrated population database) and is identical, i.e. has the same NOBC apartment identifier or the same address after SIMC, ULIC codes, building number and apartment number,
- unique address - an encoded address that appears on one of the sides (OdBR or the integrated population database),
- common address - a unique address that appears on each OdBR and the integrated population database and is identical, i.e. has the same NOBC apartment identifier or the same address after SIMC, ULIC codes, building number and apartment number.

2. Results of work.

The prepared set from OdBR contained 1582002 records, 98% of which were paired with IPD, provided that at least one field was consistent with information from the integrated population database.

The results of pairing are presented in tab. 1:

Tab.1 Number of paired and unpaired farms

Specification	Number of agricultural farms
Farms with OdBR paired with IPD	1560112
Number of farms with the NOBC identifier in OdBR and IPD	1298848
Number of holdings with different NOBC identifiers in OdBR and IPD	125412
Filled NOBC identifier in dDBR and no NOBC identifier in IPD	16259
No NOBC identifier in OdBR and NOBC identifier filled in IPD	79420
No NOBC identifier in OdBR and no NOBC identifier in IPD	40173

Unpaired farms with IPD	21890
Including:	
- completed NOBC identifier in OdBR	18332
- no NOBC identifier in OdBR	3558

Address analysis shows that there are situations where:

- NOBC apartment identifiers are the same on the OdBR and IPD sides, while addresses differ by SIMC, ULIC codes, house number and apartment number.

- NOBC apartment identifiers are different on the OdBR and IPD sides, while addresses are the same after SIMC, ULIC codes, building number and apartment number.

- the address for various farms within the family after the SIMC and ULIC codes, the building number and the apartment number is identical, while only some of them have been coded with the NOBC apartment identifier.

Tab.2 Families by number of farms and addresses	
Specification	Number of households
No family identifier	245862
1 family = 1 farm	1233519
1 family> 1 farm (at one address)	79369
1 family> 1 household (at more than one address)	23252

Of this population, 84% of records received a family identifier, the remaining part of unpaired records was subjected to detailed analysis. It should be noted that unpaired records from both sets also apply to the situation in which despite the matching of features it was not possible to assign a family identifier.

Tab.3 Number of agricultural farms as part of a family								
	Number of various addresses							
Number of farms within the family	1	2	3	4	5	6	7	8

no family id	245862	0	0	0	0	0	0	0
1 agricultural farm	1233519	0	0	0	0	0	0	0
2 agricultural farms	37045	10961	0	0	0	0	0	0
3 agricultural farms	1579	332	42	0	0	0	0	0
4 agricultural farms	122	33	8	2	0	0	0	0
5 agricultural farms	8	5	1	0	0	0	0	0
6 agricultural farms	1	1	0	0	0	0	0	0
7 agricultural farms	0	0	0	0	0	0	0	0
8 agricultural farms	1	0	0	0	0	0	0	0

A (biological) family means two or more people who are connected as a husband and wife, people in informal relationships or a parent and a child. Thus, a family includes a couple without children or a pair with one or more children, or a single parent with one or more children.

Reasons for not assigning a family ID:

- the person has no children and spouse or is not in an informal relationship,

- for a given person, there were missing data for variables that were used to determine the family identifier.

The analysis of initial pairing showed that nearly 78% of households of holders consist of 1 family.

3. Recommendations before proper merging (integration) of frames.

The result of the activities is the development of methodology for updating the OdBR based on a fixed number of holders and members of its household. Therefore, in this process, the most important activity was to check the fact of the presence of the holder in the integrated population database, compliance of the holder's residence address established in the IPD and the address in the OdBR, and in the next stage identification of the user's household composition which allowed determining the residence address of the holder and his/her family. New methodological solutions were developed, including:

a) methods of integrating two frames by working out an integration key/keys,

b) methods for determining the number of holders of a farm and members of its household by residence.

Analysis of the results obtained after the first, initial pairing of both sets, i.e. the Integrated Population Database together with established family relations and OdBR showed that the assumptions adopted for the integration process are correct. However, additional validation is necessary. The changes will apply to the following activities:

- exclusion or modification of the Name 2 variable from the identification data compliance check due to the many missing values and errors in this variable should be considered. 137 080 non-compliances with another second name was found. Only 3 468 of these cases were non-empty, i.e. Name 2 filled in on the OdBR and IPD side. The most common mistakes are the so-called typos and spelling mistakes in the variable Name 2.
- in the process of determining address consistency, assumptions should be established that resolve differences between addresses from OdBR, integrated population database and compliance with the NOBC database,
- it is necessary to refine the validation elements in the process of separating families and include them in the implemented process of parentisation in the context of identification of household members of a holder.

2. Updating of OdBR with data from administrative registers and non-administrative sources for livestock data.

2.1 Methodological part.

The administrative register of the following state institutions with information on livestock in their systems have been identified:

1. Chief Veterinary Officer register, which contains information about:
   - Agricultural producers rearing or breeding farmed wild animals as farm animals,
   - Poultry production plants in accordance with Directive 2009/158/EC, breeding and rearing: chickens, ducks, geese, turkeys,
   - Farms producing eggs for consumption, registered in accordance with Regulation (EC) No 852/2004 of the European Parliament and of the Council of 29 April 2004 on the hygiene of foodstuffs (Official Journal L 139 of 30 April 2004, pp. 1-54), hereinafter referred to as 'Regulation No 852/2004', not covered by Commission Directive 2002/4/EC of 30 January 2002. on the registration of plans breeding and rearing laying hens, covered by Council Directive 1999/74/EC (OJ L 283, 31.10.1999, p. 1). Journal L 30, 31.1.2002, p. 44-46), hereinafter referred to as "Directive 2002/4/EC". The register takes into account the number of poultry up to 349 (laying hens, quails, ostriches),
   - Farms producing eggs for consumption, registered in accordance with the Regulation No. 852/2004, covered by the Directive 2002/4/EC, with 350 laying hens and more, the register also has information about the maximum capacity of the poultry house, according to the system of rearing: organic production, free-range and cage,
   - Farms of carnivorous fur animals, in which animal by-products are used.

2. District Veterinary Officer register (pol: GLW) which contains information about:
   - Farms engaged in breeding, beekeeping and honey production,
   - Poultry production establishments (chicken broilers),
   - Farms of herbivorous fur animals.

   GLW registers are kept in one vetLINK, MARCEL S.A. system, which facilitates data transfer to the Central Statistical Office (CSO).

   Registers of District Veterinary Officers are kept in scattered, often non-unified databases.

   The system of dispersed databases makes it very difficult to obtain data for statistical purposes.

3. The Agency for the Restructuring and Modernization of Agriculture (ARMA) has in its resources, among others, the following register: Central Database of the System for Identification and Registration of Farm Animals Labelled (Pol: CBD IRZGO) i.e. cattle, sheep, goats and pigs.
   Each owner of a farm animal is obliged to report to the CBD IRZGO, through the ARMA poviat offices, a change in the size of the herd, specifying the date and type of the event and the animals it concerned within the time limits specified in the regulations.
   The information contained in the Central Database enables, among others:
   • to determine the number of farm animals labelled with accuracy to the municipality level,
   • checking the history of each registered animal in relation to the places and times of stay reported by the animal's owners,
   • checking the history of each manufacturer's registered office in relation to the animals owned and their residence time, compliance with statutory reporting deadlines, and the epizootic status,
   • making analyses, summaries, and statistics.
   In the CBD IRZGO register, identification data of agricultural producers holding these livestock (cattle, sheep, goats and pigs) are collected, which may be useful for updating the agricultural research frame. Due to the scope of information in the structure of the livestock herd, only information about the cattle, sheep and goat population can be used to replace data from the CSO survey with administrative data of the ARMA.

4. The Chief Inspectorate of Agricultural and Food Quality (Pol: GIJHARS) in its resources has a register of organic animal production (based on Article 76 of Commission Regulation (EC) No 889/2008), including, among others:

- Register of animals arriving at the farm,
- Register of animals leaving the farm,
- Grazing register.

The Chief Inspectorate of Agricultural and Food Quality keeps a register of organicanimal production, which collects data that can be used to update the OdBR.

5. The Polish Union of Horse Breeders (Pol. PZHK) runs the Central Equine Database (Pol. CBDK), which is implemented under the Act of 2 April 2004 on the system of identification and registration of animals.
The Central Equine Database is an application for district veterinarians, among others. It contains data on equidae registered in Poland.
Cooperation with administrators of administrative registers was established to make arrangements for the possibility of using the recognized data source in agricultural statistics survay. Meetings were held with representatives of the Veterinary Inspectorate, Agency for Restructuring and Modernization of Agriculture, Chief Inspectorate of Agricultural and Food Quality and Polish Union of Horse Breeders to discuss the proper preparation of data sets and their safe transfer to the CSO.

The legal basis allowing for transferring data from administrative registers to the CSO are the records in the Program of Statistical Survay on Public Statistics (Pol: PBSSP) for a given year accepted by the Prime Minister. During the meetings with the representatives of the administrators of administrative registers in the field of livestock data, the records to the PBSSP were established specifying the scope, deadline, and frequency of data provision for statistical purposes.

Results of analysis and recommendations of the possibility of using data from administrative registers to update OdBR.

1. From the methodological, organizational, and technical point of view, the IRZGO ARMA and CBDK PZHK registers are the most useful to be used by CSO to update OdBR. Data records from the above-mentioned CSV-format registers and the correct use of TERYT register identifiers is suitable for use in statistics.

2. The data set from the GIJHARS Organic Animal Production Register saved in Excel format together with the use of the TERYT register identifiers is possible to use for statistical purposes.

3. Registers of District Veterinary Officers are a dispersed register, which is not uniformly maintained. Due to the lack of a uniform data structure in the District Veterinary Doctors' registers, the lack of use of TERYT register identifiers, the Veterinary Inspection registers require a lot of IT work related to the transformation of data into a statistical set. Work is underway to prepare these sets for use by the CSO. It is planned to use this data for data imputation in Agricultural Census 2020 if needed.

2.2 Results of updating OdBR with data from administrative registers

Data from administrative records were used to update OdBR in accordance with methodological recommendations.

1. The use of the Agency for Restructuring and Modernization of Agriculture (**ARMA)** and the Central Equine Database of the Polish Union of Horse Breeders (Pol. **CBDK PZHK**) data sources was an uncomplicated operation due to the applied entity identifiers such as: ARMA manufacturer's number, PESEL registration number (ID No.) or REGON identification number (Business Registry No.). In the case of each of the sources, obtaining subjectivity was related to data aggregation. In the case of ARMA, aggregation from the level of information on herds of certain types of animals. In case of CBDK PZHK, aggregation from the level of individual entries of equidae from this register. In the case of CBDK PZHK, the adopted algorithm for use in the sampling frame stipulated the provision of entities contained in the sampling frame with features related to the number of equidae. Potentially, it is possible to make full use of this source for such a need and modify the algorithms for preparing the agricultural survey frame.

ARMA, PZHK - animal producers 2020, the scope of use of data in the agricultural survey frame 2020

Source register	ARMA - pigs	ARMA - cattle, sheep, goats	CBDK PZHK
Number of initial data records	91.951	361.331	313.167
Number of animal producers	90.487	358.668	85.269
Number of farms added in the agricultural survey frame	90.147	358.668	38.403

2. Currently, GIJHARS Organic Animal Production Register is used as an additional, subjective source of the sampling frame in terms of information on organic farms. The valuable use of by adding agricultural features, including animal-related ones,

is planned later on similar terms as ARMA. In the case of this source, subjectivity did not require aggregation of data, but only technical operations related to the selection of entities with a specific data quality for use in an agricultural survey frame. In this case, there was 100% subjective use.

GIJHARS - organic producers 2020, the scope of data use in the agricultural survey frame 2020

Number of initial data records	18.490
Number of organic producers	18.490
Number of farms added in the agricultural survey frame	18.490

3. Veterinary data were not used in their entirety due to unstructured form. Partial use was a kind of pilotage in the breeding of laying hens for eggs for consumption based on publicly available information obtained from the Chief Veterinary Officers in 2019.

GLW - producers of laying hens for table eggs in 2019, piloting the use of the data in the agricultural sampling frame 2019

Number of initial data records	563
Number of physical producers of laying hens for eggs for consumption	511
Number of farms added in the agricultural survey frame	308
Number of farms in the survey frame, producers of laying hens for eggs for consumption after the survey 2019	254

The data usage process required initial structuring (563 initial data records). Then, by means of several technical activities, an initial subjectivity was obtained (511 producers of physical laying hens for eggs for consumption). The agricultural survey frame was provided for 308 farms. During agricultural surveys in 2019, the information obtained in this way was verified. While in terms of value it was a successful experiment, the results already obtained showed that there is a problem with the correct determination of the subjectivity of GLW for laying hens for eggs for consumption according to the algorithms defining a farm in the sampling frame (decrease in the number of actual farms conducting such activity to 254). According to the methodological suggestion, further work is needed in this area to unify the process of combining data in the sampling frame with this data source.

<u>Conclusions after updating the OdBR.</u>

1. From the methodological, organizational and technical point of view, IRZGO (livestock identification and registration) registers of the Agency for Restructuring and Modernization of Agriculture (ARMA) and the Central Equine Database of the Polish Union of Horse Breeders (Pol. CBDK PZHK) are the most useful for CSO to update OdBR. Data records from the above-mentioned CSV-format registers and the correct use of TERYT register identifiers is suitable for use in statistics.
2. The collection of data from the organic animal production register kept by the Chief Inspectorate of Agricultural and Food Quality (Pol. GIJHARS) saved in Excel format

together with the use of TERYT register identifiers is possible to use for statistical purposes.

3. Due to the lack of use of TERYT identifiers, the lack of a uniform data structure in the District Veterinary Officers (pol: GLW) registers, the Veterinary Inspection registers require a lot of IT work related to the transformation of data into a statistical set. Work is underway to prepare these sets for use by the CSO.

3. Creating agricultural statistics on the basis on administrative livestock units (cattle, sheep, goats).

The aim of the action was to develop a methodology for creating statistics, performing data comparisons and assessing their quality.

1. Methodological part

- *Statistical survay of the livestock population:*

In Poland, two livestock surveys are conducted annually in June and December. These are the "Survey of cattle, sheep and poultry population and animal production (Pol: R-ZW-B)" and the "Survey of pig population and pig production (Pol: R-ZW-S)".

Data on the condition of the cattle and sheep population are compiled on the basis of the generalised results of the R-ZW-B survey, which combines a representative survey of the cattle, sheep and poultry population and animal production on individual farms and statistical reports on the cattle, sheep and poultry population and animal production on agricultural holdings of the legal entities, as well as in unincorporated entities (state, cooperative and companies with public and private sector assets).
The number of goats is estimated on the basis of information provided by regional experts. Own estimates are also used.

Until 2018 the study of the cattle, sheep and poultry population and animal production - R-ZWB - was carried out on a sample of 40 thousand individual farms keeping the above-mentioned animal species. Since 2019 the sample for the R-ZW-B test was limited to 30,000 farms of natural persons without loss of precision of the survey. The survey results concerning the above-mentioned livestock are presented according to the seat of the farm user, i.e. for individual farms - according to the place of residence of the user, and for state farms, cooperatives and companies - according to the place of residence of the enterprise (agricultural farm).

- *Central Database of the System for Identification and Registration of Farm Animals Labelled.*

The Agency for Restructuring and Modernization of Agriculture (ARMA) maintains the Central Database of the System for Identification and Registration of Farm Animals Labelled (Pol. CBD IRZGO), i.e. bovine animals, sheep, goats and pigs, called the System of Identification and Registration of Animals (Pol. IRZ). The aim of this system is to ensure the traceability of animals, thus ensuring food safety and gaining full access to the market for animal products in the EU.

The owner of a farm animal is reported to the head of the ARMA district office all arrivals and departures of animals to the herd's residence (import from EU countries, purchase,

sale, death, killing, slaughter, export to EU countries, export to third countries) within 7 days of the event (in case of cattle, sheep, goats); marking of born animals;
marking of animals imported from third countries .

CBD IRZGO collects data concerning:
- animals: identification number (individual or group no.), date of birth, the identification number of the farm, origin,the identification numbers of the farms where the animal has been kept together with the dates of these events reported by the keepers, date of slaughter or death, sex, breed, mother/father identification number.
- farms / herd's residence: the identification number of the herd's residence, name, surname and address of the owner, geographical coordinates of the herd's residence.
- status of an epizootic farm.

The administrator of the System for Identification and Registration of Farm Animals Labelled is the Agency for Restructuring and Modernization of Agriculture.

The computerized database (Central Database of the IRZGO System) is created as a network IT system, integrated with the IACS system (Pol: ZSZiK system), based on the territorial structures of ARMA, Regional Branches and County Offices.

- Deadlines for the transmission of administrative data.

The legal basis for the transfer of data, among others, on the stock of livestock from the IRZGO system by the ARMA to the CSO are the records in the Program of Statistical Research of Public Statistics (Pol: PBSSP) for a given year, and the Act of 31 July 2019 on the 2020 agricultural census,
On the basis of the PBSSP and the Act on Agricultural Census 2020, the President of the Agency for Restructuring and Modernisation of Agriculture provided, among other things, individual data from the system for identification and registration of labelled farm animals of the following species: cattle, sheep and goats, and according to the Act on Agricultural Census 2020 as of 1 June, 31 December 2018, 1 June, 1 December 2019 and 1 June 2020.

According to the entries in PBSSP for 2018, 2019 and 2020, ARMA provided CSO with individual data on the cattle, sheep and goat population by 31 July 2018 as of June 1, 2018, until January 15, 2019 as of December 31, 2018, until July 31, 2019 as of June 1, 2019, until December 30, 2019 as of 1 December 2019 and 30 July 2020 as of 1 June 2020.

According to the provisions of the Act of 31 July 2019 on the 2020 agricultural census, ARMA provided CSO with individual data on the cattle, sheep and goat population by 30 December 2019 as of 1 December 2019 and by 30 July 2020 as of 1 June 2020

The range of features from the IRZGO system was as follows:
- number of cattle in total, divided by breed: male calves under 1 year old, female calves under 1 year old, males, 1 to less than 2 years old, females, 1 to less than 2 years old males aged 2 years and older, females aged 2 years and older, cows,
- number of sheep in total and by breed: lambs under 1 year old, ewes,
- number of goats in total and by breed, including females aged 1 year and over.

- *Data availability*

On the basis of information recorded in the IRZ system, such as: date of birth of the animal, its sex, type of use, breed, occurrence of offspring, the following categories of animals can be identified:

- Cattle under 1 year old:Calves (females) ,Calves (male)
- Bovine animals, 1 to less than 2 years old: Cows (registered offspring),Heifers (no regis tered offspring), Male bovine animals
- Bovine animals aged 2 years and older: Cows (registered offspring), Heifers (noiste regred offspring), Male bovine animals.

Moreover, the category "cows" can also be divided into the characteristics of a breed: dairy type, beef type, mixed

Sheep under 1 year old: Males, Females.
Adult sheep aged 1 year and older: Males, Female sheep without offspring, Female sheep with offspring.
Goats under 1 year old: Males, Females.
Adult goats aged 1 year and older: Males, Female goats without offspring, Female goats with offspring.
On the basis of historical data collected in the IRZ register, it is possible to determine the number of female sheep and goats covered for the first time.

Information on the destination of animals, i.e. for milk production and not for milk production (animals remaining e.g. for meat production) can be taken from the base based on the animal's breed.

Based on the breed code one can also obtain information about the buffalo population (BF - buffalo*).*

- *Data available in the database to the categories required by Eurostat*

Categories of bovine, ovine and caprine statistical data, which can be obtained directly from the database and are required by Eurostat on the basis of Regulation (EC) No 2018/1091 of the European Parliament and of the Council of 18 July 2018 on integrated farm statistics.

Eurostat code	**Availability in the IRZ database**	**Category of animals**
CLVS 001	YES	Bovine animals aged under 1 year
CLVS 002	YES	Bovine animals, 1 to less than 2 years old
CLVS 003	YES	Bullocks, 1 to less than 2 years old
CLVS 004	YES	Heifers, 1 to less than 2 years old
CLVS 005	YES	Male bovines, aged 2 and older
CLVS 006	YES	Female bovines, aged 2 and older
CLVS 007	YES	Heifers, aged 2 and older
CLVS 008	YES	Cows
CLVS 009	YES	Dairy cows

Eurostat code	Availability in the IRZ database	Category of animals
CLVS 010	YES	Non-dairy cows
CLVS 011	YES	Buffalo-cows
CLVS 012	YES	Sheep (all ages)
CLVS 013	YES	Breeding female sheep
CLVS 014	YES	Other sheep
CLVS 015	YES	Goats (all ages)
CLVS 016	YES	Breeding female goats
CLVS 017	YES	Other goats

Recommendations:
The analyses of the contents of the IRZ registers show, that IRZ register fully satisfies the needs for information recorded in Poland's commitments to Eurostat in terms of cattle, sheep and goat populations

2. Results of work for production of agricultural statistics.

The aim this part of the action was to assess the quality of data obtained as a result of the adopted solutions for the production of agricultural statistics on the basis of administrative data on the livestock population, i.e. cattle, sheep and goats.

The comparison was made of data on **the cattle, sheep and goat** population from sources such as:
- administrative register, i.e. from the Central Database of the System for Identification and Registration of Farm Animals Labelled (Pol. CBD IRZGO),
- statistical survey on the cattle, sheep and poultry populations and animal production - R-ZW-B,
- Farm Structure Survey (FSS),
- field expert evaluations (Notebook).

The analysis covered the population of cattle and sheep as of 1 June and 1 December 2018, 1 June and 1 December 2019 and 1 June 2020. The analysis of the goat population concerned the state of the herd in June in the years 2015-2020.

Comparison of data on the cattle population

a. Comparison of the total population of cattle and dairy cows at national level.

After the analysis of the comparison of the total cattle population coming from two sources, i.e. from the administrative register (CBD IRZGO) and from the statistical survey (R-ZW-B), it can be stated that in case of all examined periods the total cattle population coming from the administrative register is higher than those obtained from survey by CSO.
Comparing data from the above-mentioned sources concerning the total cattle population for 2018-2020, the differences do not exceed 7%. Analysis of data obtained in June 2018 allowed to state that the total cattle population from the administrative register was higher

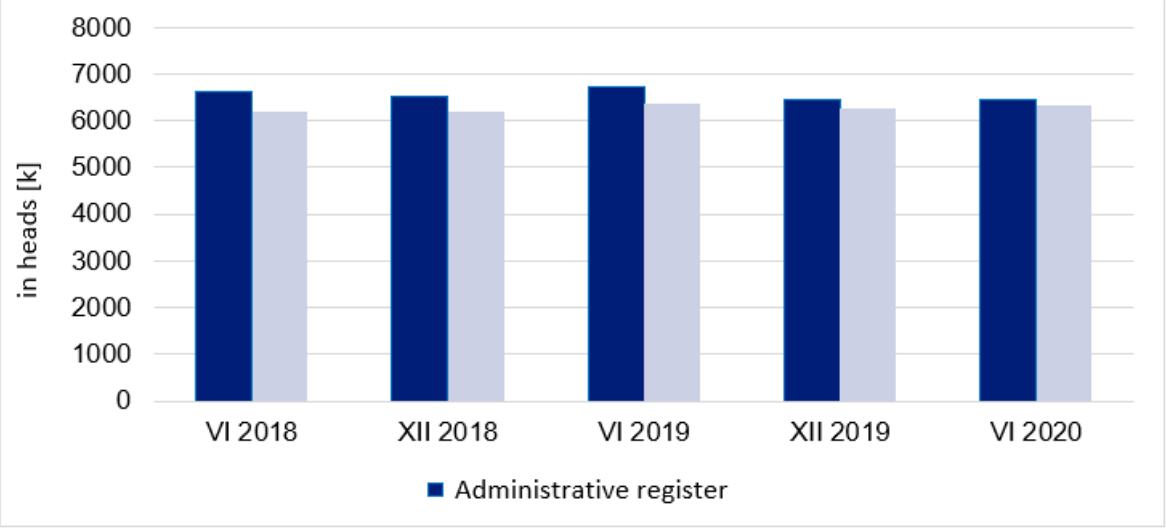
by 6.9% (425.000 animals) than in the survey by the Central Statistical Office (CSO), in June 2019, by 5.5% (353,000 of animals), while in June 2020 by 2% (125,000 of animals).
In December 2018, the total cattle population from the administrative register was higher by 5.7% (345,000 of animals) in relation to the results of R-ZW-B survey, and in 2019 by 3.1 (196,000 of animals).
The results obtained allow to conclude that the administrative data guarantee full coverage and allow to obtain reliable results.

Table 1 Total number of cattle populations.

Description	Administrative register	R-ZW-B survey	Difference (Administrative register - R-ZW-B survey)	Percentage difference (Administrative register=100)
VI 2018	6 626 366	6 201 441	424 925	106.9
XII 2018	6 533 194*	6 183 349	349 845	105.7
VI 2019	6 710 670	6 358 036	352 634	105.5
XII 2019	6 457 252	6 261 584	195 668	103.1
VI 2020	6 468 354	6 343 726	124 628	102.0

*data obtained from 31 December

Chart 1. Total cattle population, data obtained from the administrative register and R-ZW-B survey.



b. Comparison of the cow population at national level.

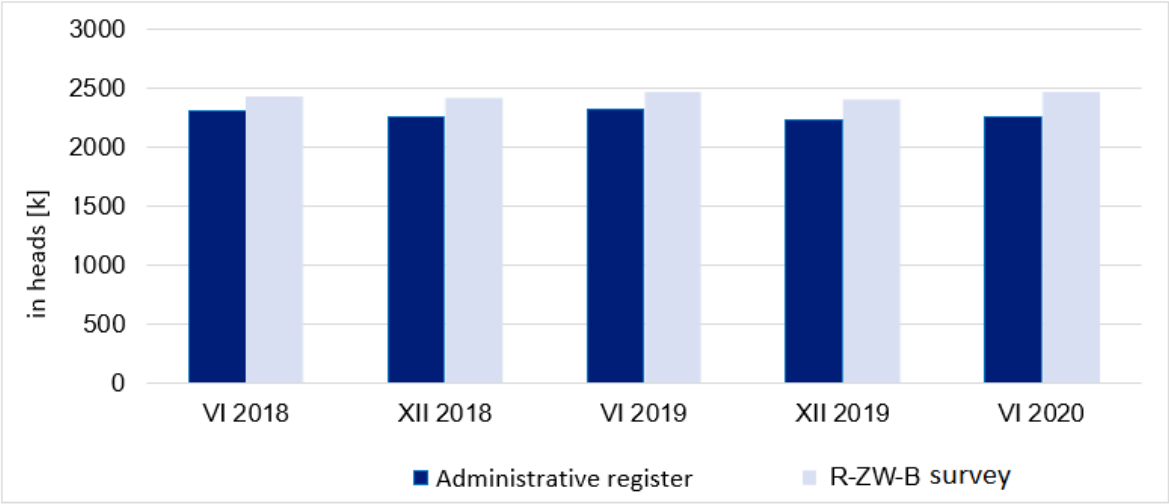
In all the examined periods the total number of cows from the administrative register is lower than that obtained from the R-ZW-B survey. The difference varies from 8.7% in June 2020 to 4.8% in June 2018, which is 214.3 thousand animals and 116.2 thousand ones, respectively.

Table 2 Total number of cows in heads.

Description	Administrative register	R-ZW-B survey	Difference (Administrative register - R-ZW-B survey)	Percentage difference (Administrative register=100)
VI 2018	2,312,978	2,429,195	-116,217	95.2
XII 2018	2,256,016*	2,417,407	-161,391	93.3
VI 2019	2,315,020	2,461,024	-146,004	94.1
XII 2019	2,226,358	2,406,293	-179,935	92.5
VI 2020	2,253,736	2,468,025	-214,289	91.3

*data obtained from 31 December

Chart 2. Total number of cows. data obtained from the administrative register and R-ZW-B survey.



c. Comparison of the percentage of cows in the total cattle population at national level

Considering the structure of the cattle population, a comparison of the percentage of total cows in the total population of this livestock species between the two data sources showed that the percentage of total cows obtained as a result of the R-ZW-B survey is higher in all examined periods than that obtained for the data from the administrative register. The

difference in the percentage of the total cow population varies between 4.0 percentage points in December 2019, and 4.6 percentage points in December 2018 (see Table 3).

Table 3. Structure of the cattle population (share of the total cow population in the total cattle population) according to data obtained from the administrative register and the R-ZW-B survey.

Description	Administrative register	R-ZW-B survey	Difference (R-ZW-B survey - administrative register)
VI 2018	34.9	39.2	4.3
XII 2018	34.5*	39.1	4.6
VI 2019	34.5	38.7	4.2
XII 2019	34.5	38.4	4.0
VI 2020	34.8	38.9	4.1

*data obtained from 31 December

Comparison of sheep population data

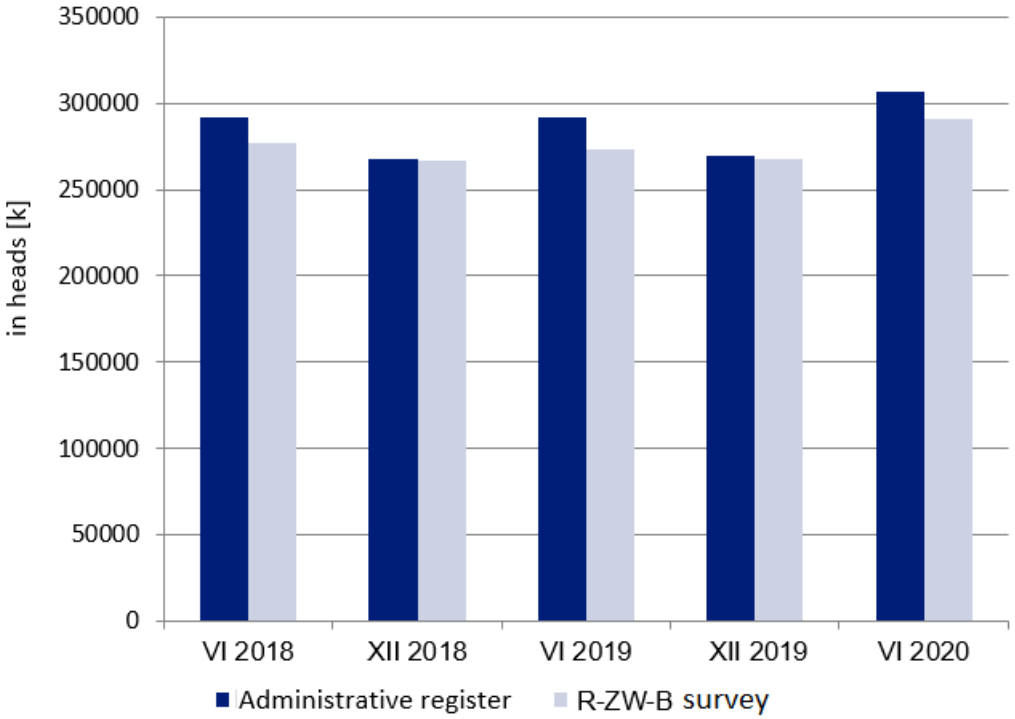
After the comparison of the total sheep population coming from two sources, i.e. from the administrative register (CBD IRZGO) and from the survey (R-ZW-B) it was found that for all the examined periods the total sheep population coming from the administrative register is larger than that obtained from the survey.
Comparing the data from the administrative register and the R-ZW-B survey obtained in December on the total sheep population, very similar results were obtained. The differences between the two data sources are small. In December 2018, the difference was 0.3% (which was 875 animals), and 0.7% (1.896 animals) in December 2019.
Analysis of data obtained in June 2018 allowed to conclude that the total sheep population from the administrative register was higher by 5.6% (the difference was 1.5558 animals) than that from the survey, and in June 2020 by 5.4% (which was 15.689 animals).
The results obtained allow to conclude that the administrative data guarantee full coverage and allow to obtain reliable results.

Table 4 Total number of sheep in heads.

Description	Administrative register	R-ZW-B survey	Difference (Administrative Register - R-ZW-B survey)	Percentage difference (Administrative register=100)
VI 2018	292,295	276,737	15,558	105.6
XII 2018	267,786*	266,911	875	100.3
VI 2019	292,368	273,097	19,271	107.1
XII 2019	269,625	267,729	1,896	100.7
VI 2020	306,431	290,742	15,689	105.4

*data obtained from 31 December

Chart 3. Total sheep population. data obtained from the administrative register and R-ZW-B survey.



Comparison of ewe populations

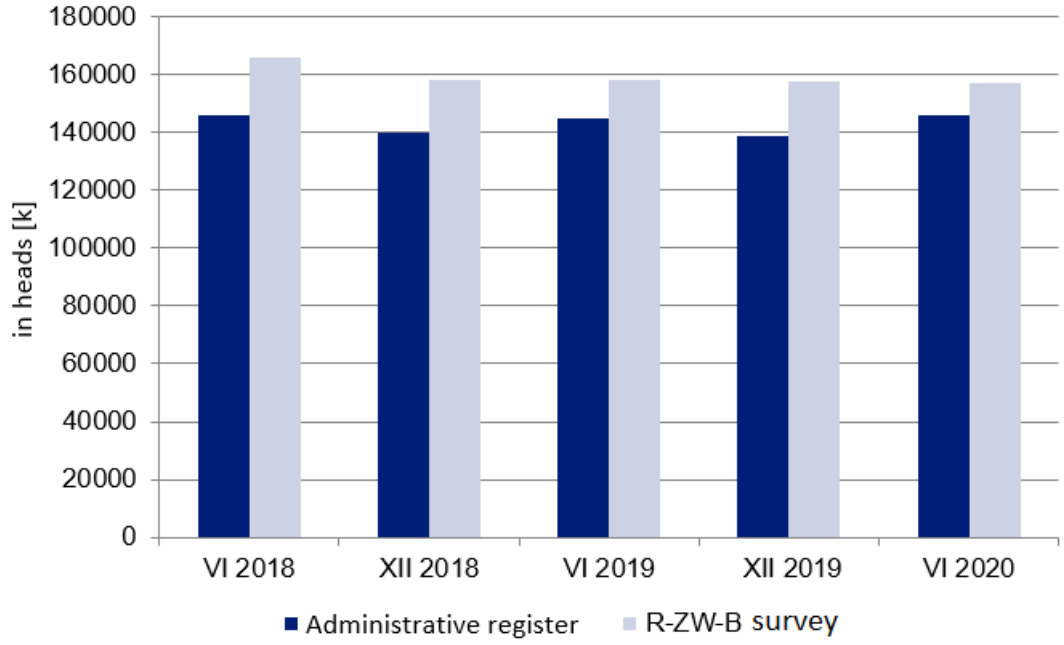
In all the periods studied the population of ewes from the administrative register is lower than that obtained from the survey. The difference varies from 12% in December 2019 to 8.4% in June 2019, which is 18.968 animals and 13.255 ones, respectively.

Table 5 Ewe population in heads.

Description	Administrative register	R-ZW-B survey	Difference (Administrative Register - R-ZW-B survey)	Percentage difference (Administrative register=100)
VI 2018	145,939	165,711	-19,772	88.1
XII 2018	140,008*	157,967	-17,959	88.6
VI 2019	144,973	158,228	-13,255	91.6
XII 2019	138,680	157,648	-18,968	88.0
VI 2020	145,737	156,843	-11,106	92.9

*data obtained from 31 December

Chart 4. Ewe population, data from the administrative register and R-ZW-B survey.



Comparison of the percentage of ewes in the total sheep population at national level.

Considering the structure of farm animals, a comparison of the percentage of ewe and the other sheep in the total population between the two data sources showed that the percentage of ewe resulting from the R-ZW-B survey is higher in all the examined periods than that obtained for data obtained from the administrative register. The difference varies between 10.0 percentage points in June 2018, and 6.9 percentage points in December 2018 (see Table 6).
The percentage of the other sheep in the total population of sheep from the survey is lower in all surveyed periods (see Table 7).

Table 6. Population structure (share of ewe population in the total sheep population) according to data from the administrative register and the R-ZW-B survey.

Description	Administrative register	R-ZW-B survey	Difference (R-ZW-B survey - administrative register)
VI 2018	49.90%	59.90%	10.0 percentage points
XII 2018	52.30%*	59.20%	6.9 percentage points
VI 2019	49.60%	57.90%	8.3 percentage points
XII 2019	51.40%	58.90%	7.5 percentage points
VI 2020	47.60%	54.0%	6.4 percentage points

*data obtained from 31 December

Table 7. The structure of the population (share of the the other sheep population in the total sheep population) according to data from the register and from the R-ZW-B survey.

Description	Administrative register	R-ZW-B survey	Difference (R-ZW-B survey - administrative register)
VI 2019	50.1%	40.1%	-10 percentage points
XII 2018	47.7%*	40.8%	-6.9 percentage points
VI 2019	50.4%	42.1%	-8.3 percentage points
XII 2019	48.6%	41.1%	-7.5 percentage points
VI 2020	52.4%	46%	- 6.4 percentage points

*data obtained from 31 December

Comparison of data on the total number of goats

A comparison of the total number of goats (as of June 2016) coming from the administrative register (CBD IRZGO) with the data obtained from the "Farm Structure Survey" (FSS) conducted by CSO in 2016, and data obtained from a notebook filled in by field experts in June 2015, 2017, 2018, 2019 and 2020.

On the basis of the comparison and analysis of the above-mentioned data it was found that in the case of all examined periods the total number of goats obtained from the administrative register is higher than that obtained from the FSS survey and notebook of field experts.

Comparing data on the total number of goats from the administrative register and from the FSS survey in the analyzed period, i.e. in June 2016, very similar results were obtained. The difference between these two data sources is small and is 0.5% more for data obtained from the administrative register, which was 204 animals.

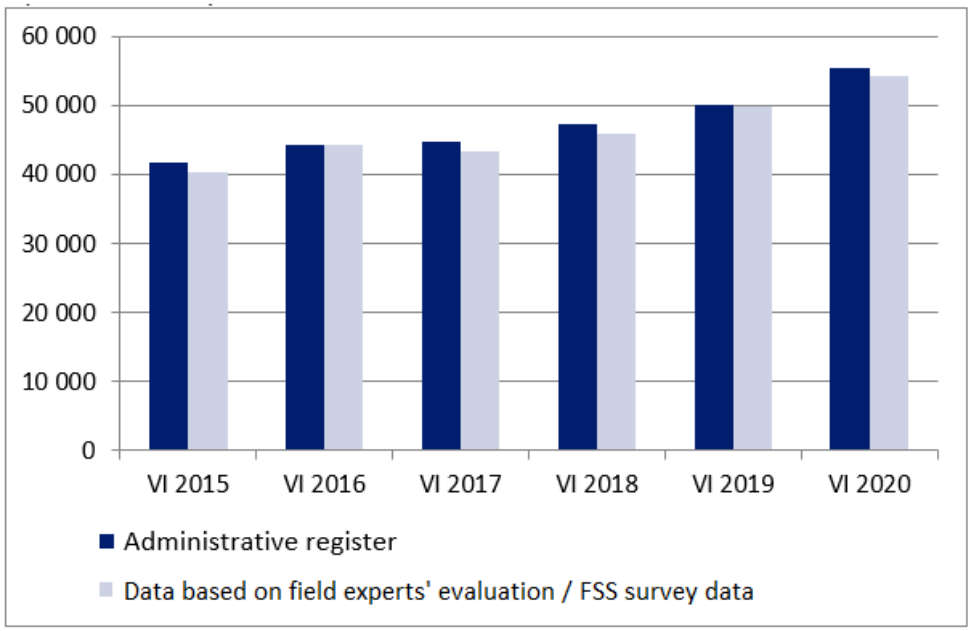
The difference between the level of the goat population obtained from the administrative register and the data from the notebook of field experts varies between 0.3% (162) in 2019 and 3.5% (1410) in 2015. This difference is decreasing systematically from 2015 onwards, as can be seen in Figure 5 and by comparing the data in Table 8.

Table 8 Total number of goats in pieces.

Description	Administrative register	Data based on field experts' assessment/Data from the FSS survey	Difference (Administrative Register - Evaluation of field experts / FSS survey data)	Percentage difference (Administrative register=100)
VI 2015	41 690	40 280	1 410	103.5
VI 2016	44 408	44 204*	204	100.5
VI 2017	44 817	43 427	1 390	103.2
VI 2018	47 258	45 926	1 332	102.9
VI 2019	50 062	49 900	162	100.3
VI 2020	55 434	54 294	1 140	102.1

*data obtained from the FSS survey

Figure 5 Goat population in total data from the administrative register and on the basis of the evaluation of field experts / FSS survey.



As in the case of sheep the results obtained also allow to conclude that the administrative data guarantee full coverage and allow to obtain reliable results.

Evaluation of the quality of the results obtained in the tests.

In accordance with Article 12 of Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics to guarantee the quality of results, European statistics data are developed, produced and disseminated on the basis of uniform standards and harmonized methods. The following quality criteria apply in this respect:

- " relevance" - Statistical information about the cattle, sheep and goat populations is used by Eurostat, FAO, state administration (Ministry of Agriculture and Rural Development, Institute of Agricultural and Food Economics), scientists, journalists, food industry, agricultural entrepreneurs, banks, pupils and students.
- "accuracy" - The data obtained from the register and concerning the number of farm animals, i.e. bovine, ovine and caprine animals in Poland, divided into groups, meet the standards of accuracy. It was found that the data from the register guarantee full coverage. All farm animals are labelled and must be reported to the register within the prescribed time limits under penalty of legal sanctions. All changes in the register are updated on an ongoing basis. The reference periods are compatible. The data from the register are constantly analyzed and the causes of discrepancies are removed on an ongoing basis.
-"timeliness" - The quality of administrative data is strongly dependent on the control system and timeliness of registration. The timeliness is an important feature for the quality of the register. For the data to be up to date, it is necessary that the information be entered in the register shortly after the events.

The delay in reporting livestock to the register was found to be insignificant, suggesting a high degree of accuracy and reliability.

- “punctuality”- The Central Database of the System for Identification and Registration of Farm Animals Labelled (CBD IRZGO) allows obtaining data at any time according to the status on a given day. This allows you to publish the data according to the schedule.

-“comparability”- After analyzing data from different sources, small differences in the level of cattle, sheep and goat population size were observed in all analyzed periods. Data from the administrative register and those from other surveyed sources are similar. On this basis, it can be concluded that if the method of data collection is changed the time series will be kept and comparable.

Both in the case of the administrative register and the Farm Structure Survey data can be obtained at the commune level, which allows them to be comparable in this respect.

- “coherence”- Coherence of two or more statistical results means the degree to which the statistical processes by which they were generated use the same concepts, classifications, definitions and target populations, as well as harmonized methods. Coherences between concepts, definitions, classifications, ranges, reference periods, geographical ranges used in the statistics obtained from the administrative register on livestock and those used in other analyzed data sources was analyzed. On the basis of the comparison, they were found to be consistent in this respect.

-“accessibility and clarity” - The statistics on the number of cattle and sheep in June and December are made available annually on the Internet as publications and in an online database.

Recommendations:

Based on the analysis carried out, CSO assessed the quality of data in the administrative register obtained from the CBD IRZGO register in terms of cattle, sheep and goat populations.

The quality of this data allows it to be used for the production of agricultural statistics.

4. Exchange of good practices in the field of conducting farm structure surveys, including agricultural censuses.

As part of the fourth objective, workshops for EU member states were held to exchange good practices in the field of conducting farm structure surveys, including agricultural censuses to be used in preparatory work and implementation of integrated farm surveys in 2020 (agricultural census), 2023 and 2026.

The workshops were conducted together with experts from statistical offices of Finland and Latvia.

During the workshop, experiences in the field of methodology, organization, modern methods of data collection, use of administrative data, construction and updating of the sampling frame for agricultural surveys were presented.

The workshop report (including the workshop programme, attendance list, copies of presentations, papers, results of the satisfaction survey and a document with conclusions) was submitted in accordance with the schedule for the implementation (point 2.(a) of the grant agreement) within 30 days from the end of the workshop.

5. Consolidation of the data collection process in IFS and representative agricultural surveys.

1. Consolidation of forms for individual data collection methods CAXI (CATI, CAII/CAWI, CAPI) and for agricultural surveys.

For several years, a system of agricultural surveys has been implemented in Polish statistics. This system combines an electronic form with individual CAXI data collection methods to enable effective preparation and execution of agricultural surveys. As part of the project work, modernization work was carried out to ensure coherence of data collection processes and to clarification of definitions used in IFS and other agricultural surveys.

For the purposes of the 2020 Agricultural Census, a form application was created using a new environment - Android. The electronic form for the Agricultural Census allows for automatic control and correction of data (at farm level) already at the stage of data collection. Experience gained in designing applications for the 2020 census will be used for other agricultural surveys.
The special data collection applications used in the form contain dictionaries, sets of validation rules with corresponding error messages and concept definitions that help to fill in the form correctly. The application of the electronic form, thanks to the use of validation and filtering questions, makes the interview conducted according to predefined paths.
Both "core" survey and “modular survey” can be found on one form.
Using the same electronic form, data are collected using such methods as CATI, CAPI and CAWI/CAII. All methods are used at the same time.

The use of modern methods of data collection is not possible without an efficient IT management system. This is particularly important when using multiple data collection methods in one survey. In Poland, such a system (CORstat) was established in 2015, based on the experience gained in Agricultural Census 2010 and subsequent years of research. The system allows automatic connection with the respondent, supporting filling in the form and sending the collected data to the databases. Completed forms are transferred to the "Substantive Module" application, which is part of the agricultural survey system. This application allows, at the stage of data collection, for experts to check, at the provincial level, whether the data from the respondent are correct. The expert control takes into account the links between the individual sections of the form, which are not or only partially covered by the automatic control. Based on previously prepared assumptions, the data is subject to logical and accounting control again. In the case of errors that cannot be corrected in the subsequent stages of data collection and control, the form is verified (contact with the respondent or interviewer). The " Substantive Module" checks, among other things, the limit values - maximum values (determined on the basis of previous years).

Both the CORstat system and the “Substamtive Module” have been adapted to the needs of the agricultural census 2020 as part of the modernization works.

As part of the work on the modernization of agricultural statistics, a proposal of "reporting bundle" was prepared. Based on June agricultural survey in 2019 – R-CzBR (June agricultural survey) – the unification was made for the entries in the boxes concerning land use and cultivation area in the forms:

- R-05 Report on crop area, yield and harvest;
- R-06 Report on cultivated area and harvest of fodder and other plants by use;

- R-08 Report on the results of horticultural production;
- Annex No. 1 to R-08 Report on size of horticultural crops under shelter.

A common crop code list was used to transfer the area sown, planted on 1 June 2019. to forms enriching these data with estimated yields, harvests.

Such a "reporting bundle" can also be used for future Integrated Farm Statistics (IFS).

Systematic mapping of fields facilitates the work of rapporteurs, enforces the correctness of transmitted data, eliminates errors that occur during data validation.

2. Introduction of harmonized code lists in agricultural statistical surveys

For the needs of agricultural surveys conducted by the Statistics Poland, 108 thematic code lists were created in the Central Statistical Metadata System. Some of them have not changed since the beginning of their creation, while others are filled up with new items every year or exclude items no longer existing.

One of the most modified code lists is *Crop List*. This list has closed items, i.e. we do not add new elements during the statistical research. Plant species currently found on farms are grouped into areas and additionally described by their use. During annual statistical surveys, new crops or ways of using plants appear on farms. In the June agricultural survey of 2019, 2413 crops (species + use) could potentially occur.

As part of the modernization work aimed at reducing the survey workload and the burden on respondents, the collection of information on plants in the most extensive groups, i.e. flowers and ornamental plants as well as herbs and spices. The currently modified *Crop List* contains 358 items for detailed description of crops in open field or under low cover and in greenhouses, tall plastic tunnels, cold frames. To eliminate the mistakes made so far due to incorrect crop selection, a change was made in the way other species of agricultural and horticultural crops were selected from the crop list. The correct choice of cultivation is made in a cascade, i.e. it starts with the name of the species (main part of the name + detail), the spring or winter form, the purpose of cultivation.
As part of the work carried out, a comparison was made between the records of declared crops in 2019 in Agency for Restructuring and Modernization of Agriculture (ARMA) applications with the CSO code list.

The data analysis carried out indicates that:

- the comparability of crops at the plant species level has definitely improved;
- In the list of ARMA plants there are groups of multispecies mixtures;
- In most cases, there is no information about the purpose of cultivated plants in the ARMA harvest;
- different records for the same crop appear in the ARMA harvest although applications are currently submitted electronically and should be used for data recording uniform crop lists.

When updating the CSO Crop Code Lists we try to use unambiguous records from the Agency for Restructuring and Modernization of Agriculture (ARMA) Crop List to make it easier for the respondents to submit data in statistical survay.

Additional analysis of the cultivated area decided to abandon the species-specific collection of information on plants in the most extensive groups (flowers, vegetables, herbs, and spices), in the future Integrated Farm Statistics (IFS). This will definitely contribute to reducing the workload of survay and the burden on respondents.

5.3 Symbolizing definitions in forms

An essential element of the preparation of a statistical survey is to develop a definition of the terms used in the survey. The definitions for the Integrated Farm Survey (IFS) have been developed on the basis of Eurostat guidelines, Handbook, explanations to FAQs (Frequently Asked Questions) on the website and, for national variables, the definitions have been adapted to users' needs.

The definitions explain more difficult concepts and issues, facilitate the correct completion of the form (questionnaire). In the case of an electronic survey, they provide quick access to substantive information, without the need to consult additional documents, and thus enable efficient conduct of the survey.

For the purposes of the General Agricultural Census 2020 definitions have been developed for all terms appearing in the survey. The definitions have been developed separately for individual sections (thematic areas) such as "Land use", "Sown area" , "Farm animals", etc.

The definition symbols are unique. The list of definitions is not closed and allows you to introduce new concepts if necessary. Thanks to this, the existing list of definitions can be extended not only to include terms that are yet to emerge in the survey, but at the same time they make it possible to ensure the coherence of definitions already used in agricultural surveys.

The definitions prepared for the purposes of Agricultural Census 2020 (same symbol, same content) will also be used in other agricultural surveys, e.g. on livestock. The most frequently used term is the definition of "agricultural area", which appears in all conducted agricultural surveys.

Results of work:

As part of the implementation of Objective 5 modernization work was carried out to ensure coherence of data collection processes and to clarification of definitions used in IFS and other agricultural surveys.
The work was carried out in parallel with the preparatory work for Agricultural Census 2020 and supported such activities as:

- preparation of the application form for the Agricultural Census 2020,modernizations of systems CORstat and the “Substamtive Module”,- preparation of "reporting bundle"; such a "reporting bundle" can be used for future Integrated Farm Statistics (IFS),
- code list for crop has been modified,
- definitions have been developed for all terms appearing in the survey; the definitions prepared for the purposes of Agricultural Census 2020 (same symbol, same content) will also be used in other agricultural surveys, e.g. on livestock

III. Expected benefits.

Work on the modernization of the sampling frame (OdBR) has significantly improved its quality. Geocoding of the sampling frame allowed to update the coordinates of address points for farms and farm's users. Moreover, based on vector plot boundaries from the Land Parcel Identification System (LPIS), farm ranges have been developed, while based on administrative boundaries and boundaries of the registration areas, boundaries of areas with natural and other specific constraints (LFA) have been developed.

The works related to the integration of the agricultural sampling frame with the sampling frame for social research made it possible to identify cases of duplicated farms in the OdBR. This has had a positive impact on the quality of the sampling frame prepared for Agricultural Census 2020.

As part of objective 2, the sampling frame has been updated with new data on farm animals. This improved the quality of the sampling frame considerably and broadened the range of features of a farm. This process will have a positive impact on the drawing of samples for future agricultural research.

A very important issue realized within the framework of the project is Objective 3, the implementation of which will allow the creation a methodology of replacing data from survey on farm animals (cattle, sheep, goats) with data from registers. The results of these works turned out to be very satisfactory. On the basis of the conducted analyses, the quality of data in the administrative register was assessed as good. Data from the register on the cattle, sheep and goat population may replace data from statistical surveys. This is important because of the rising costs of the survey and the burden on respondents.

As far as Objective 4 is concerned, the realization of which consisted in the organization of an international workshop aimed at the exchange of good practices in the field of conducting agricultural statistical surveys, including agricultural censuses, we believe that, according to the participants' assessment, the workshop fulfilled its task and broadened and equalized their knowledge in this field.

The realization of goal 5 influenced the unification of existing forms of data collection, unification of forms, symbolization of definitions in forms, introduction of harmonised code lists. All these activities should improve the quality of data from the Agricultural Census 2020 and in the future in other agricultural surveys.

IV. Sustainability of project results.

All works carried out in the project are the achievements of Polish statistics. The works related to the modernization of the sampling frame had a positive impact on the Agricultural Census implemented in 2020.

The scope of information and the quality of data in the sampling frame has definitely improved, which, in turn, contributed to the completeness of the census, and in the future will contribute to improving the quality of representative agricultural surveys.

The possibility of replacing data on farm animals from surveys with data from administrative sources is a permanent achievement of the project. It will allow to change the way of conducting research on the livestock population.

We estimate that each of the tasks in the project has contributed and will contribute to the development of agricultural statistics in the future.

V. Identified problems and opportunities/proposals for solving them.

Tasks implemented under the project were carried out in accordance with the adopted work schedule. The main supervision and monitoring of works related to the project implementation was carried out by the project manager, whose professional experience ensures that the work on the project is carried out in a substantive and timely manner.

Cooperating units performed some of the tasks on their own, reported to the project manager on an ongoing basis on the start of the tasks and provided information on the completion of work with products.

All works were monitored on an ongoing basis. Cyclical working meetings were held with the units implementing the project, in each phase of the project in order to discuss the results of the work to date and to adopt solutions to eliminate possible threats that may affect the implementation of the action.

The problems that emerged in the course of the project were mainly related to the quality of administrative data. This concerned mainly data from scattered registers kept by district vets. Unfortunately, the quality of this data did not allow it to be used to update the sampling frame concerning livestock information. However, we managed to use them to update OdBR. If necessary, they can be used for Agricultural Census 2020 data correction and data imputation.