



Co-funded by  
EUROPEAN COMMISSION  
EUROSTAT

## **Grant Agreement 2018.0211**

# **Modernisation of Agricultural Statistics Estonia**

## **Final Report**

Date:	2019.12.18
Version:	0.1
Authors:	Statistics Estonia

## Table of Contents

<b>1</b>	<b>SUMMARY .....</b>	<b>3</b>
<b>2</b>	<b>THE INITIAL PLAN OF THE PROJECT .....</b>	<b>5</b>
2.1	Aims and targeted results .....	5
2.2	Work steps during the project .....	5
2.3	Time-schedule .....	6
<b>3</b>	<b>DESCRIPTION OF WORK DONE.....</b>	<b>7</b>
3.1	Process 1. Improved sample design .....	7
3.2	Process 2. Adapting existing questionnaires .....	13
3.3	Process 3. Integration of existing administrative data sources .....	14
3.4	Process 4. Improving data processing and validation systems .....	17
3.5	Process 5. Improving the integration of data collection related to IFS .....	20
3.6	Process 6. Geocoding of farms .....	30
<b>4</b>	<b>CONCLUSIONS .....</b>	<b>33</b>
	List of main abbreviations .....	33
	List of tables .....	34
	List of figures .....	34

# 1 Summary

The project included six processes:

1. Process 1. Improved sample design.
2. Process 2. Adapting existing questionnaires.
3. Process 3. Integration of existing administrative data sources.
4. Process 4. Improving data processing and validation systems.
5. Process 5. Improving the integration of data collection related to IFS.
6. Process 6. Geocoding of farms.

In the process of “Improved sample design”, the population for Agricultural Census 2020 (AC) was analysed. The population in 2020 was estimated based on previous surveys as well as IACS land use data, the Register of Agricultural Animals and statistical data. Key variables of Farm Structure Survey 2016 were calculated, using the new IFS threshold: number of holdings, utilised agricultural area (UAA), livestock units (LSU), standard output (SO) and labour input. The results show that units above the IFS threshold cover more than 98% of UAA and LSU and therefore there is no need to extend the frame. The results show also that the number of holdings is expected to continue to decrease and it was taken as a basis for estimating the 2020 population. Based on this information, the samples for the IFS modules were designed. Also, the necessary sample sizes were estimated (taking into account that they would guarantee the required precision requirements). Possible standard errors for AC 2020 modules were estimated by calculating relative standard errors for data of FSS 2016 above the IFS threshold. The planned sizes of the population and modules were used already while compiling the IFS 2020 budget. Some changes in the sample design may occur when the frame for AC 2020 will be ready (in summer 2020).

In the process of “Adapting existing questionnaires”, it has been decided that AC 2020 data will be collected within the VVIS system (system used mainly for social statistics surveys). The same/similar questionnaire will be used during the computer-assisted as well as telephone-assisted interviews. In this process, the list of new IFS variables was prepared as an input for national metadata system iMeta. The description of variables in iMeta is a prerequisite for describing the questionnaire in the relevant software. The list of variables includes all variables – those which will be collected by the questionnaire, from administrative sources or calculated by Statistics Estonia. A draft questionnaire was worked out, which has to be further designed and then serves as input for programmers to compile the relevant electronic questionnaire. The questionnaire includes also information about the correct display/order of questions. The rules for prefilling electronic questionnaires were also worked out. The main registers which will be used for prefilling are: Agricultural Registers and Information Board land use and livestock data, Land Cadaster of the Land Board and Employment Register of Taxation and Customs Board. Some other registers will be used as well.

In the process of “Integration of existing administrative data sources”, an analysis of possible administrative sources was conducted. The main principles for processing data from administrative sources were also worked out. Consultations have been organised with the Ministry of Rural Affairs and organisations responsible for the most important registers: Agricultural Registers and Information Board (IACS land use data, Register of Agricultural Animals, rural development measures and other data), Land Board (land cadastre), Agricultural Board (organic farming register). Possibilities of using these and some other registers have been analysed and necessary improvements communicated to the registers, if necessary. It has been decided that data about supports (rural development measures, etc.) from IACS and organic farming data from Agricultural Board will be used directly and linked to the survey data later. IACS data about land use and

livestock, data about land tenure, labour force as well as some other data from different registers will be used for prefilling electronic questionnaires. Some data analysed can be used only for information texts on the questionnaires or imputation in case of possible non-respondents.

In the process of “Improving data processing and validation systems”, the validation, imputation, expanding rules and rules for forming final database were described. Three processes were described separately: data collection through the data collection system VVIS, data processing using the data processing system VAIS and data enlargement and formation of output data. The data collecting system VVIS contains collected data validation as well. In case of hard errors, it is not possible to move forward. In case of soft errors, the data have to be checked by the operator and data can be corrected on the spot with the respondent. The whole data processing will be done in the VAIS system, including merging the questionnaire based data with data collected directly from administrative sources, calculation of additional indicators, additional validation and correction, if necessary, imputation, creation of a source base and Eurostat individual data file. In addition, administrative data necessary for prefilling the questionnaire or for direct use are transformed into a suitable format using VAIS. The final source database and Eurostat’s individual data file will be created after completing the validation, data extension and trends verification. Data analysis and forming of output tables will be performed with software R.

The aim of the process of “Improving the integration of data collection related to IFS” was to improve the integration of Crop Production Survey to IFS as well as IACS. The main aim is to provide users with coherent data about areas in the years when both IFS and Crop Production Survey are conducted. Moreover, they both should be coherent with the Integrated Administrative and Control System (IACS). In this process, both the principles for the integration were worked out as well as the detailed description of the rules in special data processing software with the aim of automatic data processing. As the IFS threshold is not yet in use, three different data processing schemes were worked out: for using IACS data in non-FSS years before 2020, for using IFS data with the new IFS threshold and for using IACS data in non-IFS years after 2020. The first of them was fully implemented in the project and used for the calculation of 2018 Crop Production data (including formulas, implementation in programme R etc.). Other two calculation schemes have also been worked out, but they have to be adapted in the future to new IFS database as well as IACS land use data according to the new IFS threshold. The work has remarkably reduced the burden of statisticians, manual tasks have been reduced to minimum and efficiency has been increased through automatic data processing. Due to this development, it is possible to publish final data for Crop Production already in January n+1.

In the process of geocoding of farms, an analysis was conducted regarding the possibility and related rules of geocoding the locations of farms based on administrative sources without asking these questions during the survey. The sources used for geolocating the holdings are: IACS data of the locations given by the respondents themselves (client data), IACS land use data, data of agricultural animals and their owners, spatial data and address data. The order of applying locating rules was worked out. The first three rules take into account the location given by the client (if it is sufficiently detailed and correct), the fourth rule considers the location of animal housing and the fifth rule considers the location of the largest field (centre). Only for 21 units out of 17,350 it was not possible to find the location on the basis of abovementioned data. It was decided that for this very small part of holdings related to IACS and those few not related to IACS (the total number of which is expected to be smaller than 200), the location of the holding will be determined in the data processing process (if necessary, by asking the location directly from the respondent). Both the methodology and relevant queries were worked out and will be used for real data in AC 2020.

Overall, it can be concluded that the project works have been performed according to the plan. No major problems have occurred. It was of great help that the regulations and handbook were provided by Eurostat in good time, which enabled to follow the timetable. As a whole, the project was very useful for Estonia. A lot of preparations were made for IFS, which is a good basis for

further preparatory works of Agricultural Census 2020. Hopefully, the results described in the final report give some fresh ideas also to other Member States.

## 2 The initial plan of the project

### 2.1 Aims and targeted results

The general objective of the action was to prepare for the new framework regulation of Integrated Farm Statistics (IFS) adopted according to the new Strategy of Agricultural Statistics for 2020+.

Statistics Estonia has to conduct Agricultural Census (AC) 2020 in full accordance with the new IFS regulation. It means that the previous methodology, data collection and processing of Farm Structure Surveys (FSS) has to be adapted to new rules. On the one hand, the use of modules enables to reduce data to be collected and to reduce the burden for the respondents. On the other hand, it causes remarkable changes in sampling, data collection as well as processing. There will be new variables in IFS, which require analysis of already existing administrative data sources. Use of administrative sources had to be further analysed also in the case of “old” variables as administrative sources are developing all the time (like, for example, employment register in Estonia) and the pressure for their maximum use increases continuously. So, the early preparations for AC in the form of analysis and design were important as they guarantee the best results.

The Crop Production Survey had to be integrated with IFS in order to provide users with coherent data in the years when both surveys are conducted. Moreover, they both should be coherent with the Integrated Administrative and Control System (IACS). The areas of Crop Production Survey were adapted to the areas of FSS and/or IACS in Estonia also in previous years, but the process needed further improvement, manual tasks had to be reduced to minimum and efficiency increased through automatic processing.

Hence the aims of the action were both to prepare for the AC 2020 with the aim of maximum use of administrative data as well as to integrate the Crop Production Survey, IFS and IACS. As the similar processing will be used for both integrations (to IFS and IACS), the burden of statisticians was reduced already before IFS, as soon as the relevant data processing was improved – in the first year of the project already, i.e. 2018.

The preparations for AC 2020 and the way the Crop Production data are to be adapted to IACS and/or IFS are described in the following parts of the report and hopefully will be useful for other Member States as well.

### 2.2 Work steps during the project

The main work steps during the project were the following:

1. Process 1 (task 2 in the technical specification). Improved sample design.
  - Analysis of the population and its changes caused by the new threshold in the IFS regulation.
  - Design of the samples for the IFS modules.
  - Estimation of necessary sample sizes (which would guarantee also the required precision requirements).
2. Process 2 (task 1). Adapting existing questionnaires.
  - Description of the new variables for the metadata system (iMeta).
  - Estimation of necessary changes in the questionnaire.

- Analysis of relationships between the questionnaire and modules (display of the variables of the modules on the basis of samples).
  - Rules for pre-filling electronic questionnaires.
3. Process 3 (task 2). Integration of existing administrative data sources.
    - Analysis of the possible use of administrative sources.
    - Rules for processing data from administrative sources.
  4. Process 4 (task 1). Improving data processing and validation systems.
    - Description of validation rules.
    - Description of imputation rules.
    - Description of expanding rules.
    - Rules for forming the final database.
  5. Process 5 (task 3). Improving the integration of data collections related to IFS.
    - Integration of Crop Production Survey data to IFS and/or IACS – improvement rules and design.
    - Description of rules in special data processing software with the aim of automatic data processing.
  6. Process 6 (task 2). Geocoding of farms.
    - Analysis of the possibility and related rules to geocode the locations of farms on the basis of administrative sources without asking these questions during the survey.

The final grant report consists of the description of work done.

## 2.3 Time-schedule

The starting date of the action was 1 August 2018 and ending date 31 December 2019. The main milestones were the following:

Milestones/ Deliverables / Tasks / Results / Reports	Timetable
Improving the integration of data collections related to IFS	M + 5
Improved sample design	M + 17
Adapting existing questionnaires 2019	M + 17
Integration of existing administrative data sources 2019	M + 17
Improving data processing and validation systems	M + 17
Geocoding of farms	M + 17
Interim technical report	M + 9 <i>(expected at mid-action)</i>
Final technical and methodological report	M + 17

### 3 Description of work done

#### 3.1 Process 1. Improved sample design

The following activities were planned:

- a) analysis of the population and changes therein caused by the new threshold in the IFS regulation and design of the samples for the IFS modules,
- b) estimation of necessary sample sizes (which would guarantee also the required precision requirements).

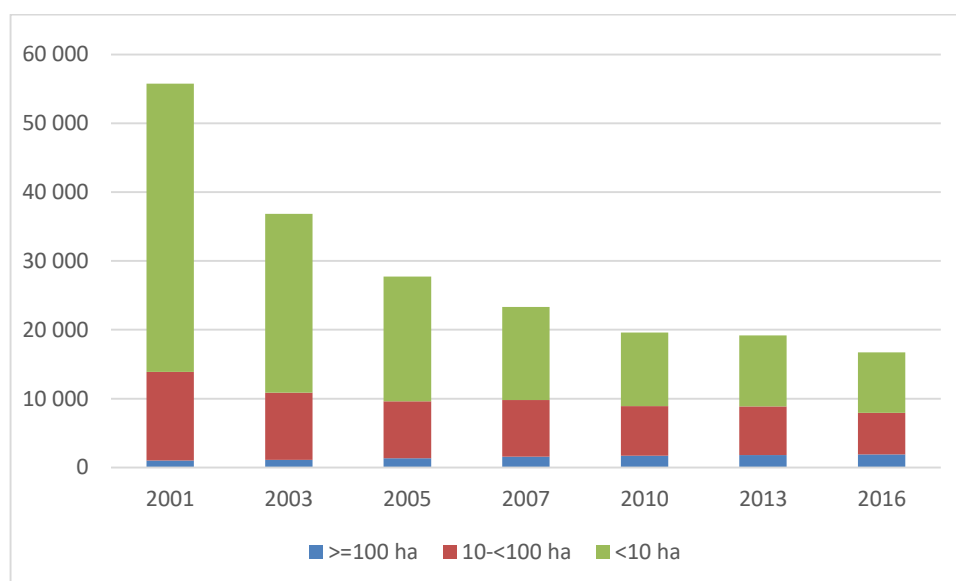
#### a) Analysis of the population and changes therein caused by the new threshold in the IFS regulation

According to the data of Agricultural Censuses and sample-based Farm Structure Surveys, the number of agricultural holdings decreased in 2001-2016 by more than 3.3 times, i.e. on an average by 2,600 holdings per year (see table 1.1 and figure 1.1).

Table 1.1. Number of agricultural holdings by size of agricultural area, 2001-2016

	<b>Total</b>	<10 ha	10-<100 ha	>=100 ha
2001	55 748	41 872	12 876	1 000
2003	36 859	26 002	9 769	1 090
2005	27 747	18 137	8 292	1 317
2007	23 336	13 540	8 247	1 549
2010	19 613	10 687	7 202	1 724
2013	19 186	10 315	7 077	1 794
2016	16 696	8 760	6 036	1 900

Figure 1.1. Number of agricultural holdings by size of agricultural area, 2001-2016



Holdings that disappear are mainly small, with less than 10 hectares of utilised agricultural area. At the same time, the number of larger holdings with at least 100 hectares of utilised agricultural area has almost doubled. As smaller holdings accounted for slightly more than half of the total number of holdings (52%) in 2016, it can be expected that the number of holdings will continue to decrease also during the next four years and this decrease can be remarkable.

The number of agricultural holdings will decrease in IFS also due to the use of new threshold. Possible changes in the number of holdings due to the change in the threshold have been analysed based on 2010 Agricultural Census, FSS 2013 and FSS 2016 data. The number of equidae has not been taken into account while determining the IFS threshold.

Table 1.2. Key variables by IFS threshold, 2010, 2013, 2016

	2010	2013	2016
Number of agricultural holdings			
Total	19 613	19 186	16 696
IFS holdings	15 155	14 082	12 676
Other holdings	4 458	5 104	4 020
Share of IFS holdings	77.3%	73.4%	75.9%
Utilised agricultural area, ha			
Total	940 930	957 506	995 103
IFS holdings	929 686	943 952	985 257
Other holdings	11 244	13 554	9 846
Share of IFS holdings	98.8%	98.6%	99.0%
Livestock, LSU			
Total	306 283	310 108	279 325
IFS holdings	305 213	309 009	278 624
Other holdings	1 070	1 099	701
Share of IFS holdings	99.7%	99.6%	99.7%



Standard output, euros			
Total	594 584 269	676 317 087	801 547 062
IFS holdings	589 999 267	671 166 717	796 691 585
Other holdings	4 585 003	5 150 370	4 855 477
Share of IFS holdings	99.2%	99.2%	99.4%
Labour input, AWU			
Total	25 116	22 063	19 876
IFS holdings	22 905	20 186	18 109
Other holdings	2 211	1 876	1 768
Share of IFS holdings	91.2%	91.5%	91.1%

Analysis shows that with the use of the IFS threshold the number of agricultural holdings will decrease by almost a quarter, utilised agricultural area will decrease by 1-1.4% and the number of livestock size units by 0.3-0.4%. Standard output will decrease by 0.6-0.8% and labour input by almost 9%. Hence, agricultural holdings above the IFS threshold will guarantee the required coverage - 98% of utilised agricultural area and livestock units.

In order to estimate the number of agricultural holdings for IFS 2020, the relevant populations were compiled for 2016 and 2018:

- Data about IACS land use and data from the Register of Agricultural Animals were linked (possible agricultural holdings). Units below the IFS threshold were excluded.
- FSS 2016 database was used to estimate the number of holdings not registered in IACS. First, IFS threshold was calculated for these data. The units which were not registered in IACS and which were above the threshold were linked to the above-mentioned IACS data (together with the areas of vegetables, strawberries, flowers and ornamental plants, vegetables in greenhouses, flowers in greenhouses).
- In the same way as units with horticultural crops, data of units with only poultry were processed (from FSS 2016 database) and linked with IACS data.
- Total population included all holdings above the IFS threshold.
- Theoretical population determined this way for 2016 was 12,633 holdings (difference from real FSS 2016 population above the threshold is only 43 holdings – 0.3%) and for 2018 it was 11,932 holdings.
- The 2020 population was estimated by taking into account that the number of holdings in 2016-2018 decreased by 701 holdings and therefore the estimated decrease for 2018-2020 may be the same. So, the estimated number of holdings above the IFS threshold for 2020 is 11,231, on an average 11,200.

### **Design of the samples for the IFS modules and estimation of necessary sample sizes**

The analysis of Agricultural Census, FSS 2013 and FSS 2016 data showed that units above the IFS threshold will guarantee the coverage of 98% of utilised agricultural area and 98% of livestock units. So there will be no need for using frame extension.

All variables of the rural development module are available in Agricultural Registers and Information Board, so these variables can be used for all units without a sample.

A sample is needed for the module of labour force and other gainful activities and the module of animal housing and manure management. For deciding the sample size and its design, the following analysis has been made.

- FSS 2016 sample was used as a basis. Units below the IFS threshold were excluded from all strata.
- Structure and extrapolation factors of the sample were not changed and theoretical populations were calculated for all strata (see table 1.3.). By using the IFS threshold the population would decrease by 24% and the sample by 14%. The sample would be 43% of the IFS population. This sample was analysed as a potential sample for the module of labour force and other gainful activities.
- As a next step, all units without livestock (LSU=0) were excluded from the theoretical sample of labour force and other gainful activities.
- Structure and extrapolation factors were not changed and theoretical populations were calculated again for all strata (see table 1.4.). When calculated this way the theoretical sample for the module of animal housing and manure management would be 41% of total population and 20% of the relevant population.
- As the theoretical samples for the module of labour force and other gainful activities and the module of animal housing and manure management included all variables covered by precision requirements of EU regulation 1091/2018, all requested relative standard errors were calculated. Relative standard errors were calculated for all variables, not taking into account their prevalence in the relevant population in 2016 (see table 1.5.).

Table 1.3. Labour force and other gainful activity module

STRATA	Population frame FSS2016	Sample of FSS2016	Extrapolation factor	Sample of IFS (LF and OGA module)	IFS population (calculated)
1	1199	1199	1	1172	1172
2	20	12	1.666666666666667	11	18.33333333
3	193	111	1.73873873873874	98	170.396396
4	8874	1677	5.29159212880143	982	5196.34347
5	132	77	1.71428571428571	77	132
6	42	25	1.68	25	42
7	405	238	1.70168067226891	238	405
8	120	70	1.71428571428571	68	116.571429
9	1440	695	2.07194244604317	676	1400.63309
10	71	40	1.775	38	67.45
11	1628	767	2.122555410691	738	1566.44589
12	24	14	1.71428571428571	11	18.8571429
13	783	367	2.13351498637602	290	618.719346
14	96	57	1.68421052631579	57	96
15	292	172	1.69767441860465	172	292
16	57	32	1.78125	32	57
17	738	432	1.708333333333333	432	738
18	473	275	1.72	274	471,28
19	109	63	1.73015873015873	56	96.8888889
Total	16696	6323		5447	12675.919

Table 1.4. Animal housing and manure management module

STRATA	IFS population (calculated) (LF and OGA module)	Sample of IFS (LF and OGA module)	Extrapolation factor	IFS subsample (animal housing and manure management module)	Population of animal housing and manure management module (calculated)
1	1172	1172	1	614	614
2	18.333333	11	1.666666666666667	3	5
3	170.3964	98	1.73873873873874	41	71.28828829
4	5196.3435	982	5.29159212880143	219	1158.858676
5	132	77	1.71428571428571	9	15.42857143
6	42	25	1.68		0
7	405	238	1.70168067226891	228	387.9831933
8	116.57143	68	1.71428571428571	43	73.71428571
9	1400.6331	676	2.07194244604317	69	142.9640288
10	67.45	38	1.775	3	5.325
11	1566.4459	738	2.122555410691	606	1286.268579
12	18.857143	11	1.71428571428571	9	15.42857143
13	618.71935	290	2.13351498637602	204	435.2370572
14	96	57	1.68421052631579	12	20.21052632
15	292	172	1.69767441860465	165	280.1162791
16	57	32	1.78125	28	49.875
17	738	432	1.70833333333333	31	52.95833333
18	471.28	274	1.72	257	442.04
19	96.888889	56	1.73015873015873	49	84.77777778
Total	12675.919	5447		2590	5141.474168

It was decided that like in FSS 2016, the one-stage stratified random sampling will be used for labour force and OGA module in 2020. Stratification variables will be standard output, type of farming and type of production (organic/conventional). Strata will be formed as showed in Table 1.5 below.

Table 1.5. Stratification plan for AC 2020 modules

Strata	Type of production	Type of farming	SO, euros	Size of sample, %
1	-	-	>=100 000	100
	-	2, 3, 5	>=25 000	100
	Organic	5	>=4000	100
	Organic	9, no data	-	100
2	Conventional	9		60
3	Organic	-	>=0-4000	60
4	Conventional	-	>=0-4000	20
5	Organic	1	>=4000-25 000	60

6	Organic	2, 3	>=4000-25 000	60
7	Organic	4	>=4000-25 000	60
8	Organic	6, 7, 8	>=4000-25 000	60
9	Conventional	1	>=4000-25 000	50
10	Conventional	2, 3	>=4000-25 000	60
11	Conventional	4	>=4000-25 000	50
12	Conventional	5	>=4000-25 000	60
13	Conventional	6, 7, 8	>=4000-25 000	50
14	Organic	1	>=25 000-100 000	60
15	Organic	4	>=25 000-100 000	60
16	Organic	6, 7, 8	>=25 000-100 000	60
17	Conventional	1	>=25 000-100 000	60
18	Conventional	4	>=25 000-100 000	60
19	Conventional	6, 7, 8	>=25 000-100 000	60

It may be necessary to adjust the sample design slightly after the compilation of the final frame for the Agricultural Census 2020.

The sample for the module of animal housing and manure management will be formed from the module of labour force and other gainful activities, from units which have livestock (Livestock units (LSU)>0.

The possible relative standard errors were calculated, which were the following:

Table 1.6. Relative standard errors

Variables	Relative standard error
Land variables	
— Cereals for the production of grain (including seed)	0.74
— Oilseeds	1.46
— Plants harvested green from arable land	1.62
— Fresh vegetables (including melons), strawberries, flowers and ornamental plants (excluding nurseries)	2.33
— Permanent grassland excluding rough grazing	1.07
— Fruits and berries (excluding grapes and strawberries)	1.85
Livestock variables	
— Dairy cows	0.28
— Non-dairy cows	1.91
— Other bovine animals (bovine animals less than 1 year, bovine animals 1 to less than 2 years, male bovine animals 2 years old and over, heifers 2 years old and over)	0.96
— Breeding sows live weight 50 kg and over	
— Piglets live weight of under 20 kg and other pigs	0.11
— Sheep and goats	3.65
— Poultry	0.27

Even though the requested limits of relative errors would allow to reduce both the sample of labour force and other gainful activities and animal housing and manure management, it is considered not to be reasonable and therefore the planned sample sizes are 40% and 20% of the population, respectively. The estimation of precision on the basis of land and livestock variables is indirect and does not give full information about the expected precision of variables of these modules. Agricultural production in Estonia is highly concentrated and a lot of variables exist only in very few

holdings. For example, in FSS 2016 there were only 14 holdings in the sample where regular employees had other gainful activity. Only 27 holdings of natural persons had regular male employees and 4 holdings had regular female employees. In Agricultural Census 2010, there were many very rare animal housing types, for example, 7 holdings that had battery cages with manure belts, 41 holdings that had battery cages with deep pit and 38 holdings that had other animal housing types for laying hens. 46 holdings held pigs on partially slatted floors and 17 holdings held pigs on completely slatted floors. As there are many scarce variables, 20% of the sample for the module of animal housing and manure management was considered reasonable and therefore sample sizes will be comparable with FSS 2016.

### 3.2 Process 2. Adapting existing questionnaires

According to the project plan, the following activities were planned:

- a) Description of the new variables for the metadata system iMETA;
- b) Estimation of the necessary changes in the questionnaire;
- c) Analysis of the relationships between the questionnaire and the modules (display of the variables of the modules on the basis of samples);
- d) Rules for pre-filling electronic questionnaires.

It has been decided that Agricultural Census 2020 data will be collected using VVIS (system used mainly for social statistics surveys) electronic questionnaires only. The same/similar questionnaire will be used during the computer-assisted as well as telephone-assisted interviews. The questionnaire will be bilingual, allowing interviews in Estonian or Russian.

Data to be collected by the questionnaire:

- a) Core structural data (excluding administrative data indicators CGNR002-CGNR005, CGNR011-CGNR015, CGNR026-CGNR029, organic farming data CLND086-CLND111, CLVS034-CLVS043);
- b) Module 1 "Labour force and other gainful activities" data;
- c) Module 3 "Animal housing and manure management" data.

Data for Module 2 "Rural Development" come from administrative data and will be merged with questionnaire data as well as "Core structural data" from the Organic farming register.

Data to be collected with Module 1 are asked only from holdings in the sample and Module 2 is mainly restricted to livestock available (LSU>0). Sample rules are provided in part 3.1. Improved sample design.

Freshly created new Agricultural Census variables and variables from the year 2016 are described according to the rules of iMETA. For the new variables, the calculation algorithm is presented, if necessary. Like the questionnaire, these data are input into the iMETA system, variable names can be adjusted.

The questionnaire prefilling rules were compiled on the basis and after the analysis of the relevant administrative registers. Agricultural Registers and Information Board's land use data for prefilling will be obtained as at 25.06.2020. All crops will be prefilled (questions B2\_1-B2\_36). Utilized agricultural area by ownership will be prefilled (B2\_38-B2\_40) and it is the sum of the crop areas (B2\_37). Since Agricultural Registers and Information Board's data may differ by definition, the areas under greenhouse vegetables and greenhouse strawberries will not be prefilled (C1 and C12), but are still shown in the questionnaire for information together with the information that crops under low cover should be excluded.

The Agricultural Registers and Information Board's livestock data are obtained for prefilling on 01.07.2020 and on 01.09.2020 at the census moment. If data provider will not change pre-filled data, they will be replaced with the Agricultural Registers and Information Board's livestock data at the census moment. The number of bovine animals (F3-F9), sheep (F12b, F13), goats (F14, F15) total number of pigs (F20) and beehives (F24) will be pre-filled.

The total area of agricultural land, total number of bovine animals, total number of sheep, total number of goats, total number of pigs and number of beehives are used also for routings in the questionnaire or in information texts.

In addition to data about land and animals to be pre-filled in the electronic questionnaires, the Agricultural Registers and Information Board also has some data about the manure storage facilities available and support paid for buying manure application equipment. Both of them will be used in the electronic questionnaires, but only as the information texts.

Wooded (E2) area and other land areas (E4) are pre-filled with data of the Land Cadaster of the Land Board.

The Employment Register of Taxation and Customs Board's data allow prefilling of holdings' male and female employees by their working time (J19 - J28) only for legal persons. In addition, it is possible to calculate on the basis of the employment register and show in the questionnaire the totals of male and female employees for legal persons and natural persons registered in the Business Register. Only very small part of family members are covered in the Employment Register and therefore, the number of family members in the Employment Register of Taxation and Customs Board will not be used in the questionnaire.

The holder's data (name, ID, contact data) is included already in the Census list and will be used also for prefilling the questionnaire. On the basis of ID, the holder's sex and year of birth will be pre-filled. These data will be showed as manager's data only if the holder is also the manager. The holder's data will also be pre-filled in the contact data.

It was analysed also whether the Estonian Education Information System could be used for prefilling agricultural training of the manager (if the manager is the holder), but it was decided that prefilling will not be used due to the complexity and very small amount of data available.

### 3.3 Process 3. Integration of existing administrative data sources

According to the project plan, two activities were planned within this process:

- a) Analysis of the possible use of administrative data sources;
- b) Rules for processing data from administrative sources.

Actions were planned as there will be new variables in IFS, which require analysis of already existing administrative data sources. Use of administrative sources should be further analysed also in the case of "old" variables, as administrative sources are developing all the time (like, for example, employment register in Estonia) and the pressure for their maximum use increases continuously.

Within FSS 2016, the following administrative sources were used (directly or for prefilling electronic questionnaires):

- a) Agricultural Support and Agricultural Parcels (together with annual land use data) - IACS
- b) Register of Agricultural Animals (including Bovine register)
- c) Organic farming register
- d) Rural Development Measures

- e) Land Cadastre of Land Board
- f) Statistical Register of Agricultural Holdings (Farm Register)

The same registers are relevant also for IFS 2020 (for direct use or for prefilling) but adaptations are needed and new sources were analysed as well.

In February 2019, Statistics Slovenia had a study visit to Statistics Estonia. A meeting of two days was held, which was full of interesting discussions, including those related with the use of administrative sources. The visit and relevant discussions gave us fresh ideas about possible use of administrative registers.

Consultations have been organised also with the Estonian Ministry of Rural Affairs and organisations responsible for most important registers: Agricultural Registers and Information Board, Land Board and Agricultural Board.

During the consultations with the Ministry of Rural Affairs, new IFS regulations and its variables were introduced. Confirmation has been received that the Ministry will support us in negotiations with registers, if necessary. The Ministry has also consulted us about the variables CGNR004 (areas facing natural constraints), CGNR014 (support from IACS), CGNR015 (young farmers), CGNR029 (environmental certification schemes) as well as rural development measures (MRDV001-MRDV015). The consultations gave us a clear understanding which data are needed for these variables from the Agricultural Registers and Information Board.

### **1. Agricultural Registers and Information Board (ARIB)**

During the consultations with the Agricultural Registers and Information Board it was agreed that the following data will be used: IACS land use data, Register of Agricultural Animals, rural development measures and some additional information.

#### **1.1. IACS land use data**

IACS land use data have been regularly used in statistical activities. Still, it has become clear that now they need some additional processing. There are some duplicates in the database as different aids have been paid for the same areas and all these data are included in the table. Therefore, spatial analysis has to be conducted before using them. Additional processing was necessary and was carried out for 2019 Crop Production Survey already. It gave us a good opportunity to work out and test the processing already before the census year. All data about crop areas and ownership of utilised agricultural area will be prefilled in the electronic questionnaires.

Additional analysis was planned also for register data about greenhouse crops. It was decided during the project that these data will not be used for prefilling (as the definition may be not exactly the same) but still they will be used in the information text of the questionnaire as it may be helpful for the respondents as well as for telephone interviewers.

#### **1.2. Register of Agricultural Animals**

The Register of Agricultural Animals is considered to be usable without major problems. It is of great help that the loading of these data into our databases has significantly improved recently. Data about numbers of livestock (cattle, sheep, goats, pigs) will be prefilled in the electronic questionnaires. Both land use data and data about livestock will be used for compiling the Census lists.

#### **1.3. Rural development measures and other data**

It has been agreed with the help of the Ministry of Rural Affairs how rural development measures and other similar data can be used for IFS 2020. Data about rural development measures will be

not asked in the questionnaires, but are derived directly from administrative files and linked to the questionnaire data later.

Analysis of other possible data showed that also some data about manure storage facilities and manure application equipment are available. About manure storage facilities, there is only information available whether the holding has it. There is no specific information about the type and capacity. For manure application equipment, the information is also available whether the holding has received support for buying the equipment. There is information available whether the equipment is for solid manure or liquid manure/slurry. There is also information available whether it is for broadcast, band spread, open-slit or closed-slit injection. In case of liquid manure/slurry techniques, there is also information whether it enables to mix the manure with soil immediately. According to the analysis, there is not enough information available to use these data for prefilling the questionnaires and therefore it was decided to use it in the questionnaire only for the information text.

It has been also clarified that the date of registration of an IACS client can be used only in the case of possible non-respondent natural persons, for the imputing year when the person started to act as the manager. The questions have been sent to ARIB to clarify possible data availability about vocational training. For data about housing of laying hens it was decided that due to the different definitions, these data will be not used.

## **2. Land Board**

During the consultations with the Land Board, it was agreed that their data (wooded area, other land) will be received and used for prefilling the questionnaire in the same way as in FSS 2016.

## **3. Agricultural Board**

During the consultations with the Agricultural Board, the use of the Organic Farming Register has been discussed. There have been some problems with the use of their data also before and so the use of the data for both "old" and new variables was discussed. Statistical data needs were explained thoroughly and improvements in the register were asked. The feedback was positive and it seemed to be good time for consultations as the Agricultural Board had IT related improvements on the way. Finally, it was agreed that statistical data needs will be taken into account as much as possible. The contract for using their data is under preparation. In the contract, the exact data structure will be stipulated.

Data about organic farming will not be asked in the questionnaires, they will be derived directly from administrative files and linked to the questionnaire data later. Preliminary organic data will be used also for compiling the census list. Data about certified seeds will be used only for compiling the Census list.

## **4. Taxation and Customs Board**

From the Taxation and Customs Board, the Employment Register will be used. The Employment Register of Taxation and Customs Board's data allow prefilling holdings' male and female employees by their working time (J19 - J28) only for legal persons. In addition, it is possible to calculate on the basis of the employment register and show in the questionnaire the totals of male and female employees for legal persons and natural persons registered in the Business Register. Only a very small share of family members are covered in the Employment Register and therefore the number of family members in the Employment Register of Taxation and Customs Board will not be used in the questionnaire.



## 5. Other registers and databases of Statistics Estonia

From the databases of Statistics Estonia, the following ones will be used: Statistical Farm Register and Statistical Business Register. From the Farm Register, legal type and personal codes for natural persons will be obtained. From personal codes of natural persons, the data of holders' age and sex will be derived and used in the questionnaires.

From the Statistical Business Register, the data on enterprise groups will be received. This information is collected based on the business registers' regulation — Regulation (EC) No 177/2008 of the European Parliament and of the Council of 20 February 2008 establishing a common framework for business registers for statistical purposes.

Additionally, the date of registration of natural persons can be received from the Business Register and this information will be used in the case of possible non-respondents for imputation of variable "When the person was qualified as a manager".

The Statistical Business Register and the Business Register have been studied also in order to clarify whether the data on managers could be available for legal persons. It was understood that the available information does not give information about which person is the manager of the agricultural holding and therefore, this information will not be used.

The holder's data (name, ID, contact data) is included already in the Census list (Farm Register) and will be used also for prefilling the questionnaire. On the basis of ID, the holder's sex and year of birth will be prefilled. These data will be showed as manager's data only if the holder is the manager. Data of the holder will be prefilled also in contact data.

It was analysed also whether the Estonian Education Information System could be used for prefilling agricultural training of the manager (if the manager is the holder), but it was decided that prefilling will not be used due to the complexity and very small amount of data available.

### 3.4 Process 4. Improving data processing and validation systems

The following activities were planned: description of validation rules, description of imputation rules, description of expanding rules and rules for forming the final database.

Data Processing of Agricultural Census 2020 is divided into three parts:

- a) Data collection using VVIS questionnaires;
- b) Data processing using VAIS (merging questionnaire based data with data collected directly from administrative sources, calculation of additional indicators, data revision, improvement, imputation, creation of source base, creation of Eurostat individual data file;
- c) Data enlargement and formation of output data.

In addition, administrative data necessary for prefilling the questionnaire or direct use are transformed to suitable format by using VAIS.

## **Data collecting with questionnaires**

According to the initial plan, data is collected in two stages:

- Computer-assisted web interviewing (CAWI) 01.09.2020-20.09.2020;
- Computer-assisted telephone interviewing (CATI) 21.09.2020-15.11.2020.

Both interviews take place using the same/similar questionnaire and data collecting system VVIS. Telephone-assisted interviews are for those holdings which will reject or discontinue web interviews. The content of the questionnaire and prefilling with administrative data is explained in chapter 3.2 Process 2. Adapting existing questionnaires. The questionnaire is prefilled with IACS land-use and livestock data, as well as with data from the Land Cadaster of the Land Board and Employment Register of the Taxation and Customs Board.

The data collecting system VVIS contains collected data verification as well. Verification is applied after input of the last indicator, direct improvement is available. In case of hard errors, it is not possible to move forward. In case of soft errors, the data will be checked by the operator and the question can be clarified with the reporting person.

## **Data processing using VAIS**

In this stage, the interview based data is merged with data collected directly from administrative sources. Additional indicators are calculated, data revised, improved and imputed if necessary.

Determining of the agricultural holdings' geographical location and corresponding settlement, municipality, county and region is explained in chapter 3.6 Process 6. Geocoding of farms.

Data of organic farming are obtained in the form of individual data from the organic farming register of Agricultural Board (authorised data processor). The rules of organic farming processing methods and software development are in progress and there is no final agreement about the format of receivable data. Most probably, the data will be delivered to Statistics Estonia by fields, similar to the format of IACS data. So, it would be necessary to create an additional organic farming data processing system to reveal detailed data about all fields of agricultural holdings and individual data by crops.

Data on rural development will be received from the Agricultural Registers and Information Board (for 3-year periods, i.e. 2018-2020). Additional processing is not needed and the data will be merged with the rest of the data in the data processing system VAIS. In the rural development module, no samples will be used, as data are directly available from the administrative source.

Other administrative data are described in chapter 3.3 Process 3. Integration of existing administrative data sources. The following data are obtained from other sources:

CGNR 003 - NUTS 3 region;

CGNR 004 - The agricultural holding has areas designated as facing natural constraints under Regulation (EU) No 1305/2013;

CGNR 005 - Natural person who is the sole holder, where the agricultural holding is independent;

CGNR 011 - Legal person;

CGNR 012 - If yes, is the agricultural holding part of an enterprise group?

CGNR 014 - The holder is a beneficiary of EU support for land or animals of the agricultural holding and thus included in the integrated administration and control system (IACS);

CGNR 015 - The holder is a young farmer or new entrant into farming who has received financial support for this purpose under the common agricultural policy (CAP) in the previous 3 years;

CGNR 016 – Year of birth;

CGNR 017 – Sex;

CGNR 029 – Participation in other environmental certification schemes;

MLFO 001 – Sex of the holder;

MLFO 002 – Year of birth.

In addition to questionnaire data and direct administrative source data, it is necessary to calculate variables that are not directly collected, but are needed according to the (EU) 2018/1874. For example, total areas are collected by questionnaires as well as received from IACS. As the questionnaire includes also data about greenhouses, outdoor areas can be calculated. Standard output and type of farming will be calculated by special software of Statistics Estonia.

The verification of all data will be described in the data processing system VAIS. Since the receivable format of organic farming data is not clear, the validation rules related to administrative data are not fully ready yet. In the VAIS system, it is possible to see only incorrect records and to correct them. All soft errors will also be reviewed and, if necessary, respondents will be contacted.

In the case of missing data in whole-coverage stratum (core structural data, module 1 "Labour force and other gainful activities" data or module 3 "Animal housing and manure management" data), imputation will be used.

Since core structural data is prefilled with IACS data and contains register-based organic farming data, there is only very few variables that require imputation. The following methods are used:

- Administrative sources that provide only indirect information and are therefore presented only in information texts of the electronic questionnaires and not used directly for prefilling. The Statistical Business Register may have information about many managers in case of legal persons (still, in case of possible non-responses, data of older persons can be used).
- Holding's own data from previous periods.
- The similar holding's data (of the same production type and economic size class).

In case of modules, holding's own data from earlier periods or similar holding's data imputation can be used. Since the number of holdings the data of which have to be imputed has been very small on the basis of previous farm structure surveys, the process is mostly manual.

Based on the validated and corrected data, the source base will be created. In the metadata descriptive system the indicators for source database are marked with corresponding id. Numbers of stratum and extension factors will be located in a separate file. Final source database and Eurostat's individual data file will be created after completing the validation, data extension and trends verification.

## Data extension and output data generation

One-stage stratified random sampling will be used. The rules and design of the sample are described in chapter 3.1 Process 1 Improved sample design. Data for the output come from the source database and will be multiplied with extension factors. Both the numbers of strata of core structural data and modules will be included in the source database. Data analysis will be performed and output tables formed with software R.

### 3.5 Process 5. Improving the integration of data collection related to IFS

The following activities were planned:

- a) Integration of Crop Production Survey data.
- b) Description of rules in special data processing software with the aim of automatic data processing.

The Crop Production Survey has to be integrated with IFS in order to provide users with coherent data in the years when both surveys are conducted. Moreover, they both should be also coherent with the Integrated Administrative and Control System (IACS). The areas of Crop Production Survey have been adapted to the areas of Farm Structure Surveys (FSS) and/or IACS in Estonia also in previous years, but within the grant project the processes were further improved, manual tasks were reduced to minimum and efficiency has been gained through automatic processing.

In 2015, register variables of FSS were described in metadata system iMeta. Variables used for both FSS and Crop Production were harmonised. The register variables are described in iMeta also for land use data that are received from IACS. These data are aggregated to the level that corresponds directly to variables used in FSS and Crop Production. All these variables are now used in the grant project to work out automatic processing and generation of automatic output tables.

The structure of the output tables is in table below. Data of agricultural holdings are nationally published by counties. Data of agricultural small units are published only as totals, they are used mainly for national needs; but data of small units has been sent also to Eurostat with Crop Production land use data. From 2020, the threshold for agricultural holdings will change and the holdings below the threshold will be included also in data of small units.

Table 5.1. Structure of the output table

Level	Code of level
Whole country	_T_EE
No distributions by county (agricultural small units and kitchen gardens of agricultural holdings)	_O_EE
Counties total (agricultural holdings)	AGRIH_EE
County 1	AGRIH_maakond1
County 2	AGRIH_maakond2
...	
County 15	AGRIH_maakond15

For automatic data processing, output variables were described in metadata system iMeta and they are presented in the following table.

Table 5.2. Output variables

Code of variable	Unit	Name
AR_C	ha	Cultivated area
AR_HA_PR	ha	Harvested area / Production area
PR	tons	Production
YI	kg/ha	Yield
SUHE	-	The share of Harvested area / Production area in the Cultivated area

In the project, three different calculation schemes were worked out for Crop Production data:

- For using IACS data in non-FSS years before 2020 (i.e. for 2018, 2019 - with old threshold for agricultural holdings),
- For using IFS data with the new IFS threshold (for 2020, 2023, 2026),
- For using IACS data in non-IFS years after 2020 (i.e. for 2021, 2022, ... - with the new IFS threshold).

The first of them was fully implemented within the project and used for the calculation of 2018 Crop Production data (including formulas, implementation in program R etc.). Other two calculation schemes have also been worked out, but they have to be adapted to new IFS database (which does not yet exist) as well as IACS land use data according to the new IFS threshold.

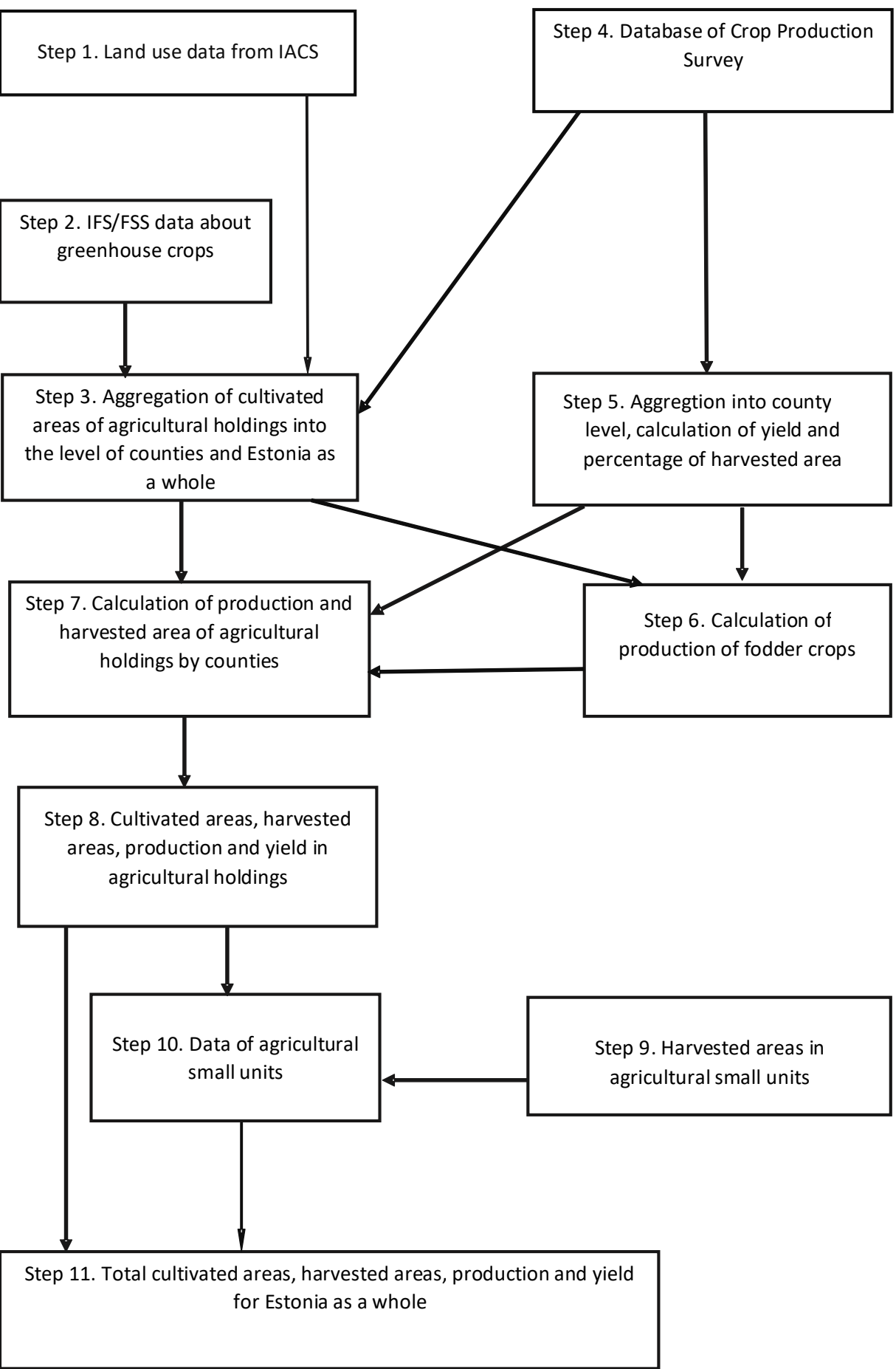
#### Data processing in non-FSS years before 2020 (in 2018, 2019)

Integration between IACS and Crop Production means that cultivated areas (AR\_C) are used from the IACS database and they are supplemented by some details about greenhouse crops from the previous FSS and horticultural crops from the annual Crop Production Survey (concerns units which are not registered in IACS). The Crop Production Survey still covers all holdings, including those registered in IACS. The aim of the integration is to provide mutually compatible data.

Based on data from the Crop Production Survey, the percentages of harvested areas (SUHE) and yields (YI) are calculated. Final harvested areas are calculated by using the formula  $AR\_HA\_PR = AR\_C * SUHE$  and production  $PR = AR\_C * YI$ . For permanent crops the production is calculated by using formula  $PR = AR\_HA\_PR * YI$ .

Figure 5.1 shows schematically the idea of data processing.

Figure 5.1. Data processing in non-FSS years before the new IFS threshold (in 2018, 2019)



Step 1. Data for calculating AR\_C are received from the IACS database. The database includes data which already have been transformed into the level of statistical register variables and where land parcels have been merged into holdings.

Step 2. As the IACS database does not include full data about greenhouse crops, they are taken from FSS.

Step 3. Data processing of land areas.

- Land areas from IACS are grouped and aggregated by counties.
- For units which are not registered in IACS, the areas of open field vegetables and permanent crops from the Crop Production Survey database are added into IACS data.
- Data of greenhouse crops from FSS are detailed based on Crop Production Survey data.
- Land areas of all crops from IACS and Crop Production Survey are aggregated into groups (cereals, dry pulses, ..., arable land, permanent grassland etc.) by counties.

Step 4. Database of the Crop Production Survey.

Step 5. Calculation of yields and percentages of harvested crops based on the Crop Production Survey (on the level of counties)

- For all crops (except for permanent crops, multiannual fodder crops, permanent grassland, flowers and ornamental plants), yield is calculated as  $YI = PR / AR\_C$ .
- For permanent crops, yield is calculated based on harvested area  $YI = PR / AR\_HA\_PR$ .
- For all permanent crops (also for cereals, dry pulses, potatoes, rape and turnip rape), the percentage of harvested area from cultivated area is calculated as  $SUHE = AR\_HA\_PR / AR\_C$ .

Step 6. Calculation of the production of fodder crops. These data are used for Economic Accounts for Agriculture and for national needs.

Step 7. Calculation of harvested areas and production in agricultural holdings (by counties).

- Harvested areas of permanent crops (and cereals, dry pulses, potatoes, rape and turnip rape) are calculated on the basis of IACS data and percentage from Crop Production Survey as  $AR\_HA\_PR = AR\_C * SUHE$ .
- Production of crops (except permanent crops) is calculated based on IACS data and yield from Crop Production Survey as  $PR = AR\_C * YI$ .
- Production of permanent crops is calculated based on harvested area and yield from Crop Production Survey as  $PR = AR\_HA\_PR * YI$ .
- Harvested areas (AR\_HA\_PR) and productions (PR) are calculated for groups.
- Yield for groups (except permanent crops) is calculated as  $YI = PR / AR\_C$ .
- Yield for permanent crops is calculated as  $YI = PR / AR\_HA\_PR$ .

Step 8. Cultivated areas, harvested areas, production and yield in agricultural holdings for Estonia as a whole.

- Cultivated areas (AR\_C), harvested areas (AR\_HA\_PR) and production (PR) of groups for all counties are summed up.
- Yield is calculated for all crops and groups of crops as  $YI = PR / AR\_C$ , for permanent crops as  $YI = PR / AR\_HA\_PR$ .

Step 9. Land areas of agricultural small units and kitchen gardens (data of agricultural small units are used in land use data of Crop Statistics and for national needs).

Step 10. Calculation of harvested areas and production in agricultural small units and kitchen gardens of agricultural holdings (for national needs).

Step 11. Cultivated areas, harvested areas, production and yield (total for agricultural holdings and small units).

Data of agricultural holdings from step 8 and total areas of small units from step 10 are used for delivering Crop Production Survey data to Eurostat.

### **Data processing in IFS years with new IFS threshold (2020, 2023, 2026)**

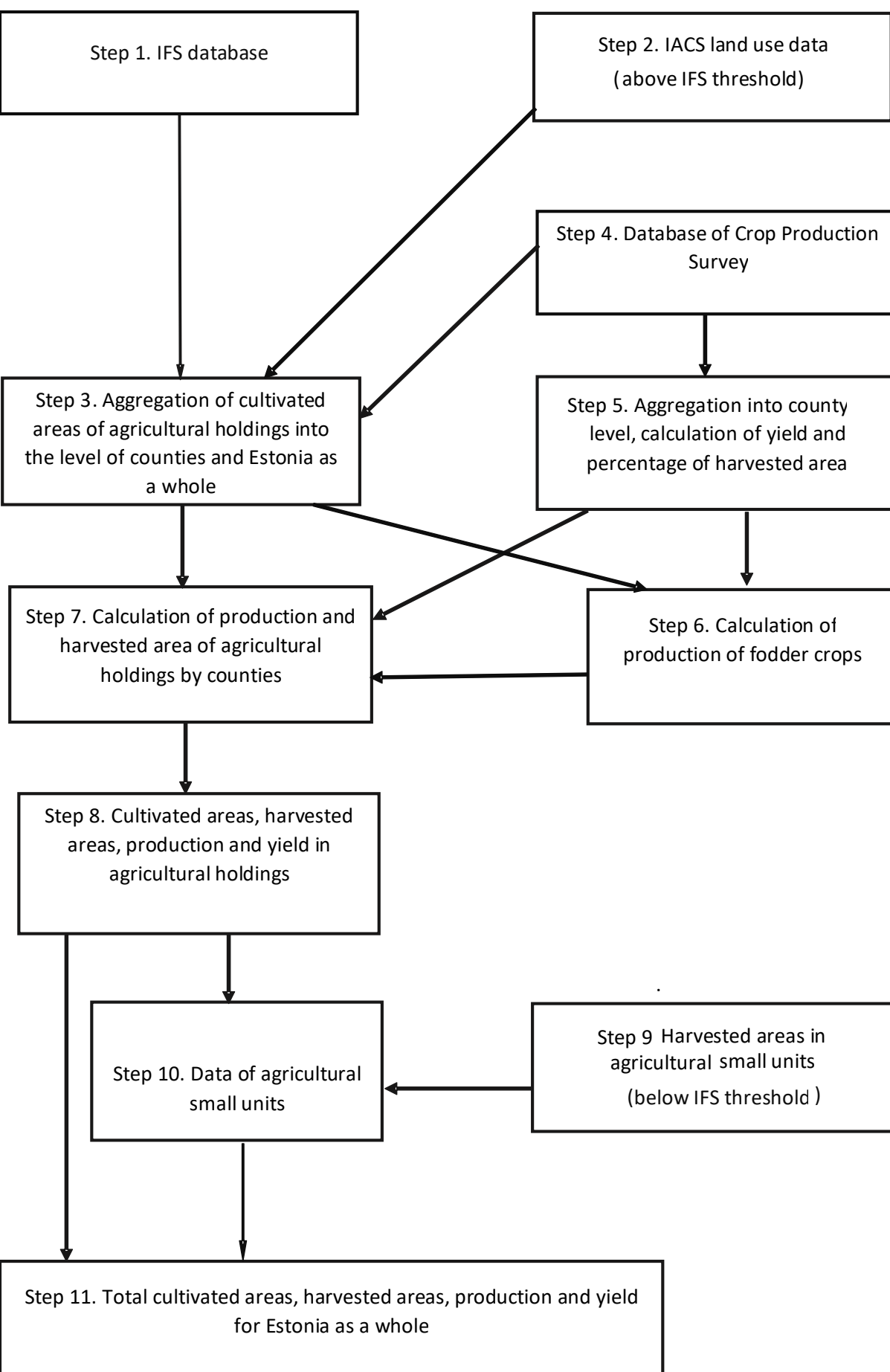
The main difference compared with the previous scheme (non-FSS years) is that cultivated areas are based on IFS data, which are detailed and supplemented by data from the IACS and Crop Production Survey. Another difference is that IACS units are divided into two groups – units which are above the IFS threshold (agricultural holdings) and units below the IFS threshold. Data of units above the IFS threshold are used for supplementing IFS data and data of units below the IFS threshold are added into dataset of agricultural small units.

From 2020, IACS units are divided into two groups – agricultural holdings and small units.

Figure 5.2 shows schematically the idea of data processing.

Figure 5.2. Data processing in IFS years with the new IFS threshold (2020, 2023, 2026)





Step 1. Data for calculating AR\_C are received from IFS databases.

Step 2. Data of IACS units above the IFS threshold (agricultural holdings) are in the database. Data have been transformed into the level of statistical register variables and land parcels have been merged into holdings. Data are used for supplementing Agricultural Census data.

Step 3. Data processing of land areas.

- Land areas from IFS are grouped and aggregated by counties.
- Data of greenhouse crops from IFS are detailed based on Crop Production Survey data.
- Data of fruit and berry plantations from IFS are detailed based on IACS data.
- Land areas of all crops are aggregated into groups (cereals, dry pulses, ..., arable land, permanent grassland etc.) by counties and data of counties into agricultural holdings in Estonia as a whole.

Step 4. Database of the Crop Production Survey.

Step 5. Calculation of yields and percentages of harvested crops based on Crop Production Survey (on the level of counties).

- For all crops (except for permanent crops, multiannual fodder crops, permanent grassland, flowers and ornamental plants) yield is calculated as  $YI = PR / AR\_C$ .
- For permanent crops, yield is calculated based on harvested area  $YI = PR / AR\_HA\_PR$ .
- For all permanent crops (also for cereals, dry pulses, potatoes, rape and turnip rape) the percentage of harvested area from cultivated area is calculated as  $SUHE = AR\_HA\_PR / AR\_C$ .

Step 6. Calculation of the production of fodder crops. These data are used for Economic Accounts for Agriculture and for national needs.

Step 7. Calculation of harvested areas and production in agricultural holdings (by counties).

- Harvested areas of permanent crops and cereals, dry pulses, potatoes, rape and turnip rape are calculated on the basis of IFS data and percentage from Crop Production Survey as  $AR\_HA\_PR = AR\_C * SUHE$ .
- Production of crops (except permanent crops) is calculated based on IFS data and yield from Crop Production Survey as  $PR = AR\_C * YI$ .
- Production of permanent crops is calculated based on harvested area and yield from Crop Production Survey as  $PR = AR\_HA\_PR * YI$ .
- Harvested areas (AR\_HA\_PR) and productions (PR) are calculated for groups.
- Yield for groups (except permanent crops) is calculated as  $YI = PR / AR\_C$ .
- Yield for permanent crops is calculated as  $YI = PR / AR\_HA\_PR$ .

Step 8. Cultivated areas, harvested areas, production and yield in agricultural holdings for Estonia as a whole.

- Cultivated areas (AR\_C), harvested areas (AR\_HA\_PR) and production (PR) of groups for all counties are summed up.
- Yield is calculated for all crops and groups of crops as  $YI = PR / AR\_C$ , for permanent crops as  $YI = PR / AR\_HA\_PR$ .

Step 9. Land areas of agricultural small units and kitchen gardens (data of agricultural small units are used in land use data of Crop Statistics and for national needs).

Step 10. Calculation of harvested areas and production in agricultural small units and kitchen gardens of agricultural holdings (for national needs).

Step 11. Cultivated areas, harvested areas, production and yield (total for agricultural holdings and small units (the latter including IACS units below threshold)).

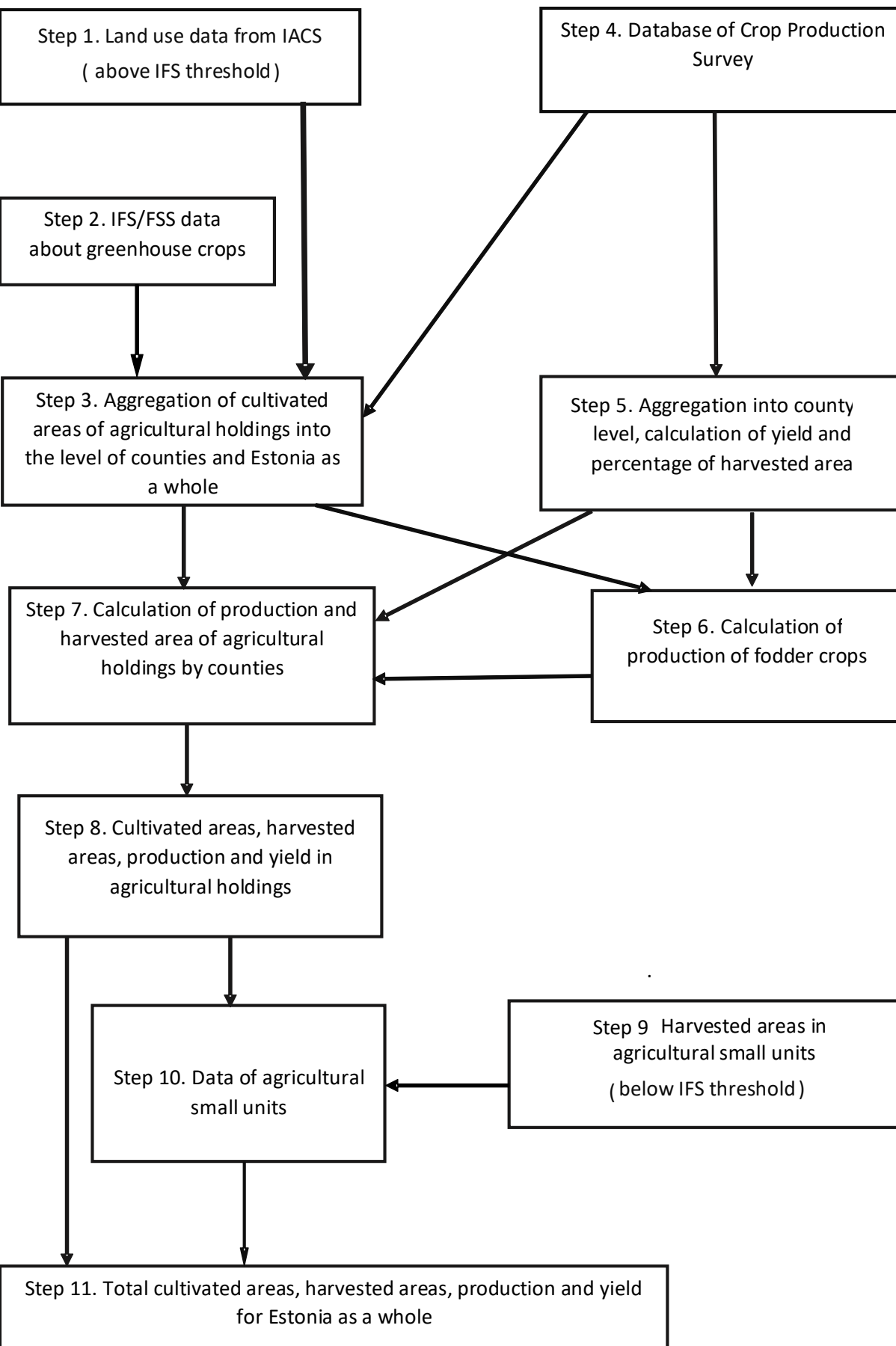
Data of agricultural holdings from step 8 and total areas of small units from step 10 are used for delivering Crop Production Survey data to Eurostat.

#### **Data processing in non-IFS years with the new IFS threshold (2021, 2022 ...)**

The main difference compared to non-FSS years before the new IFS threshold is that here IACS units are divided into two groups – units that are above the IFS threshold (agricultural holdings) and units below the IFS threshold. Data of units above the IFS threshold are used for agricultural holdings and data of units below the IFS threshold are added into dataset of agricultural small units.

Figure 5.3 shows schematically the idea of data processing.

Figure 5.3. Data processing in non-IFS years with new the IFS threshold (2021, 2022 ...)



Step 1. Data for calculating AR\_C are received from IACS database, where the data of units above the threshold are available (agricultural holdings). Data have been transformed into the level of statistical register variables and land parcels have been merged into holdings.

Step 2. As the IACS database does not include full data about greenhouse crops, they are taken from IFS (together with the expansion factor).

Step 3. Data processing of land areas.

- Land areas from IACS are grouped and aggregated by counties.
- For units not registered in IACS, the areas of open field vegetables and permanent crops from the Crop Production Survey database are added into IACS data.
- Data of greenhouse crops from IFS are detailed based on Crop Production Survey data.
- Land areas of all crops are aggregated into groups (cereals, dry pulses, ..., arable land, permanent grassland etc.) by counties and data of counties into agricultural holdings in Estonia as a whole.

Step 4. Database of the Crop Production Survey.

Step 5. Calculation of yields and percentages of harvested crops based on Crop Production Survey (on the level of counties).

- For all crops (except for permanent crops, multiannual fodder crops, permanent grassland, flowers and ornamental plants), yield is calculated as  $YI = PR / AR\_C$ .
- For permanent crops, yield is calculated based on harvested area  $YI = PR / AR\_HA\_PR$ .
- For all permanent crops (also for cereals, dry pulses, potatoes, rape and turnip rape) the percentage of harvested area from cultivated area is calculated as  $SUHE = AR\_HA\_PR / AR\_C$ .

Step 6. Calculation of the production of fodder crops. These data are used for Economic Accounts for Agriculture and for national needs.

Step 7. Calculation of harvested areas and production in agricultural holdings (by counties).

- Harvested areas of permanent crops and cereals, dry pulses, potatoes, rape and turnip rape are calculated on the basis of IACS data and percentage from Crop Production Survey as  $AR\_HA\_PR = AR\_C * SUHE$ .
- Production of crops (except permanent crops) is calculated based on IACS data and yield from Crop Production Survey as  $PR = AR\_C * YI$ .
- Production of permanent crops is calculated based on harvested area and yield from Crop Production Survey as  $PR = AR\_HA\_PR * YI$ .
- Harvested areas (AR\_HA\_PR) and productions (PR) are calculated for groups.
- Yield for groups (except permanent crops) is calculated as  $YI = PR / AR\_C$ .
- Yield for permanent crops is calculated as  $YI = PR / AR\_HA\_PR$ .

Step 8. Cultivated areas, harvested areas, production and yield in agricultural holdings for Estonia as a whole.

- Cultivated areas (AR\_C), harvested areas (AR\_HA\_PR) and production (PR) of groups for all counties are summed up.
- Yield is calculated for all crops and groups of crops as  $YI = PR / AR\_C$ , for permanent crops as  $YI = PR / AR\_HA\_PR$ .

Step 9. Land areas of agricultural small units and kitchen gardens (data of agricultural small units are used in land use data of Crop Statistics and for national needs).

Step 10. Calculation of harvested areas and production in agricultural small units and kitchen gardens of agricultural holdings (for national needs).

Step 11. Cultivated areas, harvested areas, production and yield (total for agricultural holdings and small units, the latter including IACS units below the threshold).

Data of agricultural holdings from step 8 and total areas of small units from step 10 are used for delivering Crop Production Survey data to Eurostat.

### 3.6 Process 6. Geocoding of farms

According to the project plan, an analysis was planned to be conducted regarding the possibility and related rules of geocoding the locations of farms based on administrative sources without asking these questions during the survey.

The following data were used within the analysis and for testing the planned methodology:

- 1) **Data of clients of the Agricultural Registers and Agricultural Board** (as of 31.12.2018) – 452,975 records. The active records were filtered as LQPPVERSIOON\_ID=0. After applying the filter, the number of active records is 52,258. The database includes the address of the location of agricultural holding determined by the holding (on the level of the settlement) as well as the postal address.
- 2) **IACS land use data** (as of 31.12.2018). The active records were filtered as AASTA=2018 and LQPPVERSIOON\_ID=0. After applying the filter, the number of active records was 176,792 related to 14,875 different clients.
- 3) **Data of agricultural animals and their owners** (as of 31.12.2018) – in total 9,290 records related to 6,696 owners. The file includes a reference to the livestock building where animals are kept.
- 4) **Spatial data** (as of 31.12.2018). It includes spatial shapes of fields and semi-natural grasslands. There is in total 176,517 records (spatial shapes). One client may have more than one field/semi-natural area.
- 5) **Address data** – include ADS\_OID information and their relevant x,y co-ordinates. As there may be different ADS\_OID codes, the selection has to be made (“is currently\_ind=1”), if necessary, also other ADS\_OID codes can be used. ADS\_OID is an ID number of the address object. It consists of letters and numbers. First two letters show how exact is the location of address (on the level of room, building, settlement, etc.).

All listed databases can be linked through identification number KLI\_ID. SPSS and SAS software have been used for analysis.

The following Eurostat instructions from the draft Handbook have been followed.

1. The location of the main agricultural building. For livestock farms it should be where the livestock is located. For crop and mixed farms it should be the main building of the holding, which is usually located close to the agricultural activities.

2. If there is no agricultural building to which a location of the holding could be attributed:

- a. The location of the majority of the area of the holding
- b. The location of the most important parcel chosen by physical size.

The physical size will be calculated based on the number of hectares belonging to that parcel, which also belongs to the holding located in a particular area.

c. The location of the most important parcel chosen by economic size.

3. The location of the farmer's residence if it is not further than 5 km straight from the farm

4. A combination of the above criteria

The list of potential holdings was received by linking IACS land use data, data of agricultural animals and their owners and data of clients. Duplicates were removed and common part of these

three files was taken as a basis for the analysis. The list included 17,350 clients, 6,682 of them were units with livestock and 10,668 units with land. Some of the units with livestock had also land. 1,013 clients from 17,350 did not have a location of holding determined by the holding. Still they had a postal address, which can also be partly used for determining the location of the holding. The IFS threshold was not used in the first step.

### Overview of the methodology of the agricultural census

Table 6.1. Order of applying location rules and number of clients according to rules

Location rule	Rule number	Number of livestock farmers	Number of crop farmers	Total	Aggregate number	Balance
F3 level name = name of the settlement of the place of activity / spatial data	1	4,874	8,708	13,582	13,582	3,768
P_ADS_OID = F_ADS_OID	2	497	1,279	1,776	15,358	1,992
<5 km distance between place of activity and postal address	3	596	256	852	16,210	1,140
ADS_OID of livestock housing with the largest economic output	4	692	0	692	16,902	446
Field with largest area	5	2	425	427	17,329	21
Not yet determined		21	0	21	17,350	0
<b>Total</b>		6,682	10,668	17,350		

Table 6.1 shows the order of applying rules and for how many crop and livestock farming clients a location is determined based on the respective rule. Explanations on the table rules and abbreviations are as follows:

P\_ADS\_OID – postal address ID (address type + 8-digit numerical code, e.g. ME01065633)

F\_ADS\_OID – activity location ID (address type + 8-digit numerical code, e.g. EE02090762)

Through a conversion of the above-mentioned ADS\_OIDs a coordinate pair is found, which will be the client location according to the rule applied. The abbreviations of address types and their meanings are shown in Figure 6.1.

OBJEKTILIIGID									
Jkn	Kood	Nimetus	Esitaja	Unikaalne	Initsialiseeriv tase	Minimaalne tase	Maksimaalne tase	Kehtetu	
	↕ ⤴	↕ ⤴	↕ ⤴	↕ ⤴	↕ ⤴	↕ ⤴	↕ ⤴	↕ ⤴	↕ ⤴
1	AY	asustusüksus	Maaregister	jah	3	3	3		
2	EE	elukondlik hoone	Ehitisregister	jah	678	6	8		
3	ER	eluruum	Ehitisregister	jah	8	1	8		
4	CU	katastriüksus	Maaregister	ei	678	6	8		
5	KN	kohanimi	Kohanimeregister	jah	6	6	6	18.06.2013	
6	LP	liikluspind	Kohanimeregister	jah	5	5	5		
7	LO	linnaosa	Maaregister	jah	3	3	3		
8	MK	maakond	Maaregister	jah	1	1	1		
9	ME	mitteelukondlik hoone	Ehitisregister	ei	678	6	8		
10	MR	mitteeluruum	Ehitisregister	jah	8	1	8		
11	OV	omavalitsus	Maaregister	jah	2	2	2		
12	RR	rajatis	Ehitisregister	ei		3	7		
13	TT	teeregistri tee	Kohanimeregister	jah	6	6	6	22.04.2012	
14	VK	väikekoht	Kohanimeregister	jah	4	4	4		

Figure 6.1. Abbreviations and meanings of address types. Source: Estonian Land Board (extract from the public address data application)

### Meanings and applications of rules

1. F3 level name = name of the settlement of the place of activity / settlement name in the spatial data – F3 comes from ARIB client data, names of the settlement of the place of activity and spatial data originate respectively from the spatial data files of animal housing and fields. In this case, the ARIB client data F\_ADS\_OID is applied.
2. P\_ADS\_OID = F\_ADS\_OID – if according to ARIB client data, the client has given the same ADS\_OID to the postal address and place of activity address, the client location can be determined as the ADS\_OID in the ARIB client data. It should be noted that the rule is not applied when the ADS\_OIDs begin with AY address type (i.e., the location is determined at settlement unit precision).
3. <5 km distance between place of activity and postal address – if the distance between the place of activity given by the client and the postal address is less than 5 km, the P\_ADS\_OID obtained from the ARIB client data is applied.
4. Animal housing ADS\_OID – the ADS\_OID obtained from animal housing files where the ADS\_OID of the LPK table of the livestock building with the largest economic output (previously calculated) is applied to determine the location. Beehives are also considered animal housing (taken into account last).
5. Field with largest area – all other cases for which a location is not determined based on the first four rules are set a location according to the largest field with the coordinate pair in the centre.

The first three rules take into account the location given by the client, the fourth rule considers the location of animal housing and the fifth rule considers the location of the largest field (centre).

Based on the above-mentioned location rules, it was possible to determine the location of 17,329 clients, accounting for 99.88% of the sample. The location of 21 livestock farmers was not determined.

### Other explanations

- 1) **Total** column shows how many clients a location is determined after applying the specific location rule.



- 2) **Aggregate number** shows for how many clients a location is determined after applying the previous location rules.
- 3) **Balance** shows for how many clients the location is not yet determined after applying the specific rule.

The IFS threshold was not used in the analysis, as the total number of holdings related to IACS was small. It was decided that for this very small number of holdings related to IACS and those few not related to IACS (the total number of which is expected to be smaller than 200), the location of the holding will be determined in the data processing process (if necessary, by asking the location directly from the respondent).

## 4 Conclusions

Overall, it can be concluded that the project work has been carried out according to the plan. No major problems have occurred. It was of great help that the regulations and handbook were provided by Eurostat in good time, which enabled to follow the timetable. As a whole, the project was very useful for Estonia. Many preparations were made for IFS, which is a good basis for further preparatory works for Agricultural Census 2020. Hopefully, the results described in the final report give some fresh ideas also to other Member States.

### List of main abbreviations

Acronym	Description
ARIB	Administrative Registers and Information Board
AC	Agricultural Census
FSS	Farm Structure Survey
IACS	Integrated Administrative and Control System
IFS	Integrated Farm Statistics
LSU	Livestock unit
SO	Standard output
UAA	Utilised Agricultural Area
VAIS	Data processing system
VVIS	Data collection system mainly for social statistics surveys

## List of tables

- Table 1.1. Number of agricultural holdings by size of agricultural area, 2001-2016
- Table 1.2. Key variables by IFS threshold, 2010, 2013, 2016
- Table 1.3. Labour force and other gainful activity module
- Table 1.4. Animal housing and manure management module
- Table 1.5. Stratification plan for AC 2020 modules
- Table 1.6. Relative standard errors
- Table 5.1. Structure of the output table
- Table 5.2. Output variables
- Table 6.1. Order of applying location rules and number of clients according to rules

## List of figures

- Figure 1.1. Number of agricultural holdings by size of agricultural area, 2001-2016
- Figure 5.1. Data processing in non-FSS years before the new IFS threshold (in 2018, 2019)
- Figure 5.2. Data processing in IFS years with the new IFS threshold (2020, 2023, 2026)
- Figure 5.3. Data processing in non-IFS years with the new the IFS threshold (2021, 2022 ...)
- Figure 6.1. Abbreviations and meanings of address types.