



EUROPEAN COMMISSION

EUROSTAT

Anonymisation rules for Farm Structure Survey

24 March 2020

1. INTRODUCTION

The Farm Structure Survey (FSS) provides information on the agricultural holdings (farms). It is conducted on the basis of European and national regulations (see: Regulation 1166/2008). The variables include information about:

- Characteristics of farms and their management
- Land use
- Livestock
- Special topics of interest: farm labour force, rural development issues etc.

The following data releases are available:

- 1990 – census
- 1993, 5, 7 – sample survey
- 2000 – census
- 2003, 5, 7 – sample survey
- 2010 – census
- 2013, 16 – sample survey¹

The release of FSS data for research purposes was planned since many years. In 2011 the analysis of SDC methods for release of scientific use files was conducted by ISTAT and CBS and presented to the FSS Working Group. The proposed approach was deemed too strict to be of value to researchers. Access to non-anonymised microdata in the Eurostat safe centre was considered instead. Due to technical problems the safe centre option was not followed and the whole project was abandoned for some years.

In 2017, in view of the constant demand for the FSS microdata, Eurostat decided to pursue the option of FSS scientific use files again. At its meeting on 25-26 October 2017 the Working Group on the Structure of Agricultural Holdings supported the release of the FSS in the form of scientific use files. At the same time the WG asked Eurostat to propose partial anonymisation methods for scientific use files and to analyse the related disclosure risk.

The next two paragraphs describe the general anonymisation approach for the scientific use files. Although all countries agreed to this approach, a few countries

¹ In some countries the data are always collected via census. Smallest farms are often excluded from the survey.

requested some country specific measures. These country specific measures are described in paragraph 4.

2. OBJECTIVES AND ASSUMPTIONS

The objective of the work is to develop scientific use files, i.e. the files to which methods of statistical disclosure control have been applied to reduce to an appropriate level and in accordance with current best practice the risk of identification of statistical units (definition from the Regulation 557/2013). According to this definition the risk of identification in scientific use files is reduced, not eliminated.

The users of the files are researchers. The researchers go through the eligibility procedure to get access to the data. Their organisations have to be recognised as research entities and their projects validated by Eurostat and ESS national statistical institutes. The researchers sign several commitments where, amongst others, it is stated that "they will not attempt to identify any individual record (individual, household, business etc.) in the dataset, or claim to have done so". They also commit to "preserve the confidentiality of information pertaining to identifiable individuals, households and/or organisations that are recorded in the dataset". Additional safeguards are signed at the level of research entities.

All research projects submitted by researchers are scrutinised by Eurostat and NSIs. Eurostat applies several conditions on the research proposals to be accepted, namely: scientific content of the proposal, comparative analysis of several countries, proper data security measures in place, publication of results, etc. In addition, each research proposal has to be accepted by the NSIs. If the access is not allowed by an NSI the data of that country are removed from the set.

The information about conditions of access to European microdata can be found on Eurostat website [here](#).

Farm Structure Survey data are specific because the type of production is usually visible (crops growing, elevation of cows etc.). The focus of anonymisation is therefore on limiting the additional information that the researchers will get in scientific use files.

3. PARTIAL ANONYMISATION METHOD FOR FSS SCIENTIFIC USE FILES

The partial anonymisation method for the FSS can be summarised as follows:

- Recoding of the identifying variables (legal personality of the holding, holding manager age, other gainful activities, farm typology), see table 1 for details;
- Removing of detailed geographical coordinates;
- Reducing information about the farms and their management (for example: information about farm holder/manager), see table 1;
- Merging detailed categories of some production variables or removing them completely, in general variables covering special topics of interest are removed (see excel file for details).

- Rounding to one decimal position variables expressed in hectares as unit.
- Micro-aggregation of highest metric values of all metric variables within given regional level (normally NUTS2). All numeric variables (hectares, number of heads, persons working in the agricultural holding, annual working units, tons of manure) are to be micro-aggregated (the last column in the [excel file](#) indicates if the variable is micro-aggregated).

The advantages of the micro-aggregation:

- Protection of biggest units, without indicating if few units dominate the market, nor the difference between the first and the second;
- Usability of the file: totals are preserved and numeric fields remain numeric;

Table 1 Recoding of geographical variables and variables describing the farm and farm management

Code	Description	Original categories	Final
A04a\$nutscode2	NUTS 2 region codes	NUTS 2	NUTS 2 unless impossible for the country (to be discussed bilaterally with Eurostat)
	Geographical coordinates, municipality codes, specific (protected) land categories		removed
A_2\$holdingtype	Legal personality of the holding – holding type (HT)	6 (1, 2, 3, 4, 5;6) ²	4 categories: <u>1+2+3</u> , 4, 5;6
	Holder details	Sex, age, labour, training	removed, only working hours % kept
E_1_2\$AgeCodeM	Manager: Age group	Continuous variable	4 age groups (<35, 35-54, 55-64, 65+)
	Manager: sex	2 categories	removed
F_1_2\$y_n	Other gainful activities (OGA)	11 subcategories	all "other gainful activity" variables merged without specifying their types
FARMTYPE See annex 2 for details	Farm typology calculated by Eurostat	62 (3-digit), 22 (2-digit) or 9 (1-digit)	9 categories (1-digit)
	Machinery and equipment	14 categories	removed

² Holding type categories: (1) Sole holder is also the manager, (2) Manager is a member of the holder's family, (3) Holder is not a manager who is not a member of the holder's family, (4) Holding is a legal person, (5) Holding is a group holding.

Code	Description	Original categories	Final
	Working time and sex of labour force excluding managers and holders	>10 categories	removed
	Contractor's work	2 (yes, no)	removed

Additional comments:

- Regional level: the data are consistent at the NUTS2 level that is why it is proposed to release data at this level. It is possible to discuss with countries bilaterally if the NUTS2 level is considered too detailed.
- Number of farms considered for micro-aggregation: for each metric characteristic, the top 3 farms are to be used by default. In order to adapt to national legislations, it can be discussed bilaterally the increase of this number for some variables (it needs to be duly justified).
- Stratum ID: it will be kept in scientific use files, it allows extrapolation of figures to the total population and variance estimation.
- Farm type classification: Note that release of 1 digit categories may motivate researchers to calculate the farm types on their own; this could create discrepancies with regard to publicly available data. The farm type and standard output of the farm are calculated on the basis of standard output coefficients per product, publicly available on this link: <http://ec.europa.eu/eurostat/web/agriculture/so-coefficients> . Once standard output values are calculated, farm typology can be calculated using the definitions in Commission Regulation (EU) 2015/220 of 3 February 2015 on setting up a network for the collection of accountancy data on the incomes and business operation of agricultural holdings in the European Union.
- Drawing a sample from census data and a subset of the sampled unit – Eurostat cannot do that due to lack of information about the national sampling methods/frames.

The [excel file annexed](#) presents the proposed content of the scientific use files (the second last column indicates the inclusion of the variable, the last column indicates whether or not the variable is micro-aggregated).

4. COUNTRY SPECIFIC ANONYMISATION METHODS

Although all countries agreed to the general anonymisation approach, some countries requested additional country specific measures. It concerns Slovenia, Italy, Germany and Croatia. The country specific measures are listed below.

The data of these countries are not available yet for release.

Slovenia:

- No regional breakdown (NUTS0 = country level).
- The economic size of the farm by standard output size classes (according to ANNEX II of the COMMISSION REGULATION (EC) No 1242/2008 of 8 December 2008 establishing a Community typology for agricultural holdings).
- The variable Holdingtype is removed.
- The variable E_1_2\$AWU (Actual Work Units on managers) is removed (percentile groups are kept).

Italy:

- The geographical detail is limited to the NUTS1 level.

Germany:

- Micro-aggregation of the **20** highest values of all numerical variables.
- The geographical detail is limited to NUTS0 (country level).

Croatia:

- The variable STRATUM_ID is recoded to meaningless numbers 1 to 106.

Belgium:

- The holdings included in region BE10 are classified to region BE24.

Annex 1

[The excel file presenting the proposed content of the scientific use files \(the second last column indicates the inclusion of the variable, the last column indicates whether or not the variable is micro-aggregated\).](#)

Annex 2

[Recoding of farm types](#)