Italian Farm Structure Survey: SDC Methodology for an MFR Dissemination

D. Ichim, Istat, Servizio Metodi, Strumenti e Supporto metodologico per i processi di produzione statistica (DCMT/MSS)

Italian Farm Structure Survey: SDC Methodology for an MFR Dissemination

D. Ichim, Istat, Servizio Metodi, Strumenti e Supporto metodologico per i processi di produzione statistica (DCMT/MSS)

Abstract: This document describes the statistical disclosure control methodology used for the dissemination of microdata stemming from the 2005 wave of the Italian Farm Structure Survey. Only the dissemination of a microdata file for research purposes is considered.

Keywords: Microdata File Dissemination, Statistical Disclosure Control, Farm Structure Survey

Summary

1. Introduction.

In order to satisfy the users needs, ISTAT disseminates microdata files for scientific research (MFR). The dissemination of the statistical information must be performed in full compliance with the regulations pertaining the privacy of respondents. The Farm Structure Survey (FSS) has recently been included in an amending version of the Commission Regulation 831/2002. For the FSS, a statistical disclosure control methodology was set-up in order to guarantee the confidentiality of respondents. An important key point is the fact that the microdata file would be released for research purposes, and hence subject to a signed contract. Consequently, a rigorous study of possible disclosure scenarios, including spontaneous identification scenarios, was carried out in order to define the identifying variables. Then a careful risk assessment analysis was performed. The protection was achieved using variable suppression, variable aggregation and perturbation. Data utility constraints were taken into account when the perturbation method was applied. In section 2 the survey is briefly described. The statistical disclosure control methodology is presented in Section 3. Details on the disclosure risk evaluation, disclosure limitation and data utility are also given in section 3.

2. Description of the Farm Structure Survey (FSS)

2.1 Statistical Regulation

The rules governing the farm structure surveys are laid down in a number of Council Regulations and Commission Regulations and Decisions, which are published in the Official Journal of the European Communities. These documents contain the following information:

(i) basic rules on organising the surveys:
- 2467/96/EC
- 571/88/EEC Regulation

(ii) list of characteristics the Member States are obliged to survey:
- 2004/2139/EC Regulation
- 2002/143/EC Regulation
- 98/377/EC Decision
- 97/621/EC Decision
- 96/170/EC Decision
- 94/677/EC Decision
- 93/156/EEC Decision
- 807/89/EEC Regulation
- 571/88/EEC Regulation
- 93/23 (on the statistical surveys to be carried out on pig production)
- 93/24 (on the statistical surveys to be carried out on bovine animal production)
- 93/25 (on the statistical surveys to be carried out on sheep and goats stocks)
- 93/16 (on statistical surveys of milk and milk products)

(iii) definitions of the characteristics:
- 1444/2002/EC Regulation
- 2000/115/EC Decision
- 97/418/EC Decision
- 96/170/EC Decision
- 89/651/EEC Decision

(iv) use of data sources other than statistical surveys:
- 2005/124/EC Regulation
- 98/377/EC Decision
- 97/621/EC Decision

(v) deadlines for transmission of FSS data:
- 2004/2139/EC Annex 3 Regulation
- 68/2003/EC Regulation
- 99/714/EC Decision
- 97/407/EC Decision
- 93/502/EEC Decision
- 89/652/EEC Decision

(vi) Community program of tables to be stored in the Tabular Data Bank (DBT) and of individual data of the EUROFARM system:
- 99/799/EC Decision
- 97/341/EC Decision
- 94/772/EC Decision
- 89/653/EEC Decision
- 571/88/EEC Regulation

(vii) classification of holdings to their economic size and farm types (typology):
- 2003/369/EC Decision
- 1999/725/EC Decision
- 96/393/EC Decision
- 88/284/EEC Decision
- 377/85/EEC Decision

The laws governing the Italian FSS2005 are the following:

(i) Decreto legislativo del 6 settembre 1989, n. 322

(ii) Decreto legislativo del 30 giugno 2003, n. 196

(iii) Codice di deontologia e buona condotta per i trattamenti di dati personali a scopi statistici e di ricerca scientifica effettuati nell'ambito del Sistema statistico nazionale (all. A3 al Codice in materia di protezione dei dati personali d.lgs. 30 giugno 2003, n. 196)

(iv) Decreto del Presidente del Consiglio dei Ministri del 23 aprile 2004 – Programma statistico nazionale per il triennio 2004 – 2006 (v)Decreto del Presidente della Repubblica 14 luglio 2004.

2.2 The European FSS

The FSS collects information on the agricultural holdings in the Member States at different geographical levels (Member States, regions, districts) and over periods (follow up of the changes in the agricultural sector). Thus the FSS provides a base for decision making in the Common Agricultural Policy (CAP). The FSS is organised in all Member States on a harmonised base. Whereas the characteristics are based on community legislation, the same information is available for all countries for each survey. The variables are classified in four groups:

(i) general overview (key variables)

(ii) detailed data on land use

(iii) detailed data on livestock

(iv) detailed data on special interest topics: farm labour force, rural development issues as well as management and practices.

The scope of the survey is agriculture, while the survey unit is the agricultural holding.

2.3 The Italian FSS2005

In this section, the methodology used to carry out the Italian FSS2005 is briefly described.

The Italian FSS2005 was carried out at the end of the agricultural year 2005. The reference periods of this survey were:

(i) crops and permanent crops: 01/11/2004 – 31/10/2005.

(ii) livestock: 01/12/2005

(iii) labour force: 01/11/04 – 31/10/2005

(iv) other items: 01/11/2004 – 31/10/2005

The complete list of the observed variables is given in the Annex 1. Among the characteristics requested by the Regulation 571/88 the following have not been surveyed because are irrelevant or non significant in Italy:

- Section D- arable land
  - Linseed (oil flax)
  - Flax
- Section G - permanent crops
  - Raisins

2.3.1 Survey Strategy

The target population of the survey included the farms which possess:

(i) an agricultural area utilised for farming of one hectare or more, or

(ii) an agricultural area utilised for farming less than one hectare only if

(a) their production for sale exceeds a certain threshold (2.066 euro) or

(b)if their production unit exceeds a certain physical threshold.

2.3.2 Sampling Design

The frame was the 2000 agricultural census list. The census list was updated in order to account for the merging and demerging phenomena. A stratified random sample design was used; a census was performed in the strata containing the largest farms. The selected sample contained 56542 units.

2.3.3 Data collection

A face-to-face interview was used for data collection. The interviewers directly included in the sample those farms resulting from a splitting or a merging of a sampling unit.

2.3.4 Response Rate and Handling of Non-response

The response rate was generally greater than 90%, as it can be observed in table 1. The sample size was 47780.

Table 1: *The response rate by region (NUTS2).*

Region	Response rate
Piemonte	96.8
Valle d'Aosta	87.2
Lombardia	98
Bolzano	92.4
Trento	95.4
Veneto	96
Friuli Venezia Giulia	95.6
Liguria	93.2
Emilia Romagna	97.7
Toscana	90.4
Umbria	97.5
Marche	96.9
Lazio	93.1
Abruzzo	96.5
Molise	97.8
Campania	93.4
Puglia	92.4
Basilicata	95.5
Calabria	91.4
Sicilia	83.9
Sardegna	88.7

The design weights were updated to account for unit non-response.
Missing or incorrect items on influent farms were corrected by comparison with regional administrative data (when available) or by a telephone check.
Missing or incorrect data items on non influent farms were corrected by a mixed strategy defined by hierarchical editing and imputation.

2.3.5 Extrapolation factor

The final weight of each unit was computed as a product of three factors: sampling weight, total non-response and calibration adjustment factors. The latter adjustments were performed in order to preserve some population characteristics, as derived from the agricultural census.

More details on the Italian FSS may be found in Greco (2007).

3. Statistical Disclosure Control Methodology

The microdata file described in this document should be released exclusively for scientific research purposes. It was assumed that the interested researcher would sign an agreement with Istat. It was also assumed that such agreement could be sufficient to refrain the researcher from deliberately trying to breach the confidentiality of any respondent.
However, Istat collects the data guaranteeing the dissemination of statistical information in full compliance with the national laws concerning statistical confidentiality.
In the SDC literature, many types of disclosure are mentioned, see Hundepool (2009). For the dissemination of microdata stemming from the FSS, the disclosure definition given in the Council Regulation (EC, No223/2009, March 2009) on European statistics was adopted: "confidential data means data which allow statistical units to be identified, either directly or indirectly, thereby disclosing individual information. To determine whether a statistical unit is identifiable, account shall be taken of all relevant means that might reasonably be used by a third party to identify the statistical unit". It should be emphasized that the sole existence of the above mentioned contract/agreement could not be sufficient for the release of a microdata file. Istat has the ethical and legal obligation to ensure that no statistical unit could be identified.
In section 3.1 the adopted disclosure scenarios are described. In section 3.2 the applied statistical disclosure limitation methodology is presented.

3.1 Disclosure Scenario

To assess the disclosure risk, one should make realistic assumptions about what an intruder might know about respondents and what information would be available to him to match against the microdata and potentially make an identification and a subsequent disclosure. These assumptions are known as disclosure scenarios, see Hundepool (2009). The disclosure scenarios consist of the analysis of the users and their needs and the analysis of the product to be released: the identifying and confidential variables.

3.1.1 Users

For the MFR, the potential users are known in advance: (generally) academic researchers signing an agreement to receive the file for performing their analyses. Even if the users are considered bona-fide users, the researchers might recognize/identify some units. For example, in the FSS framework, it is publicly known that the largest farms are generally included in the sample. The largest farms are also the most famous ones. Consequently, a spontaneous identification or recognition might occur. Moreover, even a bona-fide researcher might be curious about some units revealed as "particular" from the analyses.

3.1.2 Research Potential

It is the first time ISTAT releases microdata stemming from the FSS. Moreover, no researcher has required, up to now, to analyse any FSS microdata at the Italian Safe Center. At European level, the FSS has just been included in the Commission Regulation 831/2002; hence no microdata dissemination and no analysis at the Eurostat Safe Center were possible till the end of 2008. From a brief review of the scientific literature, it was observed that descriptive statistics and indicators are generally computed. The geographical detail plays an obvious important role for the phenomena under study. After several releases of microdata files stemming from different waves, it would be possible to update and better classify the analyses performed by researchers.
In the remaining part of this section the adopted re-identification scenarios are described. First, the possible ways to identify a farm are discussed. Second, the harm that might derive from such re-identification is analysed.

3.1.3 Identification Scenarios

Microdata files are released only after removing direct identifying variables, such as names, addresses, and identity numbers. Other variables in the microdata file could be used as indirect identifiers such as

geographical location, farming type, standard gross margin, etc.. An identifying key variable is defined by compounding several identifying variables and can be used by an intruder for re-identification.

External Register

The anonymized microdata file would be released only for scientific research purposes. It was assumed that the (academic) researchers would be bona-fide users. Consequently, it was supposed that the researcher would not deliberately try to identify any farm. It follows that any (record) linkage with an (external) register might not be deemed a realistic hypothesis. This is due to the huge amount of resources[1] generally involved in any (record) linkage process.

Nosy Colleague Scenario

The anonymized microdata file would be released only to researchers signing an agreement with Istat. It follows that the MFR users could not generally be colleagues, and not even competitors, of the observed statistical units. Thus, the nosy colleague scenario could not be deemed a realistic disclosure scenario.

Spontaneous Identification Based on Structural Variables

By a careful analysis of the observed variables[2], the following categorical[3] structural variables were considered identifying:

(i) *Area status* (A05) – 3 categories
(ii) *SGM region code* (A07) – NUTS2 - 21 categories for Italy (other "geographical "variables", e.g. A04A or A04D could also be used)
(iii) *Sex* (L011) – 3 categories (the holding might be a legal person)
(iv) *Age group* (L012) – 7 categories

In 4% of the 649 combinations of these 4 variables the sample frequency was equal to 1 and, in many of these combinations the population frequency[4], was not greater than 2. In 3% of combinations of the above 4 variables the sample frequency was equal to 2 and the corresponding population frequencies were not always so large. In about 2% of combinations, the population frequency corresponding to unique and double sample cases was less than 5. These latter units should/could be considered at risk of re-identification.

The variable *Utilised agricultural area* (A11) is a continuous variable and it could be used to identify a farm. Anyway, it should be noted that, except for few units, the variable A11 has a compact range of values. Consequently, there is a high degree of uncertainty associated with this kind of identification. With regard to the few extremely large values, it should be noted that they correspond[5]to *Permanent grassland and meadow* (F). For the Italian FSS2005 microdata, it was observed that all the farms with extremely large values of A11 have an unique agricultural activity: they possess only *Permanent grassland and meadow*. Since agricultural areas with such characteristics are obviously visible without any effort, it was considered that the possible intruder wouldn't increase his knowledge. Consequently these farms were not considered at risk of re-identification in this disclosure scenario.

The variable *Legal personality of the holding* (B0102) registers the relationship between the manager and the holder's family. As it can be observed in table 2, for more than 96% of cases, either the holder is the manager or the holding is a legal person. It would be hard to distinguish between these two categories without a very detailed knowledge of the farm, even if the category the *holding is a legal person* may be registered in some external register. Given also the low frequencies and the weak identification power of the other categories, the variable *Legal personality of the holding* (B0102) might not be reasonably considered a very reliable identifying variable. Moreover, if the legal personality of the holding were used as an identifying variable, the variables L011 and L012 would significantly diminish their identification power.

[1] e.g. acquisition and cleaning of the external database, time, methodology, etc.
[2] According to the suggestions of the survey experts
[3] The possible codes are provided in Annex 2
[4] computation based on *the Extrapolation factor 1* (A09)
[5] for the Italian FSS2005, this statement holds for the largest 10 values of A11, for example

Table 2: *Percentages of categories of Legal personality of the holding (B0102)*

Legal personality of the holder	**Percentage**
Sole holder is also the manager	86.08
Holder's spouse is the manager of the holding	1.78
Manager is a member of the holder's family but not his spouse	1.88
Holder is not the manager who is not a member of the holder's family	0.20
Holding is a legal person	10.06
Holding is a group holding	0.00

Spontaneous Identification Scenario

As already stated, it was assumed that the researchers are bona-fide researchers.

That is, it was supposed that they would not deliberately try to identify any unit. However, an individual re-identification might occur because of the particular characteristics of the survey. A researcher might (unintentionally) use some previous and very detailed knowledge for the re-identification of a farm.

Even if the economic dimension of the farms is not the main FSS objective, it should be noted that some of the farms are firms: some farms are well-known to the general public, not only to the researchers. Moreover, as a census was conducted for the largest farms, it is known that the most (economically) important agricultural enterprises were included in the sample. Additionally, Istat generally invests many resources in obtaining a response from the largest enterprises. This special treatment is due to the significant impact such enterprises generally have on the studied phenomenon. The re-identification of such well-known enterprises could be performed using the *Standard Gross Margin* (SGM, variable A12).

The second issue that should be taken into consideration is the territorial aspect of the survey. Since any agricultural phenomenon is highly related to the territorial characteristics, the geographical location of the farms is essential. Consequently, the release of an MFR without a detailed geographical information would significantly reduce its research potential. In conclusion, it was decided to release the geographical information at NUTS2 detail level.

Unfortunately, the territoriality has some consequences also on the risk of re-identification. The (economic) activity of the farms is expressed by phenomena that are highly visible, e.g. crops and livestock. This visibility obviously favours the spontaneous identification. For example, a farm that is specialist in field crops, e.g. cereals, oilseeds, etc., might be identified by simply observing the field. In the statistical disclosure control framework, this means that some external information is readily available to anyone. Moreover, this information might be quite detailed, as it is registered also in the microdata file. In conclusion, **EACH FARM MIGHT THEORETICALLY BE AT RISK OF RE-IDENTIFICATION** in this spontaneous identification scenario.

3.1.4 Harm –Confidential Variables

This section discusses the harm that might be produced by disclosing some confidential information. The majority of the observed variables (see Annex 1) are related to the agricultural area, including the way it is utilised, and the livestock. Among these variables there is no one that could be considered as strongly confidential. The agricultural area and livestock variables may be used to compute the total SGM and the farming type. But the total SGM expresses only the economic potential of a farm, not its real economic dimension. The farming type could be hardly considered as confidential.

The information on working time, i.e. the variables whose names start with letter "L", is registered in an aggregated form, e.g. average working units (AWU). Consequently, this group of variables couldn't be considered as confidential.

The confidential content of the file is represented by some information on resources: the way in which an enterprise gets or uses its resources.

A first confidential variable could be the *Farm work non-family members non-regularly employed* (A18). It was considered that a farm might be damaged by disclosing this information.

A second confidential variable could be represented by the *Benefits from Investment Aids* (CC05F1 and CC05F2) because such benefits are generally obtained as a result of a competition. Anyway, only the existence of such benefits is registered in the microdata file.

3.2 Statistical Disclosure Limitation Methodology

Needless to say, the users need microdata that resemble as much as possible to the original microdata file. Or, at least, the results of their analyses performed on the MFR should resemble as much as possible to the results obtained if the analyses were performed on the original confidential microdata. However, the confidentiality of the respondents should be guaranteed by Istat. By reducing the information content of the microdata file, the risk of disclosure is reduced; or, at least, the uncertainty associated to the identification is increased. The right balance between data confidentiality and data utility should be found.
This section describes the statistical disclosure limitation methodology applied to the microdata stemming from the FSS 2005. The procedure consisted of several steps: suppression of some identifying variables, variable aggregation and perturbation of some numerical variables.

3.2.1 Suppression of Some Identifying Categorical Variables

The direct identifiers should be always removed from the microdata file. Some other categorical identifying variables should be suppressed in order to eliminate the extremely rare cases in the sample. Anyway, some data utility criteria should also be considered when selecting the variables to be suppressed. For the Italian FSS2005, the following variables were suppressed for disclosure limitation reasons:

(i) *Survey District NUTS code* (A04A)
(ii) *Municipality code for objective zones 2000* (A04D)
(iii) *Area status* (A05)
(iv) *Holding identification number* (A08)
(v) *Stratum identification number* (A09A)

For the Italian FSS2005, if the variable A04D were used to give information about the geographical location of the farm, 61% of the combinations of the categorical identifying variables, see section 3.1.3, would be sample unique cases. If the variable A04A were used to give information about the geographical location of the farm, 12% of the combinations of the categorical identifying variables would be sample unique cases. In both cases, the number of unique cases was considered too high. Consequently it was decided to use the variable *SGM region code* (A07) to release geographical information. As previously stated, only 4% of the combinations of the categorical identifying variables remain unique cases in this setting. The suppression of the variable A05 produced the aimed results: the sample frequencies of the combinations of A07, L011 and L012 were greater than 3. The latter two categorical identifying variables, A08 and A09A, were suppressed because they were considered direct identifiers.
The population frequencies were not considered in this scenario because, due to their high visibility, all units were at risk of re-identification. Consequently, even if there were more than 3 units in the population (checked using the *Extrapolation factor* (A09)), it would not be sufficient because the sampled units could be visible, too.

3.2.2 Transformation of Some Numerical Variables

The regional character of the phenomenon may be observed in the table 3. For each region and for each numerical variable, the number of records with zero values was computed. For each numerical variable, table 3 shows the minimum, mean and maximum, by region, of zero values percentage. It may be observed that, except for very few variables, e.g. *Other Land* (H03), there is at least one region where

Table 3: *Summaries of percentages of zero values, computed for each region.*

Var	Min	Mean	Max	Var	Min	Mean	Max
D01	54.22	87.05	100.00	*G04A*	73.70	90.44	98.83
D02	41.29	80.65	100.00	*G04B*	68.43	81.69	96.97
D03	98.89	99.67	100.00	*G04C*	94.40	98.29	100.00
D04	66.25	85.75	99.87	*G04D*	100.00	100.00	100.00
D05	65.24	92.95	100.00	*G05*	90.75	98.46	99.90
D06	26.11	82.02	100.00	*G06*	96.60	99.28	100.00
D07	84.25	98.96	100.00	*G07*	98.39	99.82	100.00
D08	90.73	97.68	100.00	*H01*	28.10	76.22	89.77
D09	88.15	96.27	100.00	*H02*	14.23	57.94	93.55
D09E1	89.59	97.26	100.00	*H03*	3.16	22.90	45.46
D09F	96.52	99.47	100.00	*I01*	74.73	94.46	99.24
D09G	97.03	99.13	100.00	*I02*	99.64	99.88	100.00
D10	81.70	94.78	99.51	*I03A*	6.20	48.24	80.39
D11	76.34	95.31	100.00	*I03B*	10.22	57.59	83.30
D12	98.92	99.79	100.00	*I08*	58.47	87.23	100.00
D14	67.06	89.48	100.00	*I08AD22*	62.89	89.53	100.00
D14A	69.45	90.39	100.00	*I08B*	88.93	98.54	100.00
D14B	97.45	98.81	100.00	*I08C*	98.08	99.53	100.00
D15	93.57	97.89	100.00	*I08D*	96.25	98.89	100.00
D16	80.00	98.52	100.00	*I08E*	96.91	99.09	100.00
D17	89.11	98.39	100.00	*J01*	89.63	95.14	98.28
D18	46.77	71.51	100.00	*J02*	55.19	82.20	91.99
D18A	48.56	80.60	100.00	*J03*	83.21	91.52	97.97
D18B	67.36	86.69	100.00	*J04*	36.50	84.52	93.83
D18B1	80.30	95.29	100.00	*J05*	86.88	94.71	98.86
D18B23	68.47	90.65	100.00	*J06*	43.43	89.36	96.91
D19	95.40	99.03	99.87	*J07*	28.83	85.35	98.47
D20	100.00	100.00	100.00	*J08*	83.52	94.06	99.25
D21	85.67	94.82	100.00	*J09*	73.50	89.96	99.17
D23	93.33	99.21	100.00	*J09A*	73.50	90.72	99.33
D24	99.96	100.00	100.00	*J09B*	77.50	94.47	99.83
D25	99.96	100.00	100.00	*J10*	87.44	95.30	99.33
D26	99.53	99.92	100.00	*J10A*	88.34	95.80	99.33
D27	76.93	96.35	100.00	*J10B*	93.07	97.35	100.00
D28	64.64	96.97	100.00	*J11*	89.69	98.09	100.00
D29	100.00	100.00	100.00	*J12*	86.28	97.37	100.00
D30	99.37	99.93	100.00	*J13*	71.75	88.84	98.91
D31	100.00	100.00	100.00	*J14*	91.29	96.72	99.71
D32	99.86	99.98	100.00	*J15*	74.64	94.49	99.42
D33	99.93	99.99	100.00	*J16*	96.22	98.70	99.81
D34	97.76	99.72	100.00	*J16A*	98.29	99.48	100.00
D35	99.46	99.92	100.00	*J16B*	98.44	99.63	100.00
E	40.15	71.77	95.06	*J16C*	98.52	99.65	100.00
F	1.82	61.39	88.54	*J16D*	98.22	99.35	100.00
F01	2.19	64.44	89.77	*J17*	93.99	98.08	99.72
F02	36.50	91.45	99.33	*J18*	97.26	99.28	99.98
G01	60.61	84.90	94.44				
G01A	61.88	88.97	96.29	*K01A*	9.76	29.74	67.40
G01B	93.54	98.73	100.00	*K01A0*	49.55	76.58	89.81
G01C	84.11	95.40	99.33	*K01A3*	33.52	59.40	85.70
G02	80.29	97.15	100.00	*K01A41*	42.14	62.07	89.96
G03	38.10	68.49	100.00	*K01A42*	69.91	87.10	99.28
G03A	98.15	99.30	100.00	*K02A*	16.06	50.33	74.27
G03B	38.80	68.98	100.00	*K03A*	81.52	94.52	100.00
G04	58.14	72.65	88.85	*K09A*	42.34	88.42	96.51

the measured phenomenon is not significant. This is proved by the maximum percentages equal or very close to 100%. When the mean percentage of zero values is almost 100% (but not equal), e.g. *Soya* (D28), it means that, at national level, there are only few farms performing the corresponding agricultural activity. Large differences between the minimum and mean percentage of zero values, e.g. *Grain Maize* (D06) prove that the phenomenon is very concentrated in few regions. When the minimum percentage by region is also very high, e.g. *Hemp* (D32), it means that the phenomena is missing at national level. In table 3 one may easily observe which are these variables. It may also be observed that an aggregation of variables might be possible, with respect to the classification of agricultural holdings by farming type.

The standard farming type classification has three levels of types of farming, which nest:

(i) 9 *general types*, including a type for non-classifiable holdings,

(ii) 17 *principal types*,

(iii) 50 *particular types*, some of which may be subdivided, i.e. 70 *subdivision*s at the most detailed level.

The standard farming type classification is presented in Annex 2.

For each agricultural holding, the type of farming is determined by the computation of its total standard gross margin, i.e. the composition of certain predefined partial standard gross margins which make up the total SGM.

The total SGM (standard gross margin) of a holding is computed using a system of standard coefficients. In Italy, this system is developed, maintained and updated by INEA (Istituto Nazionale di Economia Agraria, www.inea.it). It should be noted that this system of standard coefficients is defined at the most detailed level of the variables needed to derive both the total SGM and farming type. Indeed, the total SGM of a holding is derived from partial SGMs computed from the coefficients and the corresponding agricultural quantities. Moreover, the farming type is derived from partial and total SGMs, according to the most dominating activity principle. Briefly, the most dominating activity is the one whose partial SGM represents more than 66% of the total SGM.

The procedure used to compute the total SGM and the farming type is presented in Annex 3.

It should be stressed once again that the exact derivation of the total SGM and farming type is possible only if the standard system of coefficients is available and if all the involved variables are registered at the most detailed level. If one of these two conditions is not fulfilled, the SGMs and the farming type could not be computed. An alternative is to guarantee that the total SGM and the farming type are registered in the MFR. This constraint is due to the fact that the data utility is highly correlated to the existence of the farming type variable (A06). In other words, the data utility of an MFR without any SGM or farming type indication would be significantly diminished.

The two strategies:

(i) release the total SGM and the farming type in the MFR

(ii) possibility to compute the total SGM and the farming type, if the standard system of coefficients were available

should be equally ranked from a data utility point of view. Instead, from a risk of re-identification point of view, the first one seems preferable. Indeed, the first strategy enables the possibility to aggregate the most detailed variables in order to reduce the risk of re-identification. For example, instead of releasing *Goats - breeding females* (J10A), *Goats - others* (J10B) and *Goats* (J10), the sole release of J10 = J10A + J10B could be sufficient without a significant loss of information. In this example, the information loss due to this aggregation of variables is proportional to (depends on) the standard coefficients difference between variables J10A and J10B. At this detailed level, the regional differences in coefficients are not generally significant. Even if this statement cannot be easily generalized, it is reasonable to assume that, for the majority of regions, the differences in standard coefficients between such detailed variables (belonging to the same group of variables) are not significant.

For the Italian FSS2005, the first strategy was chosen. Based on the system of standard coefficients provided by INEA, the total SGM and the farming type of each holding were computed and registered in the microdata file to be released.

If some variables were aggregated, the most detailed farms subdivisions (70 categories) could no more be exactly derived because **the system of standard coefficients is not released together** with the MFR. Needless to say, the *particular types* of farms (50 categories) preserve their meaning, but their identification power is lower. If the researcher would use the aggregated variables together with some (average) coefficients, only approximations of the total SGM and *particular types* of farming could be derived.
For the Italian FSS2005, several variables were aggregated in order to reduce the re-identification risk. The exact derivation of the 70 farming type *subdivisions* is no more possible. The aggregations performed on the Italian FSS2005 microdata file are presented in table 4. At the same time, the variable A06 was recoded in order to give information on *the particular types* (50 categories) of farms.

Table 4: *ItalianFSS2005. Variable aggregation - part I.*

Suppressed Variables	Released Variables	Relationship
D09E1, D09F, D09G	*D09*	*D09 = D09E1 + D09F + D09G*
D18A, D18B, D18B1, D18B23	*D18*	*D18 = D18A + D18B + D18B1 + D18B23*
F01, F02	*F*	*F = F01 + F02*
G01A, G01B, G01C	*G01*	*G01 = G01A + G01B + G01C*
G03A, G03B	*G03*	*G03 = G03A + G03B*
G04C, G04D	*G04A, G04B, G04*	*G04 = G04A + G04B + G04C + G04D*
I08AD22, I08B, I08C, I08D, I08E	*I08*	*I08 = I08AD22 + I08B + I08C + I08D + I08E*
J09A, J09B	*J09*	*J09 = J09A + J09B*
J10A, J10B	*J10*	*J10 = J10A + J10B*
J16A, J16B, J16C, J16D	*J16*	*J16 = J16A + J16B + J16C+J16D*

It is important to notice that the variables D14A and D14B cannot be summed up because they are used for the definitions of different *particular types* of farming.
From the analysis of table 4 it may be observed that only few subdivisions of farming type could no more be derived. Other aggregations of variables should/could be performed in order to limit the computation of all subdivisions. These further aggregations are presented in table 5, as they were applied to the Italian FSS2005. These aggregations are also justified by the high values of the missing values, as shown in the table 3.

Table 5: *ItalianFSS2005. Variable aggregation - part II.*

Suppressed Variables	Released Variables	Relationship
D01, D02, D03, D04, D05, D06, D08	***D01***	***D01*** *= D01 + D02 + D03+ D04 + D05 + D06 + D08*
D10, D11, D12	***D10***	***D10*** *= D10 + D11 + D12*
D23, D24, D25	***D23***	***D23*** *= D23 + D24 + D25*
D26, D27, D28, D29, D30	***D26***	***D26*** *= D26 + D27 + D28 + D29 + D30*
D31, D32, D33	***D31***	***D31*** *= D31 + D32 + D33*

Only the total SGM and the *particular types* of farming are registered in the MFR.

3.2.3 Perturbation of Continuous Variables
Up to this point, the statistical disclosure limitation methodology consisted in variable suppression and variable aggregation. In a business framework, like the FSS, it is widely recognized that the continuous variables should receive a special attention. Generally, the skew distributions of the continuous economic variables might allow a straightforward re-identification of the most dominant farms/enterprises. The most dominant farms are also the famous farms.
As already stated in section 3.1.3, the agricultural phenomenon under study is highly visible and extremely sparse, not only from a geographical point of view. This means that each continuous variable

might contribute to the re-identification of (some) statistical units. Consequently, a perturbation of the continuous variables was deemed necessary. Only the variables registered on an already aggregated form were not perturbed. These variables are A13, A14, A15, A16, A17, A18 and the variables whose names start with letter "K" or "L".
Obviously, the *Extrapolation factor* (A09) was not perturbed. Moreover, the variables *Standard Gross Margin* (A12) and *FarmType* (A06) were not perturbed in order to maximize the data utility. A06 is not exactly a continuous variable, but since it is derived from a continuous variable, it is important to remind here that the original farming type (*particular types*) is registered in the MFR.
In the SDC literature, there are many protection methods. When selecting a perturbation method, the possible usages of the MFR should be considered, too. As it was observed in section 3.1.2, the FSS data is generally used for descriptive statistics. The individual ranking (IR) is one of the simplest perturbation methods and it might be modified in order to preserve the weighted means (generating descriptive statistics). That's why it was decided to apply the individual ranking for the perturbation of the continuous variables, except for the variables are A13, A14, A15, A16, A17, A18 and the variables whose names start with letter "K" or "L".

Individual ranking (IR)
The IR depends on an a-priori defined parameter k. For each univariate variable V taking the values $v_1, v_2, \ldots, v_n$ on n statistical units, the IR proceeds in the following manner:
(i) The values $v_1, v_2, \ldots, v_n$ are sorted in increasing order. The sorting might be performed even with respect to another variable, but this option was not used in the current application. Denote by $v_{(1)}, v_{(2)}, \ldots, v_{(n)}$ the sorted values.
(ii) The sorted values $v_{(1)}, v_{(2)}, \ldots, v_{(n)}$ are classified in G groups satisfying the following conditions:
(a) each group contains at least k values
(b) the values within a group are as similar as possible. In practice, the similarity is generally measured by means of the Euclidean distance between units.
(iii) For each group $g = 1, \ldots, G$, the mean is computed:

$$m_g = \frac{\sum v_{(j)}^g}{|g|}$$

where $v_{(j)}^g$ denotes a value belonging to the g-th group while $|g|$ denotes the number of values belonging to the g-th group.
(iv) The released values $v^*_{(1)}, v^*_{(2)}, \ldots, v^*_{(n)}$ are computed by replacing each original value by the mean of the group to which the unit belongs to:

$$v^*_{(i)} = m_{g_i}$$

where g_i denotes the group to which the i-th ordered unit belongs to.
(v) The values are ordered with respect to their initial ranks [unsort the first step].
An IR illustration is presented in figure 1.

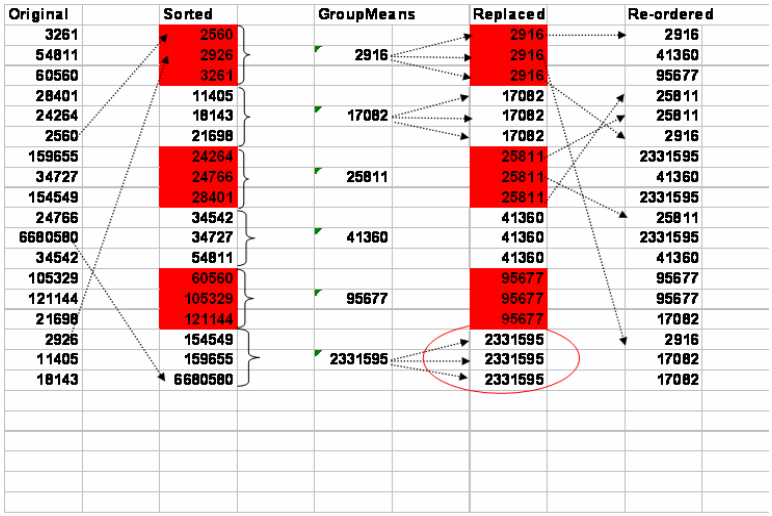

Figure 1: *Individual ranking; k = 3.*

The IR procedure was applied to the Italian FSS2005 continuous variables with the following additional enhancements:

(i) The number of units/values in a group was considered to be exactly k. The only exception to this rule was the first (sorted) group which might contain up to $2k-1$ values. Due to the sparsity of the continuous variables registered in the Italian FSS2005, the smallest values were (always) equal to zero. Hence all the values belonging to the first group were (always) equal to zero.

(ii) The group mean was replaced by the weighted mean, i.e.:

$$m_g = \frac{\sum v_{(j)}^g w_{(j)}^g}{|g|}$$

where $w_{(j)}^g$ represents the weight , *Extrapolation factor* (A09), of the j-th sorted value $v_{(j)}^g$ in the g-th group.

This modification was performed in order to preserve the weighted means. This feature is illustrated in figure 2.

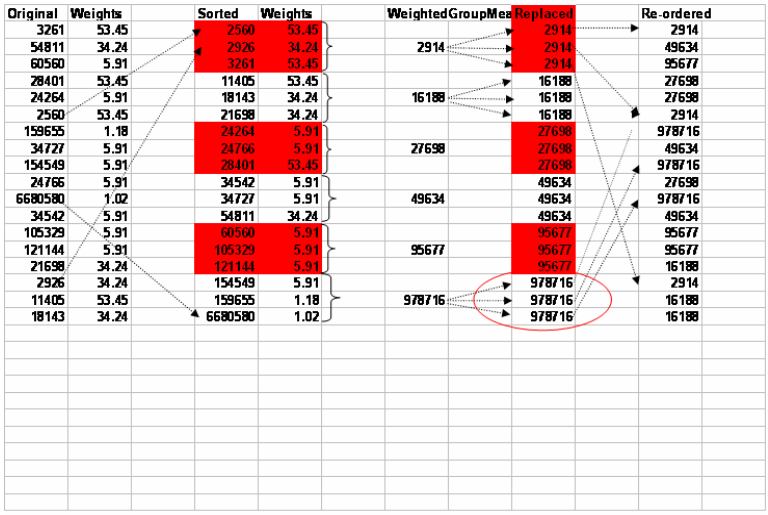

Figure 2: *Weighted individual ranking; k = 3.*

(iii) The weighted IR was applied using the blocking variable A07. This means that the weighted IR was applied independently on each block defined by the categories of the variable A07, i.e. the NUTS2 categories. This choice is justified by the regional characteristics of the

agricultural phenomena. Moreover, the weighted means are preserved at NUTS2 level, in full agreement with the already published tables. This feature is illustrated in figure 3.

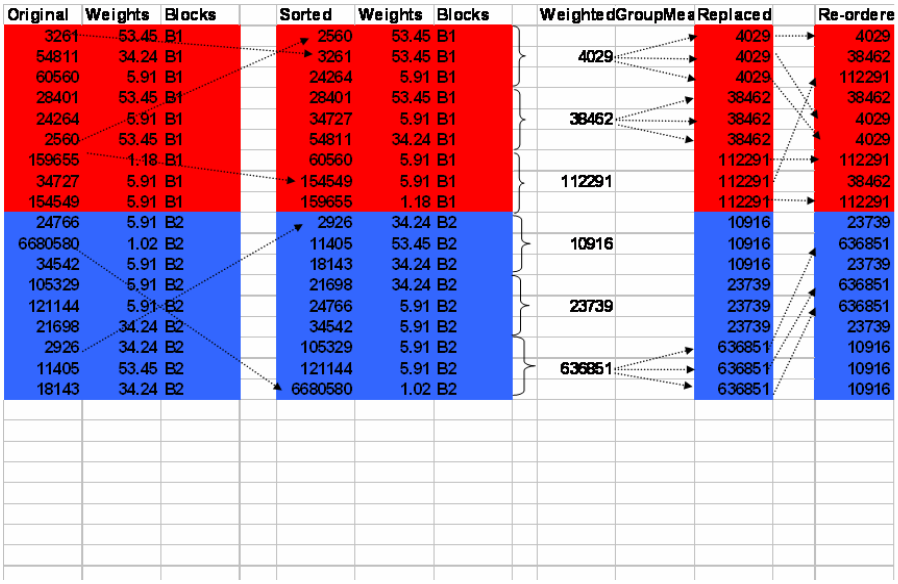

Figure 3: *Weighted regional individual ranking; k = 3.*

(iv) A further criteria was used when grouping the values. It should be reminded that a zero value means that the corresponding phenomenon is missing. When averaging together, see equation (3), non-zero and zero values, the mean (average) would obviously be different from zero. It follows that a zero value (missing phenomenon) might be replaced by a non-zero value (existing phenomenon). Consequently, an artificial phenomenon would be created where it might not really exist.
In order to control and reduce this drawback, when non-zero values were to be averaged with some zero values, the latter were searched for within the farms sharing the same farming type with the units corresponding to the non-zero values. In this way, the artificial phenomenon is created only within some farms which have similar agricultural characteristics with the original farm. When it was not possible to find candidate farms with exactly the same farming type, the immediately superior farming type level was used. For the Italian FSS2005, over the 121 cases, for 19 farms the most similar farms were searched for in the immediately superior farming type category. When several candidate farms were found, the ones corresponding to the lowest *Extrapolation factor* (A09) were selected.

(v) The weighted IR was not applied using both A06 and A07 as blocking variables. For the Italian FSS2005, it was observed that among the 918 combinations of A06 and A07, in 13% of cases, some non-zero values would be averaged with some zero values. This percentage was considered too high: the IR application using A06 and A07 as blocking variables could produce a significant information loss.

(vi) k = 3 was deemed a realistic value, in agreement with the frequency rule generally used at Istat.

3.3 Measures of Information Loss

When applying whatever statistical disclosure limitation methodology, some information loss is unavoidable; otherwise it wouldn't be possible to guarantee the confidentiality of the respondents. In this section, the impact of the protection method on some statistics is evaluated by means of some comparison with the statistics computed on original data.

3.2.1 Regional Weighted Means

As already stated, the IR was applied in order to preserve, for each numerical variable, the regional (NUTS2) weighted means. This was considered the most important data utility constraint because these weighted means are the most used statistical tools. Moreover, the released tables generally contain only information at national level.

3.2.2 Weighted Means by Region and General Type of Farming

The quartiles of the weighted means variations were computed. Namely, for each numerical variable and for each combination of A07 and *general types* of farming (A06), the weighted means were computed both for the original and perturbed data. These quantities were denoted WM_{orig} and WM_{pert}. Then, the percentage variation was computed as $v = \frac{WM_{orig} - WM_{pert}}{WM_{orig}} * 100$. Finally, in Table 6, for each numerical variable, the quartiles of v over the 186 combinations of A07 and *general types* of farming (A06) are shown. The skew distributions are a consequence of the sparsity of the represented phenomenon. Several variables are were deleted from Table 6 because of their extremely high rate of zero values.

Table 6: *The quartiles of the weighted means variations over the combinations of A07 and A06. Q1 = first quartile, Q2 = second quartile (median) and Q3 = third quartile.*

Variable	Q1	Q2	Q3	Variable	Q1	Q2	Q3
A11	-0.07	0.00	0.08	*G02*	-1.48	0.04	0.84
A13	0.00	0.00	0.00	*G03*	-0.20	-0.02	0.18
A14	0.00	0.00	0.00	*G04*	-0.26	0.00	0.21
A15	0.00	0.00	0.00	*G04A*	-0.65	0.03	0.87
A16	0.00	0.00	0.00	*G04B*	-0.45	0.00	0.35
A17	0.00	0.00	0.00	*G05*	-5.59	0.01	4.42
A18	0.00	0.00	0.00	*G06*	-18.26	1.19	13.89
CC01	-0.12	-0.01	0.07	*G07*	-6.62	-0.01	0.15
CC02	-0.28	0.00	0.25	*H01*	-0.79	0.00	0.66
CC03	-1.87	0.01	1.55	*H02*	-0.41	0.00	0.32
CC05A	-1.79	0.07	1.28	*H03*	-0.17	0.00	0.15
CC05D	-14.91	-0.27	15.31	*I01*	-1.47	0.00	1.31
D01	-0.10	-0.01	0.08	*I03A*	-0.24	-0.01	0.16
D07	-0.92	-0.01	1.02	*I03B*	-0.27	0.00	0.20
D09	-2.35	-0.17	1.88	*I08*	-0.63	0.01	0.55
D10	-1.40	0.00	1.36	*J01*	-0.71	0.00	0.57
D14	-0.85	0.00	0.90	*J02*	-0.19	0.00	0.10
D14A	-1.04	-0.12	0.35	*J03*	-0.76	0.00	1.19
D14B	-6.95	-0.03	7.04	*J04*	-0.45	0.00	0.31
D15	-3.35	-0.06	2.09	*J05*	-0.48	0.00	1.10
D16	-1.28	0.00	7.96	*J06*	-0.75	0.00	0.71
D17	-13.50	-0.04	3.87	*J07*	-0.85	0.00	0.26
D18	-0.30	-0.01	0.14	*J08*	-0.49	0.01	1.06
D19	-10.19	-0.23	11.50	*J09*	-0.50	-0.01	0.49
D21	-1.02	-0.01	1.24	*J10*	-0.63	0.00	0.51
D23	-7.31	-0.12	0.32	*J11*	-8.94	-0.06	5.56
D26	-1.09	-0.02	0.84	*J12*	-1.85	0.00	3.24
D31	-17.70	0.00	53.40	*J13*	-1.41	0.00	0.39
D34	-78.06	0.08	10.80	*J14*	-6.35	0.00	1.17
D35	-52.00	0.00	2.28	*J15*	-4.84	0.00	1.55
E	-0.09	0.00	0.04	*J16*	-24.69	-1.58	1.69
F	-0.32	0.00	0.20	*J17*	-11.30	0.00	3.25
G01	-0.75	-0.04	0.43	*J18*	-11.94	-0.36	4.32

3.2.2 Variances by Region and General Type of Farming

The quartiles of the variances variations were computed. For each numerical variable and for each combination of A07 and general types of farming (A06), the variances were computed both for the original and perturbed data. These quantities were denoted V_{orig} and V_{pert}. Then, the percentage variation was computed as $v = \frac{V_{orig} - V_{pert}}{V_{orig}} * 100$. Finally, in table 7, for each numerical variable, the quartiles of v over the 186 combinations of A07 and *general types* of farming (A06) are shown. The skew distributions are a consequence of the sparsity of the represented phenomenon. Several variables are were deleted from Table 7 because of their extremely high rate of zero values.

Table 7: *The quartiles of the variances variations over the combinations of A07 and A06. Q1 = first quartile, Q2 = second quartile (median) and Q3 = third quartile.*

Variable	Q1	Q2	Q3	Variable	Q1	Q2	Q3
A11	-1.11	0.00	1.67	*G02*	-5.63	0.00	3.26
A13	0.00	0.00	0.00	*G03*	-4.84	0.00	1.99
A14	0.00	0.00	0.00	*G04*	-1.98	0.00	2.78
A15	0.00	0.00	0.00	*G04A*	-2.53	0.10	6.78
A16	0.00	0.00	0.00	*G04B*	-5.55	0.00	3.47
A17	0.00	0.00	0.00	*G05*	-12.95	1.80	19.87
A18	0.00	0.00	0.00	*G06*	-57.24	5.18	35.90
CC01	-1.06	0.07	1.92	*G07*	-77.76	-7.10	46.46
CC02	-1.69	0.13	2.75	*H01*	-4.03	0.16	7.39
CC03	-9.11	0.18	8.91	*H02*	-3.04	0.25	5.12
CC05A	-6.40	0.38	6.12	*H03*	-4.39	0.00	2.93
CC05D	-45.09	2.95	52.85	*I01*	-7.30	0.32	9.54
D01	-1.78	-0.15	1.01	*I03A*	-1.35	0.02	2.41
D07	-1.54	0.51	10.04	*I03B*	-1.79	0.01	2.30
D09	-7.86	-0.36	12.23	*I08*	-4.42	0.27	4.06
D10	-3.90	0.70	6.83	*J01*	-9.04	0.00	8.95
D14	-4.37	0.09	6.89	*J02*	-0.81	0.00	2.28
D14A	-4.52	0.00	5.38	*J03*	-4.64	0.00	9.97
D14B	-10.34	0.49	23.38	*J04*	-2.43	0.00	3.25
D15	-9.86	0.00	13.25	*J05*	-3.99	0.00	13.05
D16	-11.11	1.12	36.45	*J06*	-2.43	0.00	5.76
D17	-28.82	2.29	18.17	*J07*	-1.35	0.00	4.07
D18	-4.18	-0.02	1.96	*J08*	-2.59	0.47	8.89
D19	-36.15	-0.24	33.86	*J09*	-2.67	0.00	3.64
D21	-5.96	0.25	8.11	*J10*	-2.98	0.00	7.56
D23	-18.03	-1.50	43.98	*J11*	-29.96	0.10	23.69
D26	-2.26	0.00	3.12	*J12*	-9.62	0.00	18.49
D31	-171.16	70.61	90.59	*J13*	-14.95	0.00	3.26
D34	-98.81	7.66	57.41	*J14*	-25.72	0.00	12.21
D35	-152.35	-109.14	84.67	*J15*	-12.48	-0.13	8.91
E	-0.38	0.00	0.39	*J16*	-95.04	-3.57	8.76
F	-1.73	0.32	2.88	*J17*	-42.57	0.00	22.31
G01	-5.53	-0.02	3.40	*J18*	-45.92	-4.78	25.45

3.2.3 Modified Zero Values

For the Farm Structure Survey 2005, the zero values represent a missing phenomenon. By applying the individual ranking, some zero values might be transformed in non-zero values. In this way, an artificial phenomenon could be created. In Table 8, for each numerical variable, the number *N* of zero values that were transformed in non-zero values is shown. It should be observed that the greatest values of *N* correspond to variables with the largest variations of the weighted means/variances.

Table 8: *Number N of modified zero values*

Variabile	N	Variabile	N
A11	0	*G02*	2
A13	0	*G03*	3
A14	0	*G04*	0
A15	0	*G04A*	0
A16	0	*G04B*	0
A17	0	*G05*	3
A18	0	*G06*	0
CC01	0	*G07*	13
CC02	0	*H01*	0
CC03	0	*H02*	0
CC05A	0	*H03*	0
CC05D	7	*I01*	0
D01	0	*I03A*	0
D07	2	*I03B*	0
D09	2	*I08*	4
D10	0	*J01*	0
D14	0	*J02*	0
D14A	0	*J03*	0
D14B	0	*J04*	0
D15	1	*J05*	0
D16	7	*J06*	0
D17	5	*J07*	0
D18	0	*J08*	0
D19	6	*J09*	0
D21	1	*J10*	0
D23	7	*J11*	2
D26	4	*J12*	2
D31	6	*J13*	0
D34	11	*J14*	3
D35	16	*J15*	0
E	0	*J16*	2
F	0	*J17*	0
G01	0	*J18*	6

3.2.4 Individual Perturbations

The impact of the perturbation method on the information loss might be evaluated using the individual perturbations, too. In Table 9, for each numerical variable, the median of the individual perturbations is presented.

Table 9: *Median of the individual perturbations*

Variable	Median	Variable	Median
A11	-0.001	*G04A*	-0.002
CC01	-0.001	*G04B*	-0.001
CC02	-0.001	*G05*	0.002
CC03	-0.001	*G06*	0.002
CC05A	0.001	*G07*	-0.028
CC05D	-0.030	*H01*	-0.001
D01	-0.001	*H02*	-0.001
D07	0.017	*H03*	-0.001
D09	0.001	*I01*	0.001
D10	0.001	*I03A*	-0.001
D14	-0.001	*I03B*	-0.001
D14A	-0.001	*I08*	0.001
D14B	-0.001	*J01*	-1.000
D15	-0.001	*J02*	-1.000
D16	-0.002	*J03*	-1.000
D17	0.001	*J04*	-1.000
D18	0.001	*J05*	-1.000
D19	-0.002	*J06*	-1.000
D21	-0.001	*J07*	-1.000
D23	-0.001	*J08*	1.000
D26	-0.001	*J09*	1.000
D31	-0.075	*J10*	-1.000
D34	-0.013	*J11*	1.000
D35	-0.303	*J12*	1.000
E	-0.001	*J13*	-1.000
F	0.001	*J14*	-1.000
G01	-0.001	*J15*	-1.000
G02	-0.001	*J16*	-1.000
G03	-0.001	*J17*	-1.000
G04	-0.001	*J18*	-1.000

4. Glossary

In this section a brief description of several terms used in the FSS is given. The full and official European FSS glossary may be found in the Commission Regulation (EC) No 1444/2002.

• **Agricultural holding** A holding is defined as a technical-economic unit under single management engaged in agricultural production.

• **Standard Gross Margin** For each activity on a holding, or: farm, (e.g. wheat, dairy cow or vineyard), a standard gross margin (SGM) is estimated, based on the area (or the number of heads) and a regional coefficient. The sum of all margins, for all activities of a given farm, is referred to as the economic size of that farm. The economic size is expressed in European Size Units (ESU), 1 ESU being equal to 1200 Euro of SGM.

• **Type of farming** In the community typology, each holding is classified by its economic size and its type of farming. The type of farming is determined on the basis of the relative importance of the individual activities carried out by a given farm. For instance, a farm where breeding sows account for more than 2/3 of the economic size is classified as specialist pig rearing (type 5011). Depending on the level of aggregation, farms are grouped into 9 to 70 types.

• **Utilised Agricultural Area (UAA)** The utilised agricultural area (UAA) is the total of arable land, permanent pasture and meadows, land used for permanent crops and kitchen gardens. The UAA excludes unutilised agricultural land, woodland and land occupied by buildings, farmyards, tracks, ponds, etc.

• **Less-favoured areas** The less-favoured areas (LFA) include mountain areas, in which farming is necessary to protect the countryside, particularly for reasons of protection against erosion. The LFA also include areas where the maintenance of a minimum population or the conservation of the countryside are not assured (OJ L142 of 2.6.97).

• **Farm work** Farm work is every type of work on the holding which contributes to the production of the agricultural products1, to the maintenance of the holding's storage capacities of these products, or to activities directly derived from these productive actions. Organisation and management of the farm (buying and selling, accounting, etc.) are also included. Work for the private household of the holder(s)/partners or the manager(s) and their families is not considered as farm work on the holding. Any forestry, hunting, fishery or fish farming operation are excluded from farm work on the holding. (A limited amount of such work carried out by an agricultural worker could be, however, included if it is impossible to measure it separately.)

• **Farm labour force** The farm labour force includes all persons having completed their compulsory education (having reached school-leaving age) who carried out farm work on the holding under survey during the 12 months up to the survey day. The figures include the holders, even when not working on the holding, whereas their spouses are accounted only if they carry out farm work on the holding.

• **Regular labour force** A category based on the availability of the labour force for the holding; it covers family labour force (even those who were working accidentally on the holding) and permanently employed (regular) non-family workers.

• **Holder and manager** The holder is the natural person (sole holder or group of individuals) or the legal person (e.g. a cooperative, an institution) on whose account and in whose name the holding is operated and who is legally and economically responsible for the holding, i.e. who takes the economic risks of the holding. On the group holdings, only the main holder (one person) is accounted.

The manager is the natural person responsible for the normal daily financial and production routines of running the holding concerned. The manager is generally, but not always, the same person as the holder who is a natural person.

• **Family labour force** Family labour force accounts the holder and the members of his family who carried out farm work (including all persons of retiring age who continue to work on the holding).

• **Non family regular labour force** Paid employees (employed by the holding) who carried out farm work every week during the 12 months preceding the survey (irrespective the length of the working week). Persons who worked only for part of that period are also considered as the regularly employed in case it was caused by: special conditions of production, holidays/sickness, commencement/cessation of employment (i.e. those who changed employment during a year), stoppage of agricultural activity of the holding due to accidental causes (flooding, fire etc.).

• **Non family non regular labour force** Short term seasonal workers (employed by the holding), e.g. labour engaged solely as fruit or vegetable pickers. In this case only the time worked is recorded (and thus, information available only in terms of annual work units). Taking into account a considerable degree of part-time work in agriculture and opportunities for part-time work in other sectors of the economy - information on employment in agriculture is given also in annual work units.
• **Work Unit (AWU)** An Annual Work Unit (AWU) is equivalent to fulltime employment. 1 AWU corresponds to the work performed by a person undertaking fulltime agricultural work on the holding over a 12 month period. The yearly working time of such a worker is 1800 hours (225 working days of 8 hours per day), unless national provisions governing contracts of employment are specified.

5. References

Greco M. (2007), *Struttura e produzioni delle aziende agricole (SPA),* available at http://www.istat.it/dati/dataset/20070613_00/.
Hundepool A. *et al.* (2009) *Handbook on Statistical Disclosure Control*, available at http://neon.vb.cbs.nl/casc.

6. Abbreviations

ISTAT	Italian National Statistical Institute
MFR	Microdata file for research
FSS	Farm Structure Survey
VariableName(VariableCode)	a generic variable called VariableName whose code in the FSS is VariableCode
SGM	Standard Gross Margin (A12)
AWU	Average working unit
SDC	Statistical disclosure control
IR	Individual ranking
INEA	Istituto Nazionale di Economia Agraria
NUTS2	Geographical location

7. Annex 1 – FSS2005 Observed Variables

Field	Label/Definition	Unitcode	Type
A01	Survey reference year	YearSrv	Code
A03	Agricultural areas with environmental restrictions	y_n	Code
A04A	Survey District NUTS Code	NutsCodeX	Code
A04D	Municipality code for objective zones 2000	NUTS5V	Code
A05	Area status	Status	Code
A06	Farmtype	FarmType	Code
A07	SGM region code	NutsCodeX	Code
A08	Holding identification number	ID	Positive Integer
A09	Extrapolation factor 1	Number	Positive Real
A09A	Stratum identification number	ID	Positive Integer
A10	Extrapolation factor 2	Number	Positive Real
A11	Utilised agricultural area	ha	Positive Real
A12	SGM of the holding	ECU	Positive Real
A13	Farm work of the holder	AWU	Positive Real
A14	Farm work of the manager of non-group holding	AWU	Positive Real
A15	Farm work of the spouse of the sole holder	AWU	Positive Real
A16	Farm work of family members of the sole holder	AWU	Positive Real
A17	Farm work non-family members regularly employed	AWU	Positive Real
A18	Farm work non-family members non-regularly employed	AWU	Positive Real
B0102	Legal personality of the holding	holdingtype	Code
B03	Agricultural training of the single manager	Traintype	Code
CC01	Agricultural area utilised for farming by owner	ha	Positive Real
CC02	Agricultural area utilised for farming by tenant	ha	Positive Real
CC03	Agricultural area utilised for shared farming or other modes	ha	Positive Real
CC05A	Farming system - Organic farming	ha	Positive Real
CC05D	Farming system - Conversion to organic farming	ha	Positive Real
CC05E	Farming system - Organic farming for animals	tot_part	Code
CC05F1	Benefiting from investment aids - productive investment framework	y_n	Code
CC05F2	Benefiting from investment aids - rural development framework	y_n	Code
CC06A	More than 50% of producion self-consumed by the holder	y_n_z	Code
CC06B	More than 50% of sales are direct sales	y_n	Code
D01	Common wheat and spelt	ha	Positive Real
D02	Durum wheat	ha	Positive Real
D03	Rye	ha	Positive Real
D04	Barley	ha	Positive Real
D05	Oats	ha	Positive Real
D06	Grain maize	ha	Positive Real
D07	Rice	ha	Positive Real
D08	Other cereals	ha	Positive Real
D09	Pulses - total	ha	Positive Real
D09E_1	Peas, field beans and sweet lupines	ha	Positive Real
D09F	Lentils, chick peas and vetches	ha	Positive Real
D09G	Other protein crops harvested dry	ha	Positive Real
D10	Potatoes	ha	Positive Real
D11	Sugar beet	ha	Positive Real
D12	Fodder roots and brassicas	ha	Positive Real

Field	Label/Definition	Unitcode	Type
D14	Fresh vegetables, melons, strawberries - outdoor	ha	Positive Real
D14A	Fresh vegetables, melons, strawberries - outdoor - open field	ha	Positive Real
D14B	Fresh vegetables, melons, strawberries - outdoor - market gardening	ha	Positive Real
D15	Fresh vegetables, melons, strawberries - under glass	ha	Positive Real
D16	Flowers - outdoor	ha	Positive Real
D17	Flowers - under glass	ha	Positive Real
D18	Forage plants - total	ha	Positive Real
D18A	Forage plants - temporary grass	ha	Positive Real
D18B	Forage plants - other green fodder - total	ha	Positive Real
D18B1	Forage plants - other green fodder - green maize	ha	Positive Real
D18B2_3	Other forage plants	ha	Positive Real
D19	Seeds and seedlings	ha	Positive Real
D20	Other crops	ha	Positive Real
D21	Fallow land without subsidies	ha	Positive Real
D23	Tobacco	ha	Positive Real
D24	Hops	ha	Positive Real
D25	Cotton	ha	Positive Real
D26	Rape and turnip	ha	Positive Real
D27	Sunflower	ha	Positive Real
D28	Soya	ha	Positive Real
D29	Linseed (oil flax)	ha	Positive Real
D30	Other oil seed crops	ha	Positive Real
D31	Flax	ha	Positive Real
D32	Hemp	ha	Positive Real
D33	Other textile crops	ha	Positive Real
D34	Aromatic, medicinal and culinary plants	ha	Positive Real
D35	Industrial plants not mentioned elsewhere	ha	Positive Real
E	Kitchen gardens	ha	Positive Real
F	Permanent grassland and meadow - total	ha	Positive Real
F01	Permanent grassland and meadow - pasture and meadow	ha	Positive Real
F02	Permanent grassland and meadow - rough grazings	ha	Positive Real
G01	Fruit and berry plantations - total	ha	Positive Real
G01A	Fruit and berry plantations - temperate climate	ha	Positive Real
G01B	Fruit and berry plantations - subtropical climate	ha	Positive Real
G01C	Fruit and berry plantations - nuts	ha	Positive Real
G02	Citrus plantations	ha	Positive Real
G03	Olive plantations - total	ha	Positive Real
G03A	Olive plantations - table olives	ha	Positive Real
G03B	Olive plantations - oil production	ha	Positive Real
G04	Vineyards - total	ha	Positive Real
G04A	Vineyards - quality wine	ha	Positive Real
G04B	Vineyards - other wines	ha	Positive Real
G04C	Vineyards - table grapes	ha	Positive Real
G04D	Vineyards - raisins	ha	Positive Real
G05	Nurseries	ha	Positive Real
G06	Other permanent crops	ha	Positive Real
G07	Permanent crops under glass	ha	Positive Real
H01	Unutilised agricultural land	ha	Positive Real
H02	Wooded area	ha	Positive Real
H03	Other land	ha	Positive Real
I01	Successive secondary crops - total	ha	Positive Real
I02	Mushrooms	ha	Positive Real
I03A	Total irrigable area	ha	Positive Real

Field	Label/Definition	Unitcode	Type
I03B	Irrigated once a year - Total	ha	Positive Real
I08	Set-aside areas under incentive schemes - total	ha	Positive Real
I08AD22	Set-aside areas under incentive schemes - fallow land with no economic use	ha	Positive Real
I08B	Set-aside areas under incentive schemes - raw material for non-food purposes	ha	Positive Real
I08C	Set aside areas under incentive schemes - converted into permanent pasture	ha	Positive Real
I08D	Set aside areas under incentive schemes - converted into wooded areas	ha	Positive Real
I08E	Set aside areas under incentive schemes - others	ha	Positive Real
J01	Equidae	heads	Positive Integer
J02	Bovine under one year old - total	heads	Positive Integer
J03	Bovine under 2 years - males	heads	Positive Integer
J04	Bovine under 2 years - females	heads	Positive Integer
J05	Bovine 2 years and older - males	heads	Positive Integer
J06	Heifers, 2 years and older	heads	Positive Integer
J07	Dairy cows	heads	Positive Integer
J08	Bovine 2 years old and over - other cows	heads	Positive Integer
J09	Sheep - total	heads	Positive Integer
J09A	Sheep - breeding females	heads	Positive Integer
J09B	Sheep - others	heads	Positive Integer
J10	Goats	heads	Positive Integer
J10A	Goats - breeding females	heads	Positive Integer
J10B	Goats - others	heads	Positive Integer
J11	Pigs - piglets under 20 kg	heads	Positive Integer
J12	Pigs - breeding sows over 50 kg	heads	Positive Integer
J13	Pigs - others	heads	Positive Integer
J14	Poultry - broilers	heads	Positive Integer
J15	Laying hens	heads	Positive Integer
J16	Poultry - others	heads	Positive Integer
J16A	Turkeys	heads	Positive Integer
J16B	Ducks	heads	Positive Integer
J16C	Geese	heads	Positive Integer
J16D	Other poultry	heads	Positive Integer
J17	Rabbits (breeding females)	heads	Positive Integer
J18	Beehives	hive	Positive Integer

Field	Label/Definition	Unitcode	Type
J19	Other livestock	y/n	Code
K01A	Tractors etc.belonging exclusively to the holding	mach	Positive Integer
K01A0	Tractors etc.< 40 kw belonging exclusively to the holding	mach	Positive Integer
K01A3	Tractors etc.40 - <60 kw belonging exclusively to the holding	mach	Positive Integer
K01A41	Tractors etc. 60 - <100 kw belonging exclusively to the holding	mach	Positive Integer
K01A42	Tractors etc. >=100 kw belonging exclusively to the holding	mach	Positive Integer
K02A	Cultivators etc. belonging exclusively to the holding	mach	Positive Integer
K03A	Combine harvesters belonging exclusively to the holding	mach	Positive Integer
K09A	Other fully mechanised harvesters belonging exclusively to the holding	mach	Positive Integer
K10	Irrigation equipment	Irrig	Code
K01B	Tractors etc. used in last 12 mths not belonging to the holding	y_n	Code
K02B	Cultivators etc. used in last 12 mths not belonging to the holding	y_n	Code
K03B	Combine harvesters used in last 12 mths not belonging to the holding	y_n	Code
K09B	Other fully mechanised harvesters used in last 12 mths not belonging to the holding	y_n	Code
L011	Holder: Sex	SexHold	Code
L012	Holder: Age group	AgeCodeH	Code
L013	Holder: Working hours % code	WorkCodeH	Code
L01A1	Manager of non-group holding: Sex	SexMan	Code
L01A2	Manager of non-group holding: Age group	AgeCodeM	Code
L01A3	Manager of non-group holding: Working hours % code	WorkCodeM	Code
L021	Spouse of sole holder: Sex	SexSpouse	Code
L022	Spouse of sole holder: Age group	AgeCodeS	Code
L023	Spouse of sole holder: Working hours % code	WorkCodeS	Code
L03A101	Number:Other members of sole holders'family: male ,worktime >0-<25% AWU, age <25 yrs	pers	Positive Integer
L03A102_3	No. of family males: worktime >0-<25% AWU, age 25-34 yrs	pers	Positive Integer
L03A104_5	No. of family males: worktime >0-<25% AWU, age 35-44 yrs	pers	Positive Integer
L03A106_7	No. of family males: worktime >0-<25% AWU, age 45-54 yrs	pers	Positive Integer
L03A108_9	No. of family males: worktime >0-<25% AWU, age 55-64 yrs	pers	Positive Integer
L03A110	No. of family males: worktime >0-<25% AWU, age >= 65 yrs	pers	Positive Integer
L03A1T	Total no. of family males: worktime >0-<25% AWU	pers	Positive Integer
L03A201	No. of family males: worktime 25-<50% AWU, age < 25 yrs	pers	Positive Integer
L03A202_3	No. of family males: worktime 25-<50% AWU, age 25-34 yrs	pers	Positive Integer
L03A204_5	No. of family males: worktime 25-<50% AWU, age 35-44 yrs	pers	Positive Integer
L03A206_7	No. of family males: worktime 25-<50% AWU, age 45-54 yrs	pers	Positive Integer
L03A208_9	No. of family males: worktime 25-<50% AWU, age 55-64 yrs	pers	Positive Integer

Field	Label/Definition	Unitcode	Type
L03A210	No. of family males: worktime 25-<50% AWU, age >= 65 yrs	pers	Positive Integer
L03A2T	Total no. of family males: worktime 25-<50% AWU	pers	Positive Integer
L03A301	No. of family males: worktime 50-<75% AWU, age < 25 yrs	pers	Positive Integer
L03A302_3	No. of family males: worktime 50-<75% AWU, age 25-34 yrs	pers	Positive Integer
L03A304_5	No. of family males: worktime 50-<75% AWU, age 35-44 yrs	pers	Positive Integer
L03A306_7	No. of family males: worktime 50-<75% AWU, age 45-54 yrs	pers	Positive Integer
L03A308_9	No. of family males: worktime 50-<75% AWU, age 55-64 yrs	pers	Positive Integer
L03A310	No. of family males: worktime 50-<75% AWU, age >= 65 yrs	pers	Positive Integer
L03A3T	Total no. of family males: worktime 50-<75% AWU	pers	Positive Integer
L03A401	No. of family males: worktime 75-<100% AWU, age < 25 yrs	pers	Positive Integer
L03A402_3	No. of family males: worktime 75-<100% AWU, age 25-34 yrs	pers	Positive Integer
L03A404_5	No. of family males: worktime 75-<100% AWU, age 35-44 yrs	pers	Positive Integer
L03A406_7	No. of family males: worktime 75-<100% AWU, age 45-54 yrs	pers	Positive Integer
L03A408_9	No. of family males: worktime 75-<100% AWU, age 55-64 yrs	pers	Positive Integer
L03A410	No. of family males: worktime 75-<100% AWU, age >= 65 yrs	pers	Positive Integer
L03A4T	Total no. of family males: worktime 75-<100% AWU	pers	Positive Integer
L03A501	No. of family males: working full time, age < 25 yrs	pers	Positive Integer
L03A502_3	No. of family males: working full time, age 25-34 yrs	pers	Positive Integer
L03A504_5	No. of family males: working full time, age 35-44 yrs	pers	Positive Integer
L03A506_7	No. of family males: working full time, age 45-54 yrs	pers	Positive Integer
L03A508_9	No. of family males: working full time, age 55-64 yrs	pers	Positive Integer
L03A510	No. of family males: working full time, age >= 65 yrs	pers	Positive Integer
L03A5T	Total no. of family males: working full time	pers	Positive Integer
L03B101	No. of family females: worktime >0-<25% AWU, age < 25 yrs	pers	Positive Integer
L03B102_3	No. of family females: worktime >0-<25% AWU, age 25-34 yrs	pers	Positive Integer
L03B104_5	No. of family females: worktime >0-<25% AWU, age 35-44 yrs	pers	Positive Integer
L03B106_7	No. of family females: worktime >0-<25% AWU, age 45-54 yrs	pers	Positive Integer
L03B108_9	No. of family females: worktime >0-<25% AWU, age 55-64 yrs	pers	Positive Integer
L03B110	No. of family females: worktime >0-<25% AWU, age >= 65 yrs	pers	Positive Integer
L03B1T	Total no. of family females: worktime >0-<25% AWU	pers	Positive Integer

Field	Label/Definition	Unitcode	Type
L03B201	No. of family females: worktime 25-<50% AWU, age < 25 yrs	pers	Positive Integer
L03B202_3	No. of family females: worktime 25-<50% AWU, age 25-34 yrs	pers	Positive Integer
L03B204_5	No. of family females: worktime 25-<50% AWU, age 35-44 yrs	pers	Positive Integer
L03B206_7	No. of family females: worktime 25-<50% AWU, age 45-54 yrs	pers	Positive Integer
L03B208_9	No. of family females: worktime 25-<50% AWU, age 55-64 yrs	pers	Positive Integer
L03B210	No. of family females: worktime 25-<50% AWU, age >= 65 yrs	pers	Positive Integer
L03B2T	Total no. of family females: worktime 25-<50% AWU	pers	Positive Integer
L03B301	No. of family females: worktime 50-<75% AWU, age < 25 yrs	pers	Positive Integer
L03B302_3	No. of family females: worktime 50-<75% AWU, age 25-34 yrs	pers	Positive Integer
L03B304_5	No. of family females: worktime 50-<75% AWU, age 35-44 yrs	pers	Positive Integer
L03B306_7	No. of family females: worktime 50-<75% AWU, age 45-54 yrs	pers	Positive Integer
L03B308_9	No. of family females: worktime 50-<75% AWU, age 55-64 yrs	pers	Positive Integer
L03B310	No. of family females: worktime 50-<75% AWU, age >= 65 yrs	pers	Positive Integer
L03B3T	Total no. of family females: worktime 50-<75% AWU	pers	Positive Integer
L03B401	No. of family females: worktime 75-<100% AWU, age < 25 yrs	pers	Positive Integer
L03B402_3	No. of family females: worktime 75-<100% AWU, age 25-34 yrs	pers	Positive Integer
L03B404_5	No. of family females: worktime 75-<100% AWU, age 35-44 yrs	pers	Positive Integer
L03B406_7	No. of family females: worktime 75-<100% AWU, age 45-54 yrs	pers	Positive Integer
L03B408_9	No. of family females: worktime 75-<100% AWU, age 55-64 yrs	pers	Positive Integer
L03B410	No. of family females: worktime 75-<100% AWU, age >= 65 yrs	pers	Positive Integer
L03B4T	Total no. of family females: worktime 75-<100% AWU	pers	Positive Integer
L03B501	No. of family females: working full time, age < 25 yrs	pers	Positive Integer
L03B502_3	No. of family females: working full time, age 25-34 yrs	pers	Positive Integer
L03B504_5	No. of family females: working full time, age 35-44 yrs	pers	Positive Integer
L03B506_7	No. of family females: working full time, age 45-54 yrs	pers	Positive Integer
L03B508_9	No. of family females: working full time, age 55-64 yrs	pers	Positive Integer
L03B510	No. of family females: working full time, age >= 65 yrs	pers	Positive Integer
L03B5T	Total no. of family females: working full time	pers	Positive Integer
L03C1T	Total no. of family males and females: worktime >0-<25% AWU	pers	Positive Integer
L03C2T	Total no. of family males and females: worktime 25-<50% AWU	pers	Positive Integer

Field	Label/Definition	Unitcode	Type
L03C3T	Total no. of family males and females: worktime 50-<75% AWU	pers	Positive Integer
L03C4T	Total no. of family males and females: worktime 75-<100% AWU	pers	Positive Integer
L03C5T	Total no. of family males and females: working full time	pers	Positive Integer
L04A101	No. non-family males: worktime >0-<25% AWU, age < 25 yrs	pers	Positive Integer
L04A102_3	No. non-family males: worktime >0-<25% AWU, age 25-34 yrs	pers	Positive Integer
L04A104_5	No. non-family males: worktime >0-<25% AWU, age 35-44 yrs	pers	Positive Integer
L04A106_7	No. non-family males: worktime >0-<25% AWU, age 45-54 yrs	pers	Positive Integer
L04A108_9	No. non-family males: worktime >0-<25% AWU, age 55-64 yrs	pers	Positive Integer
L04A110	No. non-family males: worktime >0-<25% AWU, age >= 65 yrs	pers	Positive Integer
L04A1T	Total no. non-family males: worktime >0-<25% AWU	pers	Positive Integer
L04A201	No. non-family males: worktime 25-<50% AWU, age < 25 yrs	pers	Positive Integer
L04A202_3	No. non-family males: worktime 25-<50% AWU, age 25-34 yrs	pers	Positive Integer
L04A204_5	No. non-family males: worktime 25-<50% AWU, age 35-44 yrs	pers	Positive Integer
L04A206_7	No. non-family males: worktime 25-<50% AWU, age 45-54 yrs	pers	Positive Integer
L04A208_9	No. non-family males: worktime 25-<50% AWU, age 55-64 yrs	pers	Positive Integer
L04A210	No. non-family males: worktime 25-<50% AWU, age >= 65 yrs	pers	Positive Integer
L04A2T	Total no. non-family males: worktime 25-<50% AWU	pers	Positive Integer
L04A301	No. non-family males: worktime 50-<75% AWU, age < 25 yrs	pers	Positive Integer
L04A302_3	No. non-family males: worktime 50-<75% AWU, age 25-34 yrs	pers	Positive Integer
L04A304_5	No. non-family males: worktime 50-<75% AWU, age 35-44 yrs	pers	Positive Integer
L04A306_7	No. non-family males: worktime 50-<75% AWU, age 45-54 yrs	pers	Positive Integer
L04A308_9	No. non-family males: worktime 50-<75% AWU, age 55-64 yrs	pers	Positive Integer
L04A310	No. non-family males: worktime 50-<75% AWU, age >= 65 yrs	pers	Positive Integer
L04A3T	Total no. non-family males: worktime 50-<75% AWU	pers	Positive Integer
L04A401	No. non-family males: worktime 75-<100% AWU, age < 25 yrs	pers	Positive Integer
L04A402_3	No. non-family males: worktime 75-<100% AWU, age 25-34 yrs	pers	Positive Integer
L04A404_5	No. non-family males: worktime 75-<100% AWU, age 35-44 yrs	pers	Positive Integer
L04A406_7	No. non-family males: worktime 75-<100% AWU, age 45-54 yrs	pers	Positive Integer
L04A408_9	No. non-family males: worktime 75-<100% AWU, age 55-64 yrs	pers	Positive Integer
L04A410	No. non-family males: worktime 75-<100% AWU, age >= 65 yrs	pers	Positive Integer

Field	Label/Definition	Unitcode	Type
L04A4T	Total no. non-family males: worktime 75-<100% AWU	pers	Positive Integer
L04A501	No. non-family males: working full time, age < 25 yrs	pers	Positive Integer
L04A502_3	No. non-family males: working full time, age 25-34 yrs	pers	Positive Integer
L04A504_5	No. non-family males: working full time, age 35-44 yrs	pers	Positive Integer
L04A506_7	No. non-family males: working full time, age 45-54 yrs	pers	Positive Integer
L04A508_9	No. non-family males: working full time, age 55-64 yrs	pers	Positive Integer
L04A510	No. non-family males: working full time, age >= 65 yrs	pers	Positive Integer
L04A5T	Total no. non-family males: working full time	pers	Positive Integer
L04B101	No. non-family females: worktime >0-<25% AWU, age < 25 yrs	pers	Positive Integer
L04B102_3	No. non-family females: worktime >0-<25% AWU, age 25-34 yrs	pers	Positive Integer
L04B104_5	No. non-family females: worktime >0-<25% AWU, age 35-44 yrs	pers	Positive Integer
L04B106_7	No. non-family females: worktime >0-<25% AWU, age 45-54 yrs	pers	Positive Integer
L04B108_9	No. non-family females: worktime >0-<25% AWU, age 55-64 yrs	pers	Positive Integer
L04B110	No. non-family females: worktime >0-<25% AWU, age >= 65 yrs	pers	Positive Integer
L04B1T	Total no. non-family females: worktime >0-<25% AWU	pers	Positive Integer
L04B201	No. non-family females: worktime 25-<50% AWU, age < 25 yrs	pers	Positive Integer
L04B202_3	No. non-family females: worktime 25-<50% AWU, age 25-34 yrs	pers	Positive Integer
L04B204_5	No. non-family females: worktime 25-<50% AWU, age 35-44 yrs	pers	Positive Integer
L04B206_7	No. non-family females: worktime 25-<50% AWU, age 45-54 yrs	pers	Positive Integer
L04B208_9	No. non-family females: worktime 25-<50% AWU, age 55-59 yrs	pers	Positive Integer
L04B210	No. non-family females: worktime 25-<50% AWU, age >= 65 yrs	pers	Positive Integer
L04B2T	Total no. non-family females: worktime 25-<50% AWU	pers	Positive Integer
L04B301	No. non-family females: worktime 50-<75% AWU, age < 25 yrs	pers	Positive Integer
L04B302_3	No. non-family females: worktime 50-<75% AWU, age 25-34 yrs	pers	Positive Integer
L04B304_5	No. non-family females: worktime 50-<75% AWU, age 35-44 yrs	pers	Positive Integer
L04B306_7	No. non-family females: worktime 50-<75% AWU, age 45-54 yrs	pers	Positive Integer
L04B308_9	No. non-family females: worktime 50-<75% AWU, age 55-64 yrs	pers	Positive Integer
L04B310	No. non-family females: worktime 50-<75% AWU, age >= 65 yrs	pers	Positive Integer
L04B3T	Total no. non-family females: worktime 50-<75% AWU	pers	Positive Integer
L04B401	No. non-family females: worktime 75-<100% AWU, age < 25 yrs	pers	Positive Integer

Field	Label/Definition	Unitcode	Type
L04B402_3	No. non-family females: worktime 75-<100% AWU, age 25-34 yrs	pers	Positive Integer
L04B404_5	No. non-family females: worktime 75-<100% AWU, age 35-44 yrs	pers	Positive Integer
L04B406_7	No. non-family females: worktime 75-<100% AWU, age 45-54 yrs	pers	Positive Integer
L04B408_9	No. non-family females: worktime 75-<100% AWU, age 55-64 yrs	pers	Positive Integer
L04B410	No. non-family females: worktime 75-<100% AWU, age >= 65 yrs	pers	Positive Integer
L04B4T	Total no. non-family females: worktime 75-<100% AWU	pers	Positive Integer
L04B501	No. non-family females: working full time, age < 25 yrs	pers	Positive Integer
L04B502_3	No. non-family females: working full time, age 25-34 yrs	pers	Positive Integer
L04B504_5	No. non-family females: working full time, age 35-44 yrs	pers	Positive Integer
L04B506_7	No. non-family females: working full time, age 45-54 yrs	pers	Positive Integer
L04B508_9	No. non-family females: working full time, age 55-64 yrs	pers	Positive Integer
L04B510	No. non-family females: working full time, age >= 65 yrs	pers	Positive Integer
L04B5T	Total no. non-family females: working full time	pers	Positive Integer
L04C1T	Total no. non-family males and females: worktime >0-<25% AWU	pers	Positive Integer
L04C2T	Total no. non-family males and females: worktime 25-<50% AWU	pers	Positive Integer
L04C3T	Total no. non-family males and females: worktime 50-<75% AWU	pers	Positive Integer
L04C4T	Total no. non-family males and females: worktime 75-<100% AWU	pers	Positive Integer
L04C5T	Total no. non-family males and females: working full time	pers	Positive Integer
L0506	No. of working days for non-family males and females working on non-regular basis	days	Positive Integer
L07	Other gainful activity of holder-manager	OGA	Code
L08	Other gainful activity of spouse of sole holder	OGAS	Code
L09A	No.of other family members of sole holder with other major gainful activity	pers	Positive Integer
L09B	No.of other family members of sole holder with other subsidiary gainful activity	pers	Positive Integer
L10	No. of working days by contractors	days	Positive Integer
M01A	Other gainful activity: tourism	y_n	Code
M01B	Other gainful activity: handicraft	y_n	Code
M01C	Other gainful activity: processing of farm products	y_n	Code
M01D	Other gainful activity: wood processing	y_n	Code
M01E	Other gainful activity: aqua culture	y_n	Code
M01F	Other gainful activity: renewable energy production	y_n	Code
M01G	Other gainful activity: contractual work	y_n	Code
M01H	Other gainful activity: others	y_n	Code

8. Annex 2 – FSS2005 List of Codes

Field Id	Label	Codelist	Value	Value Label
A03	Agricultural areas with environmental restrictions	y_n	n	No
			y	Yes
A05	Area status	Status	l	Less-favoured non-mountainous area
			m	Less-favoured mountainous area
			n	Normal
B0102	Legal personality of the holding	holdingtype	1	Sole holder is also the manager
			2a	Holder's spouse is the manager of the holding.
			2b	Manager is a member of the holder's family but not his spouse.
			3	Holder is not the manager who is not a member of the holder's family
			4	Holding is a legal person
			5	Holding is a group holding
B03	Agricultural training of the single manager	Traintype	b	Basic training
			f	Full agricultural training
			p	Practical experience only
CC05E	Farming system - Organic farming for animals	tot_part	n	Not at all
			p	Partly
			t	Totally
CC05F1	Benefiting from investment aids - productive investment framework	y_n	n	No
			y	Yes
CC05F2	Benefiting from investment aids - rural development framework	y_n	n	No
			y	Yes
CC06A	More than 50% of producion self-consumed by the holder	y_n_z	n	no
			y	yes
			z	Holding is legal entity or group holding
CC06B	More than 50% of sales are direct sales	y_n	n	No
			y	Yes
J19	Other livestock	y_n	n	No
			y	Yes
K10	Irrigation equipment	Irrig	0	no irrigation equipment
			b	fixed and mobile irrigation equipment
			f	fixed irrigation equipment
			i	Irrigation equipment, but not known if fixed or mobile
			m	mobile irrigation equipment
K01B	Tractors etc. used in last 12 mths not belonging to the holding	y_n	n	No
			y	Yes
K02B	Cultivators etc. used in last 12 mths not belonging to the holding	y_n	n	No
			y	Yes
K03B	Combine harvesters used in last 12 mths not belonging to the holding	y_n	n	No
			y	Yes
K09B	Other fully mechanised harvesters used in last 12 mths not belonging to the holding	y_n	n	No
			y	Yes
L011	Holder: Sex	SexHold	f	Holder is female
			m	Holder is male
			z	Holding is a legal person

Field Id	Label	Codelist	Value	Value Label
L012	Holder: Age group	AgeCodeH	24	Holder < 25 years
			34	Holder 25 - 34 years
			44	Holder 35 - 44 years
			54	Holder 45 - 54 years
			64	Holder 55 - 64 years
			99	Holder >= 65 years
			z	Holding is a legal person
L013	Holder: Working hours % code	WorkCodeH	0	Holder doesn't work on the holding
			100	Worktime holder 100% of 1 AWU
			24	Worktime holder >0 - < 25% of 1 AWU
			49	Worktime holder 25 - < 50% of 1 AWU
			74	Worktime holder 50 - < 75% of 1 AWU
			99	Worktime holder 75 - < 100% of 1 AWU
			z	Holding is a legal person
L01A1	Manager of non-group holding: Sex	SexMan	f	Manager is female
			m	Manager is male
L01A2	Manager of non-group holding: Age group	AgeCodeM	24	Manager < 25 years
			34	Manager 25 - 34 years
			44	Manager 35 - 44 years
			54	Manager 45 - 54 years
			64	Manager 55 - 64 years
			99	Manager >= 65 years
L01A3	Manager of non-group holding: Working hours % code	WorkCodeM	100	Worktime manager 100% of 1 AWU
			24	Worktime manager >0 - < 25% of 1 AWU
			49	Worktime manager 25 - < 50% of 1 AWU
			74	Worktime manager 50 - < 75% of 1 AWU
			99	Worktime manager 75 - < 100% of 1 AWU
L021	Spouse of sole holder: Sex	SexSpouse	f	Spouse of the holder is female
			m	Spouse of the holder is male
			x	No spouse
			z	Holding is a legal person or group holding
L022	Spouse of sole holder: Age group	AgeCodeS	24	Spouse < 25 years
			34	Spouse 25 - 34 years
			44	Spouse 35 - 44 years
			54	Spouse 45 - 54 years
			64	Spouse 55 - 64 years
			99	Spouse >= 65 years
			x	No spouse
			z	Holding is a legal person or group holding
L023	Spouse of sole holder: Working hours % code	WorkCodeS	0	Worktime spouse 0% AWU
			100	Worktime spouse 100% of 1 AWU
			24	Worktime spouse >0 - < 25% of 1 AWU
			49	Worktime spouse 25 - < 50% of 1 AWU
			74	Worktime spouse 50 - < 75% of 1 AWU
			99	Worktime spouse 75 - < 100% of 1 AWU
			x	No spouse
			z	Holding is a legal person or group holding

Field Id	Label	Codelist	Value	Value Label
L07	Other gainful activity of holder-manager	OGA	m	major other gainful activity of the holder-manager
			n	No other gainful activity of the holder-manager
			s	subsidiary other gainful activity of the holder-manager
			x	Sole holder is not the manager
			z	Holding is a legal person
L08	Other gainful activity of spouse of sole holder	OGAS	m	major other gainful activity of the spouse
			n	No other gainful activity of the spouse
			s	subsidiary other gainful activity of the spouse
			x	no spouse
			z	Holding is a legal person or group holding
M01A	Other gainful activity: tourism	y_n	n	No
			y	Yes
M01B	Other gainful activity: handicraft	y_n	n	No
			y	Yes
M01C	Other gainful activity: processing of farm products	y_n	n	No
			y	Yes
M01D	Other gainful activity: wood processing	y_n	n	No
			y	Yes
M01E	Other gainful activity: aqua culture	y_n	n	No
			y	Yes
M01F	Other gainful activity: renewable energy production	y_n	n	No
			y	Yes
M01G	Other gainful activity: contractual work	y_n	n	No
			y	Yes
M01H	Other gainful activity: others	y_n	n	No
			y	Yes

9. Annex 3 – SGM Computation

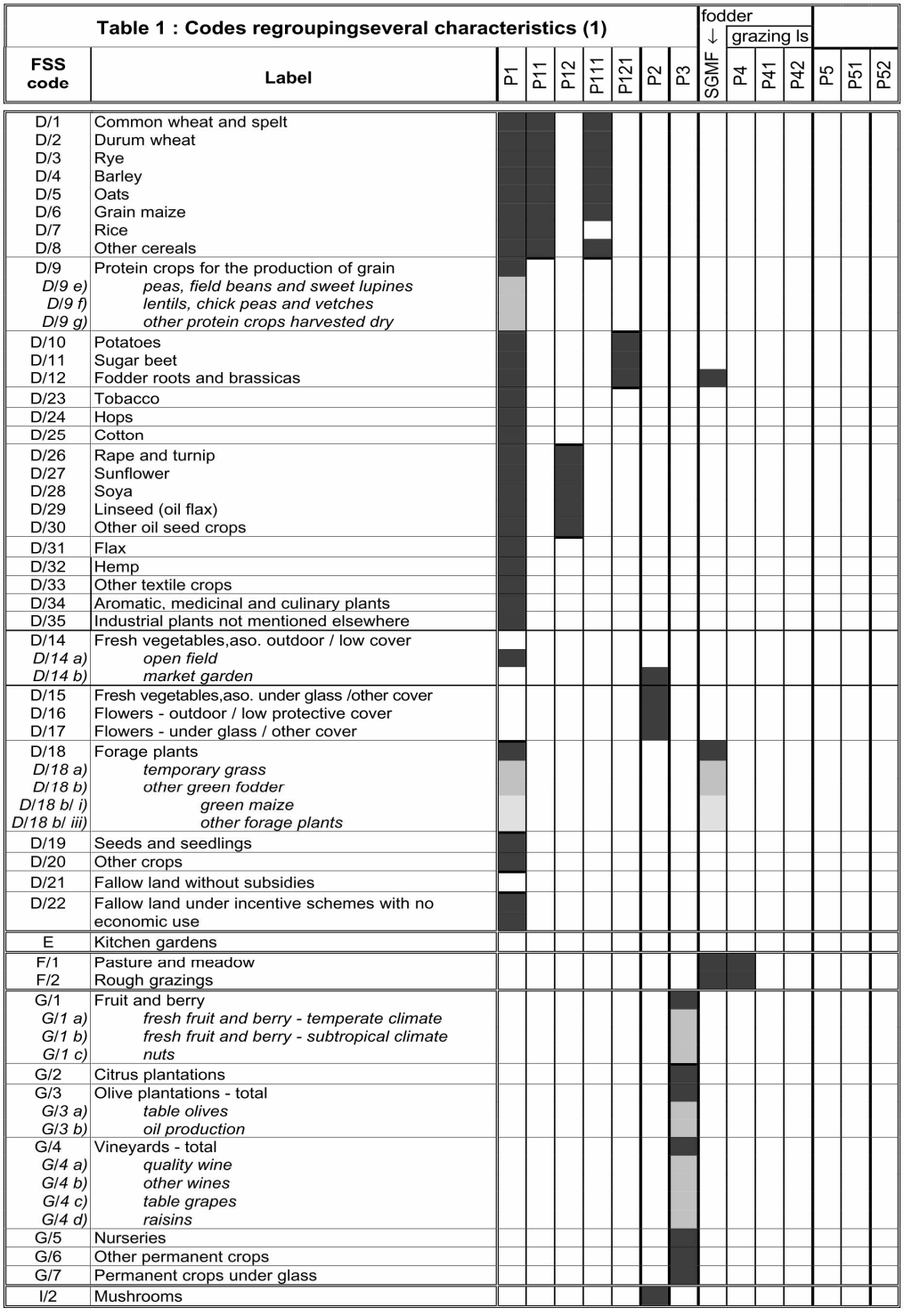

Table 1 : Codes regroupingseveral characteristics (1)									fodder ↓	grazing ls					
FSS code	Label	P1	P11	P12	P111	P121	P2	P3	SGMF	P4	P41	P42	P5	P51	P52
D/1	Common wheat and spelt														
D/2	Durum wheat														
D/3	Rye														
D/4	Barley														
D/5	Oats														
D/6	Grain maize														
D/7	Rice														
D/8	Other cereals														
D/9	Protein crops for the production of grain														
D/9 e)	*peas, field beans and sweet lupines*														
D/9 f)	*lentils, chick peas and vetches*														
D/9 g)	*other protein crops harvested dry*														
D/10	Potatoes														
D/11	Sugar beet														
D/12	Fodder roots and brassicas														
D/23	Tobacco														
D/24	Hops														
D/25	Cotton														
D/26	Rape and turnip														
D/27	Sunflower														
D/28	Soya														
D/29	Linseed (oil flax)														
D/30	Other oil seed crops														
D/31	Flax														
D/32	Hemp														
D/33	Other textile crops														
D/34	Aromatic, medicinal and culinary plants														
D/35	Industrial plants not mentioned elsewhere														
D/14	Fresh vegetables,aso. outdoor / low cover														
D/14 a)	*open field*														
D/14 b)	*market garden*														
D/15	Fresh vegetables,aso. under glass /other cover														
D/16	Flowers - outdoor / low protective cover														
D/17	Flowers - under glass / other cover														
D/18	Forage plants														
D/18 a)	*temporary grass*														
D/18 b)	*other green fodder*														
D/18 b/ i)	*green maize*														
D/18 b/ iii)	*other forage plants*														
D/19	Seeds and seedlings														
D/20	Other crops														
D/21	Fallow land without subsidies														
D/22	Fallow land under incentive schemes with no economic use														
E	Kitchen gardens														
F/1	Pasture and meadow														
F/2	Rough grazings														
G/1	Fruit and berry														
G/1 a)	*fresh fruit and berry - temperate climate*														
G/1 b)	*fresh fruit and berry - subtropical climate*														
G/1 c)	*nuts*														
G/2	Citrus plantations														
G/3	Olive plantations - total														
G/3 a)	*table olives*														
G/3 b)	*oil production*														
G/4	Vineyards - total														
G/4 a)	*quality wine*														
G/4 b)	*other wines*														
G/4 c)	*table grapes*														
G/4 d)	*raisins*														
G/5	Nurseries														
G/6	Other permanent crops														
G/7	Permanent crops under glass														
I/2	Mushrooms														

Table 1 : Codes regroupingseveral characteristics (2)										fodder ←	grazing ls						
FSS code	**Label**	P1	P11	P12	P111	P121	P2	P3	SGMF	P4	P41	P42	P5	P51	P52
J/1	Equidae														
J/2	Bovine under one year old - total														
J/3	Bovine under 2 years - males														
J/4	Bovine under 2 years - females														
J/5	Bovine 2 years and older - males														
J/6	Heifers, 2 years and older														
J/7	Dairy cows														
J/8	Bovine 2 years old and over - other cows														
J/9	Sheep														
J/9 a)	*breeding ewes*														
J/9 b)	*other sheep*														
J/10	Goats														
J/10 a)	*breeding goats, female*														
J/10 b)	*other goats*														
	Pigs														
J/11	Piglets < 20 kg														
J/12	Breeding sows over 50 kg														
J/13	Other pigs														
	Poultry														
J/14	Broilers														
J/15	Laying hens														
J/16	Poultry - others														
J/16 a)	*turkeys*														
J/16 b)	*ducks*														
J/16 c)	*geese*														
J/16 d)	*other poultry*														
J/17	Rabbits (breeding females)														
J/18	Beehives														

Total SGM of a holding =	**SUM OF**
	P1 **P2** **P3** **P4** **P5**
	D/21 Fallow land without subsidies E Kitchen gardens J/18 Beehives

Specialist holdings – crops

P1	**>**	**2/3**	**Total SGM**	**1**
P11 + P12 + D/9 + D/22	**> 2/3**	**Total SGM**		**13**
P111+P12+D/9+D/22	> 2/3	Total SGM	131	
D/7	> 2/3	Total SGM	132	
Other cases			133	
P11 + P12 + D/9 + D/22	**≤ 2/3**	**Total SGM**		**14**
P121	> 2/3	Total SGM	141	
P11+P12+D/9+D/22 > 1/3 Total SGM; P121 > 1/3 Total SGM			142	
D/14 a)	> 2/3	Total SGM	143	
D/23	> 2/3	Total SGM	1441	
D/25	> 2/3	Total SGM	1442	
Other cases	> 2/3	Total SGM	1443	

P2	**>**	**2/3**	**Total SGM**	**20** **2**
D/14 b) + D/15	**> 2/3**	**Total SGM**		**201**
D/14 b)	> 2/3	Total SGM	2011	
D/15	> 2/3	Total SGM	2012	
Other cases			2013	
D/16 + D/17	**> 2/3**	**Total SGM**		**202**
D/16	> 2/3	Total SGM	2021	
D/17	> 2/3	Total SGM	2022	
Other cases			2023	
D/14 b) + D/15 ≤ 2/3 Total SGM ; D/16 + D/17 ≤ 2/3 Total SGM				**203**
D/14 b) + D/16	> 2/3	Total SGM	2031	
D/15 + D/17	> 2/3	Total SGM	2032	
I/2	> 2/3	Total SGM	2033	
Other cases			2034	

Table 2 : scheme on typology classification (1)

1. Specialist field crops

13. Specialist cereals, oilseeds and protein crops

131. Specialist cereals (other than rice), oilseeds and protein crops
132. Specialist rice
133. Cereals, oilseeds, protein crops and rice combined

14. General field cropping

141. Specialist root crops
142. Cereals and root crops combined
143. Specialist field vegetables
1441. Specialist tobacco
1442. Specialist cotton
1443. Various field crops combined

2. Specialist horticulture

201. Specialist market garden vegetables

2011. Specialist market garden vegetables – outdoor
2012. Specialist market garden vegetables – under glass
2013. Specialist market garden vegetables, outdoor and under glass combined

202. Specialist flowers and ornamentals

2021. Specialist flowers and ornamentals – outdoor
2022. Specialist flowers and ornamentals – under glass
2023. Specialist flowers and ornamentals, outdoor and under glass combined

203. General market garden cropping

2031. General market garden cropping – outdoor
2032. General market garden cropping – under glass
2033. Specialist mushrooms
2034. Various market garden crops combined

Specialist holdings – crops

P3	**>**	**2/3**	**Total SGM**	**3**

G/4		**> 2/3**	**Total SGM**	**31**
	G/4 a)	> 2/3	Total SGM	311
	G/4 b)	> 2/3	Total SGM	312
	[G/4 a) + G/4 b	> 2/3	Total SGM;	
	[G/4 a) < 2/3 Total SGM; G/4 b) < 2/3 Total SGM]			313
	G/4 c)	> 2/3	Total SGM	3141
	G/4 d)	> 2/3	Total SGM	3142
	Other cases			3143
G/1 + G/2		**> 2/3**	**Total SGM**	**32**
	G/1	> 2/3	Total SGM	321
	G/1 a) + G/1 b)	> 2/3	Total SGM	3211
	G/1 c)	> 2/3	Total SGM	3212
	Other cases			3213
	G/2	> 2/3	Total SGM	322
	Other cases			323
G/3		**> 2/3**	**Total SGM**	**33**
Other cases				**34**

Table 2 : scheme on typology classification (2)

3. Specialist permanent crops

31. Specialist vineyards

311. Specialist quality wine
312. Specialist wine other than quality wine

313. Quality and other wine combined
- 3141. Specialist table grapes
- 3142. Specialist raisins
- 3143. Mixed vineyards

32. Specialist fruit and citrus fruit

321. Specialist fruit (other than citrus)
- 3211. Specialist fresh fruit (other than citrus)
- 3212. Specialist nuts
- 3213. Fresh fruit (other than citrus) and nuts combined

322. Specialist citrus fruit
323. Fruit and citrus fruit combined

33. Specialist olives

34. Various permanent crops combined

Specialist holdings – animal production

Table 2 : scheme on typology classification (3)

P4	**>**	**2/3**	**Total SGM**	**4**
P41 > 2/3 Total SGM;		**J/7 > 2/3 P41**		**41**
J/7	> 2/3	Total SGM	411	
Other cases			412	
P42 > 2/3 Total SGM;		**J/7 ≤ 1/10 Total SGM**		**42**
J/8	> 1/3	Total SGM	421	
J/8	≤ 1/3	Total SGM	422	
P42 > 2/3 Total SGM ; J/7 > 1/10 Total SGM ; excluding class 41				**43**
J/7	> 1/4	Total SGM	431	
J/7	≤ 1/4	Total SGM	432	
P42		**≤ 2/3**	**Total SGM**	**44**
J/9	> 2/3	Total SGM	441	
P42 > 1/3 Total SGM ;		J/9 > 1/3 Total SGM	442	
J/10	> 2/3	Total SGM	443	
Other cases			444	

4. Specialist grazing livestock

41. Specialist dairying

411. Specialist milk production
412. Specialist milk production with cattle-rearing

42. Specialist cattle – rearing and fattening

421. Specialist cattle – mainly rearing
422. Specialist cattle – mainly fattening

43. Cattle – dairying, rearing and fattening combined

431. Cattle – dairying with rearing and fattening
432. Cattle – rearing and fattening with dairying

44. Sheep, goats and other grazing livestock

441. Specialist sheep
442. Sheep and cattle combined
443. Specialist goats
444. Various grazing livestock – no dominant enterprise

Specialist holdings – animal production

Table 2 : scheme on typology classification (4)

P5	**>**	**2/3**	**Total SGM**	**50**	**5**
P51		**> 2/3**	**Total SGM**	**501**	
J/12		> 2/3	Total SGM	5011	
J/11 + J/13		> 2/3	Total SGM	5012	
Other cases				5013	
P52		**> 2/3**	**Total SGM**	**502**	
J/15		> 2/3	Total SGM	5021	
J/14 + J/16		> 2/3	Total SGM	5022	
Other cases				5023	
Other cases				**503**	
P51 > 1/3 Total SGM ;		P52 > 1/3 Total SGM		5031	
Other cases				5032	

Mixed holdings

	[1/3 < P1 ≤ 2/3 ;	**P4 ≤ 1/3 ;**	**P5 ≤ 1/3 Total SGM]**	**60**	**6**
OR	**[1/3 < P2 ≤ 2/3 ;**	**P4 ≤ 1/3 ;**	**P5 ≤ 1/3 Total SGM]**		
OR	**[1/3 < P3 ≤ 2/3 ;**	**P4 ≤ 1/3 ;**	**P5 ≤ 1/3 Total SGM]**		
P2 > 1/3 Total SGM ;		**P3 > 1/3 Total SGM**		**601**	
P1 > 1/3 Total SGM ;		**P2 > 1/3 Total SGM**		**602**	
P1 > 1/3 Total SGM ;		**G/4 > 1/3 Total SGM**		**603**	
P1 > 1/3 Total SGM ; P3 > 1/3 Total SGM ; G/4 ≤1/3 Total SGM				**604**	
1/3 < P1 ≤ 2/3 Total SGM ;P2 ≤ 1/3 Total SGM ; P3 ≤1/3 Total SGM				**605**	
P1 ≤ 1/3 Total SGM ;		1/3 < P2 ≤ 2/3 Total SGM			
P3≤ 1/3 Total SGM ; P4 ≤ 1/3 Total SGM; P5 ≤ 1/3 Total SGM;				6061	
Other cases				6062	

5. Specialist granivores

501. Specialist pigs

5011. Specialist pig rearing
5012. Specialist pig fattening
5013. Pig rearing and fattening combined

502. Specialist poultry

5021. Specialist layers
5022. Specialist poultry-meat
5023. Layers and poultry-meat combined

503. Various granivores combined

5031. Pigs and poultry combined
5032. Pigs, poultry and other granivores combined

6. Mixed cropping

601. Market gardening and permanent crops combined
602. Field crops and market gardening combined
603. Field crops and vineyards combined
604. Field crops and permanent crops combined
605. Mixed cropping, mainly field crops

6061. Mixed cropping, mainly market gardening
6062. Mixed cropping, mainly permanent crops

Mixed holdings

Table 2 : scheme on typology classification (5)

[1/3 <P4≤ 2/3; P1≤ 1/3; P2≤ 1/3; P3≤ 1/3 Total SGM] **OR [1/3 <P5≤ 2/3; P1≤ 1/3; P2≤ 1/3; P3≤ 1/3 Total SGM]**	**7**	**7. Mixed livestock holdings**
1/3 < P4 ≤ 2/3 Total SGM ; P5 ≤ 1/3 Total SGM	**71**	**71. Mixed livestock, mainly grazing livestock**
P41 > 1/3 Total SGM ; J/7 > 2/3 P41	711	711. Mixed livestock, mainly dairying
Other cases	712	712. Mixed livestock, mainly grazing livestock other than dairying
1/3 < P5 ≤ 2/3 Total SGM	**72**	**72. Mixed livestock, mainly granivores**
P41 > 1/3 Total SGM ; J/7 > 2/3 P41	721	721. Mixed livestock; granivores and dairying combined
⌈ (P4 > 1/3 Total SGM ; P41 ≤ 1/3 Total SGM)⌉ ⌊OR (P41 > 1/3 Total SGM ; J/7 ≤ 2/3 P41)⌋	722	722. Mixed livestock: granivores and grazing livestock other than dairying combined
Other cases	723	723. Mixed livestock: granivores with various livestock
Total SGM > 0* Other cases	**8**	**8. Mixed crops – livestock**
P1 > 1/3 Total SGM ; P4 > 1/3 Total SGM	**81**	**81. Field crops – grazing livestock combined**
P41 > 1/3 Total SGM ; J/7 > 2/3 P41 ; P41 ≤ P1	811	811. Field crops combined with dairying
⌈ P41 > 1/3 Total SGM ; J/7 > 2/3 P41 ⌉ ⌊ P1 > 1/3 Total SGM ; P41 ≥P1 ⌋	812	812. Dairying combined with field crops
P1 > P4 excluding class 811	813	813. Field crops combined with grazing livestock other than dairying
Other cases	814	814. Grazing livestock other than dairying combined with field crops
Other cases	**82**	**82. Various crops and livestock combined**
P1 > 1/3 Total SGM ; P5 > 1/3 Total SGM	821	821. Field crops and granivores combined
P3 > 1/3 Total SGM ; P4 > 1/3 Total SGM	822	822. Permanent crops and grazing livestock combined
J/18 > 2/3 Total SGM	8231	8231. Apiculture
Other cases	8232	8232. Various mixed holdings
Total SGM = 0*	**9**	**9. Non-classifiable holdings**

*Following the algorithm defined by the Commission Decision (CEE) No 85/377, all holdings would be classifiable. Indeed class 8 includes the 'Holdings excluded from classes 1 to 7'. But this Decision set also, in the column 'General types of farming', a category '9. Non-classifiable holdings'. In the practice, the holdings whose SGM is 0 or not defined form this category. The only holdings whose SGM is >0 are then classified in the classes 1 to 8.