



EUROPEAN COMMISSION
EUROSTAT

Directorate B: Methodology, corporate statistical and IT services
Unit B-1: Methodology and corporate architecture

ESTAT/B1/WGSC(15)1
Available in EN only

7th meeting of the
Working Group on Statistical Confidentiality

Luxembourg, 30 November - 1 December 2015

Eurostat building BECH, Room Quetelet

Start: 30 November 2015, 13:00

End: 1 December 2015, 16:00

Item 4.5

Confidentiality charter for animal production statistics

CONFIDENTIALITY CHARTER FOR ANIMAL PRODUCTION STATISTICS

1. OBJECTIVE

The objective of this document is, first, to report on the outcome of the discussions conducted by the Working Group on Animal Production Statistics and involving the Expert Group on Statistical Disclosure Control (EGSDC) and, then, to do a step towards making the confidentiality charter applicable.

The Working Group on Statistical Confidentiality is invited to endorse the proposed charter and to give its approval for transmission of the document to the next step in the ESS governance.

2. BACKGROUND

2.1. Background

Treatment of confidential tabular data in agricultural statistics has been mostly a side issue as long as Eurostat has disseminated statistics of national interest with EU total as a by-product. The Member States have a limited interest in producing confidential results as they cannot be published. Importance of milk production for the Common Agricultural Policy (CAP) and concentration in the dairy sector increased the need for EU non-confidential totals based on confidential national data. The basic method for treatment of confidential tabular data, currently used for animal production statistics, propagates the confidentiality status through the statistical tables. Eurostat disseminates thus some tables on milk statistics with almost no EU total. In December 2012, the European Dairy Association, respondents and users of milk statistics, addressed a complaint to the Commission on this matter, which highlights how serious is the concern.

During the last decade, this issue has been raised several times and a draft charter was even proposed in 2009, but interactions with the specific EU needs (national solutions cannot simply be implemented at EU level) and with other quality issues (missing national values also contribute to discard the EU results) made it inapplicable and it remained a draft.

In 2014, an amended charter was submitted to the Working Group on Animal Production Statistics which requested, for assessing the charter, the support of the Expert Group on Statistical Disclosure Control (EGSDC). This charter covers animal production statistics, as confidentiality concerns also slaughterhouses and hatcheries. The recommendations of the confidentiality experts and the comments of the Working Group have been integrated and the charter has been re-shaped. It also takes on board further harmonised internal Eurostat recommendations. The EU totals have been produced and disseminated for few tables after formal approval by the Member States as a pilot exercise and a provisional solution. The Directors Group on Agricultural Statistics (DGAS) approved this charter to the next step in July 2015.

2.2. Presentation of the charter

The charter is structured in five headings and several annexes. The purpose of the annexes is to provide technical information to be agreed between Eurostat and the Member States. The flexibility offered by the three technical annexes guarantees that the charter does not interfere significantly with any change in the legal requirements for animal production statistics. The charter is presented with the standard cases met in the domain.

WGSC(15)1		Summary presentation	
Annex A: Confidentiality Charter			
Main part	1. Introduction		
	2. References	Legislation Glossary Particular concepts Rules implemented	
	3. Rules for coherence (Member States)		
	4. EU totals (Eurostat)		
Annex I	Methods implemented		
Annex II	Parameters		
Annex III	Inventory of business rules		
Annex B: Standard cases – 12 cases (issues, scenarios, options taken)			

Figure 1: structure of the document

The charter is intended to be flexible in the design of the metadata flows: it can be implemented based on the current data collection design, but further exchanges of information on the reasons for confidentiality between the Member States and Eurostat can be developed. Such a change can be progressive, i.e. only some Member States implementing it. The purpose is to meet the assumptions stated in part 4 of the charter model (the same rules must be applied at national level everywhere) after a certain time, making the charter aligned on the Eurostat recommendations.

An additional annex presents the typical cases faced in treatment of confidentiality and discusses the possible solutions. It illustrates the purpose of the confidentiality treatments to the non-experts and the concrete cases met in the domain to the experts.

The rules for treating confidentiality refer to a frequency rule and to a dominance rule. Their parameters are provided in the charter annex and may be amended by the Working Group on Animal Production Statistics after opinion of the EGSDC.

The EU aggregates are calculated following a priority order to limit (1) the afterwards change of the disseminated data due to further enlargements and (2) the risk that the most relevant EU aggregate cannot be published due to successive aggregations. The charter also draws benefit of the missing national values by imputing non-published estimates and by using them as partners for treatment of confidentiality.

During the discussion a particular case was left aside, i.e. the charter does not cover its treatment and the Member States meeting it or suspecting it will report so that investigations can be conducted. It refers to transnational firms represented by various statistical units amongst the Member States and which would share information one with each other (case 8).

3. REQUEST

The members of the Working Group on Statistical Confidentiality are invited to:

- Take note of the annexed Confidentiality Charter
- Approve it in its methodological aspects, further to the approval by the DGAS.
- Provide comments on its application within the ESS, if any

CHARTER FOR THE TREATMENT OF CONFIDENTIALITY IN ANIMAL PRODUCTION STATISTICS

1. INTRODUCTION

One of Eurostat's main duties is to disseminate data aggregated at European Union level. These statistics are necessary for the Commission to develop and monitor European Union (EU) policies. They also meet an increasing demand from users of statistics outside the Commission, national administrations and private enterprises, including the respondents, as European integration progresses. In order for this data to be useful for the users it should be available to the maximum extent possible, while guaranteeing at the same time that no data on individual respondents is disclosed.

The main purpose of this charter is:

- to lay down practical rules in order to guarantee that the confidentiality pattern of data disseminated by Eurostat is consistent with that of the data disseminated at the national level,
- to lay down the confidentiality rules to be applied to the EU aggregates taking into account the practices and laws in the area of confidentiality in the Member States,
- to ensure a proper protection of confidential data and to encourage Member States to report them and
- to facilitate the systematic treatment of confidentiality and thus, to reduce significantly the need to consult Member States each time a change occurs either into the confidentiality pattern or in the size of an EU aggregate.

The charter will enter into force after its adoption by European Statistical System Committee (ESSC).

2. REFERENCES

2.1. Legal references

Regulation (EC) No 223/2009¹ of the European Parliament and of the Council on European Statistics refers to common principles and guidelines ensuring the confidentiality of data used for the production of European Statistics. In particular, detailed rules assuring the protection, the transmission and the access to confidential data are defined in Articles 20 – 26 thereof.

Animal Production Statistics covers especially:

- statistics on milk and milk products (Directive 96/16/EC and Decision 97/80/EC),
- statistics on livestock and meat production (Regulation (EC) No 1165/2008),
- statistics collected under Regulation (EC) 617/2008,

¹ Official Journal No. L87, 31/3/2009, p. 164-173.

or any more recent legislation replacing them, and any further agreement referring to them.

2.2. Glossary

The following concepts are used in the field of statistical disclosure control (see *Manual on the protection of confidential statistical data at Eurostat* and the related *Recommendations for treatment of statistical confidentiality in business data*).

Confidential cells: the cells of a table which have to be protected due to the risk of statistical disclosure (risk of the identification of the statistical unit).

Confidential cluster: the group of statistical units contributing to an aggregate and whose data is confidential for a particular variable.

Confidential EU subtotal: the aggregated data of the group of countries within the EU whose data is confidential for a particular variable.

Confidential statistical data: data are considered confidential when they allow statistical units to be identified, either directly or indirectly, thereby disclosing individual information (Regulation (EC) No 223/2009 Article 3).

Contributor: a statistical unit (e.g. business/enterprise) taking part in an aggregate (e.g. confidential cluster/national total/Community total/EU aggregate).

Direct identification means the identification of a statistical unit from its name, address or from a publicly accessible identification number.

Indirect identification means the identification of a statistical unit by any other means than by way of direct identification. It refers to the possibility of deducing the identity of a statistical unit other than from the direct identification means. To determine whether a statistical unit is identifiable, account shall be taken of all means that might reasonably be used by a third party to identify the said statistical unit.

Dominance occurs when the value of the variable for one or two enterprises for the countries in the EU subtotal exceeds a given percentage of the total value of that variable of the confidential subtotal. This percentage is referred to as the *dominance threshold*.

Primary confidentiality: the identification of the cell whose dissemination would allow disclosure of individual contributor (statistical unit). The two main reasons for data to be primary confidential are *too few* units in a cell or *dominance* of one or two contributors in a cell.

Primary suppression means that the values of primary confidential cells are not shown in the table, but replaced by a symbol such as "x".

Secondary confidentiality: application of SDC methods to some safe cells in order to prevent disclosure of primary confidential cells through recalculation (based on additive feature of tabular data and/or hierarchical structure of statistical data).

Statistical disclosure control methods

Methods to reduce the risk of disclosing information on the statistical units, usually based on restricting the amount of, or modifying, the data released.

Statistical confidentiality shall mean the protection of data related to single statistical unit which are obtained directly for statistical purposes or indirectly from administrative or other sources. It implies the prevention of non-statistical utilisation of the data obtained and unlawful disclosure.

Statistical unit means the basic observation unit, namely a natural person, a household, an economic operator and other undertakings, referred to by the data (Regulation (EC) No 223/2009 Article 3).

Tabular data: aggregate information on entities presented in tables. The tables may have one, two or more dimensions. A cell is defined by reference to these dimensions and provides the value of variables and possibly further information like footnotes or flags.

2.3. Specific concepts used in the Confidentiality Charter for Animal Production Statistics

Aggregates

An aggregate is a group of particulars which are comparable by their nature and which have in common at least one property. This or these properties define the aggregate.

For the purpose of the present charter, aggregates may concern:

- reference areas such as countries or groups of countries for which aggregates are called here **EU-aggregates** regardless whether the aggregate is actually EU or another aggregate of countries. Aggregates of reference areas such as regions within a given Member State are not considered as EU-aggregates, although they are geo-aggregates.
- successive time periods for which aggregates are called here **time-aggregates**;
- other items other than time-aggregates defined at country level are called here **product-aggregates**; aggregates of regions within a given Member State are also covered here; a variable which is derived from variables for an EU-aggregate instead of being derived from the national values is also a **product-aggregate**.

For **additive variables**, the value taken for the aggregate is the sum of values for the particulars. For any other variable, reference to additive variables is a way to estimate its value for the aggregate.

Geo-aggregates concern by definition different statistical units. In this charter, geo-, time- and product- aggregation are conducted independently one from each other.

EU is the acronym used for the changing economic territory of the European Union and of the former European Economic Community. The objective definition of EU has changed with the successive enlargements and therefore it depends on the reference period of the statistics. By agreement and when EU definition changed during the reference period, the EU definition at the end of the reference period is considered.

Supplementary variable for dissemination

A supplementary variable is a product-aggregate or another variable derived from available statistical information and intended to dissemination, further to those under legislation or further agreement. Non-coordinated dissemination of supplementary variables may enable indirect identification.

Information on confidentiality

Available information on confidentiality is information, further to the value of a cell, available for assessing the confidentiality of this cell and of any other cell derived from it. Indication on whether the cell is confidential is the *minimum level of information*. Detailed information including individual information on some statistical units for each cell or series of cells may be required for an accurate assessment of confidentiality.

2.4. General rules applied for treatment of confidentiality

The Charter refers to the following statistical disclosure control rules allowing identification of the confidential cells (primary confidential cells).

2.4.1. *Minimum number of contributors (threshold)*

The number of statistical units contributing to the value of a variable is critical when somebody can estimate confidential information from the published statistics. The risk is the highest for those persons aware of some pieces of information, i.e. the respondents. If only two statistical units contribute to a value, knowing it enables them to calculate easily the contribution of the other one. If three units are concerned, disclosure is possible only if some further conditions are met. The value is confidential as long as the number of contributors is critical. Another expression is that the number of contributors must be over a certain threshold m , over this critical number. Usually the threshold is the minimum value over the critical number of contributors, i.e. m is at least equal to 3.

2.4.2. *Dominance of n statistical units*

If several statistical units have a negligible contribution to the value of a variable, the sub-contribution of the n other statistical units can be reliably estimated. These n statistical units are dominant as soon their contribution is over a percentage k of the total value of the variable. A dominance rule (n, k) checks that the n highest contributions to a value are not over a percentage k . Lower is k and higher is n , lower is the risk.

3. PRACTICAL RULES FOR GUARANTEEING A COHERENT CONFIDENTIALITY PATTERN FOR NATIONAL ANIMAL PRODUCTION DATA DISSEMINATED BY A MEMBER STATE AND BY EUROSTAT

The Member States communicate the data from the animal production surveys in the form of statistical tables and not as micro-data concerning individual units. Only where a statistical cell contains data concerning only one unit, can the information transmitted be considered as micro-data.

The annexes to the Confidentiality Charter may be adapted after agreement between the Member States and Eurostat in the Working Group on Animal Production Statistics.

3.1. Rules for identification of confidential data in the Member State

These rules apply where sufficient information on confidentiality is available, even if the results are not transmitted directly to Eurostat.

3.1.1. *Primary confidential data*

Member States apply different rules for identifying primary confidential data in animal production statistics. Data may be declared confidential for the following reasons:

- They may concern less than the *minimum number of contributors*.
- *Dominance for n statistical units*. This criterion may be examined for a fixed variable, and result in hiding the complete set of variables for the group of units considered. It may also be examined variable by variable, where, for a single group of units, certain variables will be published while others are masked.
- The data are such that the **user or any respondent can estimate the reported value** of some respondent too accurately. Such disclosure occurs, and the data is declared sensitive, if the upper estimate for the respondent's value is closer to the reported value than a pre-specified percentage, p (the p% rule).

3.1.2. *Secondary confidentiality*

The Member State identifies secondary confidential data, with a particular attention to:

- **hierarchy in the classifications**, as the tables may contain several levels of sub-total,
- **consistency of the confidentiality pattern between tables**, as the value of some variables can be recalculated on the basis of variables in different tables. The Working Group on Animal Production Statistics must provide and maintain an inventory of the variables statistically connected and drawn from different tables transmitted to Eurostat.

3.2. **Rules for coordination of the confidentiality treatment between Member States and Eurostat**

The following rules may refer to particular methods explained under Annex I – Methods implemented. In such a case, the reference is indicated between squared brackets.

3.2.1. *Calculations implemented by Eurostat*

The Member State and Eurostat must agree on exchanges of information on confidentiality. Similar agreements may also be sought within the national statistical system if such information is not available in the service providing Eurostat with statistics.

Information exchanged must enable the statistics, even confidential, to be processed so that the results (other statistics, for instance EU aggregates) can be disseminated by Eurostat without risk of disclosure. The most complete dissemination in secure conditions is ensured by rules considering the level of information available in Eurostat for each statistical variable. Freshness of the statistics is important for their quality and therefore availability of information on statistical confidentiality together or before availability of the statistical values is also considered.

With the purpose of providing the users of statistics with relevant information, Eurostat disseminates the results of the following operations, listed in their implementation order:

- Calculate supplementary variables on the national data
- Calculate time aggregates

- Calculate EU-aggregates
 - Impute particular values at EU level [M 10]
 - Sum up national additive variables for the EU-aggregates [M 9],
 - Further calculation of EU-aggregate values which cannot be obtained by the previous operations

The Working Group on Animal Production Statistics must provide and maintain

- an inventory of the supplementary variables disseminated and
- an inventory of the variables which take a particular value.

3.2.2. *Limitation in the scope*

Only EU-aggregates are subject to particular rules for treatment of confidentiality. Time- and product-aggregates and other supplementary variables for dissemination can only be calculated by Eurostat when all underlying data are public: if one or more of the components used for such a calculation is confidential, the result is confidential [M 1].

4. CALCULATION OF THE EU TOTALS

4.1. Rules on confidentiality

Eurostat can compile and publish all the national tables free of any confidential value.

Eurostat can also publish the variables for the EU-aggregates compiled from national confidential values under certain conditions which must all be met:

- The number of statistical units in the confidential cluster is at least the threshold m .
- There is no dominance by n statistical units having a contribution higher than $k\%$ to the confidential cluster, i.e. breaking the rule (n, k) ; for the variables which value can be negative, the absolute value is considered for dominance assessment.
- There is no linear combination of published values leading to a disclosure of confidential result [M 2] and [M 3].
- If the value of the confidential cluster is negligible compared to precision of the EU total (lower than five times the displayed precision), it can be published [M 7].

The Working Group on Animal Production Statistics must provide and maintain a statement with limited access on the values used for m , n , and k . It may be amended by the Working Group on Animal Production Statistics after opinion of the EGSDC.

The following practical rules need to be respected:

- The variables considered are the mandatory variables defined by the legislation or by further agreements between Eurostat and the Member States.
- When a non-confidential value is not provided by a Member State (incomplete data file), it is not published as official statistics anywhere else. In such a case Eurostat may draw benefit of the missing value by imputing an estimate which can be considered as secondary confidential.

- Most of the cases where secondary confidentiality is an issue arise because of additive relations between aggregates. The few other cases (e.g. relationship between tables, non-additive aggregation) must be controlled by pertinent relevant rules (resp. validation rules and propagation of confidentiality status for non-additive aggregation).
- The Member States should not disseminate national data provided to Eurostat as confidential. Revision of the confidentiality pattern might lead to possible disclosure of confidential data for other Member States and should therefore be avoided.

4.2. Implementation based on level available information

Detailed information on confidentiality may be replaced by less accurate assumptions as long as this does not lead to underestimate the risk for disclosure.

The metrics required for confidentiality treatment (number of contributors, dominant contributions, etc.) are estimated *on the worst case* [M 4] based on available information.

The coverage of information on confidentiality considered for processing statistics can be presented at four extents.

- The confidential values are identified with the flag “C” only.
- The confidential values are transmitted and identified with the flag “C”, the number of relevant statistical units is available; this extent may co-exist with the previous case but for some particular tables only.
- The reasons for confidentiality are indicated by a flag and the necessary information on the reasons (e. g. individual units' dominance levels, frequencies) is available at the latest when the statistical values are delivered to Eurostat.
- Further exchanges of information enables coordinated dissemination by the Member State and Eurostat.

Only the three first extents are considered in this Charter and no feedback by Eurostat to the Member States after transmission of the statistical tables is foreseen for treatment of confidentiality.

- **Minimum information on confidentiality**, the confidential cells being delivered with flag "C", whatever is the reason (primary or secondary confidentiality).

The principle “*One cell, one unit*” [M 5] is implemented as there is no additional information available on the reasons for confidentiality. Each confidential value, except true zero, is considered as individual information from one statistical unit.

- **The accurate number of statistical units** is known by size class, further to minimum information on confidentiality (flag "C")

The *worst case* is calculated based on the size class limits where available and on the number of units. Otherwise the worst case refers for each cell to dominance by one unit [M 6].

- **Further exchanges of information** may be agreed by the Member States and Eurostat. The Working Group on Animal Production Statistics may agree on implementing it after assessment of expected gains in efficiency. This may involve some or all the Member States and some or all the tables transmitted by them. Such a change must be covered by the Charter.

Under such an agreement, detailed information describing the confidentiality cases is available in Eurostat, at the latest together with the statistical values (no delay for further exchanges of information on confidentiality) for higher efficiency of the confidentiality treatment. It refers to information on the reasons for confidentiality and on the way confidentiality is treated in the Member State. *Country-adapted rules* [M 8] consider these various sets of parameters.

○ **Flags on reason of confidentiality**

The following confidentiality flags could be included in the data file instead of flag “C” indicating confidentiality:

- flag A= too few enterprises
- flag O= one enterprise dominates the data,
- flag T= two enterprises dominate the data
- flag D= confidential data due to secondary confidentiality

Note: If the minimum number m of statistical units is higher than the number n of units to be checked for dominance and if the number of statistical units is between n and m , the reason for confidentiality is expected to be dominance of one or two enterprises (flag O or T) if there is dominance and “too few enterprises” (flag A) otherwise.

The coding proposed can be adapted for efficiency purpose. The Working Group on Animal Production Statistics will provide the valid list of flags to be used if it is different from the above list and will agree on the way information on confidentiality is to be transmitted.

○ **Information on how confidentiality is treated in the Member State**

The Member State will inform Eurostat in advance on the checks it implements and on the dominance percentage it uses where relevant:

- the dominance percentage for one dominant contributor (flag O),
- the dominance percentage for two dominant contributors (flag T).

Countries using the **p% rule** may use other confidentiality flags than the ones described above after agreement with Eurostat. They may also provide information on the contribution of the n largest contributors.

Some values may be **hidden** by a Member State because of **insufficient quality**. These values can be used as partner for treatment of confidentiality. Nevertheless, Eurostat can increase such values for the EU-aggregates as long as such an inflation does not meet 1% of the results but is sufficient to avoid identification of the confidential values, i.e. is over $(100/k - 1)$ times the highest confidential value in the EU-sub-total. The Member States may request implementing a different k -parameter for assessing an efficient protection of the values.

**CHARTER FOR THE TREATMENT OF CONFIDENTIALITY
IN ANIMAL PRODUCTION STATISTICS**

ANNEX I

Methods implemented

Some methods are described to check the risk of disclosure and/or to estimate the metrics required for assessing confidentiality.

M 1 Flag propagation [/Killer approach]

M 2 Checking linear combinations between rows and columns in a table

M 3 Checking secondary confidentiality within the various EU-aggregates

M 4 Worst case for disclosure (principle)

M 5 “One cell, one unit”

M 6 Worst distribution for distribution by size class

M 7 Protect negligible value by rounding

M 8 Country-adapted rule

M 9 Calculate EU-aggregate with EU priority

M 10 Particular values taken by some aggregates

M 1 Flag propagation [/Killer approach]

The so-called “killer approach” is the secure method used when no more sophisticated method is defined or applicable, for instance for product-aggregation without individual data. Every value derived from a confidential value is confidential. A cascade of derivation propagates the confidentiality status over the datasets. Therefore the process is a method of treatment “by propagation” of the flags. It is over-secured, as the number of published results is low, but simple in the principle and in its implementation.

M 2 Checking linear combinations between rows and columns in a table

Within a 2-dimensional table, each confidential cell should constitute the angle of a closed path (formed only of vertical and horizontal lines) of which each other angle constitutes a confidential cell. If this is not the case, the confidential cell can be recalculated by linear combination and an additional EU total must be hidden.

In practice, the national tables should be checked by the Member State whereas it is up to Eurostat to check it for EU tables, considering the relation between Member State and EU level data as additional dimension of the table. If the data of each Member State have a 2-dimensional structure (i.e. with rows and columns), the structure that has to be considered at EU-level is hence 3-dimensional. In that case the closed path criterion is not applicable. A sufficient (although not always necessary) criterion is then for the confidential cells to constitute a square (or rectangle). All table cells being corner points of the square must be confidential cells. As the square-criterion is not a necessary criterion, mathematical algorithms may establish for a given instance a valid pattern of confidential cells that is not a square.

The same procedure can be applied to each pair of correlated variables drawn from different tables and displayed through the various countries. In this case, some cells may have been already published for the first variable received which threaten confidentiality of the latest values received.

M 3 Checking secondary confidentiality within the various EU-aggregates

The various “squares” drawn within the table [M 2] of an EU aggregate will be possibly propagated as “cubes” amongst the various EU aggregates. However, the only variables which values can be hidden for secondary confidentiality are those drawn from confidential national values, which would have otherwise been publishable.

On the worst case this would result in the same confidentiality scheme as with the *killer approach*. But this is rare.

In practice this means that the variables resulting from national confidential values are marked so that they can possibly be flagged back as confidential, even if their calculation met the other conditions for publication.

The national confidential values are over-protected by this method. A softer method consists in checking that the combination of the EU-aggregate values does not enable to estimate accurately the national confidential values, i.e. that the expected accuracy of such a result is over 40% of their actual value. Of course this check must correctly consider the aggregation structure, given by the additive relations between variables and between Member State and EU-level data.

M 4 Worst case for disclosure (principle)

The worst case for [risk of] disclosure is an imaginary set of individual values created in order to assess the content of confidential cells when the actual individual values are not available. For checking dominance the number of statistical units is minimized, the values are maximized for fewest of them and the other values are minimized.

M 5 “One cell, one unit”

In order to protect efficiently the confidential values, when the number of statistical units contributing to the value of a cell is not known, the worst case for disclosure [M 4] is when only one unit contribute to the value, i.e. it is an individual value. The number of statistical units is considered as one, except if this value is a true zero, in which case there is no statistical unit involved. A confidential cell which value is a true zero is therefore secondary confidential.

M 6 Worst distribution for distribution by size class

A typical set of constraints applied on the *worst case for disclosure* [M 4] is when a variable is also used for defining size classes (minimum = M_{\min} and maximum = M_{\max}) and when the number of units (N) is available. After having generated this virtual distribution, the treatment of confidentiality for individual values can be implemented.

The method described hereafter enables generating all the individual values for the worst distribution for variables defined as positive or equal to zero:

- **One single unit:** the individual value is directly available.
- **More than one unit**

- **Floored semi-open classes**

If N units contribute to class i with a lower size limit P_{min_i} , no upper size limit, and total value being P_i , then the maximal production for a given unit (without knowing more) is reached when the other units have the floor value, i.e. P_{min_i} . The individual extreme value will be the discrepancy between the total for the cell for the class (P_i) and the sum of values for the $(N - 1)$ other units, i.e. $(N - 1) * P_{min_i}$. The maximal value is thus

$P_i - (N - 1) * P_{min_i} = P_i + P_{min_i} - N * P_{min_i}$. This value defines one individual value and $(N - 1)$ other individual values are defined as P_{min_i}

- **Ceiled semi-open classes**

If N units contribute to class i with a upper size limit P_{max_i} , no lower size limit, and total value being P_i , then most of the units can take a value close to zero. The maximum value will be the total value of the variable, if it is under the ceiling. If the total value is over the ceiling, q individual values are at P_{max_i} , q being the result of the Euclidian division of P_i by P_{max_i} . If q is higher than the number of units taken into account for dominance this later value should be used for q . The $(N - q)$ other individual values are thus at $P_i - (q * P_{max_i}) / (N - q)$.

- **Closed classes**

When there are floor P_{min_i} and ceiling P_{max_i} , removing P_{min_i} from all the values will generate individual values in the same way as in the previous case, for a semi-open class with ceiling at $P_{max_i} - P_{min_i}$, and which individual values have to be increased afterwards by P_{min_i} .

For instance, for class 'production <120' with five units which production is 250, the worst case is two values at just less than 120, the three other ones sharing the 10 remaining income. For class 'production >= 10' with the same data the worst case is one unit with 210 and four units with 10. And for class 'production between 10 and less than 120' the worst case is one unit with just less than 120, three with 10 and the other one with 100.

From an operational point of view, the dominance rate D of n units is thus

IF $P_i - (n * P_{max_i}) / (N - n) \leq P_{min_i}$
 THEN $D = 100 * n * P_{max_i} / P_i$
 ELSE $D = 100 - 100 * (N - n) * P_{min_i} / P_i$

M 7 Protect negligible value by rounding

A value can be rounded for hiding its accurate value. In such a case only the rounded value can be considered as publishable. This is especially important if it is used for further calculation. A rounded value is displayed with the accurate number of decimal digits required. Disseminating numbers less precise than the unit requires explicit information on their actual precision. Changing the measurement unit avoids providing such an explanation. If a value is assessed as well protected by (in-) accuracy at 40%, it should be smaller than or equal to 5 times the precision for meeting such an objective.

M 8 Country-adapted rule

Various Member States may have different way for protecting statistical information, as the level of the risk for disclosure may change with the national context, in terms of impact (e.g. sensitivity of a variable) or of likelihood (e.g. numerous small unit *vs.* few dominant units). In the extreme case, an individual variable may be public in a country and highly sensitive in another one. When compiling the figures from these various Member State, confidential clusters can thus require adapted treatment reflecting the diversity of national requirements. The parameters for a rule (e.g. the value of a given threshold) can therefore be modulated depending on the source of the contributing confidential values.

M 9 Calculate EU-aggregate with EU priority

The total of EU for the reference period (e.g. EU-15 in 2000, EU-25 in 2005 and EU-28 in 2014) is calculated first. The smaller EU-aggregates are derived from it by a chain of subtractions ($EU-27 = EU-28 - HR$, $EU-25 = EU-27 - (BG + RO)$, etc.). Similarly the larger EU-aggregates are derived from the EU reference by a chain of additions. The aggregates of second level (e.g. Euro area) need to be checked against two aggregates, i.e. derived from the smallest nesting and from the larger nested EU, the confidentiality of one of the result leading to confidentiality of both. The derivation of an EU-aggregate from another EU-aggregate requires assessing confidentiality for the group of countries specific to the larger of both aggregates.

M 10 Particular values taken by some aggregates

As an exception to the rule for EU-aggregation, the EU total for the relevant intra-Community exchanges is fixed at zero and is not confidential, whatever can be the number of statistical units involved or the national values. Several other particular values can be defined based on similar agreed logical statements.

**CHARTER FOR THE TREATMENT OF CONFIDENTIALITY
IN ANIMAL PRODUCTION STATISTICS**

ANNEX II

Parameters of the rules for confidential treatment

The present annex to the Confidentiality Charter for Animal Production Statistics is intended to be accessed only by the entitled Members of the relevant Working Group and the ESS staff committed to statistical disclosure control for the relevant domains. Its transmission to any other body would threaten protection of confidential information and would therefore be considered as disclosure of confidential information.

Minimum number of units contributing to a non-confidential value m :

Number of units considered for dominance assessment n :

Dominance threshold k (in percent):

Maximum change in dominance threshold introduced by a pseudo-random function (in percent):

The shaded values () will only be communicated at the Working Group on Animal Production Statistics, to the Expert Group on Statistical Confidentiality, and to anybody entitled for this.

**CHARTER FOR THE TREATMENT OF CONFIDENTIALITY
IN ANIMAL PRODUCTION STATISTICS**

ANNEX III

**Inventories of business rules and arrangement
to be coordinated for efficiency of the confidentiality treatment**

III.1 Inventory of the variables statistically connected and drawn from different tables

The following pairs of values are statistically connected, i.e. one element can provide a good estimate of the other one.

- Decision 97/80/EC, Annex II,
 - for the following products, between (1) Table B, part B, column 1 and (2) Table H, column 1
 - 11. Drinking milk
 - 112. Whole milk
 - 113. Semi-skimmed milk
 - 114. Skimmed milk
 - 12. Buttermilk
 - 13. Cream for direct consumption
 - 21. Concentrated milk
 - 221. Cream milk powder
 - 222. Whole milk powder
 - 223. Partly skimmed milk powder
 - 224. Skimmed milk powder
 - 225. Buttermilk powder
 - 23. Butter and other yellow products
 - 2411. Cheese from cows' milk pure.
 - 25. Processed cheese
 - 26. Caseins and caseinates
 - 27. Whey
 - Cows' milk collected from farms between (1) Table B, part A and (2) Table A, quantity (annual)
 - between (1) Table A quantity (annual), cream collected from farms (in milk equivalent) and (2) Table C, part B, cream delivered to dairies (in milk equivalent)
 - between (1) Table B, part A, Raw milk collection (items I, II.1, II.2 and II.3), column 1, and (2) Table D and E, row "Total", Column "collection", sum of both values
 - between (1) Table B, part A, Raw milk collection (items I, II.1, II.2 and II.3), column 1, and (2) Table C, part A, column whole milk, total
 - between (1) Table B, part B, 3. Skimmed milk and buttermilk returned by dairies, column 1, and (2) Table C, part A, column skimmed milk and buttermilk, 1. Returned by dairies
 - between (1) Table B, part A, II.4. Cream, column 1, and (2), Table C, part B, column skimmed milk and buttermilk, 4. Delivered to dairies
 - between (1) Table B, part A, II.5. Skimmed milk and buttermilk, column 1, and (2), Table C, part C, 2. Farm cream: of which delivered to dairies

- between (1) Table B, part A, Total availabilities (items I, II and III), column 1 and (2) Table F, row “Total”, column “Volume”
- between (1) Table B, part B, 11. Drinking milk, column 1 and (2) Table G2, row “Total”, column “Annual production”
- between (1) Table B, part B, 22. Powdered dairy products, column 1 and (2) Table G3, row “Total”, column “Annual production”
- between (1) Table B, part B, 23. Total butter and other yellow products, column 1 and (2) Table G4, row “Total”, column “Annual production”
- between (1) Table B, part B, 24. Cheese, column 1 and (2) Table G5, row “Total”, column “Annual production”
- between (1) Table C, part A, 1. Cows’ milk, column Whole milk and (2) Table I, region “Total (country)”
- Regulation (EC) No 1165/2008, Annex II, all 47 categories of livestock, reference day in November/December, between (1) Articles 3 to 5 (national statistics) and (2) Article 8 (regional statistics), national total.

III.2 Inventory of the supplementary variables disseminated

- Regulation (EC) No 1165/2008, Annex IV
 - Bulls and bullocks (as in Annex V)
 - Calves and young cattle (as in Annex V)
 - Adult Cattle (Bulls, bullocks, heifers and cows)
- Decision 97/80/EC
 - Annex II, Table B, part A, columns 1 and 2
 - Raw milk collection (items I, II.1, II.2 and II.3)
 - Total Collection (items I and II)
 - Total availabilities (items I, II and III)
 - Annex II, Table D, E, F and G, both variables, grouped by
 - Table D, “over 300 000” (the five upper size classes)
 - Table E, “over 5 000” (the four upper size classes)
 - Table F, “over 300 000” (the five upper size classes)
 - Table G.1, “over 100 000” (the four upper size classes)
 - Table G.2, “over 100 000” (the four upper size classes)
 - Table G.3, “over 20 000” (the two upper size classes)
 - Table G.4, “over 10 000” (the four upper size classes)
 - Table G.5, “over 10 000” (the four upper size classes)

III.3 Inventory of the variables which take a particular value

The only variables requested from the Member States at national level and taking a particular value (zero) at EU level are the volumes of intra-Community exchanges, as required in

- Decision 97/80/EC, Annex II, Table B, part A., items 11, 21, 31 and 41,
- Decision 97/80/EC, Annex II, Table B, part B., item 41,
- Regulation (EC) No 617/2008, Annex III, Part II, “Intra-Community trade”.

III.4 Inventory of which modality applies to each of the various statistical tables collected

All the tables are fed with minimum information on confidentiality, i.e. the confidential cells are identified by a single flag 'C', whether they refer to primary or secondary confidentiality.

III.5 List of flags to be used the initial list of the Charter

The list provided in the original proposal was not amended.

III.6 Agreement on the way information on confidentiality is to be transmitted

In the current EDAMIS web forms, field "obs_status" is intended to receive flag 'c'.

After implementation of the new design in the web forms (foreseen in 2015), field obs_conf will be intended to receive flag 'C'.

ANNEX B – STANDARD CASES

The standard cases are intended to illustrate typical theoretical cases, here about confidentiality. They help non-experts to better understand and can be handled in the discussions amongst specialists easier than virtual cases. The examples given below are true, even if the figures have been invented for the purpose of illustration. As an assumption, the entity in charge of processing an EU total knows only whether a national value is confidential or not and has no explanation on the reason for confidentiality. This is currently the situation in Eurostat regarding animal production statistics.

The cases are described, several solutions are proposed, and the solution selected for this case is explained. Some further comments help to shape the mind on the various issues raised here.

Most of the examples refer to figures on production. The main producer is called the leader and the next one the challenger.

Case 0: number of units – few dairies (basic case)

Description: only two dairies produce milk powder in a country. The objective is publishing the national total.

No solution: if the total of their individual values is published, each of them can easily estimate the value of the other one by discrepancy. Therefore there must be always at least three confidential values for being able to publish a total drawn from confidential data. This case is not specific to Animal Production Statistics. The agreed threshold may be higher if this enables simplifying the other conditions to be checked.

Case 1: extreme dominance – Dutch cream milk powder

Description: one statistical unit represents the major part of the EU production of a cream milk powder for a given period.

No solution: there is no way to provide any usable EU total for this production without disclosing the confidential value or (if another national value is confidential) giving an indirect way to estimate it with good accuracy. Publishing a EU total derived from a perturbed Dutch value would make the figure on production for the Netherlands and for EU at the same level of accuracy. This would make unusable the latter one if the first one is made unusable.

Further issue: collection of this detailed value can appear useless, especially if the nesting product (fat milk powder) becomes confidential because of it. On the other hand, if several other kinds of fat milk powders met the same issue in various Member States, EU statistics would be published on some fat milk powders and on the total for fat milk powder.

Case 2: dominance – standard case for whole milk powder

Case 2.1: dominance – National level

Description: two statistical units represent respectively 60% and 38% of the national production whole milk powder in a country. If the national production were published, one of them would estimate the production of its main competitor at less than 40% (or than 62%) of production which is, by chance, a good estimate. Therefore the national total is not published.

Comment 1: All the Member States use a rule based on the number of statistical units. If only two units contributed to the confidential value, each of them would know precisely the value of the other one by discrepancy. Here, at least three units (but their actual number is not known) contribute to the national total and therefore the leader cannot directly estimate accurately the value for the challenger, but it can have a good estimate. This is called dominance of the two main values (98% of the total).

Comment 2: The leader can only estimate that 40% of the volume is produced by the other enterprises in the sector. The quality of its estimate does not depend on the way a confidential value is estimated (confidential total minus its own value), but on its actual value: the estimate by the leader of the challenger's value is anyway 40% whether the actual value of the challenger is 39.9% or 5% of the national total.

Some Member States use the contribution of a single enterprise based on this comment, noted (1, k). Some other ones prefer implementing a rule on dominance of two enterprises (here the two major contributions reach 98%), noted (2, k). The parameter k is the limit contribution chosen for protection, expressed as a percentage.

Comment 3: The contribution of the enterprises other than the two major ones is negligible (2%). For the challenger it represents $2/62 = 3.2\%$ of the value he can guess and for the leader, it is $2/40 = 5\%$. E.g. the relative error of the estimate the challenger can derive for the leader is quite small, 3.2%. The relative error of the estimate the leader can derive for the challenger is a little larger, though still only 5%. Therefore a further rule is sometimes used for protecting confidential values, called p%, declaring cells confidential, when the relative error of the estimate the challenger can derive for the leader is below p%.

Solution: The total cannot be published with $n=2$ as the two major enterprises dominate the value. Parameter n is at least 2 and it may take various values depending on the statistical process and on the organisation in the sector.

Case 2.2: dominance – EU level

Description: The results are also confidential in 11 other Member States. The national confidential totals in thousand tonnes are respectively 95, 85, 68, 10, 4, 3, 2, 1, 1 and less than 0.5 in the three latest countries, for a subtotal of confidential values at 250. The EU total published is 560. Even if there are only single firms in the major producing countries (the leader producing thus 95 and the challenger 85), the estimate of the leader for any its main competitor is far from precise. If they estimate the confidential EU subtotal as 250 and the total production of the competitors as at most 155 ($250 - 95$), even regarding the challenger this estimate has a relative error of $(155-85)/85$, e.g. about 82%. In terms of relative precision, the best estimate can be computed by the challenger (i.e. for the leader): this estimate would be 165 ($250 - 85$) with a relative error of $(165-95)/95$, i.e. about 74 – hence still very imprecise.

Comment: The EU total is well protected in this example. The rules (1, 50), (2, 75) and p-5% make the EU total publishable.

Solution: the dominant values are 95 and 85 and represent together 72% of the confidential subtotal. They can be published if the agreed threshold is over 72%.

Case 3: limit dominance rate – concentrated milk

Description: seven national values are confidential for the production of concentrated milk. The confidential EU subtotal for these countries is at 200 (thousand tonnes). In the major producing country, there is only one firm involved, producing a bit less than 100 (thousand tonnes). This firm knows that its main EU competitor is the only firm in one of the other seven countries. This second firm produces a bit less than 50.

Comment: This case is a limit case for dominance rules (1, 50) and (2, 75) and for p%-rule with $p=50$. Indeed the maximum confidential value (just below 100) represents less than 50% of the confidential EU subtotal and the two maximum values represent less than 75% of this subtotal. By applying either rule (1, 50), or (2, 75) or p% with $p=50$, the data are published. The first firm knows then that its competitor produces less than 100 (relative error: 100%) and the competitor knows that the leaders produces less than a bit more than 150 (relative error: above 50%). Whether the second firm produces only about 20 or as much as 99 is not known to the leader. To the challenger it is unknown, if the leader produces 50 or 150.

Suppose that the rule implemented is (2, 75) and that the result are confidential a given year and published the next year. Now the leader or the challenger can do a good estimate of the value for their main competitor by estimating their own contribution at around 75% of the national total. By subtraction they can thus know the value for their main competitor.

Solution 1: With rule (1, 50) only, the EU total would be publishable, whatever is production of the challenger. If the challenger's production is almost the same size (e.g. almost 100), the two competitors can almost exactly estimate each other value.

Solution 2: With rule (2, 75), the EU total would not be publishable if production of the challenger is over 50.

Solution 2 appears as the most secure. The example of 75% gives an idea of the scale for protecting the values. For security reasons (see the above comment), the actual rate implemented is not published, as threshold effects could lead to provide good estimates of the hidden values.

Case 4: single value with low contribution – Maltese cows' milk collection

Description: Cows' milk collection for the single Maltese dairy is confidential in 2012. The value is between 40 and 45 thousand tonnes and such a proxy is not confidential. Collection of cows' milk is not confidential for any other Member States and the total for the other Member States is 140 108.11 thousand tonnes. The Maltese contribution to EU cow's milk collection is thus about 0.03%.

Solution 1: EU total is the sum of national production for the other Member States, i.e. the confidential value is interpreted as zero for calculation.

Solution 2: The Maltese confidential value is perturbed by adding or removing a certain amount. EU production published is closer to the true value. But changes from a year to another are farther from the true changes. Benefit is low compared to cost.

Solution 3: A fix amount is chosen for substitution, e.g. 42.5, and the EU total is rounded to 10 thousand tonnes, i.e. 140 150 thousand tonnes or, sounder, 140.15 million tonnes. In such a case collecting Maltese data would be questionable.

Solution 4: The true value is used for calculation and the EU total is rounded to 10 thousand tonnes, i.e. 140 150 thousand tonnes (or, sounder, 140.15 million tonnes) in any case. The stable result is due to the particular value of total for the other countries. The benefit is clear but an increase (or decrease) of the single hidden value (as EU total minus displayed results) from a year to another (the complementary EU total being known with accuracy) would threaten the confidential value.

Solution 5: An agreement is reached locally in the Member States for making publishable the confidential value. Whereas not initially suggested, this solution was successfully implemented in Malta since the reference period January 2014. This example is kept nevertheless for illustrative purpose.

Solution 3 appears otherwise as the best combination of accuracy and automation of the treatment for similar cases.

Case 5: missing significant value – Missing national cheese production (and accuracy of zero by rounding)

Description: For a given year, Luxembourg and Maltese total cheese production are the only confidential values. National total cheese production in a third Member State is not provided but production of cow's milk cheese is 185.5 thousand tonnes, the three other types of cheese being provided with values at zero, i.e. less than 500 tonnes. Total cheese production for the 24 other EU-27 Member States (none being confidential) represents 9 035.763 thousand tonnes.

Solution 1: The missing value for total cheese production of the third country is calculated at 185.5 thousand tonnes and published. Publishing the EU total would mean having an accurate idea of the production for Luxembourg and Malta together, i.e. the Maltese data provider could know the accurate value hidden for Luxembourg cheese and the same for the Luxembourg data provider with the Maltese figures.

Solution 2: The missing value for total cheese production of the third country is calculated as about 185.5 thousand tonnes plus or minus 50 tonnes (displayed accuracy of the figure for cows' milk), plus zero to 1 500 tonnes (margin of error for the three values provided as zero). The total cheese production for the third Member State is thus between 185.45 and 187.05 thousand tonnes. This range of uncertainty can be used to make inaccurate indirect estimate of one of the confidential value when the other is known with accuracy. A EU-27 total can thus be published as 9 226 thousand tonnes ($9\,035.76 + 185.45$ to $187.05 + 3$ to 5 , possible range for MT + LU). If total production of cheese for the third Member State were published anywhere else, the current Eurostat estimate for that country (non-published) would remain sound on the basis of information available in Eurostat. This estimate perturbs the EU total with a limited impact on EU figures (0.02%) protecting nevertheless the confidential cells (uncertainty on 30% on their sum, i.e. about 60% on each value).

Solution 2 appears as the most secure and performant, as it draws benefits of all usable uncertainty.

Case 6: secondary confidentiality – Austrian slaughter for poultry

Description: slaughter for poultry refer to four categories of poultry, of which chicken is usually dominant, and a category for all poultry together. Due to the few number of slaughterhouses involved, poultry categories other than chicken are always confidential. Furthermore they refer to the same enterprises, i.e. the sum of these categories is confidential although the number of categories is three. The discrepancy between slaughter for all poultry and for chicken is confidential and one of both values is secondary confidential when the other one is published. It means that either total poultry is confidential or chicken.

Solution 1: Each data provider decides of the value to be hidden. But more heterogeneous is the distribution of confidential values, higher is the risk for hiding EU totals (more aggregates are concerned, and each by fewer confidential cells).

Solution 2: A rule is defined to decide the value amongst the candidates for secondary confidentiality which should receive such a status. This rule may suggest publishing the most (or the less) aggregated level in order not to propagate confidentiality downstream (resp. upstream) in the hierarchy of products. For conflict between publishing values at the same level a rule for priority should be defined (e.g. publish the highest absolute value or the first item in a pre-agreed priority list for the products).

Solution 2 appears as the most efficient at EU level. The option chosen is publishing the most aggregated variable, as it is more likely to be re-used at an even more aggregated level and therefore this option limits propagation of confidentiality over the statistical tables.

Case 7.1: individual information to be protected – German regional milk production

Description: In region *Sachsen (DED)*, statistics on production of milk on farms are reported to be at 1.7 million tonnes. The farms from three NUTS2 sub-regions (DED1, DED2 and DED3²) contributed respectively for 37%, 47% and 16% at the NUTS1 production. Their structure as reported by FSS does not show any dominant farm which production could be confidential. Nevertheless the farms collecting milk in these regions are really few and milk production reflects well milk delivery. Furthermore data on milk delivery from farms are drawn directly from the survey on dairies in order to limit burden on the respondents. Therefore Germany considers these data confidential as they would disclose individual information on the dairies.

Confidential data refers to (Article 3(7) of Regulation (EC) No 223/2009), which refers to *disclosing information about statistical units*. Article 2 of Directive 96/16/EC refers to the dairy enterprises as well as to the agricultural holdings. Article 4 specifies that *the questionnaires must be compiled in such a way as to avoid duplication*.

Solution 1: The results are confidential because they would disclose information on the dairies collecting milk.

Solution 2: The results should be published because (1) they should be drawn from the farm questionnaire. Furthermore, (2) collection is not production, (3) collection area for the dairy enterprises does not fit necessarily with the NUTS region boundary.

Solution 3: The statistics are collected in this special case and only for the concerned regions directly from the farms.

² Actually DED2, DED4 and DED5 have been renamed as this does not interfere with the example.

Solution 4: As the accurate collection area of a dairy is not known, the dairies record the figures on milk collected broken down by region, so that the published statistics do not refer directly to their activity. This, combined with uncertainty on relationship between collection and production, could be sufficient for protecting the figures.

The Member State should assess the alternative solutions before surveying another statistical unit than the original one (the farm in this case). But solution 2 cannot be implemented as it would disclose confidential information. When facing this case, the Member State must assess whether **solution 3** would solve the issue. Otherwise **solution 4** is to be implemented.

Case 7.2: individual information to be protected – use of register of products

Description: Slaughter statistics are drawn from a veterinary register of animals, not identifying the plant or the enterprise having slaughtered this animal. Therefore only one single reporting entity provides statistics on slaughtering. Regulation (EC) No 1165/2008 does not explicitly refer to any statistical unit for these statistics, but to the coverage on the Member State's territory. As the register managers cannot identify the reporting units, they claim that the data drawn from the register should be taken as confidential.

Solution 1: the results are confidential because they could disclose information on a slaughterhouse being the only one (or one of the few) for a given category of animals.

Solution 2: structural information on activity of slaughterhouses is to be conducted regularly in order to identify sensitive cases. It should be repeated often for the slaughterhouses concerned. Finally only actually confidential information is to be concealed.

Solution 2 should be applied, or a similar solution leading to an effective assessment of the confidential status of the figures produced. Using a register as a data source without assessment of the confidential status of the figures is not a good practice as it makes such information non-publishable.

Case 8: protection of individual information across the Member States with a transnational enterprise

Description: Individual statistical units are in the best case a combination of units reporting to the relevant national administration. Multinational enterprises are split into national entities based on their plants in various countries. They may have in hands information about some other statistical units, i.e. the other plants of the same firm in EU. Considering the simplest case, firm X having three plants X_{AA} , X_{BB} and X_{CC} in three countries AA, BB and CC where the totals are confidential, Y_{AA} being its competitor in country AA, and no other dairy competes with them (for simplicity). Obviously the three national totals are confidential. Over the three countries there is not necessarily dominance and firm X can easily derive from the EU total published the total for these three countries and, by removing its own figures, for firm Y_{AA} . This issue appears with more complex situations (longer to describe) and reflects somebody having in hands confidential information about more than one statistical unit. Similar situations may arise if a group of firms put together statistical information for their private use.

Solution 1: Statistics from firm X are aggregated and collected at EU level and they do not take part to the national totals collected.

Solution 2: Statistics from firm X are collected by the Member State where the seat of the firm (or its main plant) is located and they do not take part to the national totals collected.

Solution 3: An inventory of such cases is conducted with identification of the involved firms, in order to find, case by case, the most adapted solution. An intermediate solution could be that each Member State reports on its own situation, identifying the particular cases when only one firm is involved in the national total. This would enable knowing for instance whether three countries with a single firm refer to one, two, or three different firms.

Solution 4: The statisticians assess the risk for such cases and conclude that the effort for meeting together these figures is not reasonable. Such a firm has, by usual (legal) ways, information of better quality on its competitors than an estimate based on the published statistics. Furthermore, crossed contribution of firms one to each other makes any clear conclusion on independency of firms impossible.

Solution 5: the number of firms is collected and, when a group of countries display all the results for a single enterprise, this group is supposed as representing a single firm.

Currently solution 1 is not applicable and data exchanges with actual identifiers of the firms cannot be foreseen in the current technical and legal context. Solution 2, if applicable, would provide a Member State with confidential information on dairy activity in another Member state, which is not a sound solution. Solution 3 could be envisaged as a one-shot for preparing an answer if assumptions of solution 4 are invalid.

Solution 5, proposed by a Member State, requires extra-information on the number of enterprises (but this is available for the major products), than put assumptions which are doubtful. In the most frequent case, a single firm in Luxembourg, Malta and Slovenia can hardly be supposed being a single one. Finally solution 3 would provide the same results, but more soundly.

Therefore the statisticians implement **solution 4** but do not deny relevance of the risk. An inventory of the cases (**solution 3**) could support such an option.

Case 9: contribution to changes – German chickens for fattening

Description: A German firm intends to deliver light broilers to an important third country. Therefore the placing of chick for such a purpose increases suddenly, as drawn from the number of chicks hatched and the external trade of chicks. The firm is not dominant on the national market, but it is one of the rare ones dealing with external trade of broilers and the change in the number of chickens fattened indicates preparing an important delivery abroad. The sensitive dominance does not refer to the figures published (each one for a given reference period). It refers to the contribution in the changes and they are easily derivable to every data user.

First the firm delays his answer to the questionnaire and, facing pressure from the NSI, it explains the situation.

The issue is about dominant contribution to a derived variable (change) which, even if not published as such, is easy to calculate by everybody. The value should actually be considered as confidential. By delaying its transmission, the firm expresses its right for protection of individual information, but agrees that such information can be disclosed when it becomes no longer sensitive. Such an explanation should be accepted.

Case 10: seasonal confidentiality – production of skimmed milk powder

Description: Production of skimmed milk powder is a seasonal activity and, at the beginning and at the end of the production period, only few factories are involved. Therefore, some months may be provided as confidential, the next ones as publishable, and the latest ones as confidential again. The question is whether the annual total is confidential.

The first month, two enterprises, A and B, are the only producers and the value is confidential. For 10 months, enough enterprises contribute to the production and the data are not confidential. The last month, two enterprises contribute to the production, at the same level as the first month.

The confidential data over the time periods may concern different units for different period (enterprises C and D contribute for the last month) or the same units for different periods (enterprises A and B contribute for the last month). Without considering possible dominance, in the first case the total would become publishable and in the second case it would remain confidential.

Solution 1: The annual total is transmitted together with the December values, for monthly data. This would increase the workload on the data providers and would make the data collection more complex (further values, further risk of errors, further validation checks).

Solution 2: A study is conducted for assessing whether more results could be published, i.e. whether different statistical units are involved in the production at the beginning and at the end of the production period.

Solution 3: The cumulated results (either from the beginning of the year or for the twelve last months) are published together with the monthly results, if they improve the situation.

Each solution would increase burden and complexity of the data flows and should be first assessed in comparison of the issue. Solution 2 can be seen as a further assessment which could highlight a possible longer term solution. Whatever is the conclusion, increase in complexity may appear worst regarding protection of individual information than a status quo.

Case 11: Change in flags for confidentiality – request from the data users

Description: An individual value is confidential as long as the person whom the statistics describes disagrees with its disclosure. The publication by a person of results on himself is an implicit agreement for publication. The data providers could check whether the dairies publish the statistics collected, so that fewer cells are confidential. A data user comes with the hyperlink to the website of a dairy enterprise, where some results are made public. He requests that the confidentiality treatment is re-assessed based on such information.

The enterprise is the only one producing butter in the country, and therefore production of butter should no longer be confidential for the country. But for milk collection, three dairies contributed to the total, and publishing the results of one of them discloses confidential information.

Furthermore the published results do not fit the answers given to the statistical questionnaire and illustrate a growth that the statistics contradict.

Solution 1: Eurostat corrects the figures collected and re-assesses confidentiality based on the new elements. Whoever did a query the day before has in hand the figures previously collected and can easily calculate the value of the confidential sub-total based on this. Making confidential a cell already published may thus increase the risk for disclosure.

Furthermore the values published by the enterprise are not intended to follow the legal statistical requirements (definitions, methods, etc.). Also the communication of the enterprise is not intended to be checked by the statisticians.

Solution 2: Information potentially published by the respondents is not considered when assessing confidentiality.

Considering that information published by individual persons has not to fit with their answers to statistical surveys, one cannot state that this information became public. If a data user estimates the values in a cell based on public information from private persons, he must not be able to know whether the statistical value fits with such information, in order to avoid indirect disclosure. The benefit of making public confidential information is ruined by the cost in making confidential previously published cells (if feasible) and especially the cost for managing such an information system.

Solution 3: The respondents are explicitly asked whether the individual statistics they provide are published anywhere else, or whether they agree with their disclosure. The national statistical authority considers as confidential (subject to protection) only the answers without agreement for disclosure. The national authorities check coherence and manage the time table for these publications.

Possible improvement due to such afterwards change needs to be assessed case by case. No systematic rule should be proposed in order to avoid competition between statistical activities in producing fresh statistics and in revising confidential status of former statistics. Both solutions are thus relevant and arguments about good enough estimates in the published data should not lead to data users burdening the respondents with requests for accessing their results.

Case 12: Confidential values obtained with different rule sets

Description: In a first country, statistics on milk production are public for every enterprise, i.e. all individual results concerning production are public. In a second country only those concerning less than three statistical units are protected, without taking into account dominance. In a third country, the statistics concerning less than five enterprises and those where an enterprise contribute to at least 35% have to be protected from dissemination, as they would threaten individual values. This exaggerated example illustrates nevertheless the diversity met amongst the 28 EU Member States.

Such diversity in protection of individual statistics makes sense as it results from (1) diversity in culture (whether knowing precisely something about an enterprise is good or bad) and (2) the structure of the enterprises in the national sector (whether there is no risk for dominance of one or, in the contrary, strategic information would be endangered by rough estimates). When trying to build a EU total based on such diverse rules, what should be the ones to be implemented?

Solution 1: Eurostat implements its own rules, without taking into account the national ones. This would mean that a EU total with less than five firms could be published, despite this would not be possible at national level in the third country. It is supposed that the structure of the dairy sector at EU level is intermediate between the national ones and furthermore the data from the most selective country will not be published without those from other countries, diluting the risk for meeting a case of non-publishable national data disclosed at EU-level.

Solution 2: Eurostat considers the various parameters (minimum number of enterprises, limit-rate of dominance for one, for two enterprises, etc.) used in the countries which data contribute to the confidential cluster (set of confidential values). It implements the most strict ones (the highest national minimum number of enterprises, the lowest dominance rate) and uses it for checking whether a EU total is confidential or not. This is over-protective as such, as often a higher minimum number of units compensate a less strict dominance rate.

Solution 3: Eurostat and the Member states agree on the further conditions to be applied to the national data in order to be compatible with the national set of rules.

Solution 2 takes pragmatically into account diversity in the EU and can therefore be implemented soon but the objective is going towards solution 3.