Data Confidentiality

STATISTICS DENMARK

DANMARKS STATISTIK

23 February 2017, Version 1.0

Data Confidentiality

Contents

1 Preamble 2

2 Principles 2

3 Organisation 2

4 Legislative framework 3

5 Rules about data confidentiality in general 4

a. Personal data 4

b. Enterprise data 4

6 Rules concerning external users' access to microdata 6

7 Rules concerning Statistics Denmark as a processor of data 7

8 Rules concerning internal users' access to data 8

Table of appendixes 8

Appendix 1. Concepts and definitions 9

Appendix 2. Handling confidentiality in social statistics and business statistics 10

Social statistics 10

Business Statistics 13

The link between business statistics and social statistics 16

Please note that the Danish-language version includes another five appendixes that have not been included in this translated version.

1 Preamble

Statistics Denmark's data confidentiality policy is a set of rules and guidelines applied by Statistics Denmark when processing the large volumes of data about the Danes and Danish enterprises, which is the foundation for the production of statistics.

With respect to personal data, data confidentiality is governed by the Danish Act on Processing of Personal Data. The legislation governing the data confidentiality of enterprise data is less comprehensive.

With respect to both fields of statistics, the data confidentiality policy is Statistics Denmark's specific implementation of the statutory rules and basic principles governing the processing of data in a way which ensures that citizens as well as enterprises can rest assured that their data is only used for statistical purposes or scientific surveys and that information about individual persons or individual enterprises cannot be disclosed via the statistical application of their data.

The data confidentiality policy is supplemented by a number of appendixes explaining specifically how the general policy is handled in a number of specific areas and in a number of concrete situations, e.g. in connection with the users' access to data from Statistics Denmark and in connection with the exchange of data between Statistics Denmark and other authorities.

2 Principles

As the central Danish producer of statistics, Statistics Denmark must focus primarily on documenting the development in society in terms of figures. A necessary requirement for this activity is for Statistics Denmark to have access to large volumes of data about citizens and enterprises – many of which are of a confidential nature. For this reason, it is only reasonable that citizens and enterprises are entitled to protection of their data and application of the data for only statistical or scientific purposes. Statistics Denmark's data confidentiality policy specifies how to ensure this protection.

The overall principles for the data confidentiality policy are:

- To protect the identity of the persons, enterprises and institutions[1] for which Statistics Denmark has data.

- To ensure that data in Statistics Denmark is solely applied for statistical or scientific surveys.

The principles have resulted in a number of rules and practices, which are described in the below sections. In the description, we use a number of concepts, which we explain in appendix 1.

3 Organisation

The Director General is the senior person responsible for Statistics Denmark's technical and administrative activities, including the data confidentiality policy.

[1] Statistics Denmark's rules for ensuring confidentiality for enterprises do not apply to public institutions or private and self-governing institutions solving tasks for central and local government in accordance with concluded operating agreements or in accordance with legislation.

The Director General has prepared the data confidentiality policy in collaboration with Statistics Denmark's board.

Statistics Denmark's Data Confidentiality Committee, which has been appointed by the Director General, handles the task of making decisions on matters of principle concerning the data confidentiality policy as well as the auditing of these. The Data Confidentiality Committee prepares an annual statement for the Director General of the results of the committee's discussions and informs the Board of key aspects of the statement.

The head of division of the individual statistical sections is responsible for ensuring the implementation of the data confidentiality policy in the work of the office, which includes ensuring that it is not possible to identify individual citizens or enterprises in the statistical products.[2]. The responsible head of division has immediate knowledge of product details and is consequently in a position to identify any critical issues. In cases of doubt, you must contact the Data Confidentiality Committee.

4 Legislative framework

Statistics Denmark's data confidentiality policy is a specific implementation of a number of rules in legislation and executive orders: The EU Regulation on European Statistics, the Data Security Directive, the Danish Act on Processing of Personal Data, the National Security Executive Order, the Danish Public Administration Act, the Danish Access to Public Administration Files Act, the Danish Criminal Code and the Act on Statistics Denmark.

For relevant legislation, please refer to Statistics Denmark's website
www.dst.dk/en/OmDS/lovgivning?#fortrolig

The EU legislation, the Danish Act on Processing of Personal Data and the Danish Public Administration Act lay down the overall principle that data collected for statistical purposes may be applied only to conduct statistical or scientific surveys. Once data has entered the (closed) statistical system, the Act on Processing of Personal Data and the National Security Executive Order provide extended access to the processing of data for statistical and scientific purposes. Data from the (closed) statistical system must not subsequently be used for other purposes, although it may be disclosed to other public authorities within the framework of the mentioned legislation. However, disclosure must only happen for statistical or scientific purposes, and the data must not be used for e.g. administrative, legal or tax purposes or for checking the facts of statistical entities. This does not prevent Statistics Denmark from providing services in terms of e.g. processing, troubleshooting and documentation of the data of other institutions used for administrative purposes.

The Public Administration Act emphasizes the professional secrecy of public servants with regard to confidential information that they obtain as part of their job (section 27). Furthermore, the act contains rules (sections 28 and 30, 31 and 32) about the disclosure of data to other public authority (see Statistics Denmark's Data Confidentiality Policy, section 6).

[2] International trade in goods and the manufacturing industry's sale of goods are exceptions, where passive confidentialising is applied and where indirect identification is possible in certain cases.

Section 10 of the Public Administration Act lays down that the rules about right of access to documents do not apply to material obtained for the preparation of statistical or scientific surveys.

Section 152 of the Criminal Code lays down the penalties for (present and former) public servants' disclosure of confidential information.

The Danish Act on the Central Business Register defines e.g. what is meant by a legal entity (section 3) and a production entity (section 4).

The Act on Statistics Denmark makes it clear that public authorities and institutions must give Statistics Denmark access to the data in their possession which is necessary to comply with the work programme of Statistics Denmark (section 6) – and that business owners have an obligation to provide data to Statistics Denmark (sections 8-13).

5 Rules about data confidentiality in general

As the overall principle is for data in the (closed) statistical system not to leave this system, Statistics Denmark handles all data in their possession as confidential.

a. Personal data

The main rule is that data that can be attributed to an individual person must be confidentialised so that it is not possible to obtain new knowledge about individual persons' affairs from a statistical product.

The majority of data for social statistics is collected from administrative registers in accordance with the Act on Statistics Denmark. It is absolutely fundamental that register data disclosed to Statistics Denmark is only applied for statistical and scientific purposes. Consequently, we must not return it to public data providers nor disclose it to other public authorities in a way so that data from Statistics Denmark will influence the decision of specific cases in the administrative system.

In the few cases where the underlying registers do not offer a sufficient basis to prepare the statistics, we may perform surveys where we collect data by contacting a selected group of persons directly. Such surveys are not subject to the Act on Statistics Denmark. As a result, the persons must give their acceptance on an informed basis. This means that we must inform the persons, prior to the actual data collection, of the purpose and contents of the survey, and we must respect any request not to participate. This also applies to surveys that are not included in the statistical production but performed as a service job.

There is a detailed description of the rules for data confidentiality in social statistics, including the rules for surveys and sampling, in appendix 2.

b. Enterprise data

The processing of enterprise data is regulated in the Public Administration Act, which regulates e.g. issues concerning the disclosure of data, as well as the regulation on European statistics. In addition, the processing of data concerning sole proprietorships is regulated by the Act on Personal Data, since this is considered as enterprise data as well as personal data. Besides this, no Danish legislation

requires confidentialising of enterprise data, but in accordance with the recommendation in the European Statistics Code of Practice, Statistics Denmark has decided to confidentialise enterprise data.

With reference to the first overall confidentiality principle of protecting the identity of the persons, enterprises and institutions about which information exists in Statistics Denmark, we ensure – in accordance with the data confidentiality policy adopted by the Board – that statistical output is confidentialised by two methods: active and passive confidentialising, respectively. The background for these methods is that enterprises as statistical units are much more heterogeneous than persons and households.

Active confidentialising

All business statistics – except from the statistics of international trade in goods and the statistics of the manufacturing industry's sale of goods, cf. below – apply active confidentialising.

The most essential confidentiality rule is the application of a number criterion, i.e. a minimum requirement of three observations in a table cell, for the relevant data in the cell to be published.

In addition, due to the heterogeneity of enterprise data, the so-called dominance criterion for economic variables (e.g. sales or value added) is applied. This means that if the two largest businesses together account for a dominant share of the value of a given table cell, confidentialising is applied, cf. appendix 2.

However, for all business statistics, in principle, the number of statistical units (i.e. enterprises, establishments or kind-of-activity units) in the relevant statistics may always be stated regardless of the number of observations, as long as they do not include variables with special identification potential. Variables in the statistics with special identification potential are economic variables and employment variables.

In the determination of whether data may be published, we first apply primary confidentialising where the individual groupings are considered in isolation and where the criteria are number and/or dominance. We supplement this by a secondary confidentialising, which we perform manually, and the point of which is to prevent the user of the statistics from reconstructing the confidentialised figures for a group by means of published (non-confidentialised) figures for other groups. This is typically relevant in publications of detailed statistics with combinations in hierarchical structures such as industries, size groups and geographical areas. Another term for secondary confidentialising is consequential confidentialising.

Passive confidentialising

For the statistics of international trade in goods and services as well as the manufacturing industry's sale of goods, we apply so-called passive confidentialising, where the enterprises themselves must request confidentialising. This is owing to the fact that it is the goods/services and not the enterprise which make up the statistical unit. Furthermore, the level of detail of the EU-required product classification applied in the two sets of statistics is so that active confidentialising at enterprise level would imply that the statistics could not be published in any meaningful way.

In brief, the method implies that, in principle, table results cannot be confidentialised, cf. the method mentioned above on active confidentialising. However, it is possible for the reporting businesses to have the results for a given product

kept wholly or partly secret (confidentialised) if, according to specified criteria, they have a dominant position in the market for the given product. In other words, the enterprise must make a request to Statistics Denmark for this to be effected. If the enterprise does not comply with the criteria for market dominance, the request for secrecy will be denied. We will follow-up annually to establish whether approved instances of secrecy continue to meet the criteria. Special criteria apply to the international trade in goods to ensure comparability with statistics of other countries. You will find a detailed description of these rules in appendix 2.

With respect to the other main confidentiality principle of ensuring that data in Statistics Denmark is applied only for statistical and scientific surveys, the same rules apply as for statistics about persons. As an exception, however, updates performed by Statistics Denmark in the Register of Business Statistics, on e.g. the industry of the enterprises, are replicated to the Central Business Register (CVR), whereby it is made generally available as part of the public basic data.

6 Rules concerning external users' access to microdata

Access to individual data

External users can get access to analyse anonymised individual data with Statistics Denmark, if they meet a number of conditions. Individual data means data that are associated with a single individual or a single enterprise as opposed to aggregated data, which only describes groups of individuals or enterprises.

External users may apply individual data (microdata) via one of Statistics Denmark's four service schemes:

1. The researcher scheme
2. The ministry scheme
3. The law model scheme
4. Data warehouse

The four schemes are described in more detail in appendix 3.

Access to individual data

Regardless which of the four schemes that is applied, *data is always stored in Statistics Denmark*. Users who are authorised by Statistics Denmark get access for a specified period to the individual data that they need for specified subject-related purposes, but users must never take delivery of individual data, i.e. transfer individual data to servers outside Statistics Denmark.

Individual data is only made available to the user in *anonymised form*. Full anonymization is not required, but certain restrictions in the access to data, e.g. in the form of a limited number of variables or further confidentialising, may prove necessary if the risk of recognisability of individual persons or enterprises is too obvious – e.g. in connection with small populations.

Common to all four schemes is the fact that they may be used for analytical purposes only. It is not allowed under any circumstances to use them for administrative purposes. Nor may they be used for continuous statistical production unless a separate agreement exists regarding this.
In case of doubt in connection with the interpretation of the rules on data confidentiality, all four schemes are handled by Statistics Denmark's Data Confidentiality Committee. If necessary, the committee will make a recommendation to the Director General.

Regardless which of the four schemes external users apply, it is a general condition that the users have *signed special confidentiality and non-disclose agreements* in advance.

The researcher scheme

The researcher scheme may be used by researchers/analysts who are already attached to an institution that is authorised to access data with Statistics Denmark on a project basis.

The ministry scheme

Based on their special need for analyses, central government authorities such as departments, government agencies etc. may access data on more flexible terms than others in the research and analysis environment may.

The law model

The law model is a very special model for those parts of the central administration that need to make *pre-legislative analyses on the government's economic models*. Departments requesting access to use the Law model need approval from Statistics Denmark as well as the department responsible for the law model.

Data warehouse

In connection with the establishment of a ministerial data warehouse, a development and operation agreement is concluded allowing ministerial staff access to the data warehouse in the development phase and in the subsequent operational phase.

The data warehouse will typically contain data at civil registration number (cpr) level, but the civil reg. no. will have been replaced by a neutral ident allocated by Statistics Denmark. Using the neutral ident as a key allows linking of Statistics Denmark's data with data that the ministry places in the data warehouse.

Disclosure

In very special cases, as an exception and based on a specific decision from the Director General, Statistics Denmark may disclose identifiable personal data to another public authority for the purpose of preparing statistics. This is described in appendix 4.

7 Rules concerning Statistics Denmark as a processor of data

Other public authorities may have a right of access to the same data that is included in the statistical production of Statistics Denmark. To avoid duplicate collection of information, Statistics Denmark may collect data on behalf of another authority. In some cases, the relevant other authority is the data-controlling authority, and Statistics Denmark is the data processor. Collection and disclosure of data happens in accordance with the instructions of the authority, including any disclosure to other authorities in advance of processing by Statistics Denmark. The process is concluded by transferring the data to Statistics Denmark's ordinary activity after which Statistics Denmark will become the controller of this data in the normal way.

The application of data for statistics in accordance with this model is not incompatible with an originally administrative purpose of the data collection. Data collected for administrative purposes may be disclosed for statistics – but the opposite must not be the case.

Statistics Denmark's application of the data processor model requires permission from the Director General in each single case.

The rules concerning data processor agreements are described in more detail in appendix 7.

8 Rules concerning internal users' access to data

For Statistics Denmark's employees to gain access to data, their superiors must approved it, and it must be justified by a data requirement in strict relation to the set or sets of statistics with which the employee in question is working. As a result, we grant access by a 'need-to-know' principle, and Statistics Denmark's IT department review the data access every six months in order to audit the granted access.

The employees' application of register data is logged in accordance with section 19 of the National Security Executive Order.

Table of appendixes

1. Concepts and definitions
2. Handling confidentiality in social statistics and business statistics
3. External users' access to individual data
4. Disclosure of personally identifiable data to other public authorities
5. Reporting-back to data suppliers
6. Sample-based statistics
7. Data processor agreements

Appendix 1. Concepts and definitions

Concept	Definition
De-identified individual data	Data for which all formal identification details, such as name, civil registration number, central business register (CVR) number and address, have been removed
Active confidentialising	Standard procedure for Statistics Denmark involves confidentialising on our own initiative – see also passive confidentialising
Number criterion	A certain minimum number, e.g. three observations per table cell, which must ensure that the data in the cell can be published without revealing new knowledge about the relevant persons or enterprises
Data processor	The entity that performs data processing of confidential data for a data-controlling public authority in accordance with that authority's instructions and in compliance with a signed data processor agreement
Data-controlling authority	The authority that is responsible for ensuring data confidentiality and data security, and which must report the processing of confidential data to the Danish Data Protection Agency (DPA)
Decryption	Process that makes encrypted information readable. See also Encryption.
Directly identifiable individual data	The specific individual data, which is identified by e.g. a civil registration number or central business registration number
Confidentialising	Blocking any chance of recognition of individual persons in table presentations to prevent that new knowledge is revealed about the relevant persons or enterprises
Confidentialising, active	See Active confidentialising
Confidentialising, passive	*See Passive confidentialising*
Dominance criterion	One or two enterprises constitute a dominant share of a delimited group
Indirectly identifiable individual data	The specific individual data that can be identified indirectly on the basis of a combination of characteristics about the relevant person or enterprise
Individual data	Data about the specific statistical unit, person or enterprise
Available individual data	Access to data subject to non-disclosure, e.g. research access
Encryption	Process that makes information unreadable to third party by means of an encryption algorithm. See also Decryption.
Microdata set	Data set with individual data
Model data set	Data set with anonymised individual data, possibly a sample
Passive confidentialising	Confidentialising happens at the request of the enterprises – see also Active confidentialising
Primary confidentialising	Confidentialising based on number criterion and/or dominance criterion
Pseudonymisation	*A formal identification number is converted to an inane code which can only be interpreted through the use of a given conversion key and/or procedure*
Secondary confidentialising	Confidentialising which is performed manually to prevent the user of the statistics from reconstructing the confidentialised figures for a group by means of published (non-confidentialised) figures for other groups
Truncation	Truncation of extreme values
Identification power of variable	Expresses how strongly the variable – alone or in combination with other variables – can contribute to the identification of a person/enterprise
Disclosure of individual data	Disclosure of data to another public authority which will then be a data controller
Criterion of materiality	Enterprises with a trade worth at least DKK 2m annually

Appendix 2. Handling confidentiality in social statistics and business statistics

Introduction

In this appendix, we describe how Statistics Denmark ensures confidentiality in statistical output (i.e. aggregate data) which must be published or delivered as service jobs for specific users and customers. In general, we break down the statistical output in the statistical domains social statistics and business statistics. First, we go through confidentialising of social statistics, then business statistics, and finally the link between the two. There may be supplementary requirements or minor deviations for confidentialising of individual registers and statistical domains. With respect to these, we refer to the guidelines for the relevant statistics, which can be obtained from the individual head of statistics.

Social statistics

Confidentiality requirements

The confidentiality requirement for social statistics prescribes that it must not be possible to derive new information about individual persons or households from the statistics. In practice, this means that the reader of a table must not be able to obtain new knowledge about individual persons besides gender and age, which are considered to be core demographic variables and accordingly not subject to confidentialising. As a general rule, this means that:

1. There must be a minimum of three individuals in a group before any further information is disclosed about the group[3]

Counting of individuals

This requirement is fulfilled by checking all combinations of groupings in addition to the full combination of all variables.

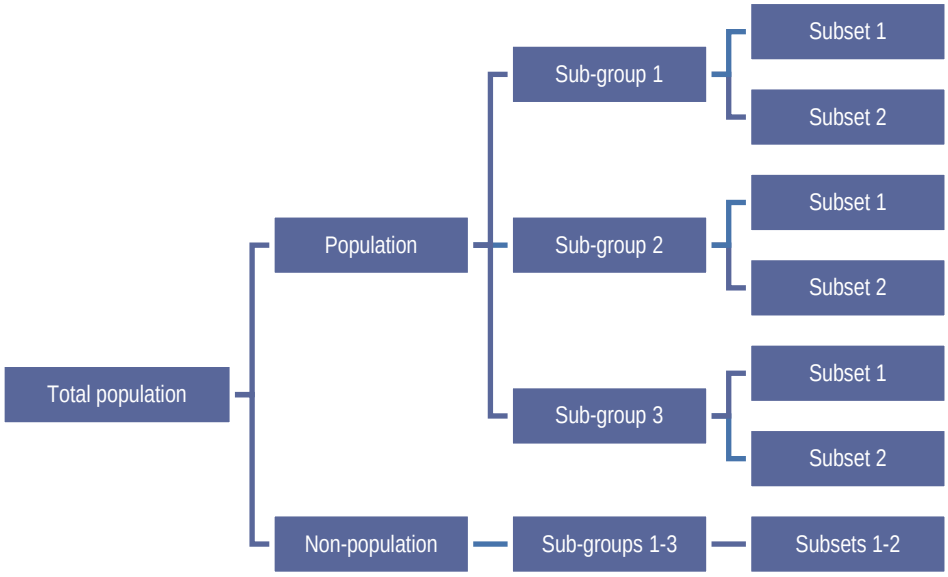
For example, if a breakdown by three variables is considered, it is checked that each combination of two variables contains a minimum of three individuals. I.e. each combination of variable 1 and variable 2 must contain at least three individuals. The same applies to the combination of variables 1 and 3, as well as variables 2 and 3.

Provided that the population in a statistical product is not the entire population, the population selection is considered as a grouping (inside and outside the population) on its own on equal terms with other groupings.

Figure 1 illustrates a selected population, to which two additional details are attached (sub-groups 1-3 and subsets 1-2).

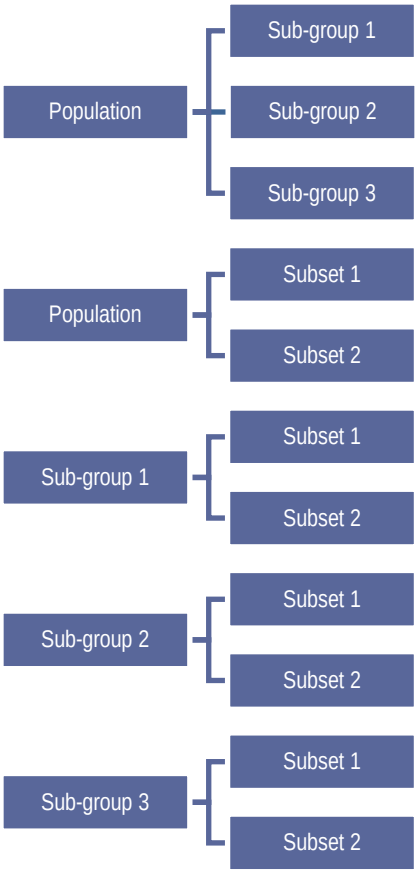
[3] In special cases, there may be a requirement for more than three individuals if the subject is considered particularly sensitive. If information is retrieved from several sets of statistics, the most strict confidentiality requirement is applied.

Figure 1 Example of confidentialising for social statistics



We ensure that the statistical output complies with the confidentiality requirement be checking that all combinations of groupings from which one of the groupings has been excluded, contain a minimum of three individuals. In the example, it means that all of the following combinations of the above breakdowns must contain a minimum of three individuals:

Figure 2 Confidentiality requirements for social statistics in connection with counting of individuals



Confidentiality requirements:

All combinations must contain a minimum of three individuals.

I.e. in the population, there must be a minimum of three individuals in sub-group 1, a minimum of three individuals in sub-group 2, etc.

Calculated values

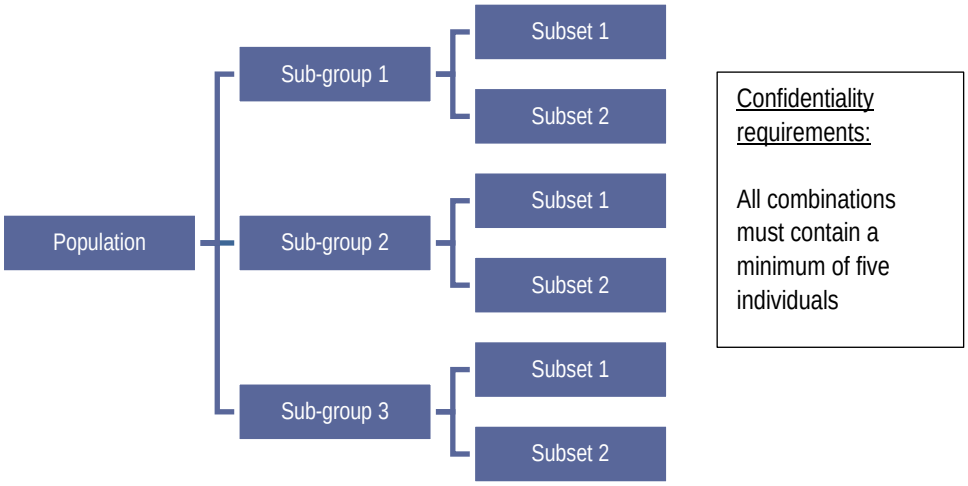
If the statistical output contains calculated values, for instance the average income for a group, rather than a counting of individuals, the confidentiality requirement prescribes instead that:

1. The calculated values must be based on a minimum of five individuals[4]

Note that it is no longer necessary to take the entire population into account, since there will always concurrently be sufficient individuals in the population.

Figure 3 illustrates a selected population, to which two additional details are attached (sub-groups 1-3 and subsets 1-2), after which a calculated value is stated.

Figure 3 Confidentiality requirements for social statistics in connection with calculated values



For the calculated value to be disclosed, a minimum of five individuals must be represented in each end node; i.e. each combination of sub-groups 1-3 and subsets 1-2 in the selected population.

Example

We consider a population defined as all immigrants from outside the Western World as of 1 January 2015. This is connected with information about education in groups and labour market status. In relation to figures 1 and 2, the sub-grouping is then education groups and the subset grouping is labour market status. To comply with the confidentiality requirement:

a. There must be a minimum of three immigrants from outside the Western World in each education group.
b. There must be a minimum of three immigrants from outside the Western World in each grouping of labour market status.
c. For the entire population in total, there must be a minimum of three persons in each combination of education group and grouping of labour market status.

If a., b. and c. are fulfilled, there is no need for confidentialising.

[4] In some cases, the requirement may be higher than five if the subject is particularly sensitive.

Handling confidentiality issues

If the above requirements have not been fulfilled, there are issues of confidentiality that must be handled in the form of confidentialising. This means that it is necessary to mask data, so that the data of individual persons cannot be identified. This can be done by:

1. Merging subsets.
2. Deleting the relevant information.

When merging subsets, we subsequently check whether the new grouping complies with the confidentiality requirements above. If instead we delete information for the problematic combinations, we ensure that it is not possible to backtrack calculations and derive the actual values. Here we pay special attention to the fact that the statement of totals and subtotals may create a need for secondary confidentialising; for an explanation, see the business statistics section.

Example

If again, we look at the example above, and condition a. has not been fulfilled, i.e. one of the education groups contains less than three immigrants from outside the Western World. Given that conditions b. and c. have been fulfilled, this can be solved by merging the education groups so that there are at least three immigrants from outside the Western World in each sub-group.

Conditions implying requirements that are more stringent

For some types of service jobs, more stringent confidentiality requirements apply. This applies e.g. if the data is owned by someone other than Statistics Denmark. For example, the Danish Health Data Agency owns the Register of Medicinal Product Statistics, and agency places demands on the minimum number of units required in the individual cells of these statistics.

In addition to this, exacting circumstances exist in connection with delivery of aggregate data where the customer has submitted the population that is merged with Statistics Denmark's register data. In these cases, it is important to take into account in the confidentialising process that the customer has his own information about the individuals for whom he has submitted information.

Research Services

For rules regarding transfer of aggregate data through Statistics Denmark's Research Services, please refer to the appendix about confidentiality requirements when working with research data sets.

Business Statistics

Confidentiality requirements

Similarly, the confidentiality requirements for business statistics prescribe that it must not be possible to obtain identifiable information about individual enterprises or establishments on the basis of statistical output. This means that it must not be possible to obtain information about the activity or other confidential characteristics for the individual statistical unit, such as sales, number of persons in full-time employment etc.

However, the same requirement does not exist for public entities, since it is not considered problematic if information about these can be identified.

All sets of business statistics, except from the international trade in goods and services and the manufacturing industry's sale of goods, apply the principle of active confidentialising, as described below.

Practice of active confidentialising

For a piece of information to be regarded as non-attributable to the individual statistical unit, the following requirement must be met for each separate combination:

1. There must be a minimum of three statistical units.

Statistical units can be the economic unit or legal entity, the central business register (CVR) number, the establishment, etc. It depends to which unit level the statistics are broken down.
In case of statistics by economic variables, it must be taken into account that a few enterprises may dominate the activity to an extent that allows identification of the size of the selected variable for the individual statistical units by means of totals stated for the industry, for example. For this reason, unlike the social statistics, there is also a need for a dominance criterion:

2. The two largest statistical units (enterprises) may account for a maximum of 85 per cent of the sales in a given table cell. However, the confidentialising is only done in case of economic variables.

If aggregate business data is delivered, for example, then the dominance criterion is fulfilled when the two highest sales figures constitute less than 85 per cent of total sales in each combination.

Note that the dominance criterion in general will always be calculated on the basis of sales regardless if another economic variable, e.g. value added, is disclosed in the table. Otherwise, a user would be able to piece together identifiable pieces of information by ordering a number of different tables with single variables.

However, the actual number of statistical units in the relevant set of statistics can always be stated regardless of the number of observations.

Example

We look at a set of aggregate statistics in which the sales for a group of central business register (CVR) numbers are broken down by main industry and sector (public/private). For the two above requirements to be fulfilled, the following must apply:

a. Within each main industry, there must be a minimum of three CVR numbers in each sector.
b. Within each main industry and sector, the two CVR numbers representing the highest sales must not account for more than 85 per cent of total sales.

If requirements a., b. and c. are fulfilled, there is no need for confidentialising.

Handling confidentiality issues

When examining whether the set of aggregate statistics may be published, we first carry out a primary confidentialising check, where we consider the individual group in isolation, i.e. check whether the two above requirements have been met for the group.

If the requirements have not been met, it means that there are confidentiality issues and it is necessary to mask data, so that the data of individual units cannot be identified.

This can be done by:

1. Merging subsets.
2. Deleting the relevant information.

When merging subsets, we subsequently check whether the new grouping of units complies with both of the confidentiality requirements above. When deleting information, we must pay attention to any need for secondary confidentialising (also called consequential confidentialising) so that the deleted information cannot be derived from totals, subtotals, etc. This is typically relevant in publications in hierarchical structures such as industries, size groups, and geographical units, such as regions and municipalities.

Example 1 If we look at the example above, where requirement a. has not been fulfilled. I.e. within one main industry, there is less than three CVR numbers in one sector. This means that there is a confidentiality issue that we need to solve. We solve this by removing the sector breakdown for the main industry in question, as illustrated in table 1 below.

Table 1 Not enough CVR numbers in one group

	Confidentialising	Number of CVR nos	Sales
Main industry 1	No	9	DKK 80m
- Public	Yes, primary	2	~~DKK 30m~~
- Private	Yes, secondary	7	~~DKK 50m~~

First, it is checked if there is a minimum of three CVR numbers in each sector within the main industry (primary confidentialising). As it appears, there are only two CVR numbers in the public sector in main industry 1, which is why sales must not be disclosed. Then it is checked if it is possible, on the basis of the remaining information, to derive the sales for the public sector event though it has been edited out. Here it can be calculated as the difference between total sales of the main industry and sales for the private sector, which means that there is a need for secondary confidentialising. For this reason, sales for the private sector are also edited out.

Example 2 If we look at the example above, where requirement b. has not been fulfilled. I.e. within one main industry and sector, there are two CVR numbers that account for more than 85 per cent of the sales. This means that there is a confidentiality issue that we need to solve. We solve this by removing the sector breakdown for the main industry in question, as illustrated in table 2 below.

Table 2 Dominance issue in one group

Confidentialising	Number of CVR nos	Sales	Sales two largest
No	25	DKK 100m	DKK 70m
Yes, secondary	5	~~DKK 30m~~	DKK 20m
Yes, primary	20	~~DKK 70m~~	DKK 65m

Table 2 shows that there are more than three CVR numbers in each group, which is why requirement a. has been fulfilled. Now we check requirement b. (the dominance criterion) for each sector within the main industry. For the public sector, the two

CVR numbers with the highest figures account for 67 per cent (20/30) of sales, so in this case there is no dominance issue. For the private sector, the two CVR numbers with the highest figures account for 93 per cent (65/70), and consequently we have a dominance issue. This is why we first confidentialise sales for the private sector (primary confidentialising), and then also confidentialise sales for the public sector (secondary confidentialising), since it would otherwise be possible to derive sales for the private sector. For the main industry in total, the two CVR numbers with the highest figures account for 70 per cent (70/100) of sales, so in this case there is no need for confidentialising.

As an alternative to removing the grouping by sector for the main industry, we could choose to state just the sales for the public sector and in this way confidentialise sales for the private sector and in total for the main industry. In that case, we must make absolutely certain that the user cannot obtain sales for the main industry from another source, e.g. published figures in Statbank Denmark or prior deliveries.

Definition of economic variables

The dominance criterion is used to find out whether aggregate data for economic variables may be disclosed. We define ”economic variables” tightly as accounting statistics variables, i.e. variables that say something about volume. This further implies that more qualitative variables, such as “active on social media”, should not be subject to the dominance criterion. With regard to employment information, as described in the section about the link between business statistics and social statistics, only “persons in full-time employment” are processed according to both criteria, i.e. other employment variables, such as number of persons with a particular education and number of persons in employment (individuals), are not processed in accordance with the dominance criterion; and in employment statistics, “persons in full-time employment” are used as the basis for the dominance calculation.

Special statistics with passive confidentialising

The statistics of international trade in goods and services and the statistics of the manufacturing industry's sale of goods have special confidentiality requirements that differ from those of other business statistics, because confidentialising is carried out only at the request of the enterprises. The argument for this is that the two sets of statistics operate with goods and services as the statistical units rather than e.g. the individual enterprises. For this reason, it is difficult to derive information about individual respondents in these sets of statistics. Statistics of international trade in goods and services broken down by enterprise characteristics are confidentialised in the same way as other business statistics, except for the fact that the dominance calculation is based on turnover.

The link between business statistics and social statistics

Confidentiality requirements

When linking between business statistics and social statistics, we account for the need for confidentialising in both domains. I.e. no information must be given about individual persons or the activity of individual business entities. The confidentialising for employment statistics and other links between business statistics and social statistics is described below.

Employment statistics

The nature of employment statistics is a combination of business data and personal data. Depending on the unit by which the employment statistics are assessed, the confidentiality requirements for either business statistics or social statistics apply.

As long as the employment statistics do not involve the enterprises' level of activity (sales, number of persons in full-time employment, etc.), the confidentiality requirements for social statistics apply. This means that e.g. an assessment of the number of persons in employment (heads) broken down by educational level and industry is a set of social statistics.

If, on the other hand, the number of persons in full-time employment is broken down by educational level and industry, a set of business statistics is concerned, and consequently it must be confidentialised in accordance with the requirements for business statistics – the difference being that the dominance criterion check here is based persons with full-time employment rather than sales.

Confidentialising according to the rules of the business statistics happens to avoid showing anything about the activity level of individual enterprises. Only when the employment concept applied measures volume (understood as persons in full-time employment or number of hours worked) does the employment say something about the activity in the enterprise. For this reason, only employment statistics broken down by enterprise characteristics and assessing employment volume must be confidentialised in accordance with the confidentiality requirements of the business statistics.

Other links between business statistics and social statistics

When linking business statistics and social statistics, data must comply with all of the outlined requirements for each statistical domain. For this reason, the combined confidentiality requirements are:

1. There must be a minimum of three statistical units.
2. The two largest statistical units (enterprises) may account for a maximum of 85 per cent of the activity (the dominance criterion).
3. A minimum of three persons in all combinations of groupings.

Note that the dominance criterion, just as it is the case for pure business statistics, is only relevant if economic variables are concerned. Otherwise, it is left out and the requirements are given only with 1 and 3.

Example

We want to assess sales broken down by industry and by the employment or not of a person with a PhD in the enterprise. In this case, we need to ensure that:

a. There is a minimum of three persons in employment with and three persons in employment without a PhD within each industry.

Once this is fulfilled, we can break down the enterprises by industry and by employment or not of a person with a PhD in the enterprise. Requirements 1. and 2. must be fulfilled before we can disclose the sales.

b. For each combination between industry and employee with a PhD (yes/no), there must be a minimum of three enterprises.
c. The two largest enterprises in the same combination may account for a maximum of 85 per cent of total sales in the relevant combination.

When requirements a., b. and c. have been fulfilled, we may disclose total sales for each industry broken down by the enterprises' employment of a person with a PhD or not.

Handling confidentiality issues

Any confidentiality issues are handled as described separately for Social Statistics and Business Statistics respectively.

If we follow the example above, the confidentiality requirements can be checked in the below table:

Table 3: Dominance issue when linking business statistics and social statistics

	Confidentialising	No. of persons	No. of enterprises	Sales	Sales two largest
Industry 1	No	500	30	DKK 100m	DKK 70m
- With PhD	Yes, secondary	30	10	~~DKK 30m~~	DKK 20m
- Without PhD	Yes, primary	470	20	~~DKK 70m~~	DKK 65m

Requirement a. is fulfilled, since there are at least three persons with and three persons without a PhD degree employed in the relevant industry. Requirement b. is also fulfilled, since there are enough enterprises employing and not employing a person with a PhD. On the other hand, requirement c. is not fulfilled, since the two largest enterprises in the industry not employing a person with a PhD account for more than 85 per cent of sales, which is why this must be confidentialised. Sales for enterprises employing persons with a PhD are confidentialised in the same way as described in example 2 in the business statistics section. Note that number of persons and number of enterprises may be disclosed broken down by employment or not of a person with a PhD.

Exceptions regarding confidentialising

Statistics Denmark's rules for ensuring confidentiality for enterprises do not apply to public institutions or private and self-governing institutions solving tasks for central and local government in accordance with concluded operating agreements or in accordance with legislation. Examples hereof are private and self-governing institutions, nursing homes and accommodation facilities, all educational institutions and private enterprises providing education and training with government approval and fully or partly paid by public funds.

It is possible for external users to have the data provided by Statistics Denmark linked with their own data, if this is necessary to realise the project. To do so, the users need prior approval from the Danish Data Protection Agency (DPA). Statistics Denmark performs the actual linking.

STATISTICS DENMARK

Statistics Denmark
Sejrøgade 11
2100 København Ø

Tlf. +45 39 17 39 17
www.dst.dk
dst@dst.dk