C.3. Statistical confidentiality

C.3. Statistical confidentiality

20.20. Statistical confidentiality refers to the protection of information of individual statistical units and must be differentiated from other forms of confidentiality under which information is not disseminated owing to other considerations, for example national security concerns. It is good practice to always strive for the full coverage of all data that are in the scope of statistics on the international supply of services, while applying appropriate methods to keep certain information confidential. The present *Guide* recognizes, however, the necessity of both statistical confidentiality and balancing it against the need for public information in cases in which the application of statistical confidentiality would limit information or make it impossible to provide sufficient or meaningful information. It is also good practice to disseminate a quantitative indicator of the amount of data subject to confidentiality.

20.21. The implementation of recommendations on statistical confidentiality depends to a large extent on each country's legislation and the general confidentiality policy adopted by its statistical system. The assurance of confidentiality is essential to securing the cooperation of reporters and to maintaining the integrity of the statistical system. Confidentiality concerns may also be a more serious issue in smaller or less developed countries in which there are fewer entities likely to be engaged in the international supply of services. An important challenge in the implementation of confidentiality rules is to ensure that confidentiality is applied across all the different classifications in which data are disseminated.

20.22. Active confidentiality is advised, which implies that the statistical agency should take the initiative to actively suppress or aggregate data whose dissemination would enable the identification of any individual person or entity.[1]

20.23. The Eurostat *Handbook on Statistical Disclosure Control* provides detailed examples and approaches to sensitivity rules when applying active confidentiality and is recommended for further reading.[2] As shown in table 20.1, common sensitivity rules used include the following: (a) the minimum frequency rule (data suppressed for cells that have less than a pre-determined number of respondents reporting, typically three)[3] and (b) the dominance rule, in which the preponderance of the value (whose exact share is to be predetermined by the compiler) is derived from the top one to three entities). It is important to keep the exact predefined parameters confidential. In some cases, the contributor with the second largest contribution to a cell that is non-sensitive according to the sensitivity rules cited above is able to derive a close upper estimate for the contribution of the largest one by subtracting his or her own contribution from the aggregate total. In such cases, the dominance rule would need to be adapted (see the p% rule in table 20.1).

20.24. If one data cell is made confidential, it is typically advisable to also suppress the next available cell with the smallest value, so that the value of the primary confidential cell cannot be calculated.

Table 20.1
Sensitivity rules[343]

A cell is considered sensitive when:	
Minimum frequency rule	The cell frequency is less than a specified minimum frequency of n respondents
(n,k-dominance rule)	The sum of the n largest contributions exceeds k% of the cell total, e.g., $x_1+\ldots=x_n>k/100 \cdot X$
p% rule	The cell total minus the two largest contributions x_1 and x_2 is less than p% of the largest contribution; e.g., $X-x_2-x_1<p/100x_1$[5]

[343] Eurostat, *Handbook on Statistical Disclosure Control: Version 1.2* (2010), para 4.2.1, table 4-1.

20.25. It is good practice to publish an overview of the confidentiality rules so that data reporters are assured that their right to confidentiality is guaranteed, while data users are informed about certain data limitations, enabling them to use the data more appropriately. It is also good practice to provide details to users on what data areas are affected most by the application of confidentiality rules and the magnitude of that effect.

20.26. Demand for access to microdata has been increasing amid the growing recognition of its value for social, economic and business analysis. Arrangements for access to microdata vary from country to country, but compilers should ensure that data are made available only for statistical purposes and that access for research purposes is granted only as long as confidentiality is protected. More information on basic principles to be adopted can be found in *Managing Statistical Confidentiality and Micro data Access: Principles and Guidelines of Good Practice*, published by the United Nations Economic Commission for Europe (ECE) and the Conference of European Statisticians (CES).[5]

20.27. It is the role of the statistical compiling agency to decide whether, how and to whom its microdata are released, based on well-established considerations, such as (a) the merits of the research proposals and the credibility of the researcher, (b) whether the risk of identification is sufficiently small and (c) whether the confidentiality adjustments made to the data have unduly damaged the microdata for research purposes, among others. Legal arrangements or some form of administrative arrangement to protect confidentiality should be put in place and made visible before any microdata are released. Such arrangements should cover what can and cannot be done and for what purposes, as well as the conditions of release and the consequences if those conditions are breached. Transparency is important to increase public confidence that microdata are being used appropriately and that decisions regarding access are made objectively. The compiling agency's website is an effective means of ensuring transparency and for providing information on research based on released microdata.

20.28. **Managing breaches** The statistical compiling agency should ensure that researchers are aware of the consequences to them and their institution if there are confidentiality breaches. Legal action could be considered if a legal offence has occurred, but at a minimum, the researcher (and possibly the researcher's institution) should be prevented from further access to microdata. For minor breaches, a warning may be sufficient.

20.29. There are a number of software products currently available for managing the confidentiality of microdata[6] and many national statistical offices also develop their own tailored processes and software specific to their legislative requirements.[7]

Next: C.4. Users and data dissemination

---

[1] It may be interesting to note that in the domain of international merchandise trade statistics, “passive confidentiality” is recommended (i.e., data are treated as confidential only when the trader requests so on the grounds that his or her interests would be harmed by the dissemination of their data and the statistical authority finds the request justified based on the confidentiality rules), unless the use of active confidentiality is already the established, desired and accepted practice (see IMTS 2010, para. 10.3); and *International Merchandise Trade Statistics: Compilers Manual, Revision 1* (IMTS 2010-CM), Studies in Methods, Series F, No. 87/rev.1 (United Nations publication, Sales No. E.13.XVII.8), para. 1.14, available from http://unstats.un.org/unsd/trade/EG-IMTS/IMTS2010-CM%20-%20white%20cover%20version.pdf.
[2] Eurostat, *Handbook on Statistical Disclosure Control: Version 1.2* (2010). Available from http://unstats.un.org/unsd/EconStatKB/Attachment474.aspx.
[3] While three is a common criteria, some statistical agencies use others. For instance, the United Kingdom sets this number at five companies (see United Kingdom, Department of Trade, Enterprise and Investment, “DETI Confidentiality Statement”, available from www.detini.gov.uk/data_confidentiality_statement__principle_5_of_the_code_of_practice_for_official_statistics_.pdf).
[4] Eurostat, *Handbook on Statistical Disclosure Control: Version 1.2* (2010), para 4.2.1, table 4-1.
[5] *Managing Statistical Confidentiality and Microdata Access: Principles and Guidelines of Good Practice* (United Nations publication, Sales No. E.07.II.E.7). Available from www.unece.org/fileadmin/DAM/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf. .
[6] Special Uniques Detection Algorithm is a system for detecting and grading special uniques. This is needed for confidentializing data sets by first identifying all special unique records and either disguising or removing them. SDCMicro is a free software for the generation of protected microdata for researchers and public use, available at www.ihsn.org/home/software/disclosure-control-toolbox. Country-specific examples of microdata procedures as presented at plenary sessions of the Conference of European Statisticians held by ECE are available from www.unece.org/stats/documents/2013.06.ces.html. More details regarding remote analysis servers can be found in a note by the Australian Bureau of Statistics presented at the sixty-first plenary session of ECE/CES, entitled “Innovative microdata access - confidentiality on the fly” (ECE/CES/2013/29).
[7] For example, a number of national statistical offices release public use files, also referred to as confidentializedu record files (CURFs), which are heavily confidentialized files that remove names, addresses, geographic information and other details. Microdata can also be made available via research data centres or data laboratories, either on-site or through virtual terminals installed in other organizations. Outputs removed from these centres should be checked manually. Finally, some statistical agencies have begun using remote analysis servers, which allow researchers to submit a query via the Internet to the agency’s server, which sends confidentialized output back to the researcher..