

Second Meeting of the 2011/2012 Bureau
Geneva, Switzerland, 2-3 November 2011

For decision

Item 8 of the Provisional
Agenda

STATISTICAL CONFIDENTIALITY

Prepared by the UNECE secretariat

I. PRINCIPLES AND GUIDELINES ON CONFIDENTIALITY ASPECTS OF DATA INTEGRATION UNDERTAKEN FOR STATISTICAL OR RELATED RESEARCH PURPOSES

1. The “Principles and Guidelines on Confidentiality Aspects of Data Integration Undertaken for Statistical or Related Research Purposes” were endorsed by the Conference in June 2009. These principles and guidelines were developed by a Task Force chaired by the Australian Bureau of Statistics. The Conference decided that the Guidelines should be tested over a period of two years, and should be reviewed in 2011.

2. In a written consultation exercise before the 2011 CES Plenary Session, countries were invited to inform the secretariat about any problems encountered with the practical implementation of the principles and guidelines. Twenty-six countries and two international organisations responded as follows:

(a) Eight countries had no comments and endorsed the principles (Armenia, Belarus, Bulgaria, Finland, Latvia, Netherlands, Sweden and Switzerland);

(b) Fourteen countries provided comments on national implementation of the Principles and Guidelines, often referring to the relationship with national legislation or practices, but did not make any specific proposals for change (Australia, Azerbaijan, Canada, Croatia, Czech Republic, Italy, Lithuania, Mexico, New Zealand, Slovakia, Slovenia, Tajikistan, Turkey and Ukraine);

(c) Four countries (Denmark, Greece, Romania and the United States), and the two international organisations (Eurostat and the IMF) made proposals for changes or clarifications to the principles and guidelines.

3. The detailed comments were presented in document ECE/CES/2011/10/Add.1 for the 2011 plenary session of the CES¹, which is reproduced in the Annex for your convenience. The Conference welcomed the comments and agreed that they were mainly seeking further clarification of the principles and guidelines, and did not indicate any fundamental problems with their implementation or use. The secretariat was requested to retain the comments for use in a future review of the principles and guidelines.

¹ See <http://www.unece.org/stats/documents/2011.06.ces.html>

4. The Conference requested the Bureau to assess the current need for such a review. The view of the secretariat is that it is not necessary, and may even be counter-productive to open the principles and guidelines for review at the current time, given the general support for the current version, the minor nature of the proposed changes, and the fact that the discussions leading up to the adoption of the principles and guidelines were lengthy and often difficult. The secretariat proposes to review the principles and guidelines after a further 3-5 years if necessary.

II. MANAGING STATISTICAL CONFIDENTIALITY AND MICRODATA ACCESS: PRINCIPLES AND GUIDELINES OF GOOD PRACTICE

5. These principles and guidelines were adopted by the CES in 2006. The Bureau has expressed concern that whilst the principles and guidelines themselves remain relevant, some of the case studies that accompany them are becoming out of date. The Bureau proposed that the secretariat should ask countries to update or provide new case studies.

6. The UNECE Secretariat has contacted authors of the case studies inviting them to provide updates. Other statistical organisations were also invited to propose new case studies to add to the range of experiences presented.

7. So far, the Czech Republic, Mexico and Eurostat have proposed new case studies. The Czech Republic has already provided a draft, and Eurostat has asked for a postponement until early 2012, following a review of their procedures for microdata access. Sweden has reviewed its two current case studies, and proposed to delete one and retain the other unchanged. Delegates to the joint UNECE / Eurostat Work Session on Statistical Confidentiality (26-28 October 2011) have also been invited to prepare or update case studies.

8. **Any further proposed updates and additions to the case studies should be notified to the UNECE secretariat by the end of January 2012.** The method of publication of these case studies will be considered, as UNECE policy has moved from paper to electronic dissemination. The most likely solution is to publish the revised version of this publication electronically via the UNECE wiki platform, thus facilitating more frequent updating in the future.

III. PROPOSAL

9. **The CES Bureau is invited to:**

(a) **Assess the current need for a review of the “Principles and guidelines on confidentiality aspects of data integration” as explained in paras. 3-4;**

(b) **Agree with the approach to updating “Managing statistical confidentiality and microdata access: principles and guidelines of good practice” as proposed in para. 8.**

ANNEX: SUMMARY OF FEEDBACK
on “Principles and guidelines on confidentiality aspects of data integration undertaken for statistical or related research purposes”

The proposals for changes are set out below in the order of the paragraphs of the Principles and Guidelines to which they refer. For ease of reference, the original text is shown in italics.

A. Introduction, paragraph 3

“These principles and guidelines apply to data integration work carried out in national statistical organizations (NSOs). In some cases international statistical organisations combine micro-data sets from different countries, but as there are unlikely to be any units in common between the national data files, no confidentiality issues arise here.”

United States: It is overstated to say that there are not commonalities between national data files, thus no confidentiality issues arise. This assumption depends on what statistical information is going to be released. People move all the time (e.g., retired and military persons) and if data from several years are integrated, people could show up on more than one nation’s registry. Even if micro-data sets from different countries did not contain common units, there could be confidentiality issues involved.

Eurostat: It has to be noted that international organisations may also match microdata from different sources, e.g. different surveys (such work is being actually carried out in Eurostat) and confidentiality issues may arise in such situations.

B. Introduction, paragraph 5(d)

“Data Matching - the linkage of micro-data from different sources based on common features present in those sources.”

IMF: Clarify whether data matching also refers to statistical matching in paragraph 6.

C. Principle 2

“NSOs should only undertake data integration activities consistent with their official statistics mandate and after completing a standard approval process (for example, a business case).”

Denmark agrees with the principles mentioned. Concern is expressed that approval of possible data integration as described in principle 2 and 5 is too heavy. It should be more general and not necessarily an approval process for each project.

D. Principle 2, guideline (a)

“Where a NSO has a mandate(s) that goes beyond statistical and related research purposes, such as involving the use of data for administrative or regulatory purposes related to natural persons, it should abstain from any data integration activities for statistical or related research purposes pertaining to these units, unless this is specifically authorized by law.”

United States: There is uncertainty regarding what this guideline meant by “units”. Perhaps it is better to say that if an NSO has a mandate to perform non-statistical work, it should be kept functionally separate from the record linkage for statistical purposes to ensure the integrity of the statistical activities and the confidentiality of the statistical data.

E. Principle 2, guideline (c)

“A standard approval process should be followed for any new data integration proposal. This may take the form of a formal business case. An example of a business case outline is given in the Annex, but each country should establish their own template for the process of endorsing data integration projects. The approval process should identify how the integration work will produce or improve official statistics or contribute to related research.”

United States: Suggest that an additional sentence be added to this guideline that addresses what policies and procedures should be followed to protect the privacy of individuals and to protect the confidentiality of the data. For example something like: The approval process should also identify the policies and procedures set forth to protect the privacy of individuals’ data and to protect the confidentiality of the integrated data.

F. Principle 3

“The public benefits of any data integration project should be sufficient to outweigh any privacy or confidentiality concerns about the use of data and/or risks to the integrity of the official statistics system.”

Greece: Which is the criterion for public benefits to be characterized as sufficient?

G. Principle 3, guideline (a)

“Data integration should occur in a secure environment and in a manner that does not pose risks to the integrity of the official statistical system.”

United States: Replace “should” with “must”.

H. Principle 3, guideline (c)

“Where appropriate, bodies with responsibility for ensuring that all benefits, privacy concerns and risks are identified and properly considered by the NSO as part of their standard approval process, should be consulted. The list of benefits should include those resulting from any intended long term retention of, or planned extension over time to, the integrated dataset.”

United States: Confidentiality of the data also needs to be ensured.

I. Principle 3, guideline (e)

“Where reasonable and practicable, consent should be obtained from the data provider(s).”

United States: The phrase “where reasonable and practicable” suggests an option to obtain consent from individuals. It may be helpful to acknowledge that informed consent is a necessary condition to data integration in some countries.

J. Principle 3, guideline (f)

“The notions of privacy and confidentiality also require careful management of the risks of indirect identification (typically for units with unusual characteristics), and the increased sensitivity of integrated data sets, which may contain a wider range of variables than any of their sources.”

United States: Proposes the guidelines should also suggest conducting disclosure analysis on the data.

K. Principle 4

“Data should not be integrated where any commitment has been given to respondents that would specifically preclude such action.”

United States: An exception could be written in this principle stating if data integration is deemed necessary, the respondents could be re-contacted to acquire their consent.

L. Principle 5

“Integrated data should only be used for approved statistical or research purposes and any significant variation in the originally approved purposes should result in the submission of a new standard approval process.”

United States: A new approval should be obtained if unanticipated events occur under the currently approved process—such as a data breach or change in policies and procedures affecting the security of or confidentiality of the data. This principle as written suggests that a new approval process is required. We think the intent of this principle is to require a new approval when significant variation of the originally approved statistical or research purpose has occurred.

Greece: The characterization “significant” for the variation is subjective. The variations in the approved purposes for which the submission of a new standard approval process is needed should be clear and specific.

M. Principle 7, guideline (b)

“The main statistical results of any data integration work should be made publicly available. When data integration work is used to improve the production of official statistics (e.g. through improving quality), the publication of that official statistic meets this requirement. Metadata of statistics published from composite databases should contain information about the original data sources used for data integration.”

Romania: Proposes the insertion of the following wording at the end of the first sentence: *The main statistical results of any data integration work should be made publicly available with respect for the legal basis for confidential data protection.*

N. Principle 8

“Access to composite unit record data resulting from data integration, but not containing any identifiers, should generally be limited to authorized staff of the NSO. As for other statistical microdata, any proposal to grant access to an external person(s) should have a clear legal basis and be consistent with the purposes of use of data for official statistics. Any person(s) granted such access should provide a legally enforceable institutional and logistical guarantee that their use will be consistent with the approved proposal and that non-authorized persons will have no access to the dataset.”

United States: It may be helpful if this principle provides greater clarity concerning access to microdata by internal NSO staff and authorized external persons. For example, files that include direct identifiers should only be accessed by a limited number of internal NSO staff. Files that include indirect identifiers only, and no direct identifiers, could be accessed by approved external persons in a restricted environment since individual unit records can still be identified. Also suggest that the phrase, “or research purposes”, should be added at the end of the second sentence. It would then read “...for official statistics or research purposes.”

O. Annex, point K

“A privacy impact assessment should be completed unless a country’s legislative and/or relevant NSO policy provides an exemption. It should also be noted that although privacy generally relates to natural persons, it may also relate to a legal person in the case of some businesses or industries. For example, certain unincorporated businesses, such as farms, may generate privacy considerations in some countries.”

Greece: The content of a privacy impact assessment should be fully described.

* * * * *