



EUROPEAN COMMISSION  
EUROSTAT

Directorate B: Corporate statistical and IT services  
**Unit B-1: Quality, methodology and research**

ACCESS TO CONFIDENTIAL DATA FOR SCIENTIFIC PURPOSES  
(SCIENTIFIC USE FILES)  
GUIDELINES FOR PUBLICATION

Luxembourg 2013

## 1. INTRODUCTION

These guidelines explain how confidential data for scientific purposes shall be used by the researchers<sup>1</sup>. The researchers shall read the guidelines before they use confidential data and respect the rules for publication laid down below. In accordance with the terms of use of confidential data signed by the duly designated representative of the research entity the researchers are bound by the present guidelines for publication:

### **Data handling (paragraph 5)**

**Researchers must ensure that any results of the research published or otherwise disseminated do not contain information that may permit the identification of individual statistical units (persons, households, enterprises, etc.).**

**In any reports, including all publications and unpublished papers, researchers must ensure the strict application of the guidelines for publication attached to the confidential data for scientific purposes.**

**No copy of all or part of the data may be made and none of the data may leave the research entity's premises.**

## 2. WHEN THE DATA ARE CONFIDENTIAL

The data are confidential if the respondents can be identified. Simply removing name and address details from the microdata files does not prevent the identification of the survey respondents. Inspire copy

The unusual characteristics of the survey respondent may lead to the respondent's recognition. The scientific use files delivered to researchers on electronic devices (CDs, DVDs) are especially prepared to make the identification of survey respondents more difficult. In particular by:

- Reducing the level of detail of scientific use files,
- Modifying some values
- Suppressing risky records or variables.

## 3. WHY PROTECT CONFIDENTIALITY (IDENTITY) OF THE RESPONDENTS

Confidential data for scientific purposes contain information provided by individuals or organisations. Each record in the microdata files represent information provided by the respondents. The individual information can only be used for statistical purposes.

---

<sup>1</sup> Acknowledgements: "Responsible Access to ABS Confidentialised Unit Record Files (CURFs); Training Manual"; Australian Bureau of Statistics; March 2005.

**Researchers granted access to confidential data may use the data only for statistical analysis for scientific purposes.**

**Researchers must not seek to identify the individuals or organisations represented on the files.**

Disclosure of individual information constitutes a breach of the law, but also a breach of the trust the respondents place in the statistical system. Such a breach could harm the reputation of that statistical system and lead to a reduction in the quality of official statistics.

#### **4. HOW PROTECT CONFIDENTIAL DATA**

Only researchers belonging to a recognised research entity may request access to confidential data for scientific purposes. The research entity's duly designated representative must have signed a confidentiality undertaking.

Access to confidential data for scientific purposes may be granted if:

- the research proposal submitted by the researcher(s) has been approved; each research proposal must be countersigned by the contact person identified in the confidentiality undertaking;
- all researchers asking for access to confidential data for scientific purposes must have signed a confidentiality declaration.

**Researchers must keep microdata files secure so that it is not accessible by anyone who is not authorized to access the data.**

**Results of the statistical analysis that may contain information on individual respondents should also be kept secure.**

The medium containing confidential data for scientific purposes must be stored in a locked room to which access is restricted to authorised persons only.

The confidential data for scientific purposes must be stored on a password-protected computer. Access to the data must be restricted to authorised researchers named in the research proposal.

The intermediate results of analysis containing confidential data must be stored in a protected environment.

After expiry or completion of the project indicated in the research proposal (or in case of termination of access by Eurostat), the principal researcher must destroy the dataset and any data or variables derived from it and sign a declaration to the effect that it has been

ensured that all data have been destroyed. This declaration applies to the original data sent by Eurostat and to all subsets of the original data set.

## **5. DATASETS AVAILABLE**

Confidential data for scientific purposes are available in two forms:

- “scientific-use files” partially confidentialised data<sup>2</sup> delivered to researchers on electronic devices (CD-Rom, DVD, etc.);
- “secure-use files” available in Eurostat's "safe centre" in Luxembourg (non-confidentialized data);

Scientific use files are available for the following data collections:

- European Community Household Panel (ECHP)
- Labour Force Survey (LFS)
- European Union Statistics on Income and Living Conditions (EU-SILC)
- Adult Education Survey (AES)
- Community Innovation Survey (CIS)
- Structure of Earnings Survey (SES)
- Community Statistics on Information Society (CSIS) – as from September 2013
- Continuous Vocational Training Survey (CVTS) – as from September 2013
- European Health Interview Survey (EHIS) – as from September 2013
- European Road Freight Transport Survey (ERFT) – as from September 2013

Secure use files are available in the Eurostat safe centre in Luxembourg for:

- Community Innovation Survey (CIS)
- Structure of Earnings Survey (SES)

## **6. SPECIFIC RULES FOR PUBLICATION (SCIENTIFIC USE FILES)**

When publishing the results of the statistical analysis for scientific purposes, the researchers will comply with the specific rules laid down below which are meant to inform the reader about quality issues and/or ensure the confidentiality guarantee<sup>3</sup>.

---

<sup>2</sup> Data on which special statistical disclosure control methods have been applied in order to reduce to an appropriate level and in accordance with current best practice the risk of identification of the statistical unit(s); see more item 2.

### **6.1. Specific rules for European Community Household Panel (ECHP)**

In any reports, including all publications and unpublished papers, two cell size thresholds will be distinguished for ECHP cross-sectional results:

- below 20 observations (unweighted sample), results may not be published;
- from 20 to 49 observations (unweighted sample), results may be published but are to be individually identified (e.g. shown in brackets).

For confidentiality reasons, those reports that include sample sizes will only mention 'less than 20 observations' and '20-49 observations' (i.e. not the actual number) for these 2 thresholds respectively.

For unweighted sample sizes below 20 observations, the actual number of observations may not be derived from (combination of) other information available in the reports, e.g. column or row totals.

Exactly the same rule applies for longitudinal results, except that the thresholds are put respectively to 10 and 30 observations linked across time.

### **6.2. Specific rules for European Union Statistics on Income and Living Conditions (EU-SILC)**

In any reports, including all publications and unpublished papers, two cell size thresholds will be distinguished for EU-SILC cross-sectional results:

- below 20 observations (unweighted sample) or if non response for the item concerned exceeds 50%, results may not be published;
- from 20 to 49 observations (unweighted sample) or if non-response for the item exceeds 20% and is lower than and equal to 50%, results may be published but are to be individually flagged (e.g. shown in brackets).

For confidentiality reasons, those reports that include sample sizes will only mention 'less than 20 observations' and '20-49 observations' (i.e. not the actual number) for these 2 thresholds respectively.

For unweighted sample sizes below 20 observations, the actual number of observations may not be derived from (combination of) other information available in the reports, e.g. column or row totals.

### **6.3. Specific rules for Adult Education Survey (AES)**

Same as for EU SILC (see item 6.5 above).

### **6.4. Specific rules for European Health Interview Survey (EHIS)**

Same as for EU SILC (see item 6.5 above).

---

<sup>3</sup> These rules apply to the users of scientific use files. The guidelines for publication for secure use files are provided separately if access to this type of data is requested.

### **6.5. Specific rules for Community Statistics on Information Society (CSIS)**

Same as for EU SILC (see item 6.2 above).

### **6.6. Specific rules for Labour Force Survey (LFS)**

In any reports, including all publications and unpublished papers, three cell size thresholds will be distinguished for LFS results:

- Confidentiality threshold: up to 3 observations (unweighted sample), results must not be published;
- Reliability thresholds: regarding reliability restrictions, Eurostat defines two limits, called 'a' and 'b'. Those reliability limits depend on the sample size and design in the individual Member States. Estimates corresponding to a (weighted) population below limit 'a' should not be published; estimates corresponding to a (weighted) population between limit 'a' and limit 'b' may be published with a warning concerning their limited reliability. This applies to quarterly data, annual averages of quarterly data, yearly data and ad hoc module results. The limits vary across Member States, years, and type of dataset. They are published on the Eurostat website: [http://epp.eurostat.ec.europa.eu/portal/page/portal/employment\\_unemployment\\_lfs/documents/Reliability\\_limits.xlsx](http://epp.eurostat.ec.europa.eu/portal/page/portal/employment_unemployment_lfs/documents/Reliability_limits.xlsx)

### **6.7. Specific rules for Community Innovation Survey (CIS)**

Any statistics (tables, graphs, textual references) on any kind of sub-population (cell) shall not be published:

- if they consist of less than 10 enterprises;
- where one enterprise represents more than 70% of the total sub-population expenditures, employment or turnover;
- where two enterprises represent more than 85% of the total sub-population expenditures, employment or turnover.

In addition, where there are primary confidential cells, the secondary confidentiality treatment is necessary to make sure that these primary confidentiality cells cannot be estimated with the help of the other non-confidential cells.

### **6.8. Specific rules for Structure of Earnings Survey (SES)**

The produced output (tables, graphs etc.) may not be possible to extract or publish in case of:

- less than 10 enterprises / local units or employees;
- where one enterprise represents more than 70% of the total sub-population employment or total earnings;
- where two enterprises represent more than 85% of the total sub-population employment or total earnings.

For cases when primary confidential cells have to be flagged, the secondary confidentiality treatment should be applied in order to make sure that these primary confidentiality cells cannot be estimated or re-calculated through by deducing from other non-confidential cells.

### **6.9. Specific rules for Continuous Vocational Training Survey (CVTS)**

Any statistics (tables, graphs, textual references) on any kind of sub-population (cell) shall not be published:

- if they consist of less than 10 enterprises;
- where one enterprise represents more than 70% of the total sub-population employment or total labour costs;
- where two enterprises represent more than 85% of the total sub-population employment or total labour costs;

In addition, where there are primary confidential cells, the secondary confidentiality treatment is necessary to make sure that these primary confidentiality cells cannot be estimated with the help of the other non-confidential cells.

### **6.10. Specific rules for European Road Freight Transport Survey (ERFT)**

The anonymised data sets may contain confidential information or information the use of which is restricted. The user of the data must therefore comply with restrictions related to dissemination of tables described in the Regulation 6/2003 of December 2002<sup>4</sup>, Art. 3:

1. Dissemination of tables to users other than the national authorities of Member States shall be subject to the condition that each cell shall be based on at least 10 vehicle records depending on the variable tabulated. Where a cell is based on fewer than 10 vehicle records, it shall be aggregated with other cells, or replaced with a suitable flag.

## **7. DATA MATCHING**

Researchers should not attempt to link the microdata to other (including public) datasets, if not expressly agreed by Eurostat. Matching two datasets may lead to identification of the statistical respondents represented in both datasets.

## **8. WHAT TO DO IN CASE OF A CHANGE IN A RESEARCH PROJECT SET UP**

The research proposal is valid for the specified purpose (research project), period, datasets and research entity(ies). A new research proposal has to be submitted to Eurostat if any of the following situations arises:

- the data are to be used for a new research project;
- a different set of data is needed;

---

<sup>4</sup> Commission Regulation (EC) No 6/2003 of 30 December 2002 concerning the dissemination of statistics on the carriage of goods by road, OJ L 1, 04/01/2003 p. 45  
(<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:001:0045:0049:EN:PDF>)

- a new research entity joins the project.

If a more recent release of the data is needed for an on-going research proposal and/or a researcher is replaced or added to the team in the same research entity taking part in the project, a principal researcher or contact person in the research entity should inform Eurostat of these changes in writing (no need to submit a new research proposal). An individual confidentiality declaration has to be signed by each researcher taking part in the project.

## **9. OBLIGATIONS OF THE PRINCIPAL RESEARCHER AND INDIVIDUAL RESEARCHERS**

Principal researcher:

- submits and signs the research proposal and the individual confidentiality declaration;
- identifies individual researchers participating in the research project;
- is responsible for the lawful access to confidential data for scientific purposes for all researchers named in the research proposal;
- protects confidential data in accordance with the confidentiality undertaking, terms of use, and individual confidentiality declaration;
- informs Eurostat of any changes to the research project;
- provides Eurostat with a copy of all reports, which have been produced using the data;
- destroys received microdata and derived files after expiration/completion of the research project.

Individual researcher(s):

- signs (each separately) individual confidentiality declaration;
- protects confidential data in accordance with the confidentiality undertaking, terms of use and individual confidentiality declaration.

## **10. YOUR RESPONSIBILITY TO PROTECT CONFIDENTIAL DATA**

The confidentiality undertaking signed by the duly designated representative of the research entity and individual confidentiality declaration signed by the individual researcher provide a basis for the legal action in cases where conditions of these documents have been wilfully ignored.

The Commission can take action in the event of a breach of confidentiality as follows:

1. by withdrawing from the offending researcher, and if necessary from his/her research entity, the possibility of accessing microdata;
2. by asking the research entity to take disciplinary action against the researcher;



3. by claiming civil-law compensatory damages from the research entity; the confidentiality undertaking includes a reference to the applicable law and competent court;
4. by filing a complaint or by reporting the breach to the police on the basis of national legislation; the Commission may participate in national proceedings as plaintiff.

Depending on the situation, sanctions may be applied to researchers or their research entities.

## **11. LEGAL BASIS**

Legal basis for granting access to confidential data can be found in the Regulation (EC) No 223/2009 on European Statistics<sup>5</sup> (Article 23) and in the Regulation (EU) No 557/2013 on access to confidential data for scientific purposes<sup>6</sup>.

## **12. ANY QUESTIONS?**

In case any issue raised in this manual is unclear, or in case of further questions on the use of confidential data, please contact us: [estat-microdata-access@ec.europa.eu](mailto:estat-microdata-access@ec.europa.eu).

---

<sup>5</sup> <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:087:0164:0173:EN:PDF>

<sup>6</sup> <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:164:0016:0019:EN:PDF>