EU-LFS and EU-SILC: legal, processing and dissemination aspects

Frank Espelage – Eurostat F2 (EU-LFS) Frank.Espelage@ec.europa.eu
Laura Wahrig – Eurostat F3 (EU-SILC) Laura.Wahrig@ec.europa.eu
EU-LFS and EU-SILC: legal, processing and dissemination aspects

1. Organisation of EU-LFS and EU-SILC
2. Microdata access: legal background
3. EU-LFS: processing and dissemination aspects
4. EU-SILC: processing and dissemination aspects
5. Microdata access: legal and technical outlook
1. Organisation of EU-LFS and EU-SILC

Regulatory framework at European level:

- Clear specifications of deliverables
- Obligation of Member States to deliver data
- Harmonisation

but

- Changes require consensus negotiations
- Procedures take time

Implementation:

- Surveys implemented nationally
- Microdata collected and processed nationally
- Transmitted to and validated by Eurostat
1. Organisation of EU-LFS and EU-SILC

- The National Statistical Institutes are responsible for the survey at national level:
  - Flexibility in the implementation (design, data sources, questionnaires …)
  - Surveys can be supplemented to serve national purposes

- Ex-ante output harmonization is reached by:
  - Definition of target concepts/variables to be measured
  - Minimum requirements (survey design, sample representativeness / sample size, fieldwork, imputation, timeliness, calculation of weighting factors, quality reporting …)
  - Guidelines and explanatory notes (updated regularly)
  - Standardized output (individual data)
  - Monitoring by Eurostat
2. Microdata access: legal background

Council Regulation (EC) No 322/97 of 17 February 1997 on Community Statistics:

- **general guidelines** for receiving, processing and disseminating confidential data obtained for Community statistics, including also the **access for scientific purposes**
- access to microdata via Eurostat may be granted by **National Statistical Institutes**
- the Commission should be assisted by the **Committee on Statistical Confidentiality** in defining the conditions governing access for scientific purposes to confidential data held by the Community authority
2. Microdata access: legal background

For the following surveys or statistical data sources access may be granted:
- European Community Household Panel
- **Labour Force Survey**
- Community Innovation Survey
- Continuing Vocational Training Survey
- **Community Statistics on Income and Living Conditions**
- Structure of Earnings Survey
- Adult Education Survey
- Farm Structure Survey
2. Microdata access: legal background

The implementing regulation (EC) No 831/2002 foresees two possible ways of access to confidential data for scientific purposes:

- **Option 1: Eurostat safe-centre**
  - Confidential data as obtained from national authorities

- **Option 2: Distribution of CD/DVDs**
  - Anonymised microdata obtained by modifying the confidential personal records to minimize the risk of disclosure

Both EU-LFS and EU-SILC use the second option.
3. EU-LFS: processing aspects

Data transmission deadlines according to regulation:

- Quarterly data: 12 weeks after the reference quarter
- Yearly data for year N (in particular yearly weights in case of a subsample for yearly variables): together with quarter 4 data in March N+1
- Ad hoc module data for year N: end of March N+1
3. EU-LFS: processing aspects

Data validation:

- Syntax: valid codes, correct routing between variables
- Logic: plausibility of relationship between members of the same household; coherence between variables
- Currently no comparison with previous results at individual record level – LFS is not a panel, even if several rotation schemes are used in the countries
- Comparison with previous quarterly/yearly results for some 50 calculated main figures, e.g. employment, unemployment, hours worked, …
- NSIs are asked for explanations of detected irregularities/correction of errors/confirmation of main results
3. EU-LFS: dissemination aspects

Availability of any aggregations of microdata using

- Quarterly data ➔ main indicators available on the Eurostat website on average 4 weeks after transmission deadline
- Yearly data ➔ annual averages: usually at the same time as quarter 4 results except for yearly variables: mid N+1 for year N
- Ad hoc module data ➔ separate processing; requires in general yearly files for countries with yearly subsamples; most data for year N confirmed mid N+1; second half N+1 special evaluation of each module; first results on the website: not before N+2
3. EU-LFS: dissemination aspects

Anonymisation of data for year N can not start before

- 3rd quarter N+1 for quarterly and yearly data
- year N+2 for ad hoc module data

Consequences:

- Release of year N+1 contains core data until N and ad hoc module data until year N-1.
- The latest available data are anonymised and disseminated, including all revisions for past years Eurostat received since the previous release
3. EU-LFS: anonymisation

General aggregation and anonymisation criteria agreed between Eurostat and the NSIs of the Member States and endorsed by the Committee on Statistical Confidentiality:

- Age (5-year groups)
- Nationality and country of birth (National / EU / outside EU)
- Income suppressed

Further aggregation of NACE (1-digit), ISCO (2-digit), marital status, professional status, number of persons working in the local unit, duration of search for employment, hours worked > 80, no tracking of persons across time possible, plus further country-specific aggregations, e.g. for geographical information
4. EU-SILC: design

- Survey and register countries (DK, NL, SI, FI, SE, IS, NO)
- Four-year rotational design in most countries (exceptions FR 9 years, NO 8 years and EE, LU and SE treated as a normal panel)
- Longitudinal and cross-sectional files cannot be linked.
- One year’s data consists of four files
  - Household register (D)
  - Household data (H)
  - Personal register (R)
  - Personal data (P) only for those aged 16+

These files can be merged.
4. EU-SILC: processing aspects

Data transmission deadlines according to regulation:

- **Cross-sectional files**: Transmission of files to Eurostat preferably within 11 months of the end of data collection. Absolute deadline is 30 November of year N+1 for data collected in year N for some countries and 30 October of year N+1 otherwise.
- The annual ad-hoc modules are transmitted with and included in the cross-sectional files.
- **Longitudinal files**: Transmission of files to Eurostat preferably within 15 months of the end of data collection. Deadline is 31 March of year N+2.
4. EU-SILC: processing aspects

Data validation:
- Syntax and logic is checked both by NSIs and Eurostat.
- Comparison of key variables and indicators with previous years’ values ➔ plausibility checked and if necessary explanation/confirmation asked for.
- NSIs are asked for explanations of detected irregularities and correction of errors where this is possible.
- Indicators are calculated on both sides and bilaterally validated before publication.
- Cross-sectional indicators are published between 15th December of N+1 and 15th of January of N+2.
4. EU-SILC: dissemination aspects

Issuing of microdata/release policy

- Cross-sectional files: have to be issued by the end of February of year N+2.
- Longitudinal files: have to be issued by the end of July of year N+2. Include the four most recent years (from 2007 onwards)
- Revisions: are issued on the CD with the new microdata for both the cross-sectional and the longitudinal component
  - 2004-2006 cross-sectional files have been reviewed in-depth ➔ will be included in the next CD release
  - For the longitudinal file revisions are only implemented in the files including the most current year.
4. EU-SILC: anonymisation

Data for the CD is anonymised as follows:

- Region: changed from NUTS2 to NUTS1
- Survey design: Most variables related to survey design withdrawn or randomised
- Time of interview: day of household/personal interview withdrawn, month of household/personal interview recoded into quarter
- Age: month of birth variables recoded into quarter, other age variables top-coded to 80+.
- Housing: dwelling type=other type of accommodation recoded as missing, number of rooms top-coded to 6+.
- Origin: country of birth and citizenship 1 regrouped in four categories: survey country, EU, non-EU, world; citizenship 2 withdrawn
- NACE classification regrouped.
- Additionally there are some country-specific provisions.
Dissemination aspects – access to microdata

Details on conditions of microdata access can be found on the Eurostat website: [http://epp.eurostat.ec.europa.eu](http://epp.eurostat.ec.europa.eu)

**EU-SILC:**
- One-off purchase of cross-sectional or longitudinal data: € 500 (€ 250 for any subsequent purchase)
- Annual subscription to both cross-sectional or longitudinal data: € 1,000 (€ 750 for any subsequent subscription)

**LFS:**
- DVD containing a set of quarterly/yearly files covering available data for all 28 countries and all available years: € 8,000
- CD containing a set of quarterly/yearly files referring to the most recent year covering available data in 28 countries: € 2,000 for new users (€ 1,000 for users who bought the full data set for the previous year)
5. Microdata access: legal outlook

New regulation of the Parliament and the Council on European Statistics will replace Council Regulation (EC) No 322/97 in 2009:

- Access to microdata for research purposes mentioned under:
  - principle of statistical confidentiality (principle of protection)
  - principle of accessibility and clarity

- Statistical disclosure control: obligation to ensure protection, to harmonise

- Mentions both public use files (anonymised records, neither direct nor indirect identification) and access for research. Both need approval of the national authorities

- Current implementing regulations remain in force for the time being; are to be replaced in the future to correspond better to the new legal framework and emerging needs
5. Microdata access – technical outlook

Recent discussions between Eurostat and Member States on the way forward focus on decentralised access also for EU microdata:

- Network of safe centres run by National Statistical Institutes
- Analysis of synergies with data archives (CESSDA)
- Remote access as part of the medium term strategy
EU-LFS and EU-SILC: further information

- General background information can be found in the EU-LFS and EU-SILC domains on CIRCA:
  

  http://circa.europa.eu/Public/irc/dsis/eusilc/library

- Dedicated sections for both surveys exist on the Eurostat website.
Thank you for your attention