Microdata collections available at Eurostat. A focus on EU-SILC

Emilio Di Meglio
Outline

1. Microdata collections available
2. Types of microdata
3. Microdata access workflow – researcher perspective
4. EU-SILC
ESS microdata for scientific purposes

1. European Community Household Panel
2. European Union Labour Force Survey
3. Community Innovation Survey
4. European Union Statistics on Income and Living Conditions
5. Structure of Earnings Survey
6. Adult Education Survey
7. European Road Freight Transport Survey
8. European Health Interview Survey
9. Continuing Vocational Training Survey
10. Community Statistics on Information Society
11. Micro-Moments Dataset
12. Household Budget Survey 2010 (since September 2016)

- EU-SILC, 697, 37%
- EU-LFS, 534, 28%
- ECHP, 147, 8%
- CVTS, 28, 2%
- EHIS, 68, 4%
- AES, 63, 3%
- SES, 116, 6%
- CIS, 137, 7%
- CSIS, 28, 2%
- MMD, 19, 1%
- ERFT, 17, 1%
- HBS, 28, 1%
Data anonymisation

- De-identification
- Partial anonymisation
- Full anonymisation

- Statistical use files
- Secure use files
- Scientific use files
- Public use files
Anonymised datasets and modes of access

- **De-identification**
  - Statistical use files
    - Access by statistical office only

- **Partial anonymisation**
  - Secure use files
    - On-site access
  - Scientific use files
    - Data transmitted over the internet

- **Full anonymisation**
  - Public use files
    - Data available on the website with or without subscription
## Types of microdata

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Secure use files (CIS, SES, MMD)</th>
<th>Scientific use files (all except MMD)</th>
<th>Public use files (EU-SILC, EU-LFS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approved research proposal</td>
<td>Approved research proposal</td>
<td>Approved research proposal</td>
<td>—</td>
</tr>
<tr>
<td>Access</td>
<td>In Eurostat safe centre</td>
<td>At researchers' workplace</td>
<td>Public</td>
</tr>
<tr>
<td>Data sensitivity</td>
<td>Only direct identifiers removed – highly confidential</td>
<td>Partially anonymised data – confidential</td>
<td>Data fully anonymised</td>
</tr>
<tr>
<td>Identification of respondent</td>
<td>Possible</td>
<td>Possible but difficult</td>
<td>Impossible</td>
</tr>
</tbody>
</table>
3. Microdata access workflow – researcher perspective
Recognition of the research entity (institutional eligibility)

Step 1
Institutional level (1 month)

Step 2
Researchers (2 months)

Is the entity doing research? (eligible to ask for access to microdata?)

Is the request for access well justified?

Research proposals

Eurostat

National statistical authorities (4 weeks)
Eligible research entities

- Doing research
- Making their results public
- Independent
- Safe
Eligible research proposals

- Scientific purpose well described
- Need for microdata justified
- Research results made public
- Data security measures in place
Assessment of research project proposal (RPP)

- **Eurostat:**
  - Microdata access team: initial check
  - Technical units: is the microdata appropriate for the planned research

- **MS:** 4 weeks
Research project proposals (RPP) received (2013-2017)
Main research subjects: EU-SILC

- Economic studies
- Impact of crisis on households situation
- (income, ethnic, geographical) inequalities
- Labour market studies
- (in work) poverty
- Mobility in Europe
- Euromod / tax-benefits models
- Wages, incomes
- Well-being, housing
- Youth
Public use files

- Available for:
  - EU-SILC and EU-LFS data available for some countries
What is EU-SILC

- The EU statistics on income and living conditions, abbreviated as EU-SILC, is the reference source for comparative statistics on income distribution and social inclusion in the European Union (EU).
- It is a sample survey conducted in 35 countries.
- It has a cross sectional and a longitudinal component.
- Every year more than 200,000 households and 500,000 persons are interviewed for EU-SILC.
Scope of the SILC Instrument

- Annual data: cross-sectional and longitudinal (4 year trajectories)
- Information on both households and individuals (micro level)
  - Income and tax
  - Material deprivation
  - Housing conditions
  - Employment, Childcare
  - Health, Education
- Output harmonization
  - Definition of target concepts/variables to be measured
  - Standardized output (format and content)
  - Common guidelines monitored by Eurostat
Specific topics

- Topics covered in the annual modules
  - 2005 and 2011: Intergenerational transmission of poverty;
  - 2006 and 2015: Social participation;
  - 2007 and 2012: Housing conditions;
  - 2008: Over-indebtedness and financial exclusion;
  - 2009 and 2014 Material deprivation;
  - 2010: Intra-household sharing of resources;
  - 2013: Well-being;
  - 2016: Access to services
  - 2017: Health and children health

- Variables included in the cross-sectional data

- EC-Regulation for each module
Organization of EU-SILC

- National flexibility in the implementation:
  - Source: survey and/or registers (DK, NL, SI, FI, SE, IS, NO)
  - Fieldwork: one-shot (≠ length) or continuous survey
  - Income reference period: fixed or moving 12-months (UK, IE)
  - Imputation method for income variables
  - Survey design

- Structure of the data files:
  Linkable 4 files per country and year:
  - **D-file** => Region, Deg_Urba, HH-weight, …
  - **H-file** => Income, Material Deprivation, …
  - **R-file** => Age, Gender, Childcare, Personal weights, …
  - **P-file** => Income, Health, Labour, Education, …
EU-SILC revision

In the framework of the modernization of social statistics, EU-SILC is undergoing a revision process to make the instrument more in line with users' need

Revision concerns:

- Content
- Timeliness
- Methodology
- Longitudinal component
Thank you for your attention!

Emilio.di-meglio@ec.europa.eu