

Access to European Microdata from Official Statistics: A Researcher's Perspective

Christof Wolf

Workshop: Decentralised Access to
European Microdata, June 17-18, 2010

Scope of presentation

- Confined to integrated OS microdata from Europe
- Confined to population/household surveys
- Only my personal view

Why do researchers want access to European microdata?

They want to understand social and economic structures across Europe

- descriptive
- analytical
- methodological

Advantages of OS microdata

Data were collected for policy and administrative needs, but:

- Large sample sizes
 - EU-LFS: 7,600 to 320,000 households
 - EU-SILC: 3,000 to 22,000 households
- Mostly individuals can be matched to households

Advantages of OS microdata

- High response rates
- Rich time series
 - EU-LFS microdata are available from 1983 onwards
- Longitudinal character of EU-SILC

Advantages of OS microdata

- Harmonized data over time and across countries
- International standard classifications [NACE, ISCO 88(Com), ISCED]

What do researchers want?

- Harmonized and integrated or easily linkable microdata
 - across countries
 - across time points
 - across surveys
- Rich structured documentation, metadata
- Timely

Currently available

- ECHP
- EU-LFS
- EU-SILC
- AES
- ICT (+SDS)

Eurostat's role

- Develop database schemes
- Develop harmonization schemes
- Collect and verify microdata from MS
- Collect and verify metadata from MS
- Integrate into common database
- Distribute data

Desirable improvement

- Data documentation
- Data harmonization
- Data anonymisation

Documentation, status quo

- User guides including descriptions of basic concepts, definitions and variables in the database are available in English
- On the internet: description of national sample designs, quality reports, EU-LFS: national questionnaires

Documentation, problems

- Information is not structured and scattered
 - Information is not available in English
 - Often Information is missing
- Quality of data cannot be fully assessed

Examples of country specific questions asking for supervisor-responsibilities in EU-LFS

Country	Supervisor - Question
Norway	Is it <u>part of your responsibility</u> to supervise or manage other employees?
Ireland	Do you supervise the work of other people <u>on a regular basis</u> ?
Austria	Do you work in a <u>leading position</u> ? (also applying to less qualified jobs!)
Sweden	Do your <u>tasks include managing and supervising</u> the work of other employees?
United Kingdom	In your job, do you have <u>formal responsibility</u> for supervising the work of other employees?
Czech Republic	Does s/he have <u>subordinate(s)</u> in the job?

Source: http://circa.europa.eu/irc/dsis/employment/info/data/eu_lfs/LFS_CORE_National_Questionnaires/index_LFS_CORE_Questionnaires_by_years.htm

Anonymisation

- High degree of anonymisation
- Considerable differences in the way EU-LFS and EU-SILC are anonymised:
EU-LFS is much more affected by aggregation than EU-SILC is

Anonymisation, examples

Age: 5-year age bands (year of birth is suppressed)

Marital status: widowed, divorced or separated are combined

Nationality and country of birth are aggregated in 3 groups

Anonymisation, examples

Occupation (ISCO):

aggregated to the 2-digits level

Professional status:

self employed with and without employees
are combined in a single category

Harmonization

- Output harmonization requires higher levels of documentation
 - Source question(s), natl. language & Englisch
 - Conversion rules
- Is often problematic for factual information
- Is impossible for subjective indicators

Harmonization

- Desirable to move towards input harmonization
- Provide exemplary blueprints for questions (see the ISSP experience)
- Desirable to move towards standardization of procedures between surveys

Conclusions

- Researchers recognize the value of OS microdata
- Urgently needed research on EU social and economic integration is reliant on OS microdata

Conclusions

- Data should be integrated or linkable
- Data must be harmonized
- Metadata could be better structured and more complete
- Anonymization procedures should be reconsidered

Conclusions

- Data should be available in a more detailed, less anonymized form; possibly through remote access or safe centers
- A closer collaboration between OS and researchers would be desirable

But:

- There has been tremendous improvement over the last years
- the progress seems to even accelerate

Many thanks to Eurostat and NSIs!