

European Union Statistics on Income and Living Conditions (EU-SILC)

This is a best practice example of data dissemination illustrating how sensitive data can be safely shared using a combination of measures which restricts access to the anonymised data and also restricts the content of the data.

When reading this case study it is important to remember that the anonymisation assessment, recommendations and actions are specific to the data examined here although they will have relevance to other similar data and data context.

- **Summary**

Organisation disseminating the data: Eurostat, via their 'User Database' (UDB).

Data: Annual data which is cross-sectional, longitudinal and multidimensional.

Data type: Microdata and tabulations released under access conditions.

Features of risk: This is a large dataset its combinations of several hundred variables make unique persons more common. Unique persons are potentially identifiable by spontaneous recognition. The dataset's longitudinal aspect also raises the risk of disclosure since household changes can be linked with data in public domain for example, births and deaths.

Risk type: Identification in sample under spontaneous recognition.

Disclosure Control Methods (SDC): Access to the data is only via research contracts and the data is restricted using variable suppression and global recoding.

Disclosure risk checks: Sample uniques.

- **Background**

The European Union Statistics on Income and Living Conditions (EU-SILC) is a cross-sectional and longitudinal multidimensional micro-dataset collected on income, poverty, social exclusion and living conditions. It is a pan-European survey made up of national samples in all the 27 Member States (MS). The member states individually prepare the EU-SILC data sets and supply these to Eurostat. In the UK, they are prepared by the ONS. The datasets are transmitted to Eurostat without any direct identifiers (e.g. name, address, fiscal numbers). In the UK, the source data for the EU-SILC has been the General Lifestyle Survey (GLF) but from the 2012 survey year it will be the Family Resources Survey (FRS), supported by a new longitudinal follow-up study, the Survey on Living

Conditions (SLC). These are both continuous surveys collecting information on people living in private households. The data are mainly at the person level with a few components at the household level.

The main purpose of the EU-SILC project is for the European Commission to assess and monitor poverty and social exclusion conditions across the member states. Structural indicators of social cohesion are compiled from the data.

Collection and dissemination is governed by European Commission regulation.

- **The Anonymisation Problem**

Eurostat has a strict policy on data disclosure and confidentiality. In addition, ONS has its own national safeguards agreed with Eurostat. The sample provided by the UK was small at less than 1% (8,000 households per sample out of up to 25 million in the UK) however there are very many variables provided in the microdata. The EU-SILC study is a four wave longitudinal survey with a sample size of approximately nine thousand cases per wave (minus attrition). Each wave is followed up for four years.

The spontaneous recognition scenario is used as the initial basis for determining individuals at risk of identification. Key variables are determined under this scenario and then Mu-Argus¹ used to identify records where there are fewer than three with that particular combination of key variable characteristics. This is done for the first wave only of individuals in the survey.

- **Anonymisation Practices**

Access to the anonymised microdata is provided by means of research contracts and generally restricted to universities, research institutes, National Statistical Institutes, central banks inside the EU and EEA.

Eurostat apply disclosure control to the data and publish it via their User Database (UDB). Together with the disclosure rules, the UDB is protected by licence which is equivalent to the UK Data Archive's (UKDA) Special Licence. There are several hundred variables on the database covering benefits, work status, household descriptors, health and illness indicators as well as responses to questions on income, crime, bills. There are two types of release:

- A cross sectional UDB containing the most recent wave of EU-SILC

¹ <http://neon.vb.cbs.nl/casc/mu.htm>

- A longitudinal UDB containing all 4 years trajectories available at the time of release.

Data are weighted from the samples. Disclosure control is then applied in the form of local suppressions applied to the records calculated at risk in Mu-Argus. Income is perturbed and replaced with a weighted mean. Occupation is banded due to very broad categorisation since when combined with geography it is possible to identify uniques. Extensive banding of other variables is applied and top-coding, some of these are country specific.

In addition to the rules specified by Eurostat, ONS considered the possibility of extra protection. This was because of the extra risk of disclosure posed by longitudinal data e.g. individuals whose marital status and/or other demographic variables change over the period of the survey. However, this was deemed not necessary due to the additional level of security given by the strict licensing conditions.

Estimates are also produced from the microdata which require a minimum number of observations.

An End User License file has been produced for the UK Data Archive. This file contains the anonymised cross-sectional element of the EU-SILC data but is released at higher levels of geography i.e. NUTS1 (country). In addition to this, month of birth is not included whilst other demographic detail is sufficiently protected, either by grouping variables (e.g. marital status) or by top/bottom coding (e.g. age, highest salaries). This file is not considered personal information and therefore appropriate for access under End User Licence.

- **Data sharing, Dissemination, Publication**

Anonymised EU-SILC data are available to approved researchers (UDB). A subset is available for training purposes. Also the cross-sectional elements of the EU-SILC data are lodged at the UKDA for access under End User Licence.

For further information please contact: Elaine Mackey at admin@ukanon.net