

## **Anonymisation method for SES microdata**

The methodology applied in the anonymisation of the SES microdata ('scientific use files') was as follows:

- ✓ Recodifying the categorical quasi-identifiers NACE, NUTS and SIZE of the enterprise in order to yield ratios between the number of sensitive combinations NACE-NUTS-SIZE (defined as those for which less than 3 enterprises exist in the member state's samples) and the total number of combinations in the SES lower than the threshold of 10% in a high majority of cases. The resulting codification mixes NACE sections, sub-sections or divisions and NUTS 0 or 1 levels as well as a maximum of 3 size classes (<50, 50 to 249 and 250+), depending on Member States;
- ✓ Removing Citizenship and performing global recoding on the age variable (2.2) to restrict its values to six intervals 14-19, 20-29, 30-39, 40-49, 50-59, 60+;
- ✓ Removing variable Key\_B identifying the enterprise and suppressing the grossing-up factor for local units;
- ✓ Ensuring an additional protection of employees using unconstrained individual ranking micro-aggregation for SES metric variables (absence days and earnings) by groups of (at least) 3 employees. This means that the latter variables are averaged for categories/combinations that include less than 3 employees in order to hide the information relating to individuals;

In addition, further specific anonymisation has been applied for some Member States (see: 'Additional aggregations of the SES anonymised microdata by country').

Finally, the Netherlands (2002 and 2006), Germany and Italy (all vintages) used their own national anonymisation methods as described in document attached to each respective country anonymised microdata.

Eurostat, Unit F3 "labour market statistics"