WORKSHOP ON LABOUR FORCE METHODOLOGY SURVEY

7 th WORKSHOP ON LABOUR FORCE SURVEY METHODOLOGY

DATA PROCESSING AND DATA QUALITY

Madrid, Spain, 10 – 11 May 2012

B. Data processing: Coding of economic activity, occupation and educational attainment

B3– Coding of economic activity, occupation and educational attainment using a new set of questions and new system of automatic codification

Hendrika Lautenbach – The Netherlands

Coding occupations
Hendrika Lautenbach, Statistics Netherlands
Workshop LFS methodology, May 2012

Why this subject?
Introduction of internet interviewing – partly dictated by the imposed cost reductions - forced Statistics Netherlands to look for ways of coding of variables, such as occupation, which does not require assistance of an interviewer nor a lot of memory space of the computer. Besides, it was considered desirable that coding of occupation should always be carried out in a similar way, irrespective of the mode of interviewing. This as to avoid differences in results by mode due to differences in coding device.

In 2011 Statistics Netherlands has carried out a pilot on a new system to code occupation. Currently this system is integrated in parallel run of the LFS which is carried out in advance of the introduction of internet data collection from the end of 2012 onwards.

Contents
A few years ago Statistics Netherlands started a project aimed to redesign the social surveys more efficiently in order to reduce costs. This was achieved in several ways. With respect to measuring occupations the following decisions that were taken in this project are relevant:

- A web based interviewing technique will be used next to the personal and telephone interviewing with the same questionnaire in all three modes;
- The total interviewing time had to be reduced;
- It was decided to give priority to the data that we are legally obliged to deliver;
- IT-policy aimed to use standard tools in our statistical process and minimize the costs for maintenance and use of custom-made software applications.

The Dutch LFS is in the process of change. For about a decade we used CAPI interviewing in the first wave, and CATI interviewing in all of the four consecutive waves. In 2010 the first wave has become multi-mode, that is, CAPI as well as CATI is used to collect the data. This year we intent to extent the first wave with internet data collection. The LFS is therefore a prime subject for testing and introducing a new way of coding of occupations.

In our current system, occupational codes are assigned during the interview interactively with the respondent in a semi-automatic way. Interviewers need a one-day training to get to know this system. Besides the fact that the current system requires the help of trained interviewers, the system is also unsuitable for implementation in the internet-mode as it requires a lot of memory space and time. Adjustment of the current system to make it suitable for internet interviewing would be very costly.

In the new system, coding will no longer be done during data collection. It takes place after the data has been collected. Coding takes place in 4 steps, and with every step it uses more information to determine the code. The first three steps

are performed automatically. Only if required, the fourth step that is manually coding of occupation is carried out.

The data on occupation gathered during interviewing – irrespective of the mode – is stored in a special database. Also data from other surveys which include data on occupation will be stored here. In an automatic way, the new coding system will run the coding program, using the data in the database, and assign ISCO codes to all the records for which it has sufficient data. Also the ISCO codes will be stored in the database. Only for the records which the system has not been able to assign a code automatically, these records will be dealt with manually. Still, the coding experts use the information available in the database and store the assigned ISCO codes also in the same database. When the coding process is finished, the codes are added to the rest of the survey data.

A great advantage of the new coding system is that it is applicable in all type of modes. This reduces the differences in results due to mode effects. Besides, it can be used across surveys and across countries. This harmonizes the cross-country results.