

The main objectives of the Working Group in question are as follows:

- exploring those administrative data sources which could be used for statistical purposes;
- developing a meta-data base to store the most important documentations, instructions and description of methodological procedures of the data sources;
- continuous analysis of the information received;
- submission of proposals to the management of the HCSO for using the administrative data for statistical purposes;
- elaboration of methodology, testing, proposals for linking of statistical and administrative data;
- clarification of data protection issues from legal and technical point of view;
- consultations with the management of the administrative data sources to consider statistical needs in maintaining their data.

Markedly there is a backlog in the field which has two reasons: one is the legal regulation of data transfers in the public administration, the other one is an internal concern of the statisticians as in the past the HCSO was not compelled to exploit administrative data sources to a greater extent. In parallel with this process the matching of data sources was also a neglected area. Even in the case of statistical files as own, matching has almost never been applied.

The literature distinguishes exact and statistical matching (Federal Committee on Statistical Methodology, 1980 and Winkler: Matching and Record Linkage, Business Survey Methods, 1995, John Wiley & Sons). An exact match is a linkage of data for the same unit from different files; linkages for units that are not the same occur because of error (e.g. a mistyped address). Statistical matching attempts to link files that have no or few units in common and the linkage is based on similar characteristics.

The paper gives an overview about the state of affairs, laying emphasis on matching. The HCSO is at the beginning of a long way to go on and many efforts need to be done to improve this situation. It must be mentioned however, that in the field of economic statistics there is well-established cooperation between the major agencies of the public administration in this respect.

In the lack of personal identification number no way at all to exactly match persons between the Census and another data source. An exact matching is limited to addresses, what is known as a potential source of error. In the case of certain social surveys the situation is somewhat better. Here the target sampling unit is a household within an address. Those countries which are in the position to use PIN codes, matching persons can be a daily routine. This is the case in economic statistics in Hungary also because for each business has an unique identification number (it is really the tax number) which is widely used in the public administration.