Towards a methodological framework for the integration of mobile phone data in the production of official statistics

Dr. David SALGADO

Head of Unit, Dept. Methodology and Development of Statistical Production, INE, *Spain*

Dr. Bogdan OANCEA

Director, Dept. of Innovative Tools in Official Statistics, INS, *Romania*

Abstract. Mobile phone data are widely recognized as a potential data source for producing official statistics in several fields such as population estimations or tourism statistics. In this paper we proposed some first elements for a methodological framework for the integration of mobile phone data in the production of official statistics, starting from an adaptation of the two-phase life-cycle model for statistical microdata used in the case of administrative data sources. Thus, our proposal considered all phases of the statistical production, beginning with the processing of raw telecommunication microdata into appropriate microdata sets, followed by an aggregation procedure according to a well-established methodology into data sets for each territorial cell and time instants, and a final inference stage from these aggregated data sets to the target population. We emphasized the differences between the treatment of administrative data and mobile phone data sets and we showed how GSBPM could be used to describe the statistical production with this new data source. For the final inference stage, we designed a hierarchical model borrowed from ecological sampling techniques, which is based on a Bayesian approach. This model combines the aggregated mobile phone data sets with other data sources such as a population register to produce estimates for the target population. We implemented this model in a software package and conducted experiments that allowed us to identify those areas that need further developments. Investigating the quality of the estimates we concluded that several principles from the European Statistics Code of Practice are affected since (i) an important part of the production process is carried out, at least at this moment and in the near future, by MNOs and (ii) the inferential paradigm is changed.