## New data sources for official statistics

Jan van den Brakel<sup>1</sup>

## Abstract

National statistical institutes are under increasing pressure to reduce administration costs and response burden for the production of official statistics. This could potentially be accomplished by using large data sets - so called big data. However, there are problems that must be addressed when using such data source for the production of official statistics. Two different research lines can be identified on how big data sources can be used in the production of official statistics. The first approach to be presented is to combine big data sources with sample data in a model-based inference approach. This implies that big-data sources are used as covariates in models used for small area estimation and time series models, where cross-sectional and temporal correlations are used to improve precision and timeliness of sample statistics. The second approach is to use big data sources as a primary data source for the compilations of official statistics. This can be considered if a big data source covers the intended target population and not suffer to much from under- and overcoverage, e.g. the use of satellite and areal images for deriving statistical information on land use. In most cases, however, adjustments for selection bias are required.

Keywords: Survey sampling, Small area estimation, now casting, non-probability samples.

<sup>&</sup>lt;sup>1</sup> Department of statistical methods, Statistics Netherlands. Department of Quantitative Economics, Maastricht University.