

# BIG DATA AS A DATA SOURCE FOR OFFICIAL STATISTICS

**Piet J.H. Daas<sup>1</sup> and Bart Buelens<sup>1</sup>**

<sup>1</sup> Statistics Netherlands, Heerlen, Netherlands

More and more data are being produced by an increasing number of electronic devices and sensors surrounding us, and on the internet. The large amount of data and the high frequency at which they are produced have resulted in the introduction of the term 'Big Data'. These data reflect many different aspects of our daily lives, they are abundant and often easily accessible and readily available. Therefore, Big Data sources are very interesting from an official statistics point of view. First experiences with analysing Big Data at Statistics Netherlands were obtained studying large amounts of Dutch traffic loop detection records, call detail records of mobile phones and Dutch social media messages. This research has revealed that a number of challenges need to be addressed to enable the application of these data sources for official statistics. In particular, the potential selectivity of Big Data is one of the major concerns. The data generating mechanisms underlying Big Data vary widely, but have in common that they are very different from probability sampling, the data collection strategy ordinarily used at NSIs. Assessment of selectivity of Big Data sets is generally not straightforward, if at all possible. Some approaches will be discussed. It is argued that the degree to which selectivity—or its assessment—is an issue, depends on the way the data are used for production of statistics. The role Big Data can play in that process ranges from minor over supplementary to vital. Methods for inference that are in part or wholly based on Big Data need to be developed, with particular attention to their capabilities of dealing with or correcting for selectivity of Big Data. The current views on these matters at Statistics Netherlands are discussed including further considerations and research.

**Keywords:** Alternative data sources, selectivity, inference.

## References:

- Buelens, B., Daas, P.J.H., Burger, J., Puts, M.J., van den Brakel, J. (2014). Selectivity of Big Data. Discussion Paper 2014—11, Statistics Netherlands, Heerlen/The Hague.
- Daas, P.J.H., Puts, M.J., Buelens, B., van den Hurk, P.A.M. (2013). Big Data and Official Statistics. Paper for the 2013 New Techniques and Technologies for Statistics conference, Brussels, Belgium.