

It is widely recognized that big data have a high potential for official statistics. As a recent worldwide UN survey has shown, many initiatives have been started at the national as well as the international level to realize this potential. Most initiatives are aimed at exploring big data sources. So far only a few have led to actual dissemination of statistics that are based on big data. Referring to Dutch experiences with big data in explorative research as well as official statistics, this overview lecture aims at showing the main issues and possible ways of dealing with the challenges.

The following aspects of dealing with big data will be discussed: getting access to big data sources; data exploration; privacy aspects; data processing and IT aspects; methodological issues such as selectivity, integration of big data with other data sources, and the use of model-based techniques; and data visualization. The challenges also have an organizational dimension, for instance regarding the launching and fitting of big data activities in an existing organization, acquiring the knowledge and skills needed, and collaboration with private and non-private partners. There may be important cultural and strategic implications. The presentation will cover these aspects as well.

The examples used as a reference concern the use of (1) traffic sensor data for statistics on road traffic intensities, (2) mobile phone location data for following the spatial distribution of a population, (3) social media data for estimating sentiment, (4) scanner data for price statistics and (5) use of website data for statistics production.