



## Global Conference on Big Data for Official Statistics

20-22 October 2015, Abu Dhabi, UAE

# Classification of Big Data

*Peter Struijs*

*Statistics Netherlands*

*[p.struijs@cbs.nl](mailto:p.struijs@cbs.nl)*

# Approach

- Subject of the classification
- Scope of the classification
- Characteristics / classification criteria
- One or more levels of classes

Design depends on **intended uses**

# Uses of Big Data Classifications

- Uses of the 2013 UNECE classification
  - as an extensive definition of Big Data
  - 2014 UNECE Big Data project
  - 2014 UNSD/UNECE Big Data survey
- Uses of other Big Data lists, e.g.
  - list used for 2015 UNSD Big Data survey
  - list used for ESSnet on Big Data
  - ad hoc lists of Big Data sources
- Foreseen uses
  - in context of SDG indicators
  - provide structure for repository of Big Data projects

# Scope and Subject

## ➤ Definition UNECE (March 2013):

*Big Data are data sources that can be – generally – described as: high volume, velocity and variety of data that demand cost-effective, innovative forms of processing for enhanced insight and decision making.*

## ➤ Issues:

- the definition does not provide decision rules
- what is one data source?
- what data is included in the source?
- what does this mean for heterogeneity / homogeneity?

# Classification Criteria (1)

## 1. Big Data characteristics:

- high volume
- high velocity
- high variety
- high veracity
- selectivity
- (lack of) structure
- high population dynamics
- event based

# Classification Criteria (2)

2. Local versus global sources
3. Regulatory framework
4. Is Big Data by-product?
5. Purpose and subject of the data
6. Original versus derived sources
7. Type of relationship with data
8. Public versus private sources

# Classification Criteria (3)

9. Human or machine sourced
10. Stability of the source
11. Accessibility
12. Real-time versus accumulated data
13. Applicable statistical methodology
14. Statistical domains of use
15. Usability for SDI indicators

# The Future

- User assessment of approach and classification criteria
- Construct new classification, building on 2013 UNECE classification
- Investigate possibilities of a flexible, evolving classification system



# What is Your View?



<http://unstats.un.org/unsd/trade/events/2015/abudhabi/gwg-agenda.asp>

Thank you for your attention!