

# Introduction to Topic: Data

---

## *Which data?*

Census, admin, geospatial, big, and other alternative data

“This topic aims to highlight specific features of various data other than that from sample surveys, and how to address their challenges to editing and imputation.” — Information notice No. 1

## *Which features and associated challenges?*

- Innovative data collection methods
- Big volume
- Combination of sources
  - Multiple types, complicated, created units
  - Overlapping/missing, latent variables
- Quality management and assessment

# Introduction to Topic: Data

---

## Innovative data collection methods

- mode: web, tablet, telephone, paper, interviewer...
- type: number, text, html, map, image, speech...
- key: id, address, name, coordinates...

## Big volume

- automatic processing methods
- software development / adaptation
- macro editing: geographic & demographic divisions
- hierarchy of aggregation
- ...

## Introduction to Topic: Data

---

### Combination of sources: Units

- record linkage techniques and associated errors
- alignment of units: e.g. persons and multi-type adrs
- composite/created units, e.g. household, enterprise

### Combination of sources: Variables

- definitional/numerical inconsistency between sources
- creation of latent variables

### Quality management and assessment

- Process: GSDEM flow models
- Output: uncertainty at individual or aggregated level