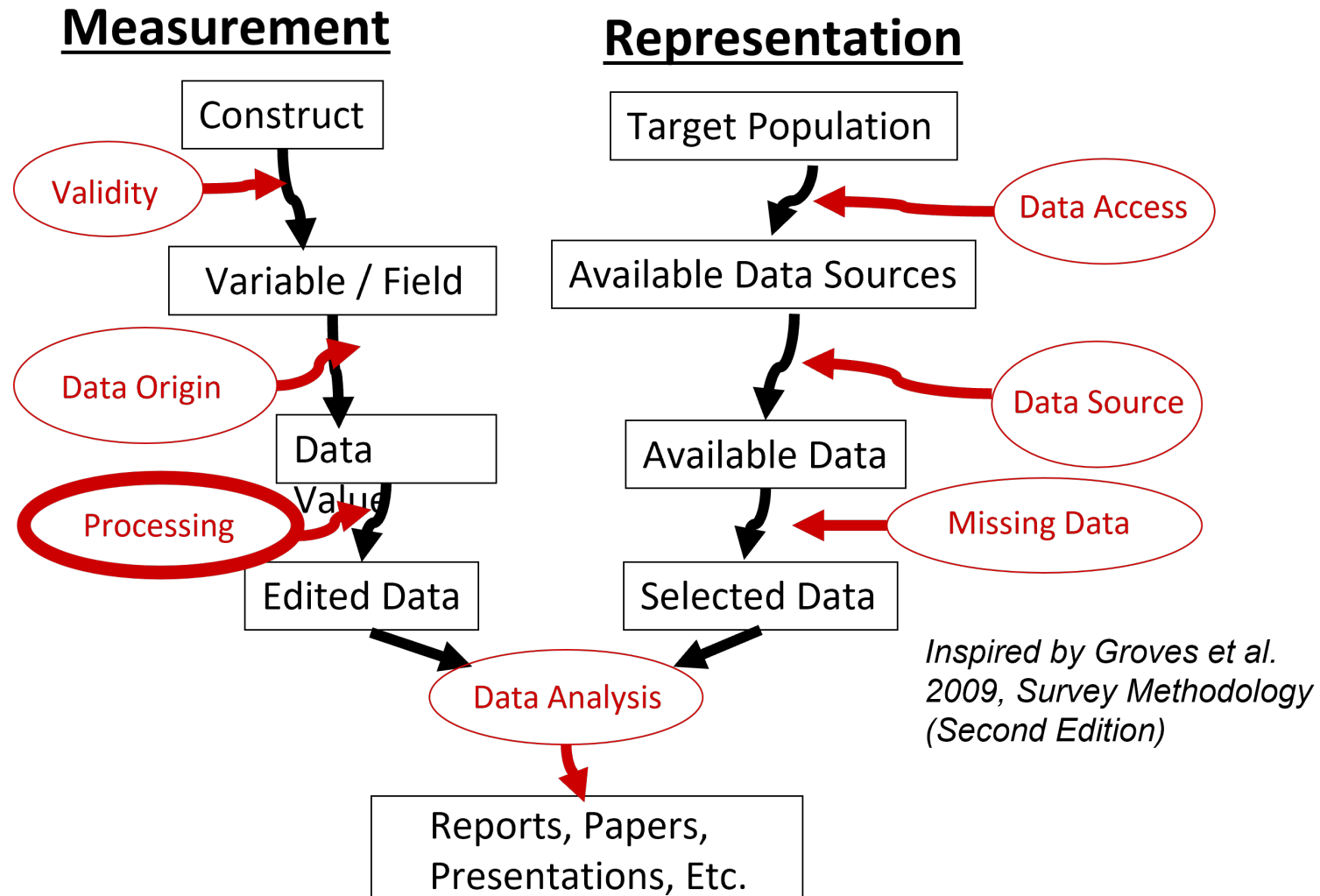


Defining Data Processing

By Jinseok Kim

Dimensions of TDQ: The Big Picture!



What is Data Processing?

- **Data processing** is a series of operations to collect, convert, and manipulate data into formats that are usable for extraction of meaningful information.
- In a broad sense, it includes **data analysis**.
- But usually refers to the early stage between data collection and data analysis.
 - In computer and information science, **data pre-processing** is used interchangeably.

Data Processing in Designed Data (1)

- Major data processing operations include:
 - Data editing
 - Recoding
 - Cleaning
 - Data linkage
 - Imputation of missing values
 - Calculation of weights

Data Processing in Designed Data (2)

- Major data processing operations include:
 - Data parsing
 - Character Encoding
 - Cleaning
 - Data integration
 - Transformation (e.g., Rescaling, dichotomization, normalization, etc.)
 - Feature engineering (e.g., Combining existing features into one, transforming an existing feature into a new one, etc.)

What's Next?

- In the next lecture, we will take a look at threats concerning data processing for designed and gathered data.



© Faculty Presenter

**Except where otherwise noted, this
work is licensed under CC BY-NC 4.0**