The Total Data Quality Framework
By Brady T. West

So What Is Total Data Quality?

- For both designed and organic/gathered data, **Total Data Quality (TDQ)** refers to a **scientific framework** for measuring and optimizing the overall quality of the data collected, along ***multiple dimensions***.
- This framework allows for ***careful thinking*** about how to maximize the quality of the data collected in a given research study, and how to clearly measure the quality of the data collected.

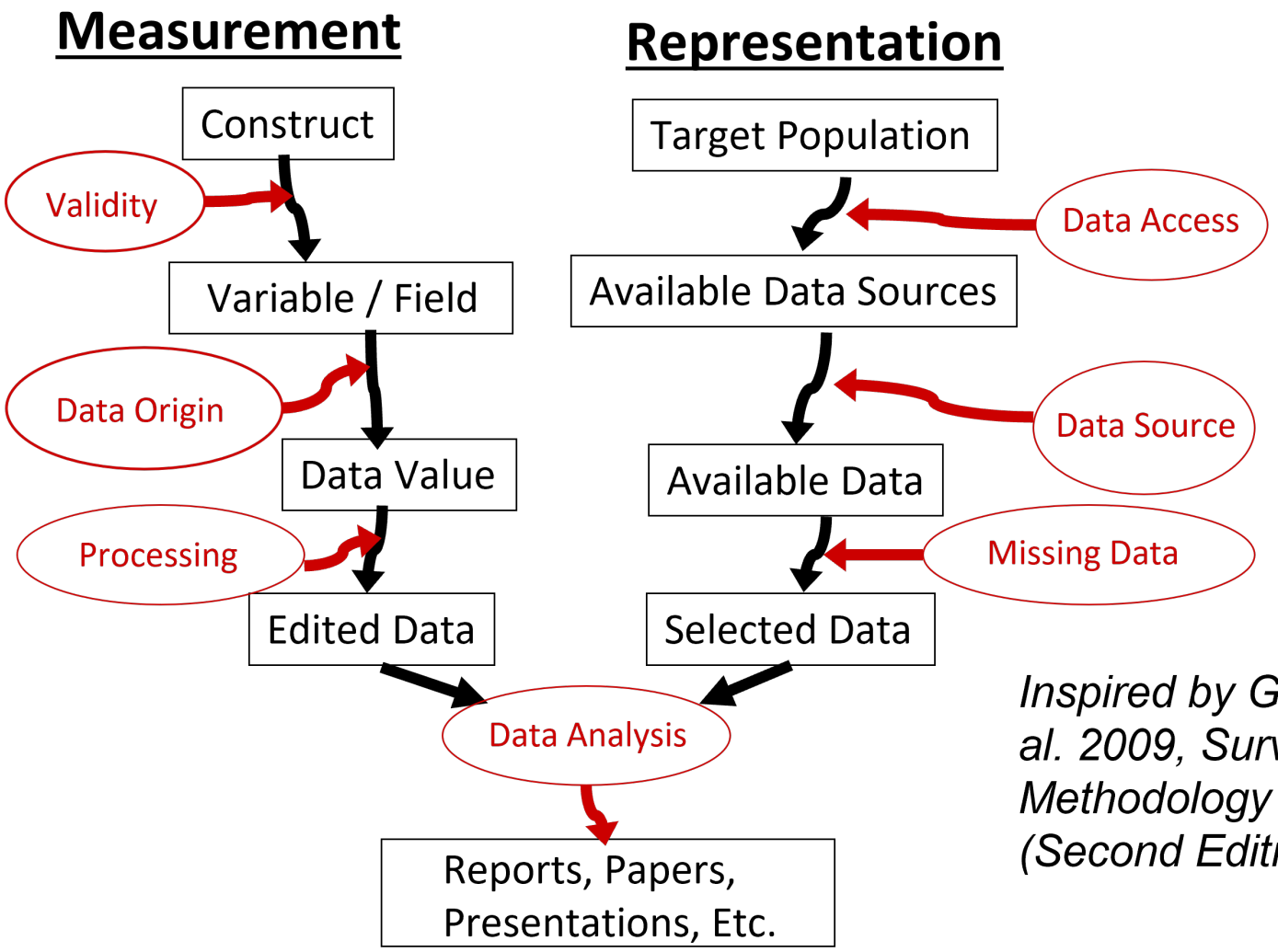
Dimensions of TDQ (1)

- We generally divide dimensions of TDQ into two categories: **measurement** and **representation**
- **Measurement** dimensions concern the quality of the values on the variables that one wishes to analyze, and these include:
  - the **validity** of the variables of interest,
  - the **origin** of the values for the variables of interest, and
  - the procedures used to **process** the variables in preparation for analysis

Dimensions of TDQ (2)

- ***Representation*** dimensions concern steps in data collection / gathering that may result in the ultimate data set looking different from the population of research interest.
- These include:
  - the process used to **access** the data,
  - the **source** of the data,
  - mechanisms producing **missing data**, and
  - techniques used for **data analysis**.

Dimensions of TDQ: The Big Picture!



Inspired by Groves et al. 2009, Survey Methodology (Second Edition)

What’s Next?

- Next week, we will take a deep dive into the concepts of **validity**, **data origin**, and **data processing** as important measurement dimensions of TDQ!

UNIVERSITY OF MICHIGAN