Dimensions of TDQ: The Big Picture!
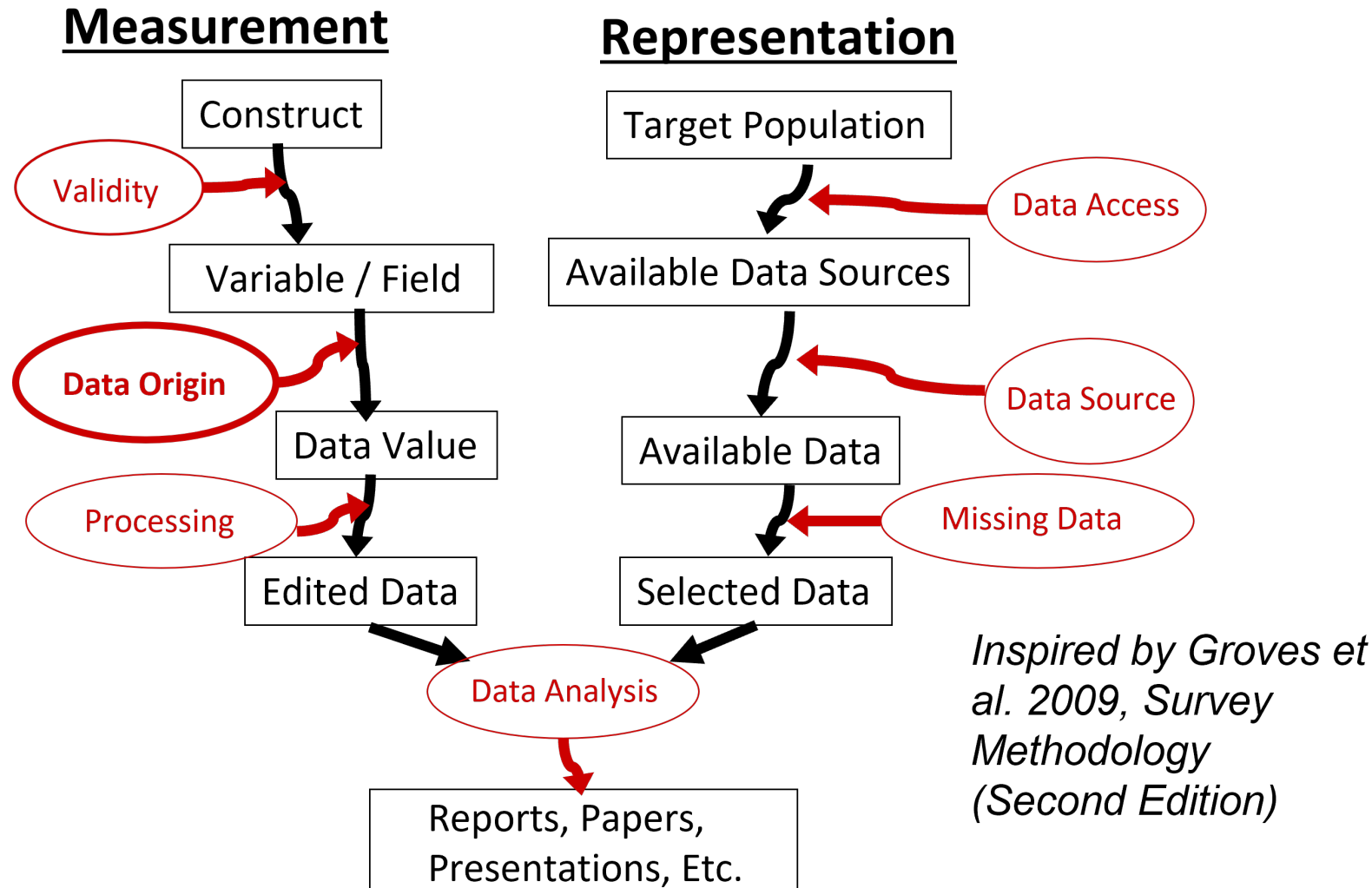
Data Origin: Definition

- **Data Origin:** How were the individual values / data points for a given variable (or field):
  - recorded,
  - captured,
  - gathered,
  - computed, or
  - represented

  …for the units of analysis?

- Were there any errors in the process that ultimately produced the value for the variable to be analyzed?

Examples: Designed Data

- **Interviewer Effects:** Different interviewers collecting survey responses may ask questions differently or not follow protocols, ultimately affecting the responses collected

- **Social Desirability:** Survey respondents may be less likely to provide honest answers if they are not socially desirable (e.g., history of excessive drug use, large # of sexual partners)

- **Reporting Error:** Web survey respondents enter incorrect information for factual questions intentionally or by accident

Examples: Gathered Data

- Generally, data origin problems arise in **gathered data** due to issues with the process of **creating** or **retrieving** existing values for fields of interest:
    - Computing the average of several organic measures (what if some values are missing?)
    - Extracting data from an existing database to populate a field (what if the incorrect information is extracted?)
- We will explore other types of threats this week!

What's Next?

- We will talk about threats to quality regarding the origin of the data for both designed and gathered data

- We will examine case studies, looking at **interviewer effects** and **measuring COVID-19 exposure** in Georgia

- Then, we will turn to the **data processing** measurement dimension of the TDQ framework

UNIVERSITY OF MICHIGAN