

# Metadata Management Variables and Datasets

Jon Johnson

12 September 2019

CLOSER, UCL Institute of Education

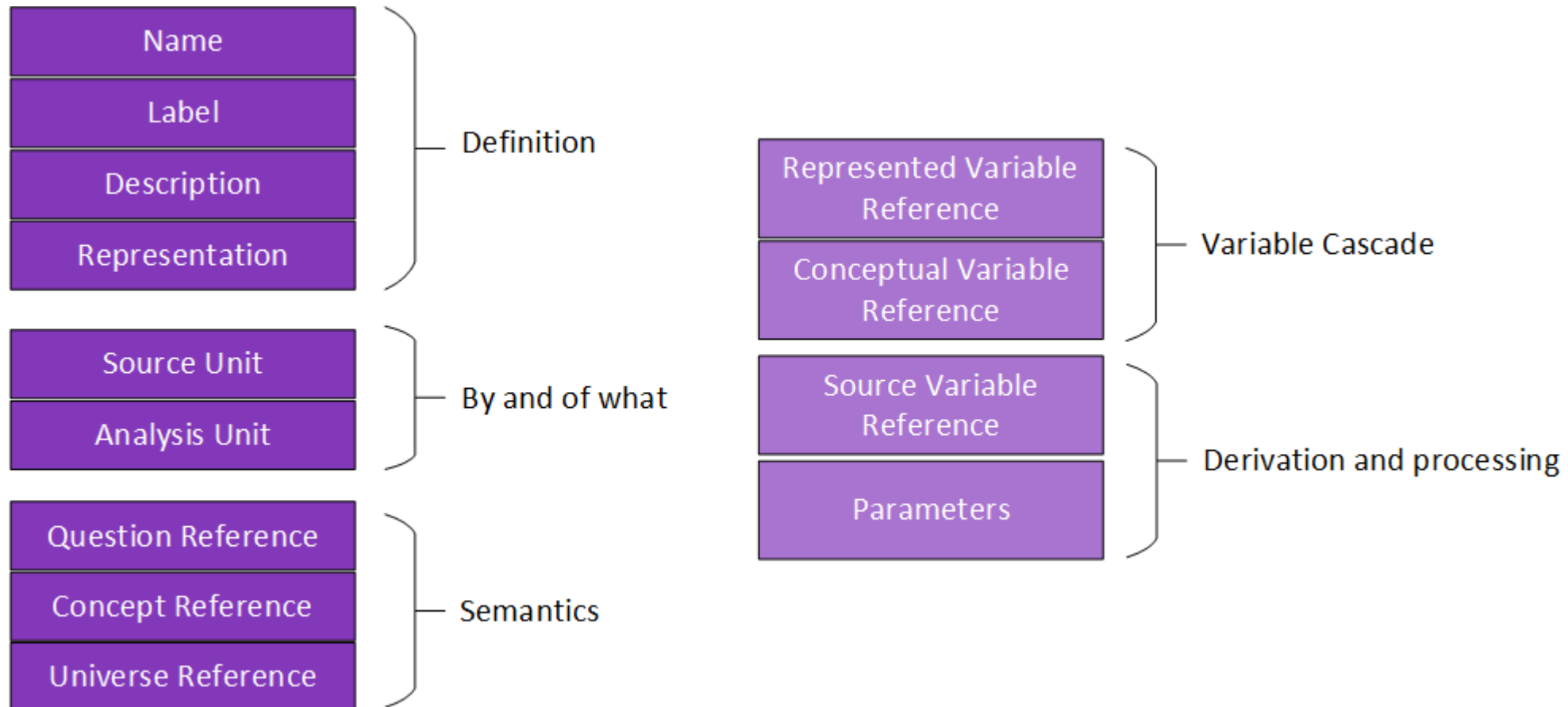


# Overview

- What is a variable
- Variable organisation
- Datasets

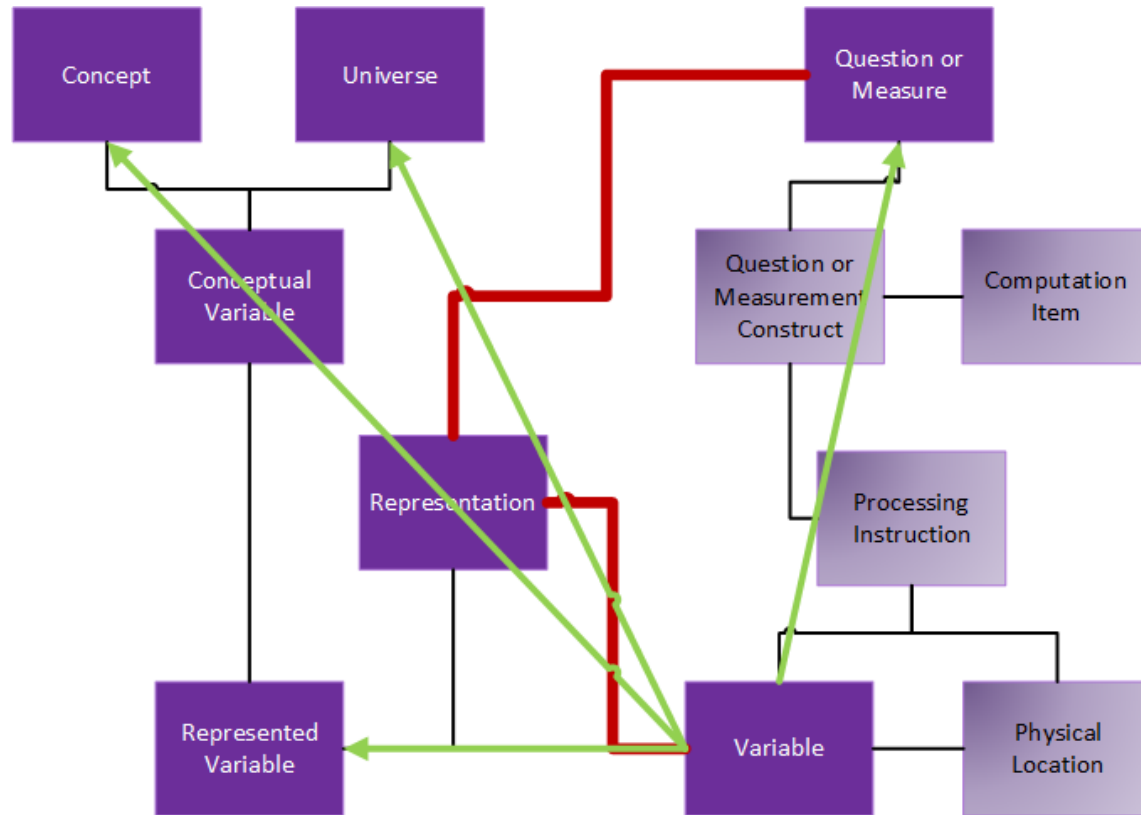


# What is a Variable





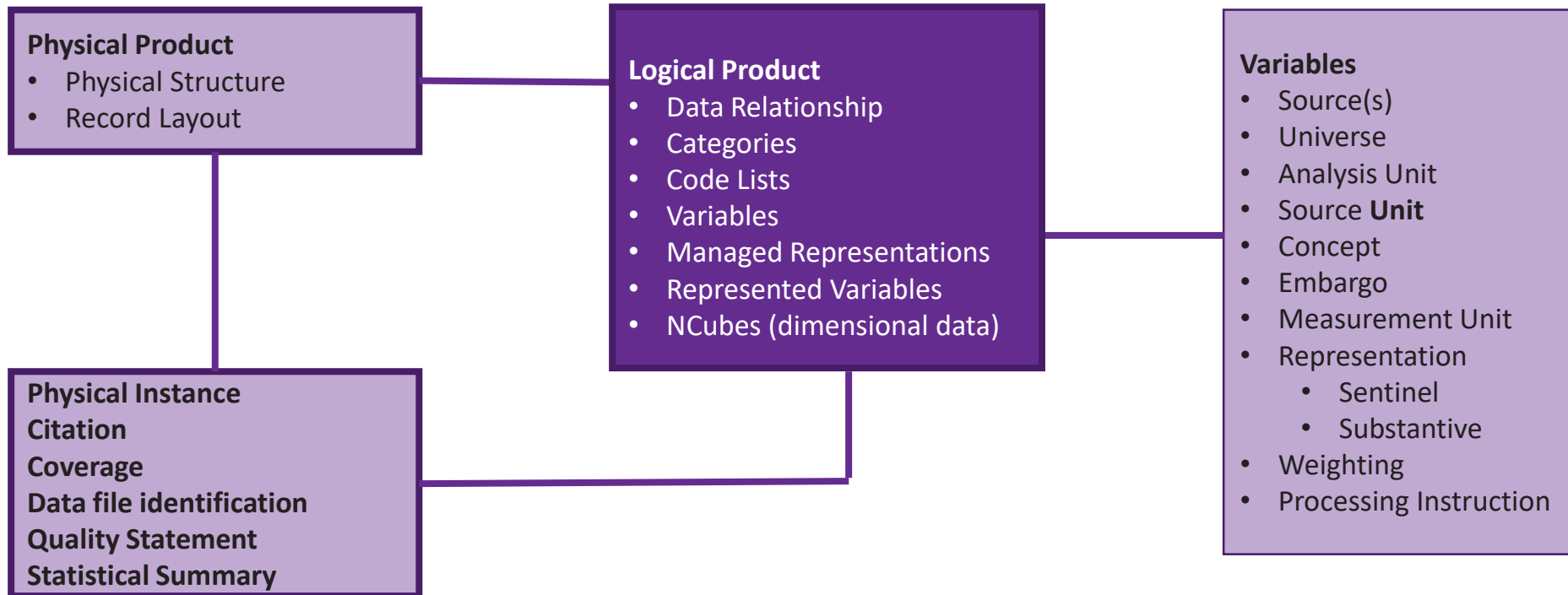
# Variable Relationships



- Variables share **representation structures** with questions, e.g. code, numeric, strings, scales (not the same response domains)
- It inherits the concept and universe through the variable cascade. If the variable is ‘created’ from a question or measure – these will be identical.

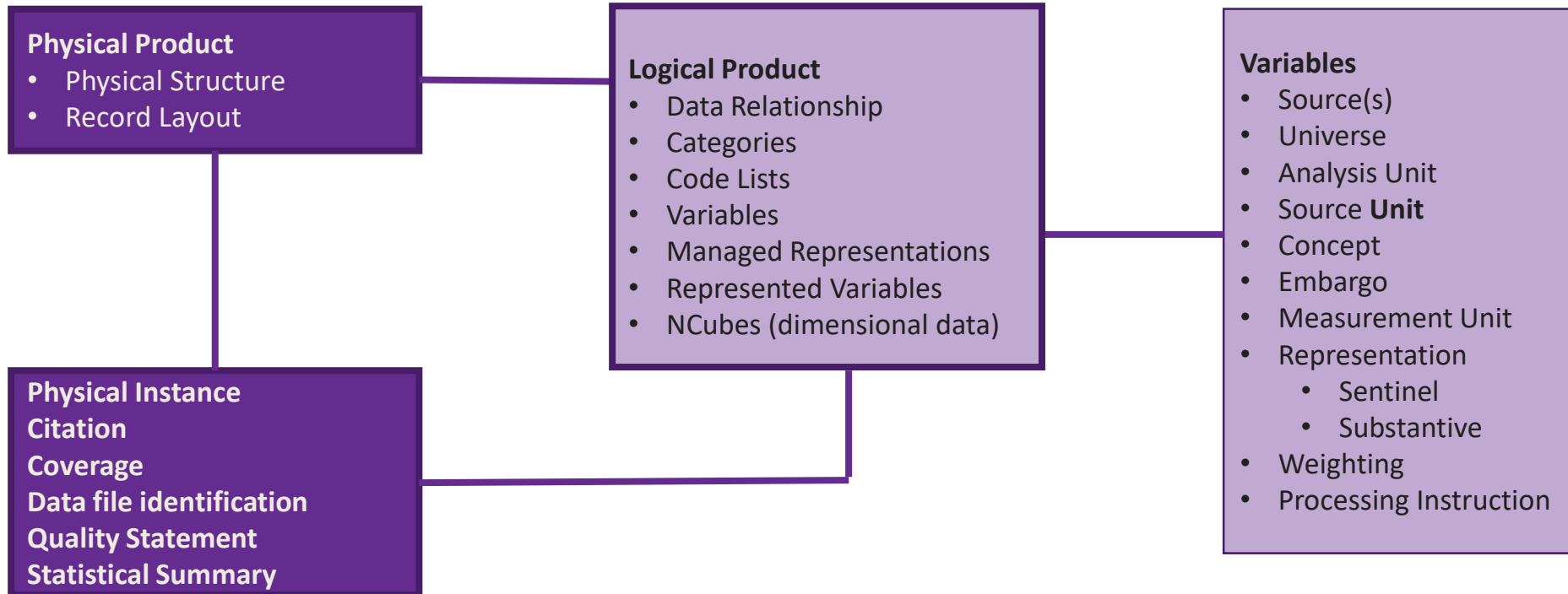


# Variable organisation





# The data itself





# Why is it so complex?

- Historically, data has been organized and formatted in many different ways
  - With multiple formats of the same data
  - Large numbers of files with identical structures/record layouts
- This implies a lot of potential reuse
- DDI provides a single, flexible way for describing a large variety of complex data files
- Typically, these constructs are created by automated tools from existing proprietary formats (SAS, SPSS, etc.)
- You do not need to master this complexity in many cases



# Benefits

- Flexible reorganisation of data that is format agnostic
- Standardised machine actionable metadata that is software independent
- Allows you to perform automated analysis, processing and/or transformation
- Attaches semantics to the data which could support 'intelligent' discovery and processing