Q 2016
European Conference on Quality in Official Statistics

Big Data: from noise to evidence

Session 40

Friday 3 June

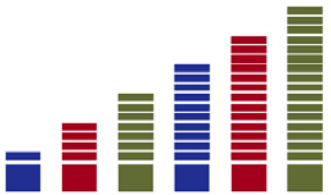
Patricia O'Hara
Poharaca@gmail.com
Pieter Everaers
Pieter.everaers@ec.europa.eu

INE Instituto Nacional de Estadística
eurostat
Madrid, May 31 - June 3

Q 2016
European Conference on Quality in Official Statistics

Key Points

- Big Data will require changes in the way official statistics are produced
- Implications for quality control and monitoring - relationships with data providers
- Modification to ES-CoP
- Adaption of role of ESGAB



Theory, operationalisation and empirical generalisation

Theory - Hypotheses

Operationalisation - Observations

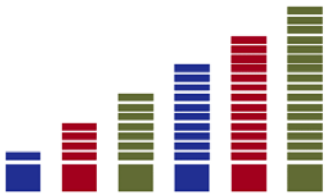
Empirical generalisation – ‘Reality’

Theory and Data

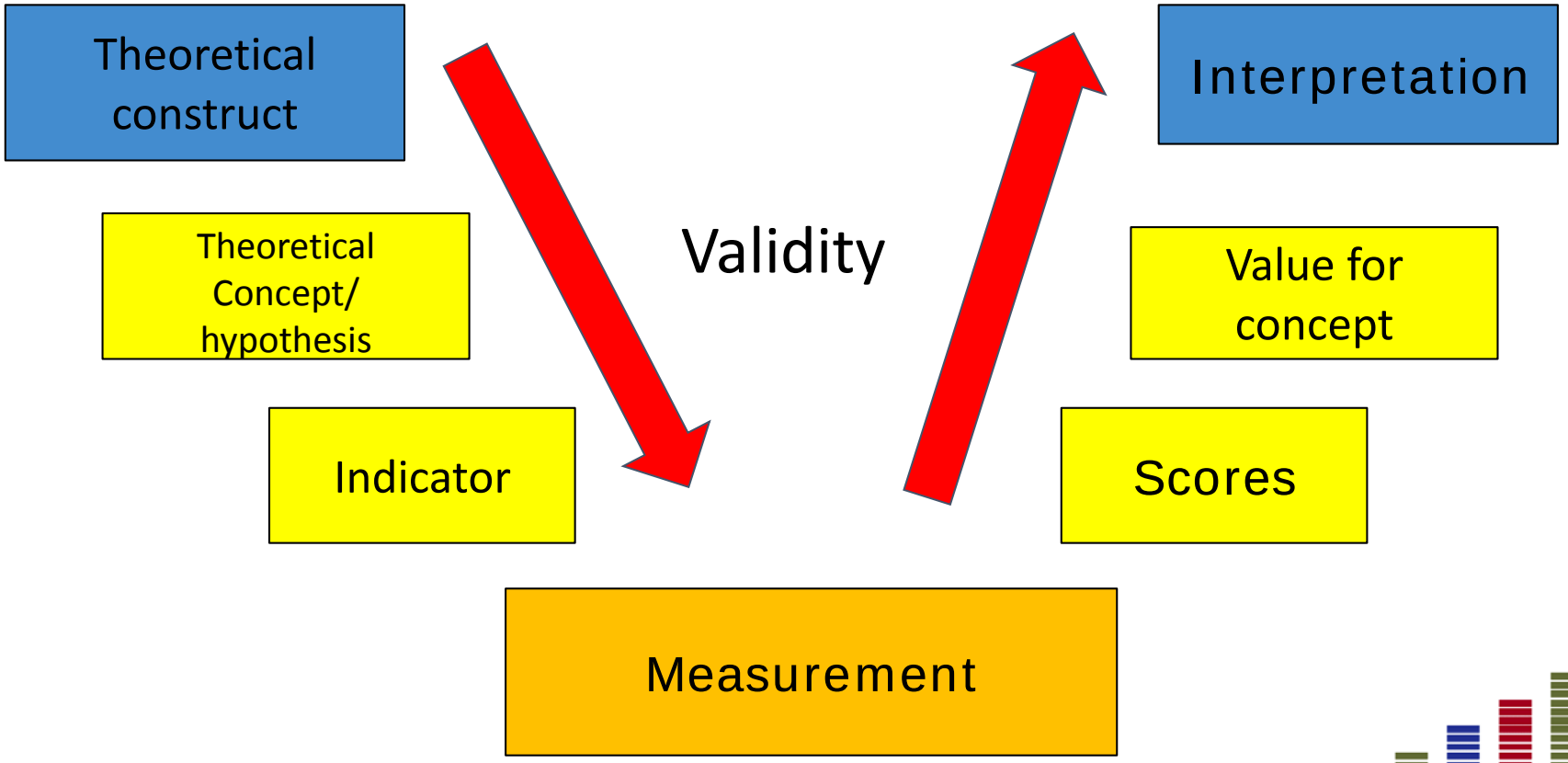
Theory: idea of reality involving concepts

How do we know that the data are **valid** i.e. a 'true/genuine' reflection of reality?

Data: reality as perceived and measured



Logic of traditional socio-economic/statistical research process



Q 2016
European Conference on Quality in Official Statistics

From Big Data to concepts: from noise to valid statistics

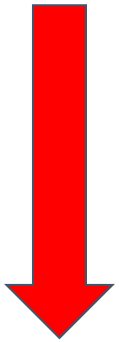
Big Dataset: Noise

Traditional dataset

Scores for specific items

Correlation

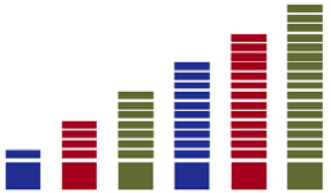
Scores for well known variables



Concept based on Big Data

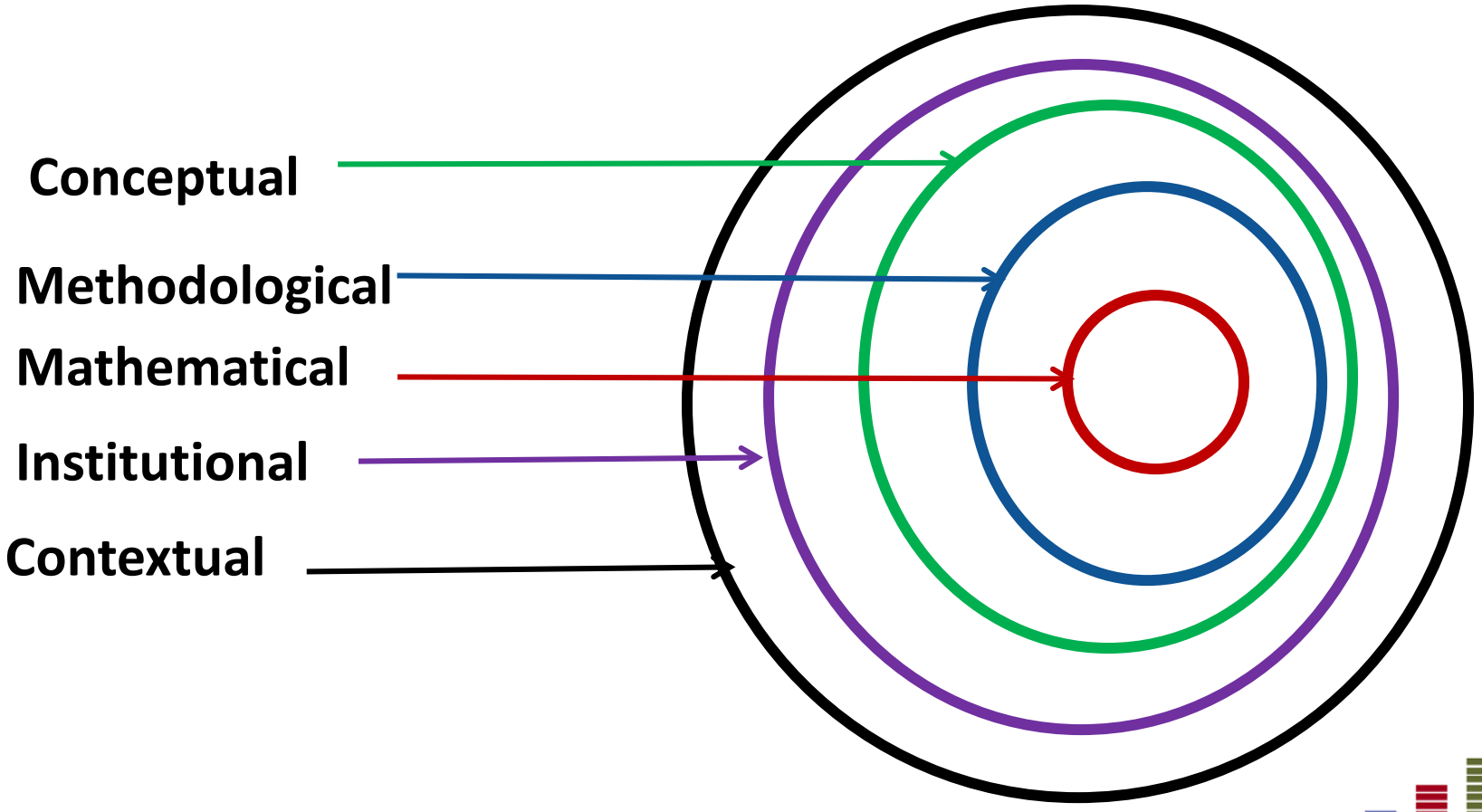
Existing concept

Similar



Q 2016
European Conference on Quality in Official Statistics

Quality levels in statistics



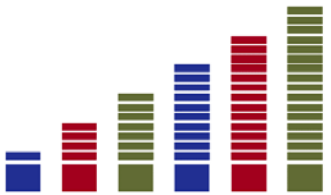
Q 2016
European Conference on Quality in Official Statistics

Big Data and compliance to ES Quality and CoP requirements

Conceptual quality – Need to prove that Big Data based variables are valid representations of agreed/accepted concepts

Institutional quality – Adequacy of governance structures of the institutions handling Big Data-based variables

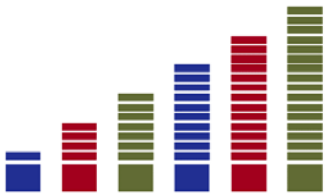
Contextual quality – Compliance with CoP and acceptance by highest authorities



Q 2016
European Conference on Quality in Official Statistics

Implications for ES-CoP and ESGAB

Conceptual quality – analytical work needed to assure 'validity' of measures/data i.e. ex post validation

Institutional quality – expansion of ES-CoP principles for big data handlers particularly regarding cooperation with producers

Contextual quality- ESGAB to assess compliance with governance structures and revised/updated ES-CoP