Alternative sources: Best practice for deciding fitness for use

Scott Kilbey, Stephen Babos, Ben Bradbury & Ben Jones Office for National Statistics

What do we need?

To produce Statistics for the Public Good, analysts must be able to access information on all available data sources

What data sources are currently used?

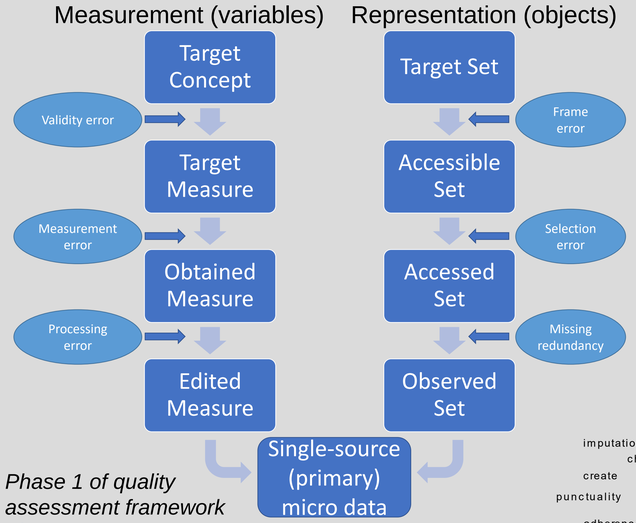
Which sources are of highest quality?

Transparency and targeted acquisitions

Development of a data source quality framework

Problem: There is no internationally recognised method for assessing the quality of administrative data sources

What is current best practice?



Phase 1 of quality assessment framework
Zhang 2012

Their framework provides 25 numerical metrics to assess data sources in isolation within the six categories above

In 2018, Stats NZ published a data quality framework for both survey and admin data. This builds on the 'Total Survey Error' framework (Groves et al. 2004) and Statistics Norway's work on 'statistical theory for register-based statistics' (Zhang, 2012)

Visualisation of the 25 metrics

What is next?

- Apply phase 2 of the framework to derived data sets
- How do we weight aspects of quality?
- Consider use case of sources: benchmark vs pattern
- Greater transparency of data source journey
- Use of quality metrics to inform data source confrontation

How will we move things forward?

Source of Error: Category	Data Quality Metric	Description	Trend
Validity	Percent of inconsistent records	Errors identified that can't be reconciled	
Measuremer	Percent of units which fail checks	Percentage of items that fail automated checks	
Procesing	Modification rate	Percentage of returns manually cleared after identified errors - no indication of changes	
Frame	Overcoverage	Zero tolerance for duplicates in 'batch tests'	
Selection	Adherance to reporting period	Percentage of returns identified as unacceptable reporting period	
Missing/ redunacy	Unit non-response rate	Factored into planning	

At the ONS we are using this framework to assess data sources that feed our Supply and Use Tables. Above shows the data quality trends in relation to one data source over 7 years. We aim to analyse multiple sources and display the quality of different sections of this complex output (see mock up below)

Dashboard of data sources. Colour scale indicates quality and tree map indicates monetary contribution to SUT. [Note: several values here are demonstrative placeholders at present time]

Office for National Statistics

ESG.Data.Sources@ons.gov.uk