



United Nations Economic Commission for Europe
Statistical Division

Workshop on the Modernisation of Statistical Production and Services

November 19-20, 2014

The Role of Big Data in the Modernisation of Statistical Production and Services



Results of the Task Team Quality

Peter Struijs

Statistics Netherlands



Members of the Task Team

- David Dufty, chair, Australia 
- H el ene B erard and Laurie Reedman, Canada 
- Sylvie Lefranc, France 
- Marina Signore, Italy 
- Juan Mu oz and Enrique Ordaz, Mexico 
- Peter Struijs, Netherlands 
- Jacek Ma lankowski and Dominik Rozkrut, Poland 
- Boro Nikic, Slovenia 
- Ronald Jansen and Karoly Kovacs, UNSD 



The tasks

- Quality framework(s) for Big Data.
- Testing the framework(s).
- Indicators and associated metadata requirements.

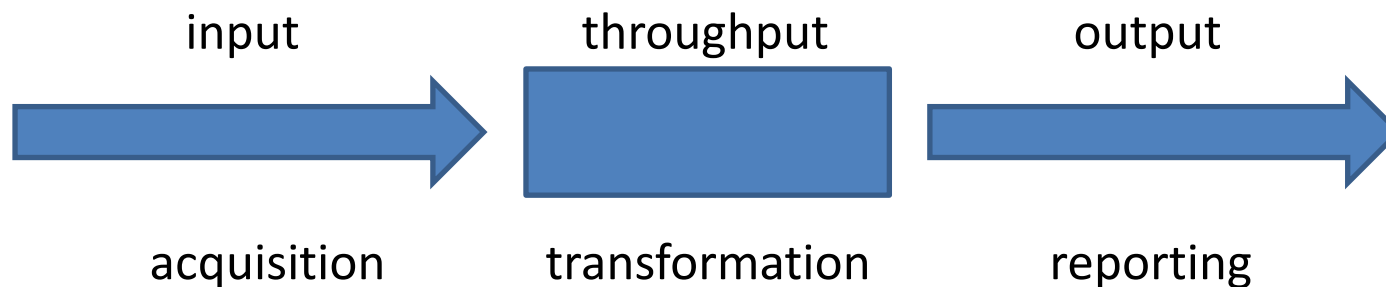


General approach

Taks: *To provide a framework for NSOs for assessing the quality of Big Data*

Quality: *To be evaluated in light of intended use ('fitness for use')*

Business process:



Framework: *For each phase, define appropriate dimensions and indicators*



Hyperdimensions

The concept of **hyperdimension** was taken from the administrative data quality framework.

Source: *Related to the type of data, the entity from which the data is obtained, and how it is administered and regulated.*

Metadata: *Description of concepts, file contents, and processes.*

Data: *Relates to quality of the data itself.*



Input dimensions

Source

Institutional factors
Privacy & security

Metadata

Complexity
Completeness
Usability
Timeliness

Data

Total survey error approach (esp.
coverage and measurement error)
Linkability



Principles of processing

- **System independence:** *Result of data processing is independent of hardware and software systems used.*
- **Application of the quality hyperdimensions:** *Full range of dimensions of source, metadata and data may apply, depending on context.*
- **Steady states:** *Stable versions of data that can be referenced by future processes and multiple parts of the organisation.*
- **Quality gates:** *To be employed as a quality control business process.*



Output dimensions

Source:	Institutional environment <i>Confidentiality</i>
Metadata:	Accessibility Clarity Relevance <i>Complexity</i>
Data:	Accuracy Coherence and consistency Timeliness and punctuality <i>Validity (of concepts, of models)</i>



Conclusions

- There is a need for quality assessment covering the **entire business process**.
- **Input quality** can be explored and assessed by using and elaborating existing input quality frameworks.
- **Throughput quality** can be maintained by following quality processing principles
- **Throughput quality dimensions** need to be further developed for Big Data processing.
- Additions have been proposed to **output quality dimensions** from existing frameworks, to make them suitable for Big Data applications. This needs further testing.

