Multisource Statistics
Overview of new approaches

Thomas Burg
Head of Quality Management, Methods and Classification

Workshop on methodologies for Official Statistics
Session 2: **Methodologies for multisource processes**

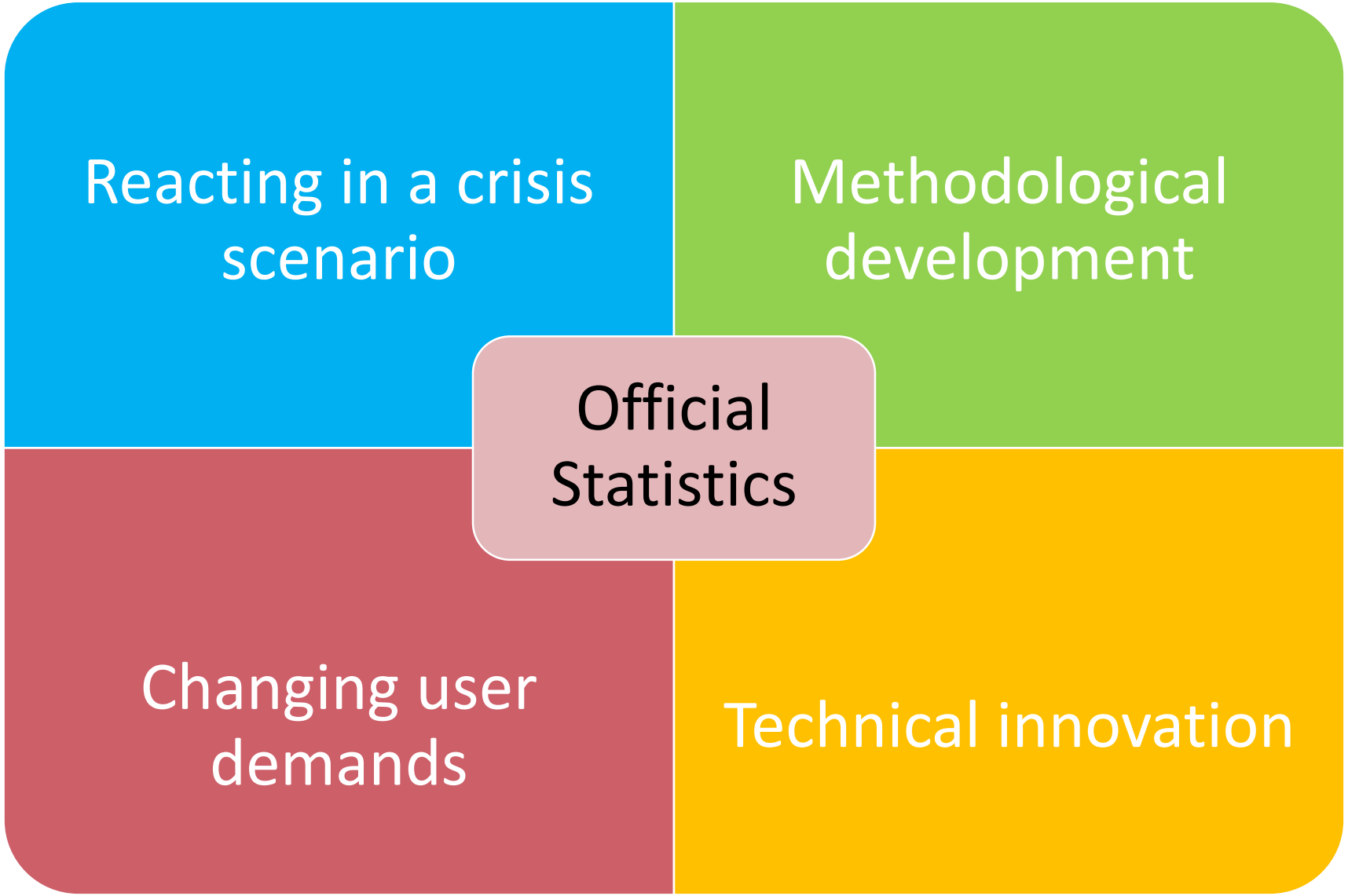
December 5th 2022

www.statistik.at

Content

- New challenges when producing official statistics
- Multisource scenario
- Approaches
- Research questions

Challenges for official Statistics



New challenges –Rreacting in crisis scenario

- Emerging data needs
- New data sources have to be deployed and combined urgently
- Playing a new role in the national data eco-systems
- Challenging situation regarding confidentiality
- Problems in data collection

New challenges – Methodological development

- Making use of all available data sources
- New methods for using privately held data
- Develop new methods in the multisource scenario
  - Estimation and calibration
  - Quality assessment
  - Estimation
  - Data integration

New challenges – Increasing user demands

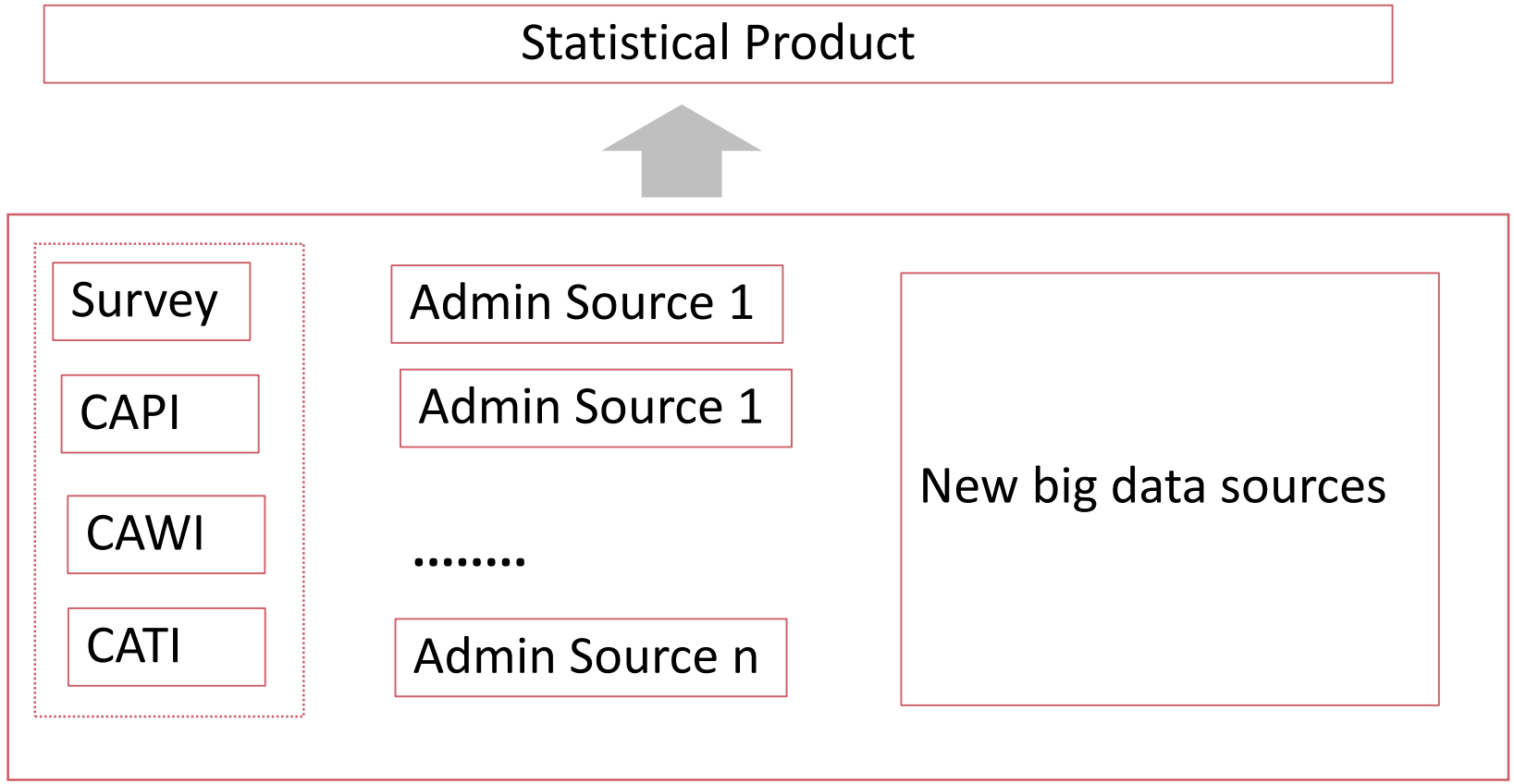
- Users want to make use of data a soon as possible
- There are other producers of statistical figures
- The perception on quality has changed
- Desire to make automated use of data has increased

New challenges – Technological innovation

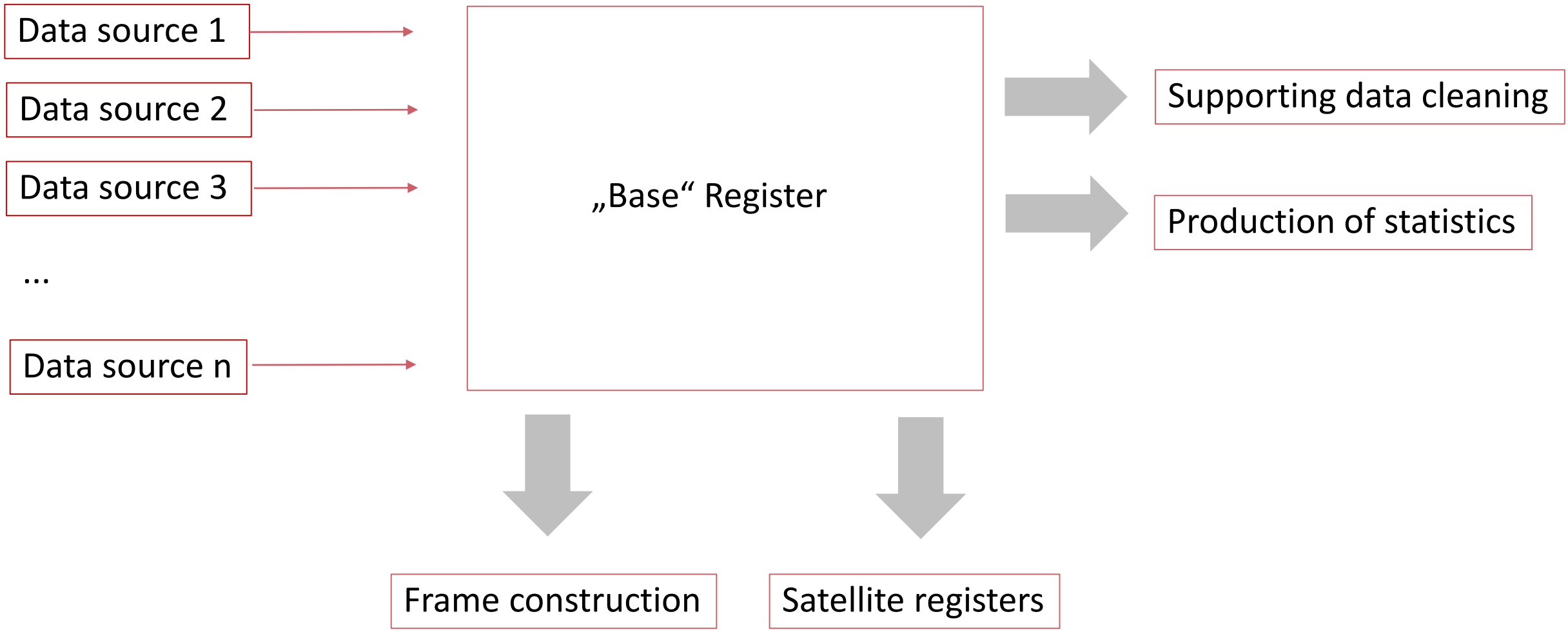
- Improved IT Standards (security)
- Improved micro data access
- API technology
- Visualization tools
- New forms of data presentation

Multisource Scenario

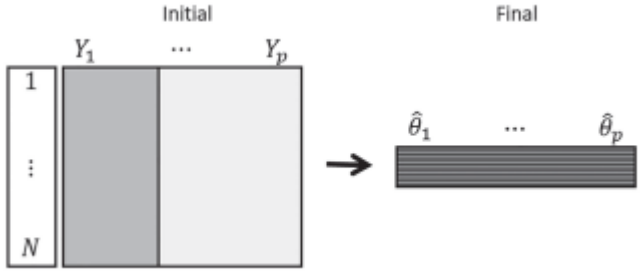
We talk about a multisource scenario when the production of figures for a certain statistical product **is based on more than one source**!

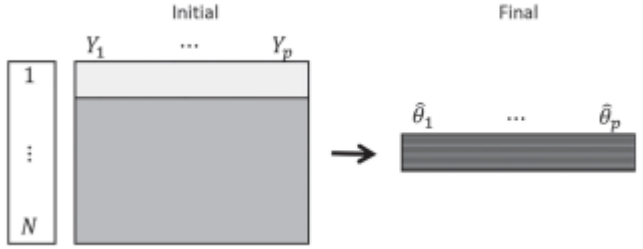


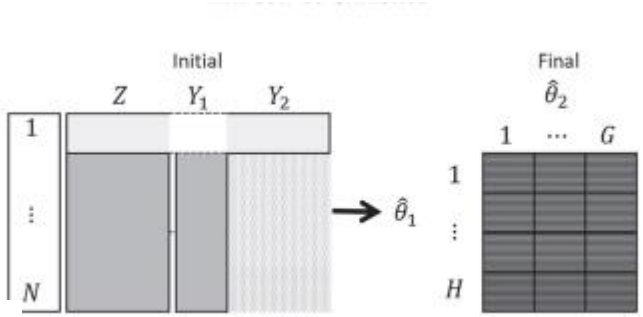
Approaches – Register integration

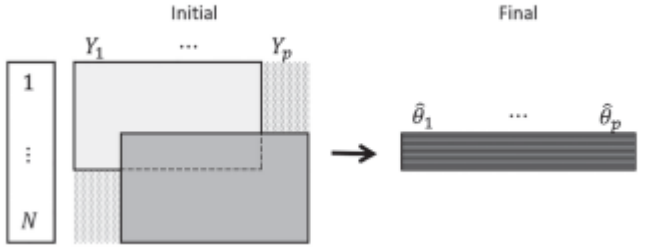


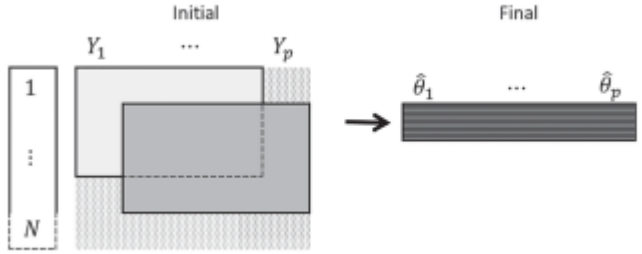
Approaches – Different configurations

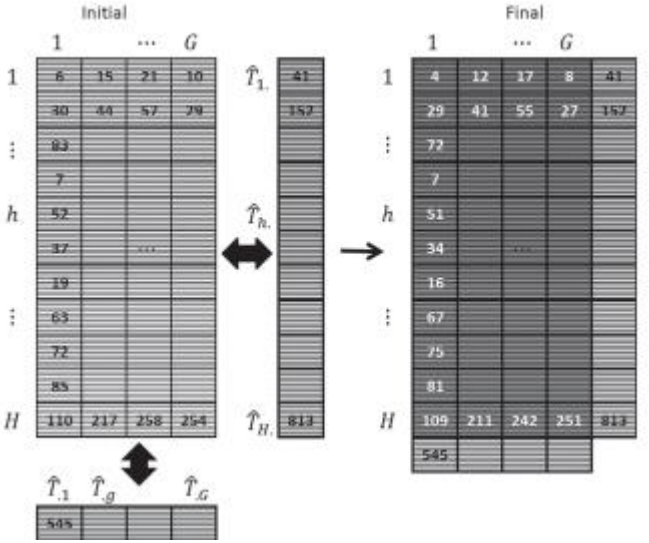


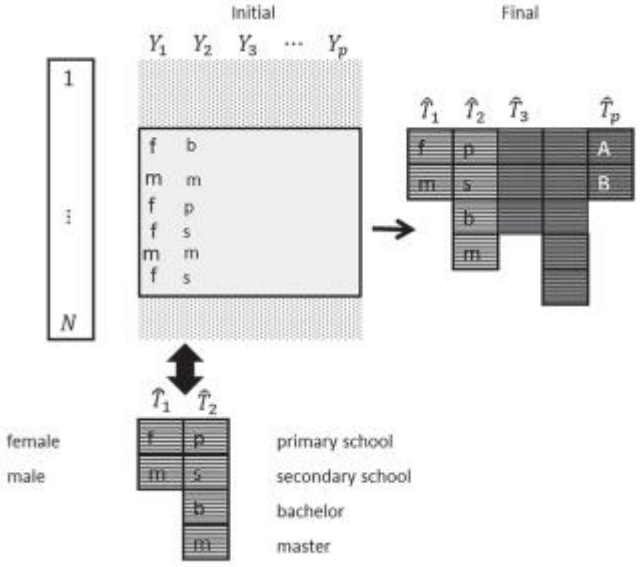


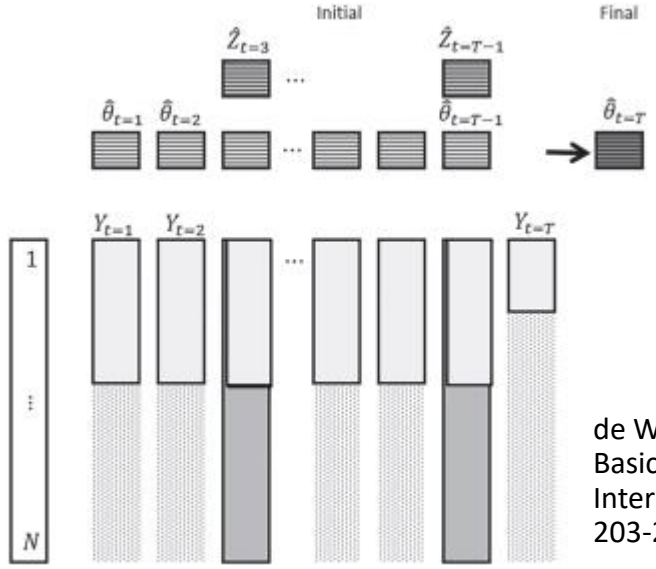












de Waal, T., van Delden, A., & Scholtus, S. (2020). Multi-source statistics: Basic situations and methods. International Statistical Review, 88(1), 203-228. https://doi.org/10.1111/insr.12352

Approaches – Assessing Quality (1)

The idea of a checklist

Indicator	Direct		Indirect			
	1. Direct tabulation	2. Substitution and supplementation	1. Creation and maintenance of registers	2. Editing and imputation	3. Indirect estimation	4. Data validation / confrontation
ESSnet 9: Item non-response	x	x	x	x	x	x
ESSnet 10: Misclassification rate	x	x	x	x	x	x
ESSnet 11: Undercoverage	x	x	x	x	x	x
ESSnet 12: Overcoverage	x				x	
ESSnet 14: Size of revisions from the different versions of the admin data	x				x	
ESSnet 15: % of units in admin data which fail checks				x	x	x
ESSnet 16: % of units for which data have been adjusted				x	x	x
ESSnet 17: % of imputed values (items) in the admin data				x	x	x
ESSnet 4: Periodicity			x		x	
ESSnet 18: Delay to accessing / receiving data from Admin Source			x		x	
ESSnet 5: % of common units across two or more admin sources	x	x	x	x	x	x
ESSnet 21: % of relevant units in admin data which have to be adjusted to create statistical units	x			x	x	x
ESSnet 19: Discontinuity in estimate when moving from a survey-based output to an output involving admin data			x			
ESSnet 7: % of items obtained from admin source and also collected by survey		x	x		x	
CBS 2009, Source 4.1: Cost of using data source			x		x	
ESSnet 2: % of items obtained exclusively from admin data		x	x		x	
ESSnet 3: % of required variables which are derived using admin data as a proxy		x	x		x	

Komuso SGA 1 WP 1 final report

Approaches – Assessing Quality (2) – ESSnet KOMUSO

Quality Indicators and Quality Measures

- Exhaustive list
- Examples

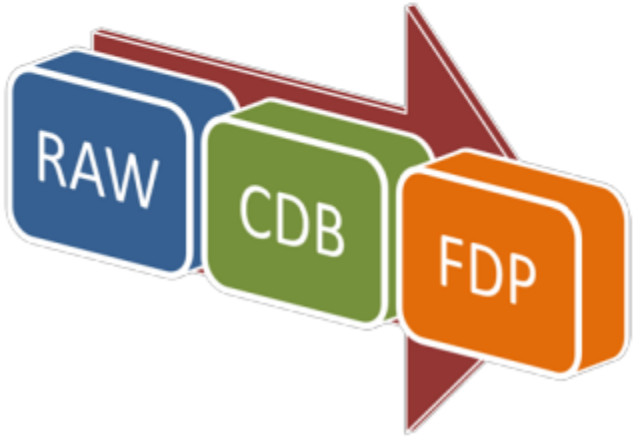
Quality guidelines

- Quality Guidelines for Multi Source Statistics
- Quality Guidelines for Frames in Social Statistics

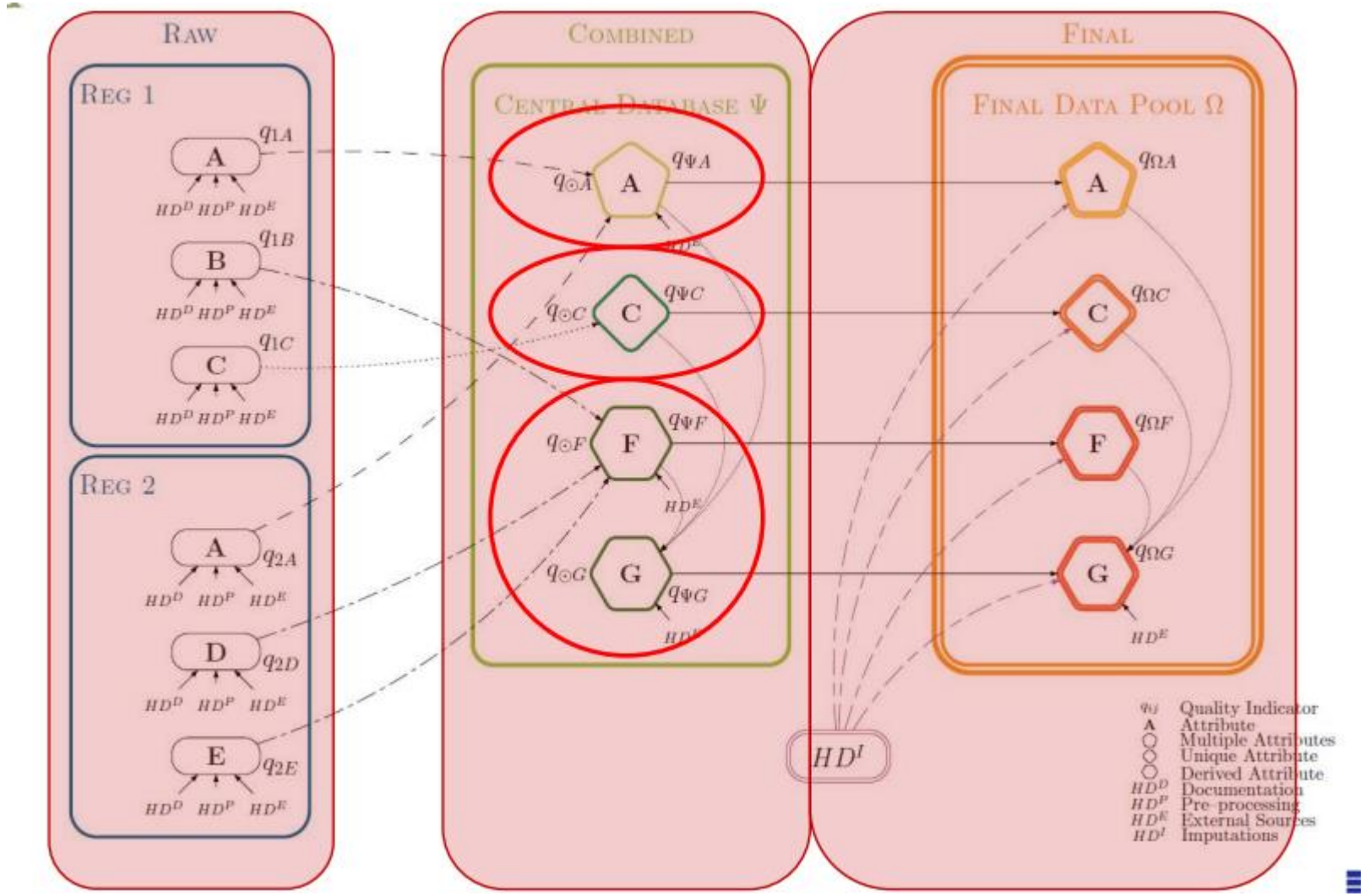
Approaches – Assessing quality - another example(1)

Three stages of quality evaluation

- Raw data
  - Registers provided by the data holders
- Central Database (CDB)
  - Combined information from the registers
  - Data is merged by a unique key
- Final Data Pool (FDP)
  - Final data including imputations

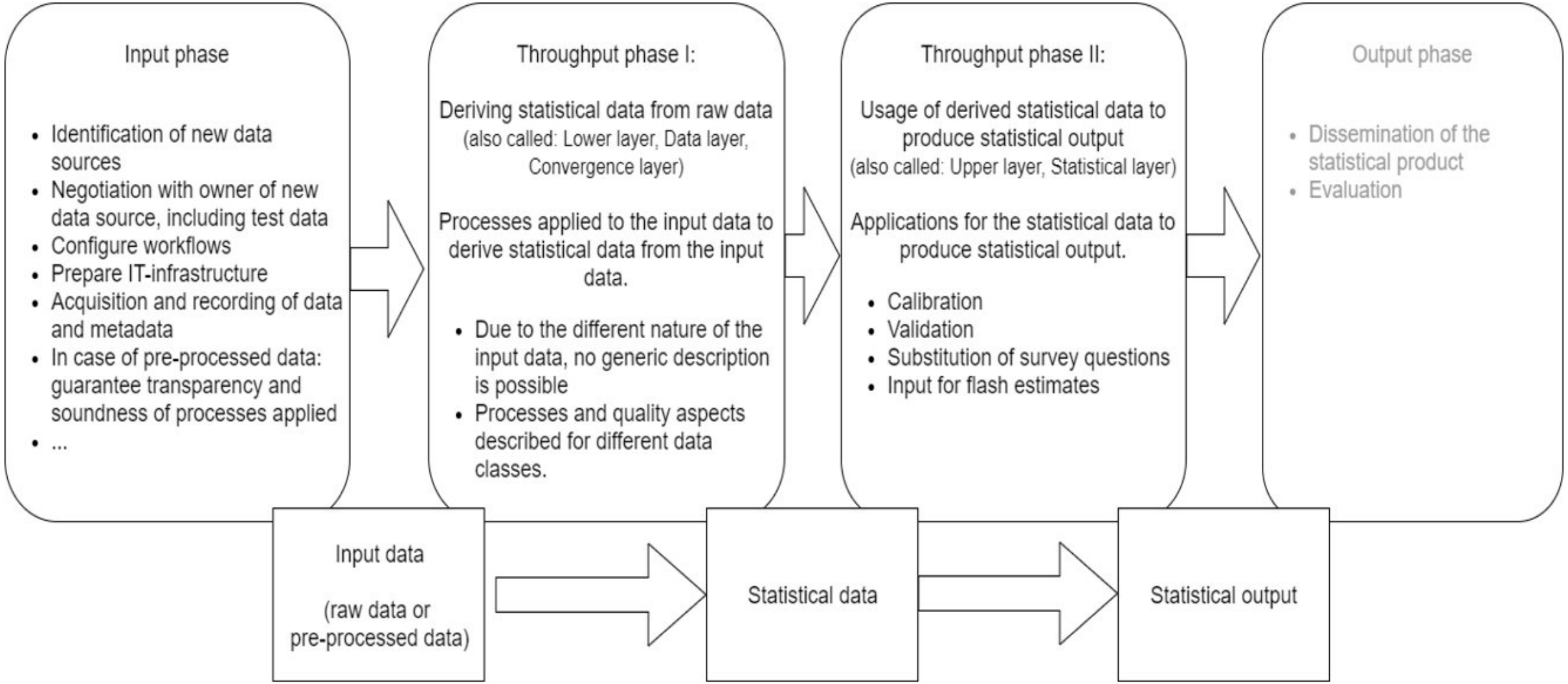


Approaches – Assessing quality another example(2)



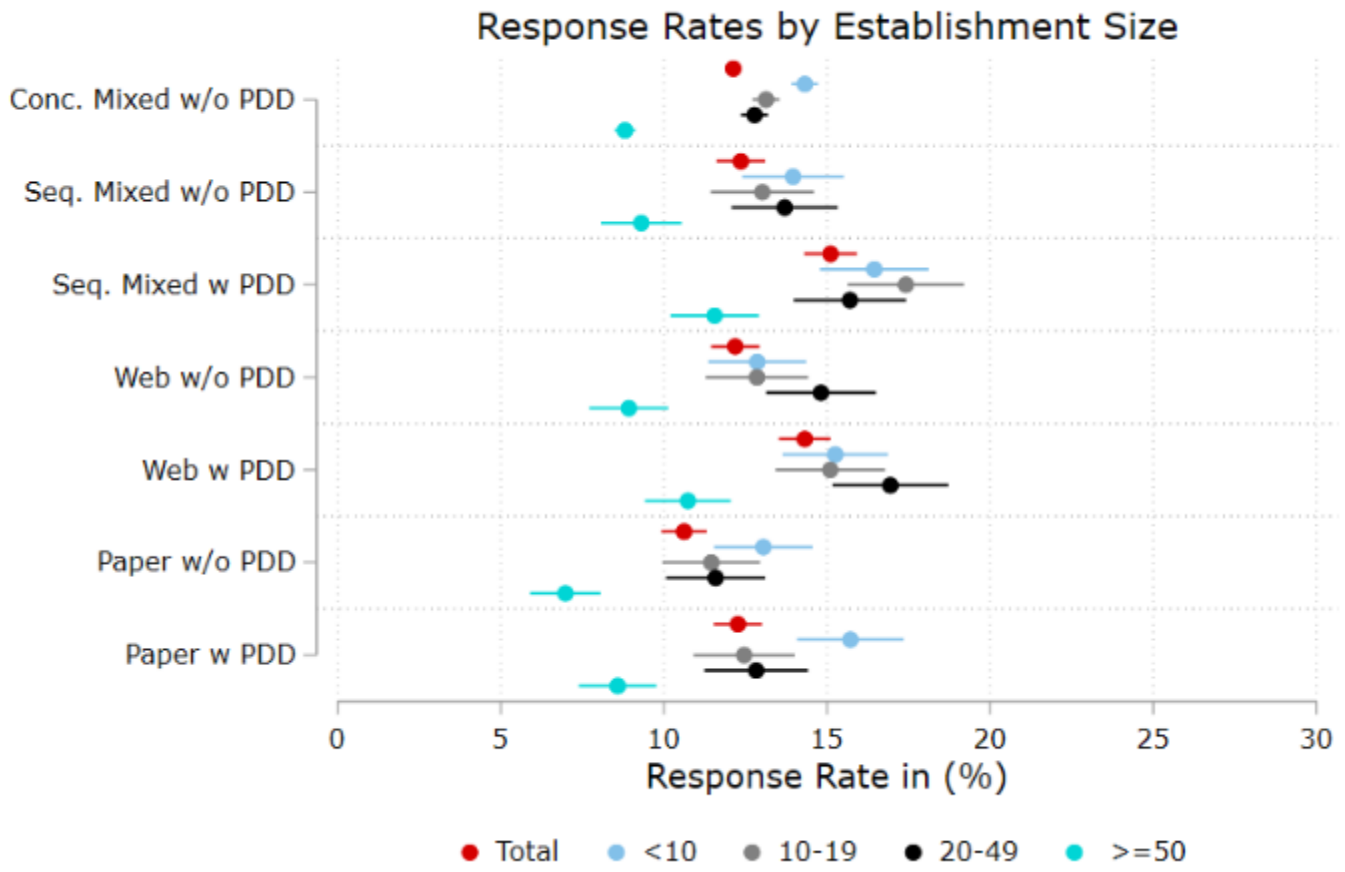
Approaches – Integrating new (big) data sources (1)

Quality guidelines for BIG Data acquisition and usage - ESSnet Big Data 2



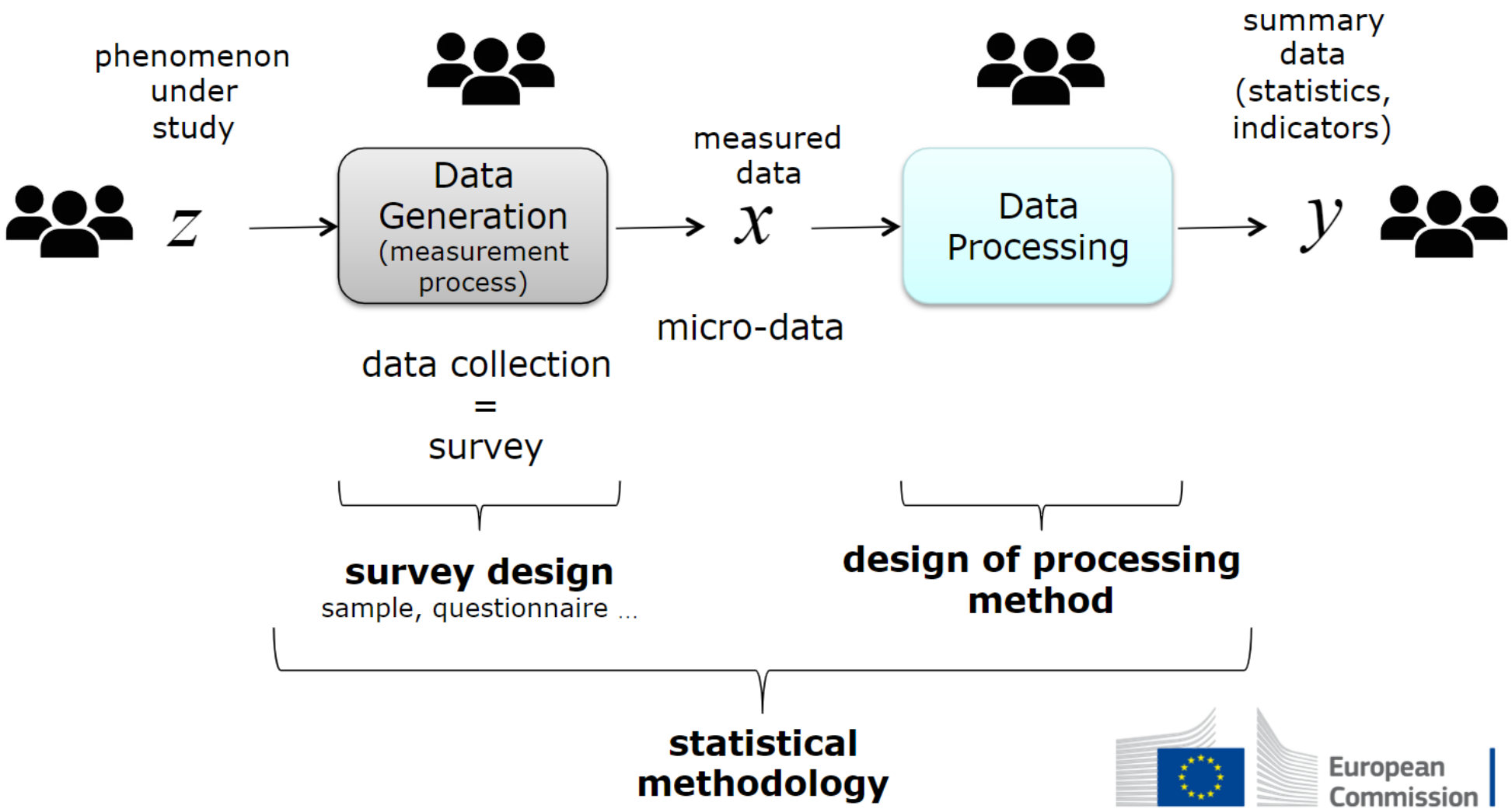
Approaches Mixed Mode as special form of Multi Source

Benjamin Küfner, Joseph Sakshaug, Stefan Zins, Q2022



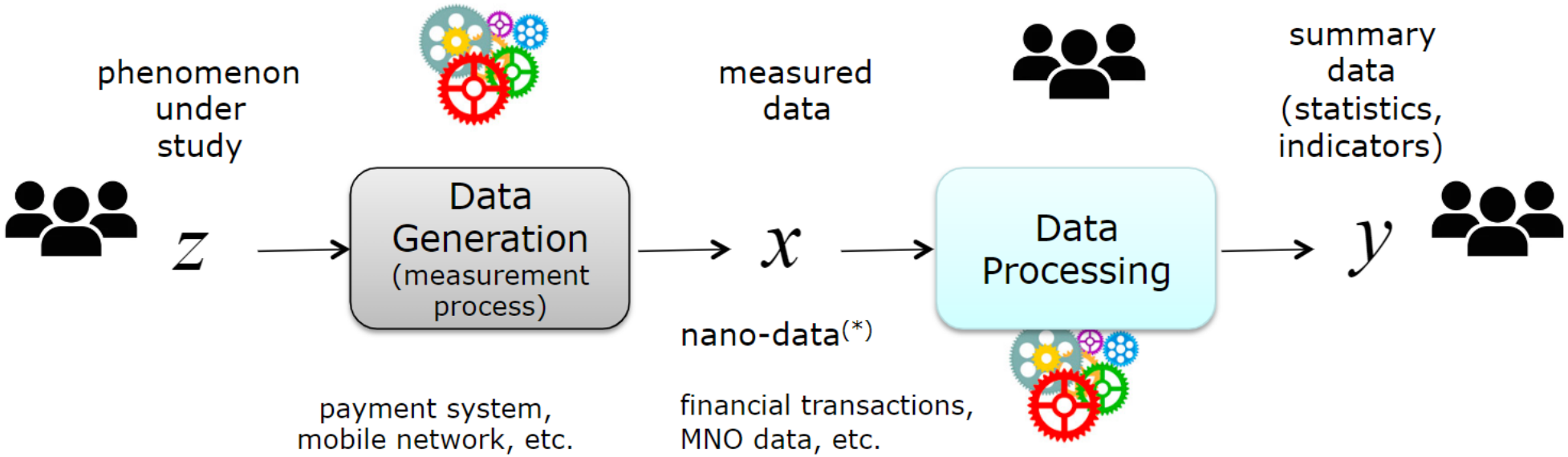
Approaches – Integrating new (big) data sources (2)

Fabio Ricatti, Q2022, Survey data



Approaches – Integrating new (big) data sources (3)

Fabio Ricatti, Q2022 – new big data sources



Data generation is now "given":
it can be "known"
but not actively controlled
by the statistician

design of processing method

statistical methodology

(*) Trusted Smart Statistics:
How new data will change official statistics
https://doi.org/10.1017/dap.2020.7

European Commission

Points for discussion

- Is the set of methodological instruments sufficient for handling multi source scenarios?
- Can we integrate new data sources sufficiently?
- How can we organize common research in this field?
- Would data stewardship help to improve administrative sources for statistical use?

Please address queries to
Thomas Burg
Statistics Austria
thomas.burg@statistik.gv.at
Tel: +43 1 711 28 7877
Mobile: +43 664 6181240

STATISTIK AUSTRIA
Guglgasse 13, 1110 Wien