Essnet Data Integration WP4. Overview of case studies

Johan Fosen, Statistics Norway

Essnet Data Integration workshop, Madrid, 24-25 November 2011

Case study 1: Norwegian register-based employment statistics

- by Johan Fosen and Li-Chun Zhang, Statistics Norway
- Micro-integration
- Statistics disseminated annually since 2005
- 1st part: describing the micro-integration. Some parts of it:
  - First: several registers are linked
  - Example of harmonisation: start and stop time of job in Employee register (main register) harmonised with Unemployment register, sickness absence register,Parental benefit register
  - At the end:
    - ranking of the register-information by its ability to separate between 'employed' and 'not employed' → classification rule
  - Level of employment determined by national level according to LFS.

Case 1, contin.

- 2nd part: Quality evaluation
  - on small areas
  - comparing MSE of
    - register-based employment tables, and
    - LFS-based employment tables
  - use small area model for reg-based bias (linear mixed model)
  - micro-linkage not required

Case study 2. Combining data from administrative sources and sample surveys; the single-variable case. Case study: Educational Attainment

- by Frank Linder, Dominique van Roon and Bart Bakker, Statistics Netherlands (CBS)
  - also major contribution by Vincent de Heij (CBS)
- Micro-integration
- Deriving educational attainment by combining administrative sources and sample surveys
- Several adm. registers on exams, not covering all exams
- Earlier years even less exams were covered

Case study 2, contin.

- A survey, LFS, used as additional data source
  - Some LFS-records used in addition to register-based records
- Aim: for each person in population, a value of educational attainment
- Some topics:
  - register information may be outdated. Survival analysis used to "approve" education information
  - Weighing method was selected (records from LFS has to have weight > 1 to cover all population

Case study 3. First Steps in Profiling Italian Patenting Enterprises

- by Daniela Ichim, Giulio Perani and Giovanni Seri, ISTAT
- Record linkage
- Linking international patent database to Italian official business register
- Aim: analyse connection between patenting activities and enterprise profile
- Record linkage based on comparing applicant names with enterprise name
- Giovanni Seri will present more on this tomorrow

Case study 5. Statistical matching: Polish case study

- by Wojciech Roszka, Statistical Office in Poznan
- Statistical matching
- Simulation study matching microcensus with LFS
- Microcensus: sample of 2.2 million (5% of Polish population). Economic and demographic variables
- LFS: sample of 220 000. On economic activity
- One LFS-unit will be donor to many microcensus unit
- Method: nearest-neighbour
  - for each microcensus-unit, the LFS-unit with shortest distance, becomes the donor.

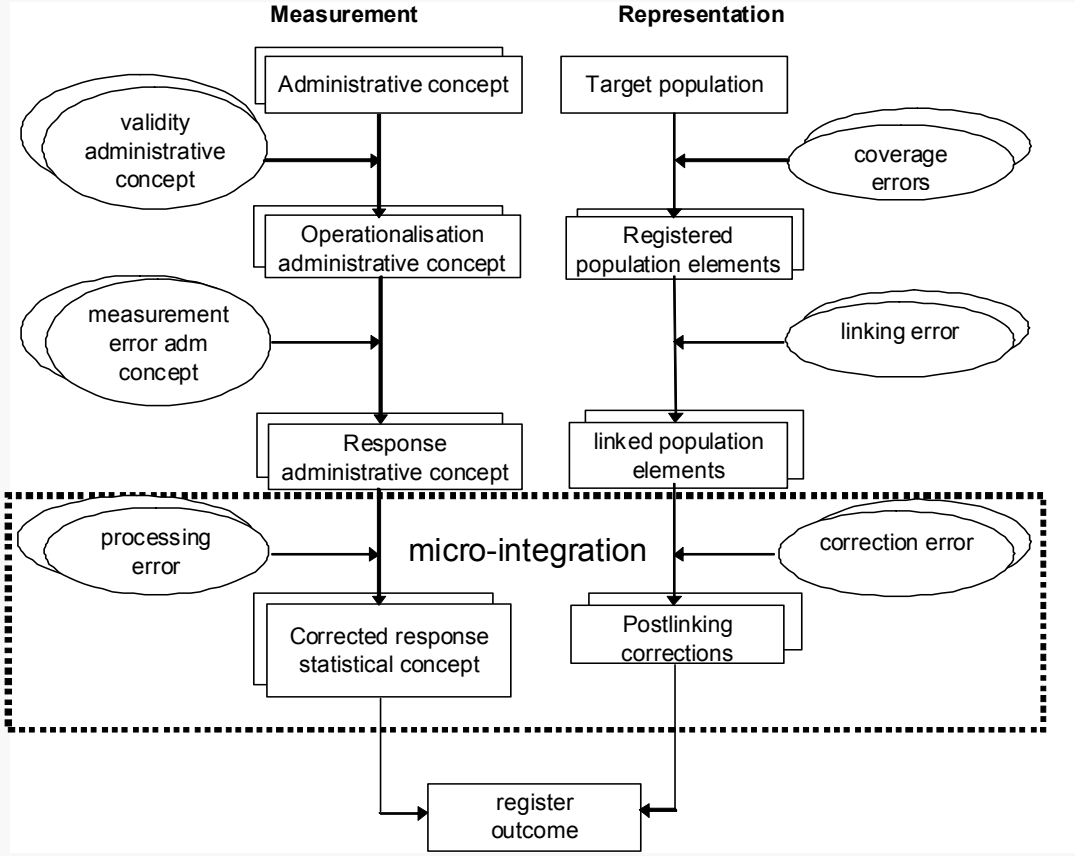
Case study 4. The framework for error in an integrated survey

- by Francesca Romana Pogelli and Filippo Oropallo, ISTAT
- Using administrative registers for imputing for non-response in Survey on Small and Medium enterprises
- Up to 2007: only use of register of Financial statements
- Now: also
  - tax return data
  - fiscal authority survey
- Selection between registers from Kolmogorov-Smirnoff test
- Focus in case study on life-cycle from a quality perspective

Life-cycle from a quality perspective

- Groves et al. (2007) constructed a life-cycle for surveys
- Bart Bakker constructed a life-cycle for administrative data including micro-integration (Bakker 2011)
  - All three micro-integration case studies (1,2 and 4) has related their micro-integration process to this

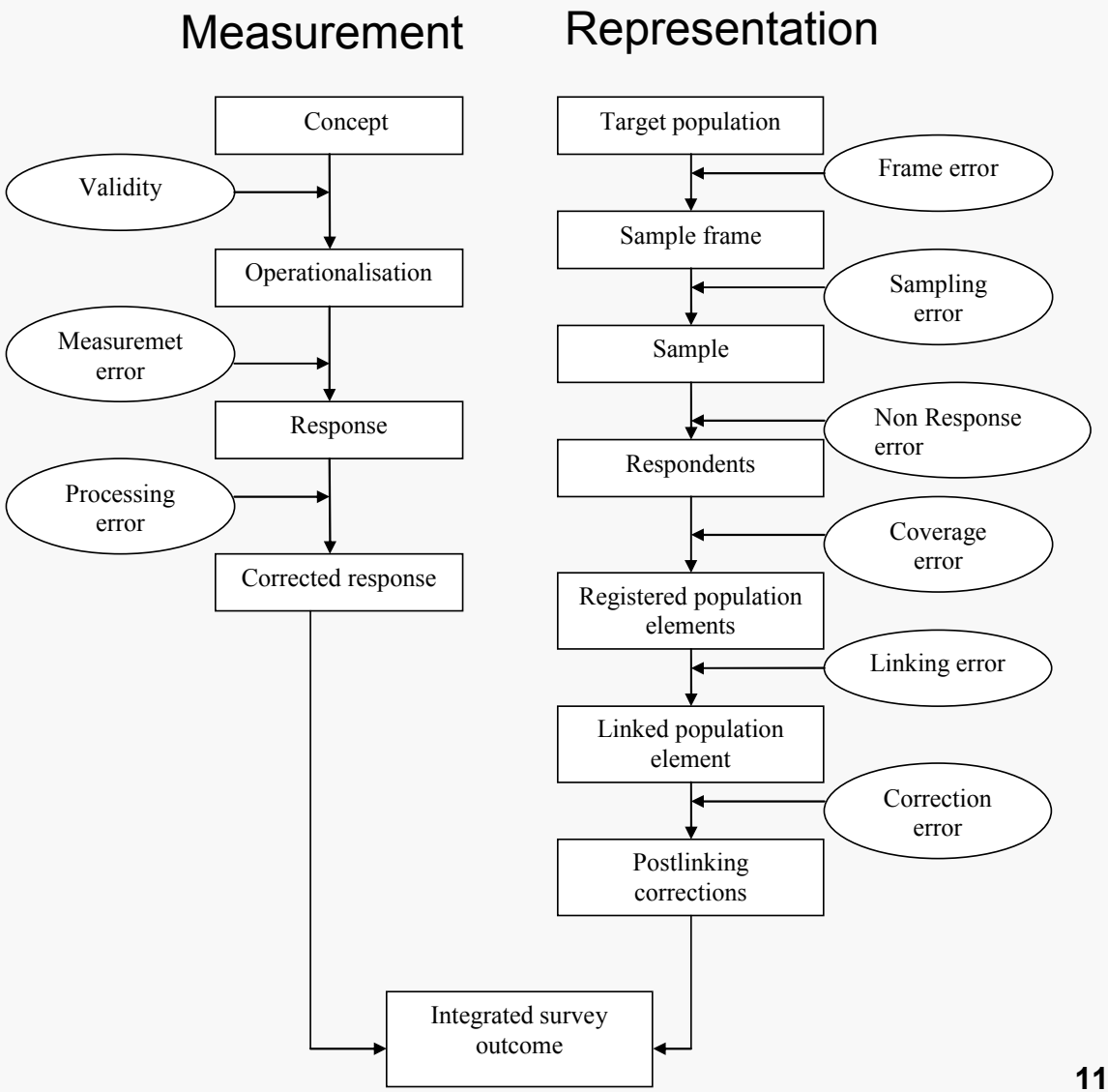
Life-cycle from a quality perspective, contin.



The life cycle suggested by Bakker (2011)

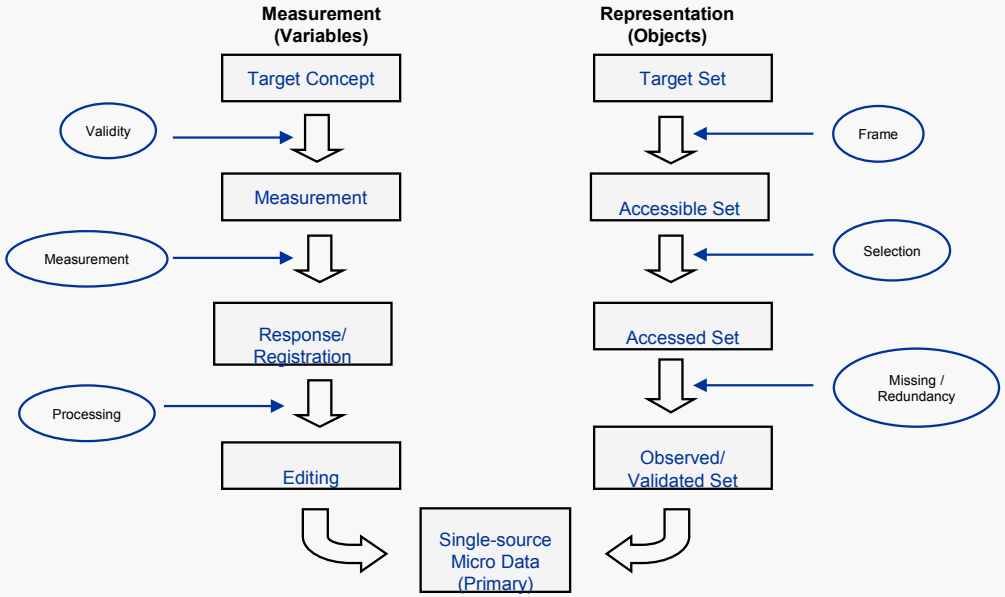
Life-cycle from a quality perspective, contin.

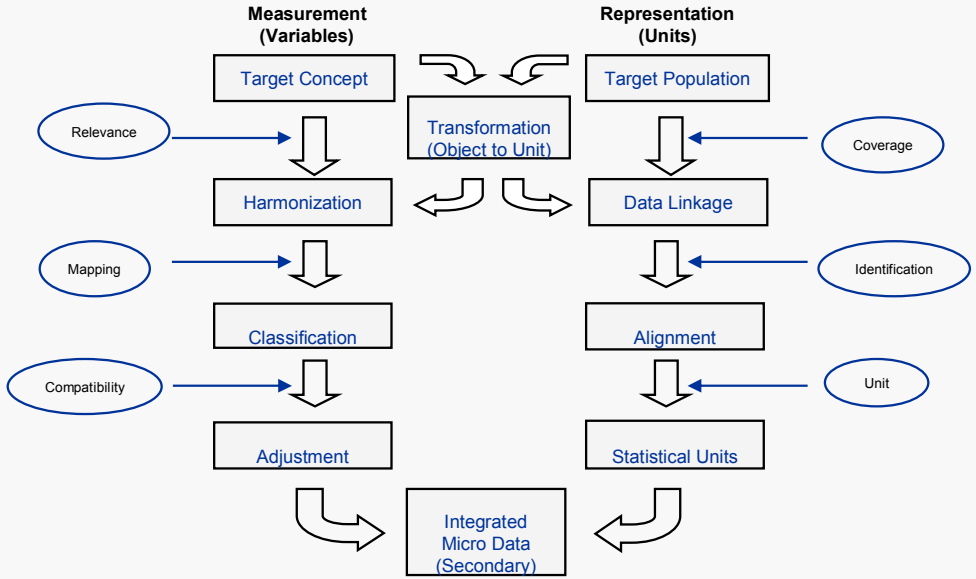
- Case study 4 suggests another life-cycle, combining the Bakker's life-cycle with Groves' on surveys (Groves 2004)
  - since a survey is the main focus, but uses administrative registers



Life-cycle from a quality perspective, contin.

- In case study 1 (employment), also a two-phase life-cycle suggested by Li-Chun Zhang (Zhang 2011) has been tried
- illuminating, e.g. fits well with "employment"
  - representation issue for adm. registers (1st phase)
  - measurement issue for micro-integration (2nd phase).





WP4 report

- describes the case studies in detail
- Marcin Szymkowiak (Statistical Office in Poznan) is formatting the WP4 report by LaTeX
- will be made available on the internet

References

- Bakker, B.F.M. (2011), 'Micro integration: State of the art', in 'ESSnet on Data Integration Report of WP1. State of the art on statistical methodologies for data integration'.
- Groves, R.M., Fowler Jr., F.J., Couper, M., Lepkowski, J.M., Singer, E. and Tourrangeau, R. (2007). *Survey Methodology*. New York: Wiley.
- Zhang, L.-C. (2011), 'Topics of Statistical theory for register-based statistics', paper for the conference ISI 2011 in Dublin, Ireland.