Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Statistical Office FSO

n|w

Analysis of the data preparation process of the structural survey of the Swiss population census

UNECE Work Session on Statistical Data Editing, WP. 27, Budapest 15 September 2015

Daniel Kilchmann, Statistical Methods Unit, Swiss Federal Statistical Office
Beat Hulliger, University of Northwestern Switzerland

Contents

Introduction

Swiss population Census 2010-

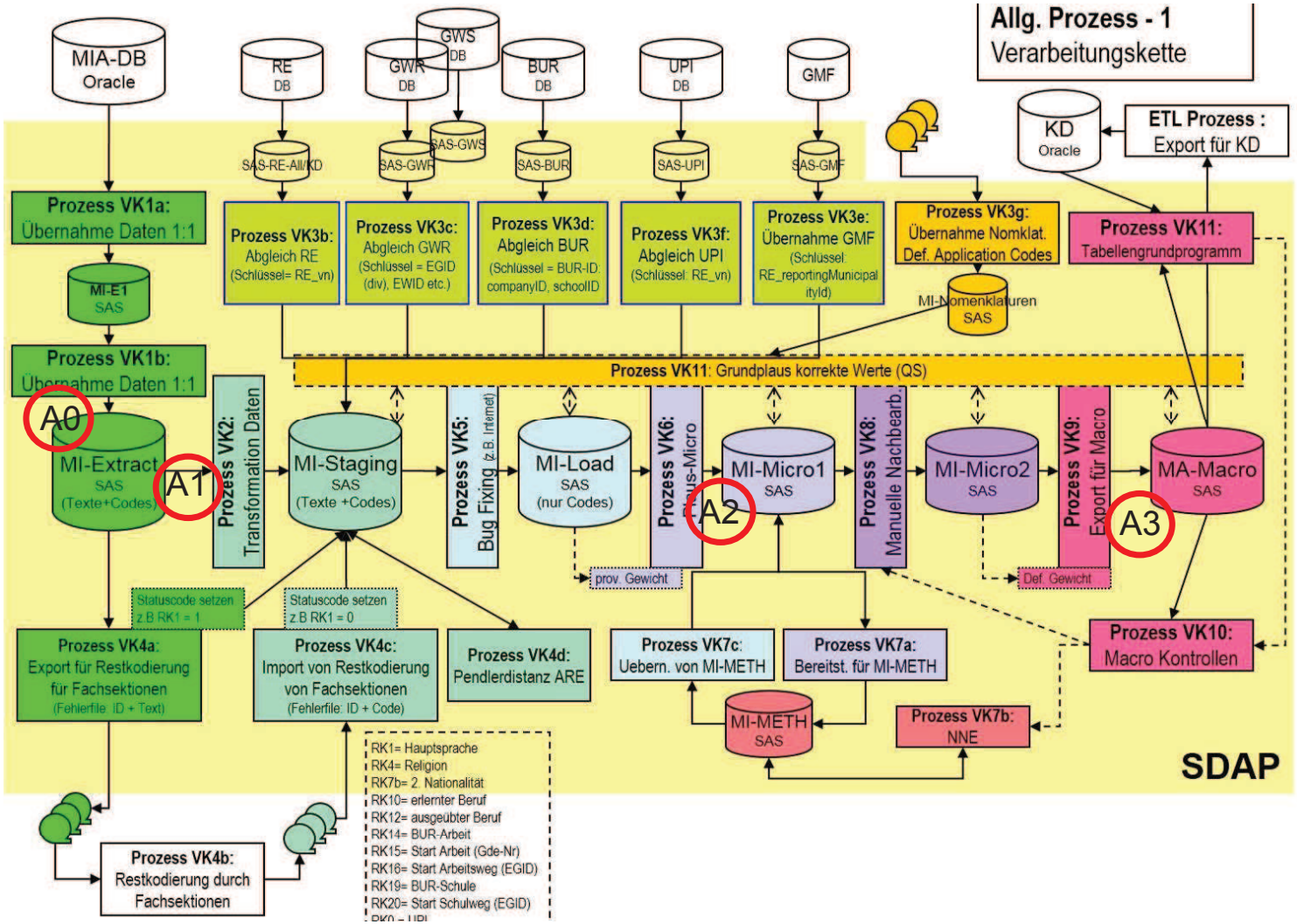
- Register based census combined with sample surveys:
  - municipalities' registers
  - federal housing and dwelling register
  - **census' structural survey (CSS)**
  - two annual sample surveys on specific themes.
- CSS, sample size 250'000 persons, covering
  - Labour market, language, religion, education, migration and commuting of persons.
  - Household composition, household member characteristics and dwelling variables.
  - Paper (75%), internet (25%).

Purpose of analysis project

- SFSO launched a project to analyse the statistical data preparation process (SDPP) of the CSS to
  - gather deeper knowledge about impact of the SDPP on results,
  - monitor the impact during the SDPP.
- Better understanding whether conceptual framework (EDIMBUS, [Luzi, O. et al.(2007)]) and process design (SFSO-SDPP) are appropriate.
- Selection of useful indicators, calculated during SDPP.

Data preparation process of the CSS

Data preparation process of the CSS



E&I methods

- Automized edit rules for missingness and inconsistencies.
- Call backs between $A_0 \rightarrow A_1$.
- Deterministic imputation rules $A_1 \rightarrow A_2$, $A_2 \rightarrow A_3$.
- Outlier detection for `rent` $A_2 \rightarrow A_3$.
- Nearest neighbour imputation based on NIM, [Bankier, M., Lachance, M. and Poirier, P.(2000)], $A_2 \rightarrow A_3$.
- Outlier detection and comparisons $A_2 \rightarrow A_3$.
- Ad-hoc analysis scripts based on indicators in [Luzi, O. et al.(2007)].

Loops only during implementation phase $\Rightarrow$ not included in the analysis.

Analysis project of the CSS-SDPP

Aims

1. Evaluation of the impact on results of individual treatments or whole phases.
2. Potential improvements to the process design.
3. Highlighting of possible questionnaire design problems.

Challenges

- Mostly categorical variables, each response category coded by a binary variable (multiple responses) $\rightarrow$ single question = group of binary variables (*response group*).
- Categ. variables not prominent on indicator lists, e.g. [Ehling, M. et al.(2007)], [Luzi, O. et al.(2007)], better in [Quality team of Eurostat(2014)] $\Rightarrow$ enhancement of use/interpretation of indicators for categorical variables.

- No baseline SDPP for comparison $\Rightarrow$ outcome of the study: baseline.
- No 'truth' available $\Rightarrow$ no indicators requiring the 'truth'.
- Development of a R-package.

Levels of indicators

1. Global indicators: for the whole data set (all observations, all variables)
2. Subset indicators: for subsets of the data (all variables)
3. Group indicators: for all observations (groups of variables)
4. Observation indicators: for single observations (all variables)
5. Variable indicators: for single variables (all observations)
6. Subset-group indicators: for subsets of the data and groups of variables

Edit rule indicators are under discussion.

Core set of indicators

Description	Global	Subset	Observation	Group	Variable
unit response rate	x	x		x	x
item response rate	x	x	x	x	x
item response ratio		x		x	x
imputation rate (responded*)	x	x	x	x	x
imputation ratio (responded*)		x			x

* Indicators for respondents only might be seen as a proxy for the impact of edit rules.

Outlook

Outlook

- Formulas and meaning will be checked by implementation.
- Indicators under discussion:
  - edit rules
  - changes in structurally missings
  - distributional.
- SDPP 'optimization'-criteria under investigation.
- Thresholds, indicating anomalies?

References

Bankier, M., Lachance, M. and Poirier, P.

2001 canadian census minimum change donor imputation methodology.
Working paper, UNECE work session of Statistical Data Editing, Cardiff, 2000.
URL http://www.unece.org/stats/documents/2000.10.sde.htm.

Ehling, M. et al.

Handbook on Data Quality Assessment Methods and Tools.
European Commission, Eurostat, 2007.

Luzi, O. et al.

EDIMBUS-RPM.
Eurostat, August 2007.
URL http://ec.europa.eu/eurostat/documents/64157/4374310/30-Recommended+Practices-for-editing-and-imputation-in-cross-sectional-business-surveys-2008.pdf.

Quality team of Eurostat.

ESS Guidelines for the Implementation of the ESS Quality and Performance Indicators (QPI).
European Commission, Eurostat, 2014.
URL http://ec.europa.eu/eurostat/documents/64157/4373903/02-ESS-Quality-and-performance-Indicators-2014.pdf.