Using the Business Process Model and Notation (BPMN) standard for the automatisation of survey fieldwork within an integrated data collection systems for social surveys

**Keywords:** social survey; survey fieldwork; case management; BPMN

# Introduction

Although administrative data and registers are used more and more, data provided by respondents directly remain important in social statistics. As in other areas of statistical production, the technical, organisational and social environment for survey statistics changed a lot over the last decades. Quality standards increase whereas a considerable decrease in participation rates could be observed in many countries. One strategy to deal with these challenges is to use more complex survey designs, including mixed-mode designs, sophisticated contact strategies and adaptive approaches. However, combining these strategies and approaches in a single survey results in rather complex business processes during the data collection phase of a survey. Running several such surveys simultaneously poses a considerable management burden on data collection units in Statistical Offices.

During the last decades much time and effort was spent to develop electronic tools for designing survey questionnaires. Despite the growing importance there is obviously much less attention paid to develop adequate solutions for case management systems. However, automated case management systems become essential to run surveys in an efficient and cost effective way. In addition to efficiency and cost effectiveness National Statistical Institutes have to consider other quality dimension too. Therefore, you end up with the demand for a tool combing as far as possible the automatisation of the data collection process with the feasibility for non-automated, individual case management. This non-automated case management should allow fieldwork management staff to overrule or complement the automated processes to be able to react on individual preferences of respondents as well as of interviewer in case of surveys including CAPI as a data collection mode.

In 2013, Statistics Austria started developing a new integrated service infrastructure called STATsurv for running social surveys. From the beginning of 2018, all social surveys including the Labour Force Survey and EU SILC are carried out successfully using the new service infrastructure. An essential element of STATsurv is the integration of automated and non-automated case management in a comprehensive survey management tool. The presentation will focus on this specific aspect.

# Methods

## Modularising survey fieldwork

In the Generic Statistical Business Process Model [1] the “Collect” phase focusses on all activities related to the implementation of a statistical survey. The phase is split up into four sub-processes. The actual fieldwork of a survey relates to the third sub-process labelled “run collection”. To automatise activities of the “run collection” phase it is necessary to reflect about some conceptual questions first, the most important one being: How to structure the relevant processes and tasks involved in survey data collection in a way that is as simple and generic as possible. Two major building blocks or modules are identified as essential in that respect:

1. Data Collection Episode
2. Communication Request Task

It is argued that for any survey design, the “run collection” sub-process for a single sample unit could be modelled as a pre-defined sequence of 1 to n Data Collection Episodes (using different or same modes) and Communication Request Tasks.

## Using a global standard for process modelling

To model a sequence of tasks, a global standard for process modelling is available, the Business Process Model and Notation (BPMN) [2]. This standard is widely used in the business world to model business processes in form of workflows. Therefore, the BPMN is also an appropriate model to design survey fieldwork workflows.

# Results

Design requirements and contact strategies of a fictitious CAWI-CAPI mixed-mode survey are translated into a modularised fieldwork workflow using the BPMN standard (see figure 1). The survey and case management application of STATsurv could upload the xml representation of a BPMN workflow. To execute and control this workflow, a BPMN process engine is part of the service layer of STATsurv.

Figure 1: Example of a fieldwork workflow for a fictitious CAWI-CAPI mixed-mode survey – BPMN representation

For each single sample unit of a survey, STATsurv executes the standard fieldwork workflow automatically. However, if requested by respondents, case management staff has several options already implemented in the survey and case management application of STATsurv to allow for some individual case treatment.

# Conclusions

STATsurv is an integrated data collection system for social surveys. One important element of STATsurv is the generic modularisation of the survey fieldwork. According to the concept, each survey fieldwork is always a pre-defined sequence of two tasks, (1) communication requests and (2) data collection episodes. A sequence of tasks is a workflow and could be modelled using the Business Process Model and Notation (BPMN) standard. After the BPMN workflow definition is uploaded to STATsurv, each sample unit is put to the start event of the workflow and pass through to the end event automatically. As a result, when reaching the end event, a sample unit has a final outcome category (disposition code) indicating what was finally achieved during the fieldwork phase, data from the questionnaire (statistical data) depending on questionnaire status, and all kind of process data (paradata). The sample unit is now ready to proceed to the next phase of the GSBPM.

# References

1. https://statswiki.unece.org/display/GSBPM/GSBPM+v5.0
2. http://www.bpmn.org/