ESSnet Smart Surveys

Grant Agreement Number: 899365 - 2019-DE-SmartStat

Link to our Confluence website

Workpackage 2
Smart Survey Pilots

Deliverable 2.7: Template for functional and technical descriptions of tools

Version 0, 24-06-2020

Prepared by:

Barry Schouten (CBS, Netherlands)

Workpackage Leader:

Barry Schouten (CBS, Netherlands)
Jg.schouten@cbs.nl
telephone   : +31 70 3374905

SUMMARY: WP2 performs four diverse pilots to inform WP3 on the specifications of a smart survey platform in the European Statistical System (ESS). This deliverable is a template describing how the outcomes of theses explorations are fed back to WP3. A prominent role is played by the ESTAT CSPA inventory. This inventory is supplemented by a methodological level, a logistics level and a legal-ethical-policy level.

INTRODUCTION

The creation and maintenance of a smart survey, like any survey, implies plan-do-check-act steps. The ESSnet Smart Surveys performs a detailed plan step and a first cycle of a do, check and act step. This is done for two reasons. The first reason is the desire to migrate ESS surveys to smart surveys with optimal comparability but also sufficient flexibility. The second reason is that smart surveys have a number of promising but also challenging features that distinguish them from traditional surveys.

The plan step consists of a conceptual design phase (phase 0), a tool inventory and exploration design phase (phase 1), a methodology test phase (phase 2) and a IT/infrastructure phase (phase 2'). This is followed by a first do-check-act step (phase 3). After phases 2/2', a first set of specifications is delivered to WP3, which are updated after phase 3. This deliverable describes what information is provided, more specifically what template is used.

In order to describe smart survey tools, it must be made clear what a tool is and what it is not in the ESSnet context. If the statistical process of a (smart) survey is divided into the three high-level stages input, throughput and output, then a smart survey tool is most of the input stage plus part of the throughput stage. To understand the latter, a smart survey employs one or more of the following features:

1. Device intelligence: It can use the intelligence (computing, storage) of the device on which it runs
2. Internal sensors: It can employ the sensors that are available in the device
3. External sensors: It can communicate through the device with other sensors close by
4. Public online data: It can go online and extract generally available data
5. Personal online data: It can go online and request access to existing external personal data
6. Linkage consent: It can ask consent to link external personal data already in possession of the survey institute

Smart survey tools, as piloted in WP2, thus collect data, possibly combine these data with existing data, and may perform part of the processing, either locally using device intelligence or in-house at the NSI. A smart survey tool is not the entire input stage as sampling of population units is generally not part of it. It performs only part of the throughput stage as data editing, data imputation and nonresponse adjustments take place afterwards in a separate statistical database. The output stage is obviously affected by the preceding two stages, but is entirely outside the scope of a smart survey tool. WP2 phase 2/2' deliverables describe consequences for the input and throughput stages for WP3, while WP2 phase 3 deliverables give insights in the throughput and output stages. Phase 3 outcomes may be the starting point for changes in the tools and, thus, also feedback to input and throughput stages.

A smart survey tool consists of the following (not exhaustive) elements:
- IT frontend:

- o Questionnaires and diaries
- o Linkage to internal and external sensors
- o Data donation
- o Data processing
- IT backend:
  - o Web server
  - o Data models
  - o User database
  - o Case management
  - o Linkage to existing data
  - o Data processing
- Methodology:
  - o Recruitment material
  - o Data collection strategy
  - o Plausibility checks and validation rules
  - o Sensor data machine learning models
- Logistics:
  - o Interviewer recruitment/motivation
  - o Monitoring and analysis
  - o Helpdesk (email, phone)
  - o Landing page website

WP3 will consider the full throughput and output stages and will widen the scope from a smart survey tool to a smart survey statistical process.

DESCRIPTION OF THE TEMPLATE

Before describing the template, it is important to note that it will be dynamic and may be changed during the ESSnet based on new insights in WP2 and/or WP3. For this reason, the deliverable will have version labels.

The template consists of four parts:
1. CSPA conceptual, physical and logical levels
2. Methodology level
3. Logistics level
4. Legal-ethical-policy level.

Since it concerns ESS smart surveys, these four levels are crossed with a country dimension:
A. Generic: These concern features and functionality that all NSI's share. They are considered to be MUST haves;
B. Country-specific majority: These concern features and functionality that the majority of NSI's prefer and that do not impact strongly comparability. They are considered to be SHOULD haves;
C. Country-specific minority: These concern all other features and functionality that are preferred by a few NSI's and do not affect comparability. They are considered COULD haves;

The focus on comparability of statistics is crucial. Features/functions that, when added, lead to method effects in statistics are to be avoided.

WP2 will deliver a matrix of the four levels against the country dimension for each of the four subprojects (consumption, time use, health/physical activity, living conditions). The four levels are discussed separately.

CSPA conceptual-logical-physical levels

For the IT frontend and backend, the existing Eurostat initiatives to create and maintain an inventory following CSPA (Common Statistical Production Architecture) are mostly sufficient. The inventory can be found at
https://webgate.ec.europa.eu/fpfis/wikis/display/ISTLCS/INVENTORY+of+Tools+and+Sources+for+HBS+and+TUS

Annex A and Annex B give the background to the inventory and the inventory questionnaire, respectively.
The descriptions will be modified based on the country dimension.

Methodology level

The methodology level is new and refers to the most important methodological design decisions. The following are distinguished:
- User interface of frontend: The most influential design choices in the user interface.
- Data collection strategy: This concerns the use of contact modes, contact and reminder strategies, incentive strategies, recruitment materials, the use of "non-smart" modes.
- Data quality checks: Soft and hard checks of plausibility of entered data and notifications of missing data.
- Machine learning models in processing or mixing/fusing sensor data

It is customary that NSI's make their own choices in some of the methodological decisions. This holds especially true for data collection strategy choices. In the template, such choices are merely identified and not prescribed. The relevant conclusion is whether they affect implementation of the smart survey.

Logistics level

The logistics level is new and mostly refers to parts of the input stage that require human interaction or intervention. Four functions are distinguished:
- Recruitment: Many of the ESS surveys have an interviewers-assisted data collection such as doing the starting questionnaire, recruiting and assisting respondents in a diary, motivating respondents during data collection, and/or picking up closing questionnaires.
- Monitoring: Sample units may not participate, drop-out or deliver insufficient data quality. Monitoring dashboards may be inspected on a frequent basis in order to determine whether interventions are needed at overall level.
- Assistance: Respondents may be assisted in starting and using smart survey tools passively through a helpdesk and website or actively through interviewers and or technical experts.
- Human-in the-loop machine learning: In sensor data applications, models seldom reach 100% accuracy. Certain population subgroups or certain survey statistics may require manual inspection. In ESSnet Smart Surveys where feedback of statistics to respondents is deemed important, such human-in-the-loop processes may even occur during data collection.

Legal-ethical-policy level

Also the legal-ethical-policy level is new and refers to what is legally allowed, ethically accepted and policy-wise decided. The legal level in essence refers to Data Protection Impact Assessments (DPIA) and describe minimal legal requirements.

Apart from the legal level, there are two more levels that need to be considered, namely the ethical level and the policy level. The ethical level concerns commonly accepted norms in a society about social interaction. The policy level concerns norms that national statistical institutes have about how they like to operate in society. Much more than the legal level, the ethical and policy level may be culture/country-specific. They may be viewed as nested: Legislation is less restrictive than ethical boundaries and ethical boundaries are usually less restrictive than policy boundaries.

Some statistical institutes have ethical committees that judge whether a survey data collection is ethical. The main evaluation that is made is whether the relevance and importance of the collected data outweigh the burden to respondents, to environment and, if applicable, to interviewers. A second evaluation that is made is whether the data collection itself is designed and organized in a proper and diligent manner. So even if data collection is legal and even if participation is voluntary, it may not be accepted, as the costs are deemed higher than the profits to society. Underneath this lie complex privileges as the right to participate, the freedom not to participate, the individual ability to judge the consequences of participation, the access to data and informed consent.

In the ESSnet, the focus is on smart surveys and it is sufficient to look at what is new, relative to existing survey data collections. In order to understand what is new, it is useful to distinguish the different steps in a survey: sampling, contacting, recruiting, interviewing, collecting, processing and analyzing. For each of these steps there are legal, ethical and policy boundaries in place. So what is new in smart surveys? The differences are mostly in the interviewing, collecting and processing steps:

- The data collection may use the computing and storage options of personal devices
- Part of the data collection may be passive
- Part of the data collection processing may be shifted to the personal devices rather than in-house at the institute
- Part of the collected data may be of specialist content of which part of the respondents has no knowledge, even under informed consent
- In order to be able to participate, one may have to possess a specific device or to accept that one needs to use a device that is provided

This is a tentative list and should be supplemented. The ethical and policy levels can (most likely) not be assessed by the NSI's nor the ESSnet project team themselves. In the ESSnet, it is foreseen that only first recommendations are given.

ANNEX A – EUROSTAT STATISTICAL SERVICES DEFINITIONS
See attachment

ANNEX B – CSPA INVENTORY QUESTIONNAIRE
See attachment

ANNEX C – WP2 PILOT EVALUATION FORM
See attachment