












Smart Survey Implementation

Grant Agreement Number: 101119594 (2023-NL-SSI)

Work package 4

Logistics

Deliverable 4.1: Review stage

Version 1.1, 2023-11-24

Prepared by:

Marc Houben (Statistics Netherlands, Netherlands)

Joeri Minnen (hbits, Belgium)

Fong Yee Wong (Statistics Netherlands, Netherlands)

Anne Elevelt (Statistics Netherlands, Netherlands)

Tim de Jong (Statistics Netherlands, Netherlands)

Work package Leader:

Marc Houben (Statistics Netherlands, Netherlands)
e-mail address : mhf.houben@cbs.nl

Index

1. Introduction

This document is a first intermediate result of work package 4 of the SSI project.

The document is primarily intended for Eurostat so that they can see that we – as the SSI project – are on the right track, and also for staff in an NSI that are considering implementing smart surveys. But for an NSI, the end product of the SSI project is probably of more interest than this intermediate report.

=> An important goal of this document is to get feedback from the reader, as input for the next two project stages.

The main goal of WP4 – at the end of the project – is to deliver **concrete guidelines** that will help NSI's to extend their business processes to adopt smart features in their surveys. Results should be helpful not only for the parties participating in the consortium for the SSI project, but for all ESS NSI's.

The main goals are:

1. Describe the statistical **business process** regarding the specific aspects concerning smart solutions. By that we mean all the process activities that need to be carried out preparing and executing a smart survey. The GSBPM and BREAL will be used as a framework. The GSBPM will to that end be elaborated where necessary.
2. Describe a **maturity model** with maturity criteria. Maturity here refers to the maturity of an NSI to use smart surveys, but also the maturity of a smart solutions to be used.
3. Develop a **benchmark** that can be used by NSI's in practice. This benchmark helps NSIs to assess themselves in terms of maturity and provides guidelines for further application of smart surveys.

In order to achieve this, we need to address some other topics:

- We need a **taxonomy** of smart solutions. That is because the business process can look different for different types of smart solutions.
- We need to identify the different **actors** in the business process that will be affected by smart solutions. Actors should be mapped to the process.
- **Microservices** and **machine learning modules**[1] will support the business process. So, we need to make clear how they fit into the business process.
- We need to describe a **PDCA-cycle** for development of the business process and for the development of microservices and ML-modules. This is because development is not a one-time effort, NSI's should be able to maintain it in the future.
- In the end we need to **demonstrate** that **the benchmark** actually works in practice, by applying the benchmark to (at least two) end-to-end solutions existing within NSI's.

[1] The microservices and ML modules themselves, will not be developed within WP4

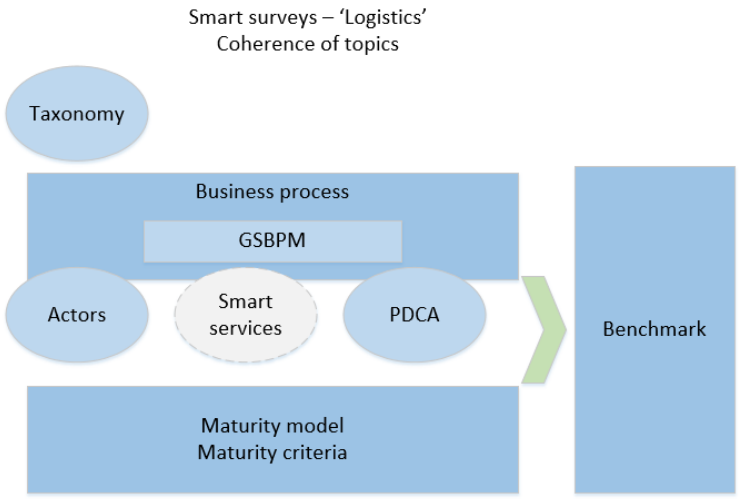


Illustration: The description of the business process and the maturity model will both be an important input, and even part, of the benchmark.

Concerning the business process we refer to the enterprise architecture proposed by BREAL[2], especially the Business Layer.

BREAL states that when looking at the business layer, four main topics/views are important: principles, business functions, business process and actors.

In the context of enterprise architecture, a business function refers to a specific, well-defined, and often distinct activity or capability that an organization performs to achieve its strategic objectives and deliver value to its stakeholders. These functions are a fundamental part of an organization's business model and are typically grouped into categories that reflect the various aspects of its operations.

A Business process cuts across different business functions and are typically designed to deliver value to customers, improve efficiency, and support the overall business goals. The business process consists of process activities. Human actors and IT applications (e.g. micro services) are needed to execute the activities. Both human actors and IT application need knowledge to perform their task. These are all aspects to look into when talking about a business process.

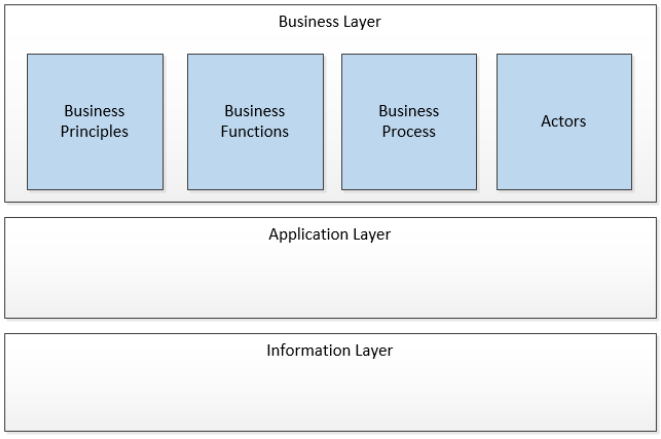


Illustration: BREAL Layers

[2] Big data Reference Architecture and Layers

The tasks for the **Review stage** of the SSI project where the following:

1. Create a first version of the **taxonomy**
2. Identify the **actors**
3. Identify missing **GSPBM** parts. Based on revisiting existing GSBPM framework from ESSnet TSS.
4. Develop ideas how to create a **benchmark**
5. Create a first version of a **maturity model** with maturity criteria

=> As said, an important goal of this document is to get feedback from the reader. Are we heading in the right direction? Is it comprehensible? Is it useable? Et cetera.

In this document there is a chapter for each of these five topics. The different topics all have a different 'degree of readiness'. The focus of the reader should take this into account. This focus is mentioned below.

In chapter 2 the **taxonomy** is described. This should be seen as a first version. In the next project stage, we will supplement and improve it. So, focus of the reader should primarily be focused on 'are we heading in the right direction?' and 'is this practically useable?'.

In chapter 3 the **GSBPM** is addressed. It is good to mention that in this Review stage we took a somewhat different approach than the task as defined above. We did focus on the first few GSBPM stages an actually tried to elaborate these stages. So, it is more than just identifying missing parts. On the other hand: we did not yet look at all of the GSBPM stages. So, this should be seen as a first version. We will supplement and improve this in the next SSI project stages. Focus of the reader should primarily be focused on 'is our approach good?', 'are we heading in the right direction?' and 'do the provided examples have added value?'.

In chapter 4 the **actors** are described. These is an inventory of all actors that can be affected when an NSI is going to use smart solutions. This should be seen as a fairly definitive inventory. So, focus of the reader should be on the completeness and accuracy of the list of actors.

Chapter 5 describes a first version of a **maturity model**. The model consists of five levels and per level, maturity criteria are listed. The input from the different work packages is embedded. This model should really be seen as a first version. We will improve and elaborate it in the next two project stages. Reading should primarily be focused on 'is our approach good?', 'are we heading in the right direction?'

In chapter 6 there is a description of what we mean by a **benchmark**. The benchmark itself will be created in the next two project stages. So, this chapter is used for framing a benchmark. And it tells how we would like to demonstrate that the benchmark will be practically useable. This is 'ready', so the focus of the reader can be on the details. Is it clearly described?

In the final chapter, chapter 7, we briefly look ahead to the next project phase.

2. Taxonomy

Introduction

The need for clear terminology and in particular a taxonomy of smart features has two motives. One is to be able to generalize to other applications of smart surveys not in scope of SSI. Another is to facilitate communication between different design and analysis disciplines underlying to smart surveys. The goal of a taxonomy is, thus, to create categories of smart features that are sufficiently homogeneous to develop generic methodology (e.g. push-to-smart recruitment and motivation strategies, active-passive trade-offs in respondent interaction, machine learning predictions and other data science methods), logistics (e.g. monitoring and analysis, interviewer training, (re)training of classification models), IT (e.g. clustering of tasks within services, input-output specifications) and legal-ethical procedures (e.g. privacy risks, mitigation measures).

Smart Survey Terminology

Let's start by delineating the definition of 'smart surveys', because that term is open to multiple interpretations. In short, a smart survey is a survey that employs one or more *smart features* in data collection. Schouten (2021) defines smart features as following:
"Smart devices offer attractive options to collect more traditional types of data (e.g. survey questions), along with new forms of data. A smart device offers the following features for collecting, linking or processing data:

1. *Device intelligence: It can use the intelligence (computing, storage) of the device, e.g. it can apply pre-trained machine learning models for image recognition;*
2. *Internal sensors: It can employ the sensors that are available in the device, e.g. the location sensors, camera or motion sensors;*
3. *External sensors: It can communicate through the device with sensor systems close by, e.g. a smart watch or an indoor climate system;*
4. *Public online data: It can go online and extract publicly available data, e.g. open street maps data;*
5. *Personal online data: It can go online and request access to existing external personal data, e.g. bank transaction data or shop loyalty card data;*
6. *Linkage consent: It can ask consent to link external personal data already in possession of the survey institute, e.g. shop scanner data or public transport data."*

In the previous ESSNET Trusted Smart Surveys project (see; ESSnet Smart Surveys | CROS (europa.eu)), upon which the SSI project builds, the link is made as well to smart devices:
"The term smart surveys has been used to refer to surveys based on smart personal devices, typically the smartphone. Smart surveys involve (continuous, low intensity) interaction with the respondent and with his/her personal device(s). They combine (inter)active data provided explicitly by the respondent (such as responses to queries, or shared images) together with passive data collected in the background by the device sensors (e.g. accelerometer, GPS) on the same device or within other devices within the personal sphere of the respondent."

In order to classify smart features one more step is needed, namely what follow-up actions, termed *smart tasks*, are evoked by the smart feature. A smart task is a processing step on data generated by the smart feature, which are termed *smart data*. Six types of smart tasks are distinguished: cleaning, editing, enriching, imputation, transformation and classification. The ability to perform such tasks during the survey, and often in-device, is typical of smart surveys. Project SSI develops smart services, also termed smart solutions, that are clusters of smart tasks.

We start by providing a taxonomy for smart features and next discuss how clusters of resulting smart tasks can be implemented.

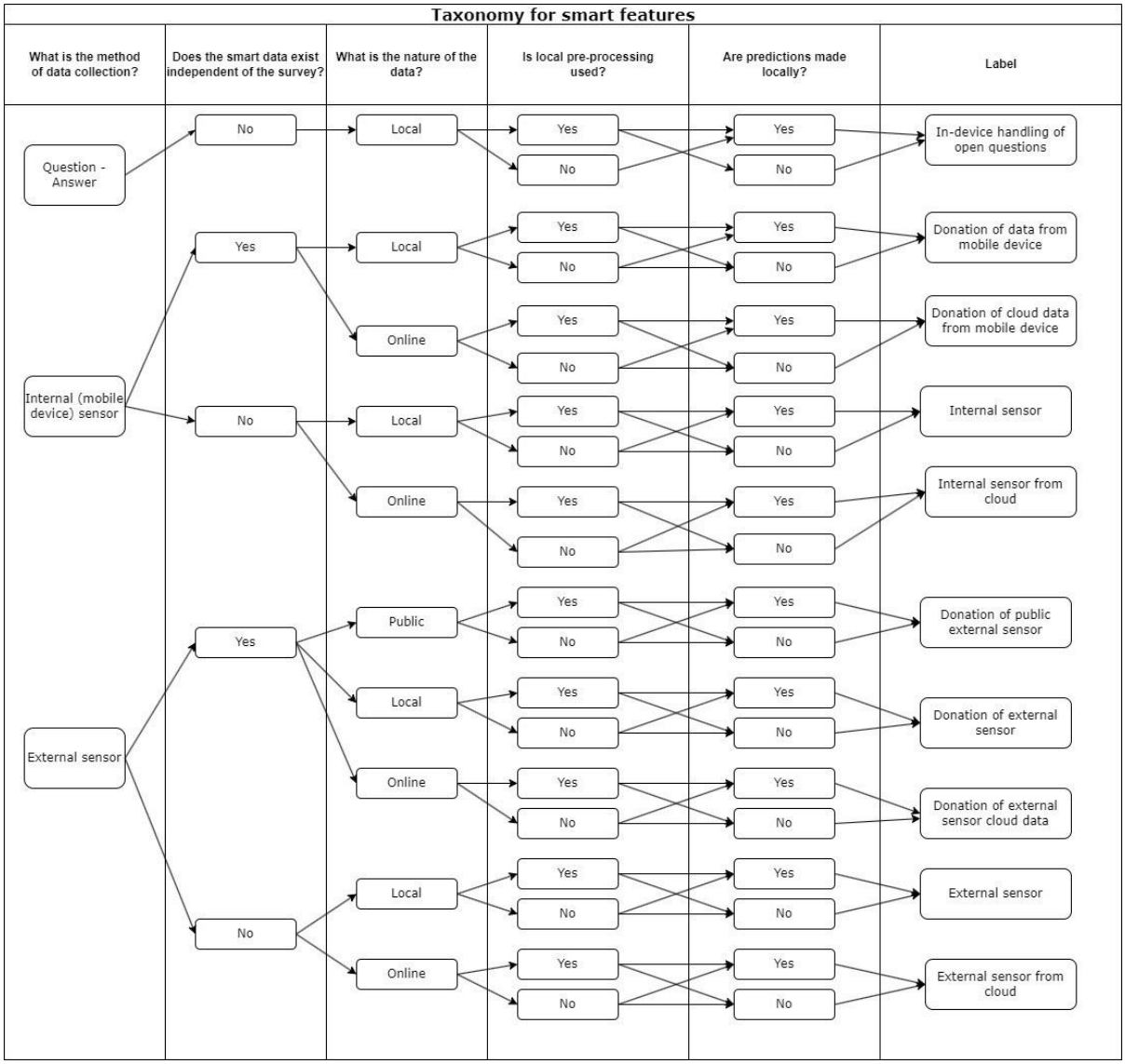
Taxonomy for Smart Features

We again note that the purpose of the taxonomy is to create homogeneous groups in all design levels (methodology, IT, logistics, legal-ethical). To determine the type of feature, for now we see five conceptual questions:

1) What is the method of data collection? *Question – answer, internal mobile device sensor, external sensor system.*
2) Does the smart data exist independent of the survey? *Yes, no.*
3) What is the nature of the data? *Public, local, online*.
4) Is local pre-processing used? *Yes, no.*
5) Are predictions made locally? *Yes, no.*

Some explanation is needed. The first question determines the 'mode' of smart data collection. This could be an internal mobile device sensor, an external sensor system or a traditional question-answer procedure. Obviously, a question-answer approach is not smart, but still the processing actions, i.e. the smart tasks, may involve functionality of a smart device such as local processing and respondent interaction. The second question refers to the existence of data which can be dependent on the survey or not. The third question then determines the nature of the data, namely public data, personal online data and local/in-device data. Public data can be used without authorization whereas the other two types of data can be accessed only with help of the respondent. The fourth and fifth questions refer to smart tasks being executed on the data in-device/locally. The fifth question limits focus to classification and the fourth question to cleaning and editing. The first, fourth and fifth questions may give rise to further subclassifications based on the type of sensor, the type of pre-processing and the type of classification. For now, further subclassifications (and, consequently determination questions) are considered too detailed, but during project SSI it may be concluded that more detail is needed to warrant generic methods and procedures.

The order of the questions is not fixed, but the proposed order is most efficient in separating smart from non-smart features. With all possible combinations, we end up with 72 combinations or options. With these 72 options, we ask ourselves 'Is this a smart feature?' and 'Does this function exist in practice (as opposed to purely hypothetical)?'. If the answer to both questions is 'yes' it is included in the taxonomy below. We can group the resulting 'answers' into 10 main smart categories, which are labelled here.



Implementation of Smart Features

Next to a taxonomy of smart features, we discuss the archetypical approaches to implementing smart features. These are internal (mobile device) sensors, external sensor systems and data donation. These three actually demarcate a subset of the smart feature categories, namely based on the type of data collection and the type of data.

1. **Internal (mobile device) sensors**

Smartphones incorporate a large number of sensors (e.g. accelerometers, GPS, light and proximity sensors) which can be logged passively, providing a large and detailed set of measurements about respondents and their environment. All these sensors can be used for research purposes. This enables researchers to collect high-intensity data passively, that is, there is no respondent activity needed after giving permission to share sensor measures with the researchers. Additional sensor data can also be collected actively, for example when respondents are asked to take pictures. Sensor data collection may be even more valuable when the data are validated or context about them is given by the respondent, but this is not always a prerequisite.

There are two main ways to collect internal sensor data; apps or browser surveys.

Apps are probably the most common and obvious type of smart service that comes to mind when hearing the term smart services. Research apps are pieces of software that can be used for data entry and sensor data collection. An app is not smart by itself: it is possible to build an app-based questionnaire that does not incorporate any of the smart features. Most research apps however, combine an interface to communicate with the respondent with sensor measurements. Apps can easily collect and store data over a longer period of time, making them particularly interesting for surveys with longer data collection periods. Possible applications are the household budget app or time use survey app, which are fielded in this SSI project, or a mobility app that passively collects respondents' geolocations for several days to learn about mobility behaviour.

Browser surveys can also combine a survey with internal sensor measurements. But in this case, it is an online survey rather than an app. When an online survey is filled out on a smartphone, it is namely possible to access the sensors in the smartphone. But nowadays sensor data collection is not limited to smartphones; computers also contain some sensors and can thus be used for conducting smart services although the options are more limited. The advantage of browser surveys is that the respondent does not have to download anything: sensor data collection is incorporated in the JavaScript code of the survey. It is only possible to collect sensor data while that specific webpage is open, making this type of smart survey suitable for one-time data collection. Possible applications on smartphones are getting context on how or where respondents fill out the survey by tracking GPS location or the acceleration (and thus the movements) of the respondent. Possible applications for computers are also tracking a GPS location or possibly eye tracking via a webcam.

2. **External sensor systems**

External sensors are stand-alone devices that collect data on a specific subject, usually about their environment or the way they are used. Smart services fall in this category, and not under data donation, when there is a direct relation between the collected data and the smart survey. The data is collected for the survey institute and sent directly to the survey institute. Examples of datatypes that can be collected are water usage, temperature, energy usage or indoor environment quality. A specific type of external sensors are consumer wearable health devices, or physical activity trackers, which can be used for health or fitness research.

Some external sensors save data on the device itself whereas other sensors can make a connection with an internet network. In the first case, the data stays on the device until it is read out at the survey institute. In the second case, the data is sent to the survey institute at a pre-specified interval (e.g. per minute, day or week) or uploaded in a system.

External sensors need to be sent out to the respondent (or distributed by interviewers). After the data collection, the external sensors need to be sent back to the survey institute. Furthermore, sensors may need to be installed and/or recharged during data collection, which may be complicated for the respondent. The exception to this is when respondents' personal (fitness or health) devices are used as external sensor system, but the data then also needs to be transferred back to the survey institute. Usually a form of data donation, as described below, is used for this.

3. **Data donation**

In the case of data donation, auxiliary data is available to the respondent but not to the survey institute. The data thus already exists but the respondent acts as intermediary to actually combine the data with the survey data. Smart services fall in this category, and not under external sensors, when the data is already collected, outside the smart service, and is available in a data source that

can be used. This data is thus generated for purposes other than research but can be used for research.

There are many types of data that can be donated. For example, bank statements of the last year can be downloaded via online banking platforms, activity data from personal fitness trackers or smart meter energy data. A specific type of data that can be donated are digital traces. Thanks to the European Union's General Data Protection Regulation's (GDPR) right to data access and data portability, all data processing entities (e.g. Google, Instagram, WhatsApp) are required to provide individuals with a copy of their personal digital data upon request. This means that as an individual, you can request a copy of all your digital data, which they typically receive in the form of .zip files.

A specific type of data donation is data that is available in the cloud or an online system which can be used to fill out the survey. For example, John Deere collects agricultural data from sensors that farmers use to carry out their work. This data is saved in an online system, where the farms have access to. When these farms are invited for a survey to produce agricultural statistics respondents can log in to the John Deere system (when that specific farm is using John Deere machines). Once logged in, the agricultural data can be extracted and filled out in the online questionnaire automatically.

Furthermore, it is important to mention data integration here, as this is very closely related to data donation. Data integration is not considered a smart service though, as the respondent plays no central role. For data integration, the data are already available in a register (where the statistical institute has access to) and the respondent has to give (active or passive) permission to link its data to a survey. Subsequently the register and survey data are linked behind the screens. Therefore, we need a personal identifier, like a citizen science number or an address. The register data are used as auxiliary data, which are used to complement the survey data collection without increasing respondent burden. Examples of registers can be: smart meter data from grid operators, educational data or railroad travel data.

Generally, the respondent needs to download its own data and upload it through a portal (which can be integrated into a survey). This might be a complex process for the respondent, and especially for digital traces as the request for the data may take a few days. In these few days, respondents may drop out.

Another way is that respondents are asked to log into a cloud or online system that has the requested information available. The data is then filled out automatically in the questionnaire. This is still technically very complicated though.

Taxonomy for Smart Services

Below you will find a taxonomy in which we try to go one step deeper and classify and order the three types of smart services mentioned above. We divide these three into the mode of data collection, namely app or browser/online. It has to be said that this distinction is particularly interesting for internal mobile device sensors, not so much for external sensor systems or data donation though. Lastly, we make a distinction between active and passive forms of data collection as this has a large impact on the (burden placed on) the respondent.

		Internal (mobile device) sensor	**External sensor system**	**Data donation**

App	Active	Respondents have to perform an action with their sensors. E.g. scanning receipts in the household budget app.	Personal devices from respondents are used as sensor data source. Data needs to be linked, copied in a survey, or in another way transferred to the survey institute. E.g. using respondent's personal fitness trackers.	Respondents download their own data and upload it via a secure portal in the app.
	Passive	Sensor data is collected passively in the background. This type of data becomes (more) active when respondents are asked to validate and correct the collected data, e.g. GPS travel data.	Respondents install or wear an external sensor, device or data is sent back to the survey institute.	Reading out data from another app installed on the same device or the cloud. This could be done via an API.
Browser	Active	Respondents have to actively perform a (one-time) action, e.g. take a picture of their energy meter.	Personal devices from respondents are used as sensor data source. Data needs to be copied in a survey. E.g. using respondent's personal fitness trackers.	Respondents download their own data and upload it via a secure portal.
	Passive	Sensor data is collected passively in the background. E.g. the GPS location of the survey fill out location is tracked.	Respondents install or wear an external sensor, device or data is sent back to the survey institute.	Give permission to get data from the cloud to fill out the questionnaire, e.g. John Deere case study.

SSI examples

Let us look into the two case studies of the SSI project and see where they would fall into this taxonomy.

Household Budget Survey App

Household Budget Surveys focus on respondents' purchases and expenditures, both small and large. A standard design is a mix of a recruitment survey (for all regular expenditures, like mortgage and insurances) plus paper and/or online expenditure diaries for all small(er) purchases. In these expenditure diaries respondents have to fill out all their purchases (type, products, total price and price per product). The survey usually takes several weeks.

In the SSI project microservices for an HBS app are developed to replace the expenditure diaries. In the app respondents can use the camera to take a picture of/scan a receipt, instead of manually entering all products themselves. Receipt scanning is mostly useful for purchases that involve many products because these purchases are burdensome to insert into the diary.

The HBS, as employed in the SSI Project, is thus an example of an active sensor survey via app with internal sensor data collection. In some countries the HBS is conducted via a web-app, in that case it is an example of an active sensor survey via browser with internal sensor data collection. One survey can thus have multiple characterisations depending on how it is conducted.

Time Use Survey (TUS)

In a TUS respondents receive a time diary in which they record their time use for 10-minutes intervals. Respondents are asked to record the most important activity and the other activities they were doing at the same time. For each activity they are asked to report where they were, whether they were there alone or in the company of someone they know. Traditionally, this survey is filled out on paper and afterwards coders go through the diary and manually assign a code from a code list to each activity.

In the SSI project, this time diary is filled out by the respondent in an app. During the fieldwork period, geolocation data is collected passively by sensors incorporated in the smartphone. These locations give context to the activities and help the respondent as a mnemonic.

The TUS, as employed in the SSI Project, is thus an example of a passive sensor survey via an app with internal sensor data collection.

3. GSBPM

Introduction

GSBPM in general

The Generic Statistical Business Process Model (GSBPM) is a widely used framework within NSI's and governmental bodies that provides a structured approach for organizing and managing the statistical production process. It breaks down the statistical production process into a series of interconnected phases or steps, guiding statisticians through the entire lifecycle of producing statistical data.

The current version is v5.1 (January 2019).

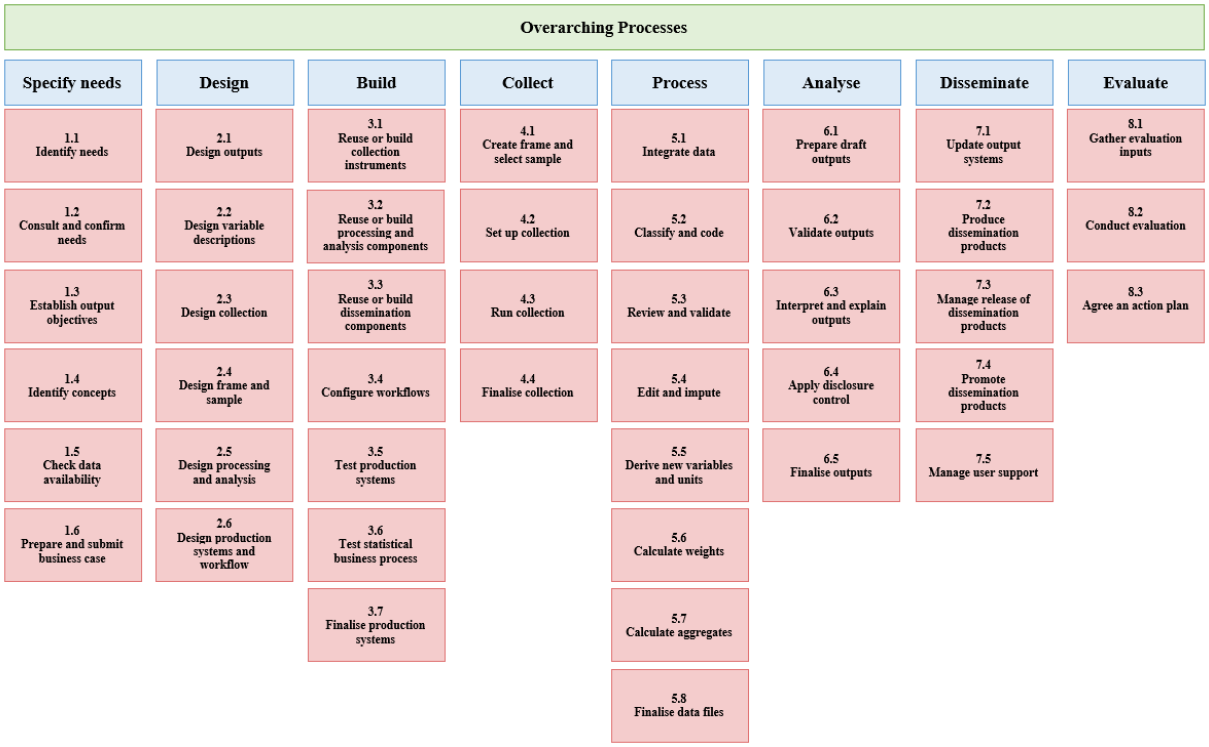
A statistical business process is a collection of related and structured activities and tasks to convert input data into statistical information. In the context of the GSBPM, organisations or groups of organisations perform statistical business processes to create official statistics to satisfy the needs of the customers/users of the statistical output.

It can be used to describe the process as-is, but also the process to-be.

It provides a standard framework and harmonised terminology to help statistical organisations to modernise their statistical production processes, as well as to share methods and components.

The total model consists of a couple of processes:

- one main process consisting of **8 phases** (from 'Specify Needs' to 'Evaluate') and within each phase several **sub-processes.**
- several **overarching processes** concerning:
    - Quality management, concerning product and process quality
    - Metadata management. This includes process-independent considerations such as metadata custodianship and ownership, quality, archiving rules, preservation, retention and disposal.
    - Data management. This includes process-independent considerations such as general data security, custodianship and ownership, data quality, archiving rules, preservation, retention and disposal.
    - Process data management. Process data can aid in detecting and understanding patterns in the data collected, as well as in evaluating the execution of the statistical business process as such.
    - Knowledge management. This ensures that statistical business processes are repeatable, mainly through the maintenance of process documentation.
    - Provider management. This includes cross-process burden management, as well as topics such as profiling and management of contact information.



Although the presentation of the GSBPM follows the logical sequence of steps in most statistical business processes, the elements of the model may occur in different orders in different circumstances. Also, some sub-processes will be revisited, forming iterative loops. Sometimes GSBPM is said to be a function model, where the functions can be put together to a process.

The GSBPM is a reference model. It is intended that the GSBPM may be used by organisations to different degrees. An organisation may choose to either implement the GSBPM directly or use it as the basis for developing customised version of the model. It may be used in some cases only as a model to which organisations refer when communicating internally or with other organisations to clarify discussion.

GSBPM and GAMSO

GSBPM is part of GAMSO. The Generic Activity Model for Statistical Organisations (GAMSO) describes and defines the activities that take place within a typical organisation that produces official statistics. It extends and complements the GSBPM by adding additional activities needed to support statistical production. The GAMSO describes activities – that is, what statistical organisations do – to ascertain that they are able to produce statistics that comply with current and future desires, insights, ways and means. The GSBPM describes the process – that is, how statistical organisations undertake the activity of statistical production.

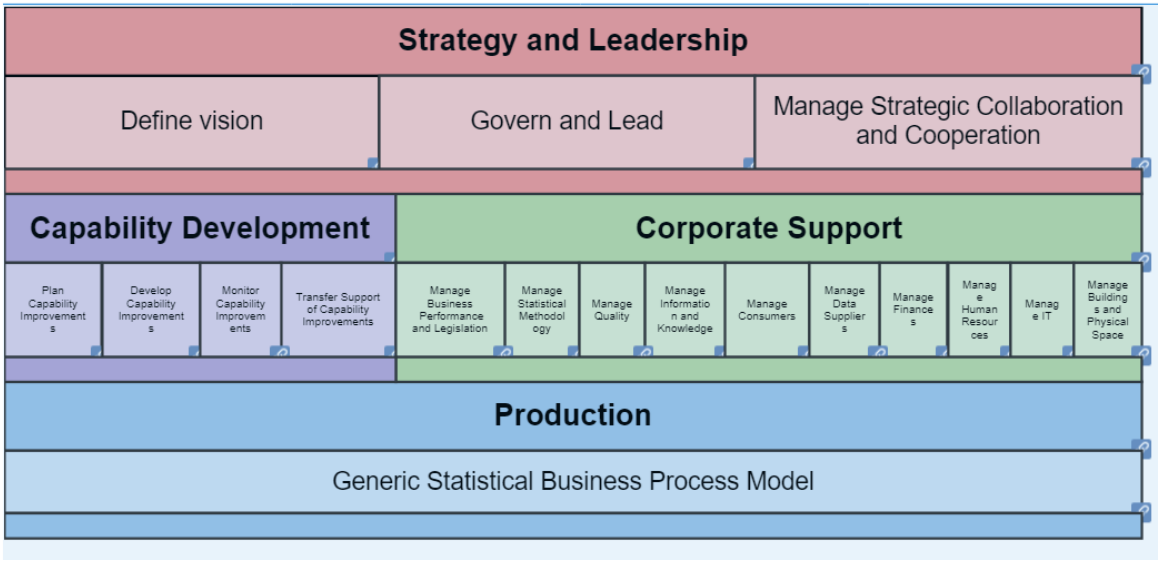


Illustration: the GAMSO model

=> The **scope** of this document is however pure the GSBPM, not the GAMSO.

In the next project stage, we will also dive into GAMSO. We see namely a relationship between the GAMSO on the one hand and actors, GSBPM and the maturity model on the other side.

Then it may also turn out that some of the things we described below in GSBPM, do more belong to, have a better place in, the GAMSO.

Goal of WP4

The goal is to elaborate the GSBPM concerning smart solutions. Earlier, in the ESSNet TSS project, the GSBPM has been extended concerning smart solutions. We are building on that here. For each GSBPM phase and sub-process, we describe what needs extra attention the moment a smart solution will be used in a survey.

On the one hand, it will be a conceptual addition, in generic terms. On the other hand, we think examples will add value. That is why we enrich the description with examples. The examples come from the three case studies in scope of the SSI project.

In SSI three case studies are considered:

1. HBS: Adding and processing purchases of goods and services reported through scans of paper tickets or uploads of e-tickets;
2. TUS: Providing a tentative timeline of stops and tracks in time use diaries;
3. Energy: Adding details on energy usage by data donation of smart energy meter data.

Scope – WP4 approach

The SSI project is divided into three project phases. The elaboration of the GSBPM is spread across these three phases.

The version in this chapter, is the end result of the Review stage. In the next two project stages we will complement this.

At the start of SSI the idea was that the work would be divided across the three project stages as follows:

- Review stage: Identification of missing GSPBM parts. This is necessary as input for the next project stage where the missing parts are elaborated.
- Smart baseline stage: Elaboration of all GSBPM stages, resulting in an expanded GSPBM supplemented with smart survey subjects
- Smart advance stage: Updating the deliverable of the smart baseline stage with new insights, resulting in a definite version of the deliverable.

However, gradually, during the review stage we decided to use a different approach:

- We did not limit ourselves identifying missing parts. We also started actually by describing – not just identifying – the missing parts.
- We restricted the scope – for this SSI project stage - to the first three GSBPM phases: 'Specify needs', 'Design' and 'Build'. This does not mean that, for these three phases, this is the end result of the SSI project. In the next two project stages we will still add new findings and ideas to these three GSBPM phases.
- And, we also added the 'machine learning' perspective. We did this for all GSBPM phases.

Reader's guide

There is a separate paragraph for each of the 8 GSBPM phases (for this version restricted to the first 3 phases). Within, each sub process has a separate sub paragraph.

Each paragraph starts with a short description of the sub process, marked in blue. This is taken verbatim from the GSBPM.

Examples (from HBS, TUS and/or Energy) are put in a box, to distinguish it from the general text.

At the end there is a separate chapter from the perspective of machine learning. Machine learning relates to different phases and sub processes. Therefore, we decided to put it in a separate chapter.

Sources

The following sources where used.

[1]: Generic Statistical Business Process Model (GSBPM) v. 5.1. January (2019). GSBPM v5.1 - Generic Statistical Business Process Model - UNECE Statswiki

[2]: essnet_smartsurveys_wp3_del_3.1_25022021.pdf (europa.eu)

[3]: essnet_smartsurveys_wp3_del_3.2_310122_0.pdf (europa.eu)

[4]: wp3_deliverable_3.3_-enhanced_framework_08_04_2022.pdf (europa.eu)

[5]: https://cros-legacy.ec.europa.eu/system/files/220406_wp2_-_deliverable_2.2_-_modularity_consumption_0.pdf

[6]: https://cros-legacy.ec.europa.eu/system/files/essnet_smart_statistics_-_wp2_2_-_dl2_10.pdf

GSBPM phase: 1. Specify Needs

Remark: In source [4] the GSBPM was elaborated with two new sub-processes: "Identity smart data requirements" and "Identify legal requirements". In the first draft version of this document, we had adopted that unabridged. But for many reviewers that was confusing, because it was not clear why that deserves separate sub-processes and why it was not just considered as another type of needs. That is why we did not take that over from [4].

1.1 Identify needs

A trigger to start exploring smart solutions can be – among other things – the decline of responsrates and the adopting in society of smart appliances. So, this triggers the process of identifying needs.

Needs can be very broad. On one hand it is about needs concerning the output of the statistical process for the study in question. But it is also about needs for the (business) process, the IT/systems to be used, and the methodology to be used. When identifying the needs also legal needs are a point of attention. That is because legal constitutes a frame that will restrict the solution / the design; at least the design should take the legal requirements into account. When a smart solution is part of the survey- and process design, extra attention is needed for the legal aspects of consent, privacy and information security.

When looking at smart surveys different aspects need extra attention. From all these aspects needs can come forth.

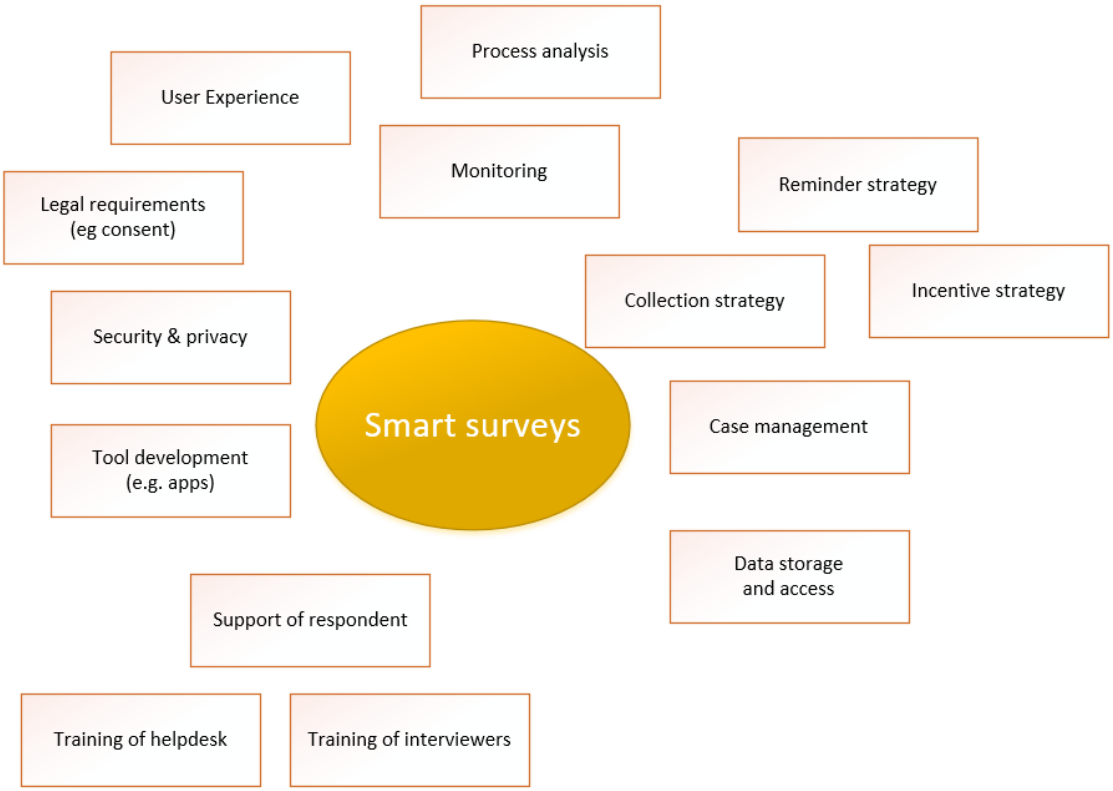


Illustration: Aspects that need extra attention in case of smart surveys (this is not per se exhaustive)

Example HBS

The Household Budget Survey (HBS) will be conducted in accordance with a new survey design. The design should become consistent with the Integrated European Social Statistics (IESS) regulation. A 'smart solution' for scanning purchase receipts will be used. Primary aim is to observe expenditure (spending) of households. Besides a general questionnaire, the survey also consists of the 'household book' ('diary') and 'recurring expenses' sections. The new design will use a 'smart' application. This includes being able to scan purchase receipts and convert them into processable information via OCR. Each expense in the statistical process should result in a COICOP coding.

When looking at HBS there are some specific needs concerning the application of the smart solution. Needs can be from different perspectives: business/process, IT/system, methodology and legal/privacy.

This list is meant as an example, it is not an exhaustive list and not all needs are applicable to all the specific situations at the different NSI's. It is also not said that these needs are 'correct'. It is just to give the reader an idea of needs to think of.

- There are specific needs for the functionality of the app.
- The respondent should be able to switch between devices ('app-mode') and web browser (cawi-mode) at any time, with the information synchronised.
- The respondent should be able to use the same login code for the cawi questionnaire as for the app;
- The respondent should be able to use the app offline as well;
- The respondent should be able to scan a ticket, but also to add a ticket by manually entering the receipt information;
- The system (app) should be able to read the different items of information on the ticket to digitalize them and to store the information in the corresponding variables (OCR);
- The respondent should be able to modify the information scanned from a ticket;
- Every expense should result in a COICOP coding;
- Attention should be paid to the privacy aspect;
- The respondent must have control over the data shared with the NSI;
- The respondent must recognize the data that is fed back to him;
- The app should be as easy as possible for the respondent to use;
- As few actions as possible should be performed by the respondent and the respondent should not have to do any unnecessary work (perceived administrative burden);
- Checks performed on the device should not impede the respondent, in the sense that, for example, they do not consume a lot of computing capacity of the device or drain the battery, making the functionality of the device unacceptably slow.
- An expenditure can be obtained either manually, via a scan of a receipt or via a digital receipt;
- The respondent does not have to check every time he/she digitises a purchase receipt to see if the receipt has been processed properly. The user may also do this once after he/she has digitised all purchase receipts;
- Unrecognised receipts or items can possibly be captured by manual typing group at the NSI;
- We should be able to monitor the usage of the app;
- Interviewers need to be trained to help respondents with installing and using the app;
- The helpdesk should be able to answer questions from respondents concerning the usage of the app;
- For receipt processing, machine learning is necessary to extract text through OCR, to interpret texts through language processing and to classify extracted products to COICOP classification. This also needs a process to train (and retrain) the machine learning methods. Each of these machine learning algorithms also need specific training data that needs to be labelled by human annotators.

- Machine learning performance needs to be monitored over time; while the performance of OCR algorithms is expected to stay quite stable, especially the COICOP classification algorithms have to deal with a large fluctuation in products. An algorithm that performs well now, may not be acceptable in one or two months from now.
- For the Process phase of GSBPM there is the need to couple the data from the questionnaire(s) and from the dairy.
- etcetera …

Example TUS

TUS is an example of a survey that uses a diary to record activities for a certain time period (few days up to a week or longer). A 'smart solution' for determining geo-locations is used.

In TUS there is the need to passively collect geo-location data. And to contextualize it with additional information from other sources (such as OpenStreetMap). The geo-location information needs to be imputed into the diary.

Several needs are of the same nature as in HBS. E.g. needs concerning:

The coupling of the data from the questionnaire(s) and from the dairy in the GSBPM Process phase.

The knowledge needed by the helpdesk and interviewers.

A process for training machine learning algorithms.

Adding smart elements to a research design increases its complexity, almost by definition. This creates new (type of) needs. The above examples give an idea what type of needs you can think of.

1.2 Consult and confirm needs

This sub-process is also focusing on the needs. Concerning smart surveys, the same aspects need attention as mentioned in 1.1. The sub-process itself is however not different from a non-smart survey.

1.3 Establish output objectives

In general, when talking about output, there should be some idea of data sources available. New sources could give more/new opportunities for the output. In fact, there is a close relationship with sub-process 1.5 ‘Check data availability’. This means that staff involved in this sub-process should have some knowledge of smart solutions and the new data sources and new opportunities that come with it.

On the other hand: Talking about output objectives, the main focus is on the end output of the entire statistical process. But you could also focus on the output of the separate (GSBPM-)phases of the statistical process, e.g. the output of the Collect phase. Seen from that angle you could specify specific output objectives regarding the smart solution to be used.

Example HBS

The output objectives of the data collection phase should be specified. There is e.g. the question where in the process the classification to the COICOP takes place. Is that within the data collection phase (pre- or in-survey)? In that case the COICOP is output of the data collection. Or is it in the Processing phase (post-survey)? In that case the output objective of the data collection is different; in that case 'only' the article description that can be classified is needed in the output. Answering this question depends partly on the need whether the COICOP-code is fed back to the respondent. These kinds of decisions have impact on the process but also on IT system-functionality. And thus, have impact on the business case (see 1.6).

1.4 Identify concepts

The concept that the user is interested in, will affect the decision whether to opt for the use of smart solutions within a survey.

1.5 Check data availability

In this case, smart solutions provide new sources of data. That means that staff involved in this sub-process should be aware of these sources. They should be aware of possibilities of data donation, data collection through apps, etc and in what cases these data are suitable to use (e.g. within HBS and TUS) and what limitations/restrictions there are on these data sources.

From a business process view, this means extra knowledge is necessary for the involved staff.

The legal framework that is mentioned in the GSBPM, is to assess the legal implications (consent management and legal/ethical analysis) and privacy issues related - in this case - to the type of smart data provider, sensor, storage and processing environment.

As part of such a framework, a Data Protection Impact Assessment (DPIA) should be conducted. That is a process to help to identify and minimise the data protection risks of a project.

Also paying attention to privacy-by-design and privacy enhancing technologies (PET) is needed.

HBS example

Source: [5]

Respondents are encouraged to crop images and remove parts of the receipt that are irrelevant to the HBS. They are informed about this through e.g. an in-app instruction movie. However, they can ignore, forget or overlook this and may submit receipts carrying potentially identifying information. If it is demanded by legal constraints, then there is no other solution than to first ask respondents for explicit consent.

Another issue is the anonymization of individual persons in the household. Is it okay that they see each other's purchases? When this is not the case, the only current solution is to provide each member with their own credentials. This puts requirements on the backend of the solution.

1.6 Prepare and submit business case

In the sub-process 1.1 to 1.5 the focus is on the needs. Primarily the needs on the output of the statistical process, so what output should be provided, what are the output objectives, what are the concepts to be measured and what data is available. This gives a first notion of feasibility: if the needed data sources are not available, the output is not feasible.

There is no sub-process in this phase where alternative solutions are described. So no 'to-be' solution is proposed. That is in fact also not yet possible because the design of e.g. the survey could have impact on the (statistical) business process to be used. And designing the production system is done in 2.6.

So, we assume that this is an iterative process and that first a very high-level business case is described, that then part of the design is done after which a more detailed business case can be described.

Because of this we state here that the activity of 'preparing a business case' in itself, is an activity that is not different when smart solutions are involved. Of course, the input for describing the business case is different, but this input is generated in other sub-process. So, the impact of smart solutions on GSBPM is described in these other sub-processes.

GSBPM phase: 2. Design

2.1 Design outputs

=> This was not in scope of the Review stage. It will be addressed in the next project stage.

2.2 Design variable descriptions

Regarding smart solutions it is necessary to have knowledge about what concepts are suitable for measuring by means of a smart solution. E.g., the use of sensors adds value in three cases:

- the survey requires information that a respondent does not know or understand,
- the survey requires a lot of effort and/or time from a respondent, and/or
- the survey aims to measure concepts that do not lend themselves to a question-answer instrument.

Example HBS

In HBS the smart solution is used to scan and interpret purchase receipts. Providing this information manually would take the respondent lots of time. And by scanning we can better measure the concepts.

Example TUS

In TUS the smart solution is used to measure locations, travel time and time spent at a certain location. This is information that is hard for a respondent to provide accurately. In this case the quality of the data can be improved.

So, in both examples, these smart solutions contribute to measuring these concepts.

A characteristic of smart survey solutions is that

a. Overall data obtained: more data is obtained than is strictly necessary to create the output of the statistical production process.
b. Undefined source data: it is data from a source (such as a sensor) whose definition we cannot determine ourselves.

c. Natural language data: data can be obtained in "natural language" where it is necessary to encode it to a code of a classification.

Having a metadata management system to a greater or lesser extent is a maturity criteria. Having a central administration within the NSI, describing all relevant resources with their variables, would be a high level in terms of maturity.

In GSBPM, there is a relationship between sub process 2.2 and the overarching process "Metadata management". The premise is that metadata administration is the responsibility of the "Metadata management" process. The sub process "Design variable descriptions" should in principle draw on the information already available in the "Metadata management" process. If a variable is already described there, it does not have to be described again. And vice versa, if there is a new variable, the meta should be routed to the "Metadata management" process.

a) Overall data obtained

We can make the following distinction:

- Variables from the source that are not used at all,
- Variables from the source that are passed on from the data collection process to the next process stage
- Variables from the source that are only used in the data collection process. For example, to provide feedback to the respondent.

Of the 2nd and 3rd categories, it is relevant to have the metadata of the variables. Of the 1st category, the meta will be needed in the "Specify needs" phase - to determine whether the variable is relevant and useful - but no longer in the "Design" phase. In general, the data of the 1st category is not available for later processes as there is no statistical purpose.

This point also has a clear relationship with privacy. For example, how does this relate to "data minimization". That point is further described under sub process "2.3 Design collection".

b) Undefined source data

The variables from the source are a given. In itself, this is not specific to smart surveys, as we already use external sources (administrative data), but it does require attention. We will have to determine whether the source's available metadata meets our requirements. If necessary, an addition of metadata or a translation of metadata will be needed that is usable for us.

c) Natural language data

An example of data in "natural language" is the scanning of purchase receipts at HBS. The description of the purchased item is described in natural language. This has to be translated into a code of a classification (e.g. COICOP). The metadata of the classification must be known and must therefore be captured in this sub process. How the translation should take place should also be described, of course.

2.3 Design collection

In this sub-process, it will have to be determined whether and which type of smart survey solution will be used. The type determines the aspects that need extra attention in the design.

Smart surveys have one or more of the following features/functions:

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

Example: At HBS, a questionnaire is used in combination with (or next to) an app, where the app is used to scan purchase receipts. The app can use one of the internal 'sensors' of the device, namely the camera.

Aspects that need attention:

Consent

The question that needs to be answered is how and at what point the respondent gives permission to use data. On the one hand this concerns the use of the data source (e.g. a sensor). On the other hand, this concerns the consent to use the observed data as input for the rest of the statistical process.

When using an external source, such as a sensor, the respondent will have to give prior consent. This can e.g. be done by explicitly asking permission or by stating that use is going to be made unless the respondent explicitly refuses. This is a choice to be made deliberately and may depend on the law/regulations in the relevant country of the NSI. This also involves considering whether and how detailed the respondent should be told what data is being used and for what purpose. This too may be partly determined by law/regulation.

With a 'normal' cawi questionnaire, the respondent gives permission for observed data to be used for the statistical process the moment he clicks the 'send' button. This makes sense because the respondent basically fills in the questionnaire from front to back. However, in some forms of smart surveys, data is collected over a certain period of time - say, a few days. Can the data recorded during that period (e.g. from the first day) then already be considered as data that may be used? Is consent sought after the end of each sub-period (e.g. day)? Is consent asked in advance? Or is permission asked only after the end of the entire period? These are questions that need to be answered in the design.

Feedback of data

When using data from external sources (such as sensors), it should be determined whether, to what extent and in what form that data is fed back to the respondent. Is this done at a detailed level, for instance, or is data aggregated? Data should be recognizable for the respondent. And what is the purpose? For instance, is the purpose to let the respondent check the data and complete/improve it? Or is the goal some kind of incentive?

The level of detail can also be determined by the features of the used software. Location is an example, with some apps/devices it is possible to ask the user to share his/her location 1) once, 2) when the app is used, 3) never.

Example HBS

With HBS, the choice can be to feed back all items on the purchase receipt and give the respondent the option to change them. Another option is to feed back only those items that are unclear / cannot be classified, to let the respondent give input so that it does become clear. Another option could be to give no feedback at all to the respondent or allow to change everything.

Example TUS

TUS is a household survey. Respondents use a log or a time use diary of at least twenty-four consecutive hours to self-report their daily behaviour in a chronological and open-ended fashion on an activity-to-activity basis. In the time use diary, respondents specify – for each new activity – the start and end time as well as some contextual information like the place of occurrence and the possible presence of others. So, there are multiple registration moments per day. Activity coding is necessary. One diary concerns a weekday, and one diary concerns a weekend day (the same two days for all household members).

Clearly, the choice has an impact on the process and on the functionality that the collection instrument (e.g. the app) must have. But it is also influenced by what is technically possible and supported by software APIs.

Data collection tool

With smart surveys, an additional type of instrument will have to be developed. For example, an app.

This requires different knowledge than, say, a regular questionnaire. NSIs will have a certain process and standards and templates for questionnaire design. This will need to be extended to include a process for the new type of smart instruments.

An app can be deployed together with a regular questionnaire. Then the cooperation/integration of the two instruments needs extra attention. For example, is data from the app included in the questionnaire?

Version management of an app is a separate issue. Because if several versions of the app arise over time, how do you make sure that a respondent uses the right version of the app?

In this respect also the type of device and the available sensors. For example, Apple vs. Android. Samsung vs. Google etc. Which type of accelerometer or location sensors do these devices have? What is the resolution of this sensor? What is the recording frequency of this sensor, etc.

Data model

Data from a regular questionnaire is captured according to the questionnaire's data model. And that data is the output of the data collection process. But the data obtained from a smart application, is it stored in a separate data model and delivered from the data collection? Or will that data be included in the questionnaire data model? The latter has the advantage that no extra database management is needed.

A choice will also have to be made there. The choice has implications for the "Process" phase. If data is delivered in multiple data models, extra activities are needed in the "Process" phase.

Monitoring

A smart application will basically generate additional requirements/requirements for monitoring. Basically, you want to gain insight into the use of the smart application, the quality of the data obtained and the quality of the translation rules when classifying.

For the former, think about:

How many respondents install the app, is it used/not used by respondent, where does the respondent drop out, how much time does the respondent spend in the app. But also: how often does the respondent need to correct the data from the machine learning model, how much time is spend using that. This can also be subject of a usability study performed during the design phase.

Monitoring of the quality of the classification rules will need to happen at several points in time. First, when training the classification model. There, you split the data you already have into a dataset for training and one for testing. Then you test the trained model on the test data you have. That is an approximation of the quality of the model. The quality that is achieved in production can only be determined during the use of the app/smart survey and then only by approximation. You could then think about how often a respondent has to adjust the scanned receipt.

Data collection strategy

"Smart" refers to data collection that combines passive or sensor data from personal smart devices (e.g., GPS, accelerometer) with active data explicitly provided by the respondent (e.g., responses to queries). So, the data collection strategy should define what part of the data is collected passively and what part actively, and also how these data can be coupled so that the data can be merged.

Mix of modes: In data collection often a mix of modes is used. When 'going smart' a new mode is added that includes the passive collection of data. This will often be combined with the cawi-mode, where questions are asked through a cawi questionnaire.

Communication strategy and tools

One of the challenges is how to persuade a respondent to participate in a survey where an app or other smart solution is used. Communication strategy and tools are essential in this.

When using a smart application, the respondent will have to be informed about it. How will this be done? Will additional information be included in the invitation letter, reminders, etc? For example, will additional credentials need to be communicated, e.g. to access the app?

If an interviewer/outreach worker helps respondents with the app, should additional communication tools be created that they can leave with the respondent? Consider a manual.

Is it desirable to place communication tools such as a manual on the NSI website where respondents can read/download it?

Also, the use of incentives is a separate aspect to think about as a different type of incentive could become available. An example could be the showing of statistics (graphics) on the expenditures of the respondent related to participating households.

So, these are all aspects that need to be considered in case of smart solutions when designing the collection.

2.4 Design frame and sample

Sampling methodology should take into account the differences in the use of smart devices within particular subsets of the selected sample. However, the sampling methodology as such is not special for smart solutions. Methodology is still focused on how to increase representativeness and prevent distortions.

However, some NSI's might consider special activities for composing the frame or sample. This has to do with how to persuade a respondent to participate in a survey where an app or other smart solution is used. Some NSI's might consider to 'go out on the streets', e.g. with posters or marketing ads, to ask people to sign up for a study. Thereby mentioning that a smart solution is used in the study.

2.5 Design processing and analysis

=> This was not in scope of the Review stage. It will be addressed in the next project stage.

2.6 Design production systems and workflow

Production process - in general

A smart survey application will have an impact on the design of the production process. Under "Design collection", a number of aspects have been named that will have an impact. Grossly, we can distinguish the following process parts where there will be impact:

- The process of providing the respondent with the app and ensuring that they can work properly with the app.
- The process of providing respondents with devices: E.g. in the case of activity trackers also the devices need to be sent to respondents.
- The front-office process: The "respondent journey". When using an app, the respondent will have to do additional activities.
- The back-office process:
    - On the one hand side the design is done in the back-office
    - On the other hand, part of the production process is done in the back-office. E.g. ensuring that the data obtained is passed on to the "Process" phase, and monitoring that process.

Classify

In certain types of smart survey applications, 'natural language' needs to be translated into a code in a classification. For example, in HBS item description to COICOP. Then the question is where that takes place in the process. There is also the question of whether classification is done fully automatically or whether a manual activity is also provided. Is it seen as a responsibility of the data collection process or of the "Process" phase? So, in the former case, the data collection process produces classification codes; in the latter case (also) text/input to be classified. The choice has impact on organization, process, human workers and IT-tools.

Apart from the coding itself, there is also a task of maintaining the quality of the coding methodology. A process will have to be provided for that too. One choice could be to design this entirely separate from the production process (although real life data can serve as input (training data)). Another choice is to (partially) intertwine this with the production process; one of the

questions is what role the respondent plays in that process. The third option is to do both. First train the model with data you already have. Then use the respondent's feedback to observe the quality of the coding and possibly retrain the model with the data collected.

Those choices and arriving at a process for this are part of "Design production systems and workflow".

GSBPM phase: 3. Build

3.1 Reuse or build collection instruments

App development

When using apps, the organization will have to make a choice about app development and management. There are essentially four choices for an NSI:

- Develop entirely in-house
- Develop together with one or more other parties (e.g. NSIs)
- Outsource the development. Where the choice can be made to do the management internally or to outsource that too.
- Reuse or purchase an existing app.

The question is of course how much can you control the data collection when the app is purchased? Do you get access to all the data and metadata about the data collected? Same is true when using external machine learning models.

If done wholly or partly within the NSI, the NSI will need to have the knowledge and resources to develop and manage apps to a greater or lesser extent. A capability will be added for the organization.

In the Common Statistical Production Architecture (CSPA) is stated that it is good to explore the reuse / sharing of solutions and services and the standardization of processes, to reduce costs. Concerning smart solutions, you could think about developing shareable micro services (e.g. the scanning of receipts and coding text to a classification).

The question is whether the development of an app is seen by an NSI as a software development activity or more as a 'questionnaire' development activity.

The creation (or purchase) of software is not a part of the statistical process. Software development is a separate business process. If an NSI perceives it like that, the development of the app is not located in the 'reuse or build collection instrument' sub process. But if an NSI perceives the app as a questionnaire then it would be within this activity. This is an important choice to be made by an NSI. It probably also determines where the app development and the app developers are located within the organizational structure. E.g. at CBS this was an explicit choice that had to be made. At CBS was decided that it is 'software development' and the responsibility for developing apps was assigned to an agile team[1].

And this concerns not just the development of the app itself. Probably there should also be a new 'app channel'. This 'app channel' also needs to be developed and integrated with the rest of the process and application landscape.

So, app development raises more questions, more concerns, than someone might think of.

Questionnaire

An app will normally be used together with a questionnaire. This requires extra attention. Questionnaire developers will also need to have some knowledge of how apps work technically, to achieve proper integration. In addition, of course, the content must work as established in "Design".

3.2 Reuse or build processing and analysis components

=> This was not in scope of the Review stage. It will be addressed in the next project stage.

3.3 Reuse or build dissemination components

=> This was not in scope of the Review stage. It will be addressed in the next project stage.

3.4 Configure workflows

Setting up the workflows will involve extra work. This is because the additional activities in the process have to be tuned in. This is not only collection sec but also, for example, monitoring. A specific point of attention is the scheduling of interviewers/outreach staff if they are used to guide respondents in obtaining and using the app. This is a scheduling activity on the one hand but, on the other, they also need to be given the right resources.

3.5 Test production systems

In the case of smart solutions, there will be additional testing. For instance, the operation of the app and, in particular, its integration with the questionnaire and interaction with other applications will have to be tested. Testing these smart services/apps will also need specific test infrastructure that may not be available in all NSIs (or difficult to implement). It will also have to be established that microservices work correctly. If microservices already exist and are 'shared', it can be assumed that they work correctly. In that case, the test should focus on integration with the rest of the applications.

Machine learning routines also have to be tested, so that it is established that they work according to specifications. A separate focus here is having good and sufficiently comprehensive test data available.

3.6 Test statistical business process

This test checks whether the set-up workflow works in accordance with the specifications from the design phase. Because with smart solutions the workflow is generally more complex, this sub process is also more complex. For the test, among other things, the app will have to be available. This needs to be installed on a device, or perhaps on several types of devices, to be able to test the flow completely.

3.7 Finalise production systems

This also needs extra attention regarding smart surveys. There should be specific training for e.g.

- Helpdesk members: They should be prepared that they can get questions about installing the app, the internal working of the app or technical problems with the app
- Interviewers/outreach workers: they should be trained to help the respondent with the app.

For these same aspects, also extra user manuals, other documentation, training videos etcetera could be necessary.

When moving the process components to the production environment, the app should be made available in the app store where the app can be downloaded. This is a separate part of the process that is not in place at an NSI when apps are used for the first time (*comment: some SSI project members even suggest defining a separate GSBPM sub process for this*).

Special attention is needed when at some point a new version of the app is developed. Will that 'overwrite' the app in the app-store? Will there be two versions in the app store? Will the new app work with all devices, even those that are a couple of years old?

So, this needs a good version management process. Especially when the new version is necessary at the moment a survey is running.

Machine Learning in the GSBPM

In the following section, we will shortly identify the opportunities of using machine learning in official statistics. First, we will look at the opportunities of machine learning for official statistics in general. Then, we will focus on using machine learning in surveys and connect it to the Generic Statistical Business Process Model (GSBPM). Last, we will apply the GSBPM to three smart survey case studies.

Machine Learning in Official Statistics

In recent years, machine learning has become more and more important in society. It should not come as a surprise that machine learning has been the subject of several studies in official statistics (Beck et. al., 2018). These studies identify machine learning as a promising direction of research in a number of areas. First of all, machine learning can automate statistical processes that previously could only be done by humans. The imputation of missing records from similar records could previously only be done by human analysts but is an area that now can be largely automated using machine learning algorithms. Second, machine learning can open up new sources and types of information. Big data sets that were too unwieldy for human researchers to analyse can now be mined for patterns using automatic machine learning techniques. What's more, while official statistics was previously limited to tabular data sources coming from surveys and register data, machine learning can help analyse new forms of data like free text (survey answers or websites), image data (receipts or satellite imagery), and signal data (accelerometer data). Last, machine learning can be used to implement smart assistant software that can help the user or respondent accomplish a task that previously would have been difficult or cumbersome to complete. These machine learning assistants enable also the use of so-called smart surveys in official statistics, which we will dive into more in the next section.

Machine Learning in Survey Statistics

The role and opportunities of machine learning in survey statistics can analysed by using the GSBPM[1] as a framework. The GSBPM is a model describing the production of official statistics using primary data (see figure 1). The GSBPM consists of eight phases with each phase further divided into sub-processes. Two previous studies have already identified the phases and sub-processes of the GSBPM that would be promising for machine learning (Beck, et.al., 2018 and Yung, et. al., 2018). Both studies differ slightly in the sub-processes that they identify as being promising. Figure 1 shows a combination of both studies; the sub-processes marked with a red rectangle were identified as promising areas for machine learning. Summarising, the following sub-processes were identified as promising:

- **Sub-process 2.4: Design frame and sample**, using *clustering algorithms* in the record-linking when preparing the survey sample from administrative data, and *classification algorithms* and *clustering algorithms* in coding activities when preparing and assessing the quality of sampling frames.
- **Sub-process 4.1: Create frame and select sample**, *classification algorithms* could be used to stratify the population.
- **Sub-process 4.3: Run Collection**, machine learning can be used in collection management strategies by estimating the response probability of individual units by using information about the entire sample. In addition, machine learning can be used for data verification during collection, i.e. identifying outliers, of process free-text answers or comments.
- **Sub-process 5.1: Integrate data,** machine learning can be used to integrate data from multiple sources and afterwards to clean the data.
- **Sub-process 5.2: Classify and code**, *classification algorithms* can be used classify and codes data to standard classification schemes.

- **Sub-process 5.4: Edit and Impute,** machine learning can be used to impute for missing or incorrect data. The machine learning can be used to identify groups of similar units.
- **Sub-process 5.6: Calculate weights,** machine learning can be used to estimate response probabilities, imputation classes, and determine calibration groups.
- **Sub-process 6.2: Validate outputs & Sub-process 6.3**: Interpret and explain outputs, machine learning can assist in identifying outlying estimates.

Since 2018, when both the cited resources were written, machine learning and artificial intelligence (AI) has become increasingly more powerful. Generative AI can generate life-like images (Midjourney[2], Dall-E[3], and Stable Diffusion[4]) and can answer free-text questions truthfully with generated free-text (Bard[5], ChatGPT[6]). While there are still many challenges, in our opinion, these techniques will become more and more powerful and will gradually be able to assist in more complex tasks that were previously the domain of human professionals. Therefore, in the near future, we expect the role of AI in the later phases of the GSBPM (phases 6 and 7) to become larger. We think AI will be able to help interpret results (sub-process 6.3), help produce and promote dissemination products (sub-processes 7.2 and 7.4) and manage user support (sub-process 7.5). Using the tools available right now, we already are able to improve existing survey processes. In the next section, therefore, we will use the GSBPM to analyse the practical application of machine learning in three smart-survey case studies.

Three smart-survey case studies

Within the SSI project we focus on smart surveys, i.e. the use of machine learning to augment traditional surveys. Especially, we focus on making the life for the respondent easier in attempt to battle non-response. In the SSI project, we focus on a couple of case studies, two of which are highlighted here, the time use survey (TUS) and the household budget survey (HBS). A third case study, using data from wearables to augment the health survey is another example currently developed at Statistic Netherlands.

Time Use Survey

The time use survey aims investigation the behaviour of the respondents. On the one hand, it aims to capture what people do, on the other why they do it. To this cause, respondents are given an activity diary that they have to fill out every day. Previously, the time-use survey would use paper-based

diaries, but more recently software platforms like MOTUS (**insert reference here**) have been used. These software platforms provide the respondent with a taxonomy of pre-defined tasks that they can choose from. Machine learning would have a supporting role, making it easier and more user friendly for the user to fill in their daily activity diary (integration of sensor data). As an example, machine learning techniques could use location data from GPS sensors to derive concepts like home, work, shop, restaurant, and provide a list with suitable activities accordingly. In this way, it is made easier for the respondent to fill out the survey, and also the quality of the answers is improved. The following GSBPM sub-processes are influenced by the introduction of a Time use survey app:

- **Sub-process 2.3 Design Collection:** by introducing machine learning in the survey, the collection process is affected and needs to be re-designed. In this sub-process, a choice should be made for the type of machine learning model used.
- **Sub-processes 3.1 Build Collection Instrument:** the machine learning model chosen in sub-process 2.3 needs to be trained and tested to be part of the collection phase. For this, training data should be collected and a model pipeline for processing the data during collection should be built.
- **Sub-process 4.3 Run Collection:** the machine learning process in the time use survey assists the respondent to fill out the survey more quickly and more accurately during the collection phase by suggesting activities to the user, for instance based on the time of day, their location, or observed patterns of activity.
- **Sub-process 6.2 Validate outputs:** a machine learning model used in production should be monitored as often, over time, models will encounter input that they do not see during model training. The results of the model can therefore degrade over time and influence the results of the survey. Monitoring should detect model degradation on time and make sure the machine learning model keeps up to date with the current situation.

Household Budget Survey

The household budget survey in many respects is similar to the time use survey, as both aim to gather information about daily lives and their allocation of resources. While both surveys provide insights into different aspects of respondents' lives, their primary focus differs. The time use survey focuses on understanding how respondents allocate their time across various activities throughout the day. On the other hand, the household budget survey aims to collect detailed information about respondents' income, expenditures, and savings. That being said, the data collection of both surveys has many similarities as they both ask the respondents to fill out a diary. Where the diary in the time-use survey contains respondents' activities, the diary in the household budget survey contains daily expenses. In this sense, a software platform supporting the respondent to fill out the household budget survey looks different from one for the time use survey.

For instance, the household budget survey app used in this SSI project allows the user to scan receipts, edit the entries in those receipts, and classify the products on those receipts using the COICOP taxonomy. In several stages in this app machine learning is used. First, the photo of the receipt is analysed using segmentation algorithms to detect the presence and the location of the receipt. When a receipt is present, the text on the receipt is extracted using OCR software which often also uses machine learning. The text of the receipt is then sent to a server where it is further analysed and classified into the COICOP taxonomy. This classification also uses machine learning. Summarising, the following GSBPM sub-processes are influenced by the introduction of the Household Budget Survey app:

- **Sub-process 2.3 Design Collection:** by introducing machine learning in the survey, the collection process is affected and needs to be re-designed. In this sub-process, a choice should be made for the type of machine learning model used.
- **Sub-processes 3.1 Build Collection Instrument:** the machine learning model chosen in sub-process 2.3 needs to be trained and tested to be part of the collection phase. For this, training data should be collected and a model pipeline for processing the data during collection should be built.
- **Sub-process 4.3 Run Collection**, during collection, instead of a paper and pencil approach, the users can now use an app. Instead of manually writing down the products, by scanning receipts they can now add a long list of products instantaneously. While machine learning can greatly simplify the data entry process for the respondent, it can also cause new challenges and unexpected behaviour. For instance, light conditions, photo background, and the quality of the receipt (wrinkled/unwrinkled) can affect the quality of the scan greatly. In the worst case, this could also lead to a decrease in data quality or even non-response. In contrast to standard surveys, respondents need to be prepared and assisted to deal with the machine learning challenges during the data collection phase.
- **Sub-process 5.2 Classify & Code**, after the receipt scans have been processed to text, the text needs to be classified using the COICOP taxonomy. Machine learning is used to perform this classification.
- **Sub-process 6.2 Validate outputs:** a machine learning model used in production should be monitored as often, over time, models will encounter input that they do not see during model training. The results of the model can therefore degrade over time and influence the results of the survey. Monitoring should detect model degradation on time and make sure the machine learning model keeps up to date with the current situation.

Health Survey

The health survey aims at investigating the health of a countries' population. While the survey can capture some health areas quite accurately, other areas are prone to subjectiveness. Especially, the questions about daily activity patterns, and whether they are of moderate or intense, are highly subjective. It has become clear from previous surveys, that daily movement intensity is largely overestimated by most of the participants. Because of this, several institutions, amongst others Statistic Netherlands and Norwegian School of Sport Sciences in cooperation with Statistics Norway, started pilot studies to collect more objective measures of daily movement patterns and especially moderate to vigorous activity (MVPA). Next to the traditional survey, respondents are also given activity trackers that they have to wear for a week. The activity trackers are worn on the thigh and record the acceleration, or increase in speed, of the respondent in three axes (sidewards, forwards, and upwards), giving a detailed view of the respondent's daily activity.

The sensor data can be analysed by machine learning algorithms to provide insights into (1) type of activity (walking, running, sedentary etc.) and (2) the intensity of the activities. Using this data, we expect that questions about the amount of moderate or vigorous activity, the amount of sedentary behaviour, and amount of sleep, can be replaced by more objective measures. In this sense, the following sub-processes in the GSBPM are involved in this smart survey:

- **The Design Phase, especially sub-processes 2.3 Design Collection, 2.5 Design processing and analysis, and 2.6 design production systems and workflow**, introducing a wearable device into the survey data collection will influence its design, data collection will run different and also the processing of the data will have to be adapted. The processing of signal data from

activity trackers uses different techniques than the tools normally used to analyse survey results.

- **The Collect Phase, especially Sub-processes 4.2 Setup Collection, 4.3 Run collection and 4.4 finalise collection**, next to the survey, respondents should be equipped with the activity trackers, which after use should also be sent back to the NSI. In addition, the data from the activity trackers should be collected separately from the standard survey data. The logistics of each of these steps is therefore heavily influenced by the addition of an extra source of data.
- **The Process Phase, especially Sub-process 5.1 Integrate Data,** where data from the activity trackers needs to be integrated with the survey data (5.1) and **5.2 Classify and Code,** where the signal data from the activity trackers is classified into an activity taxonomy such as sleep, walk, run, sedentary, etc. which will serve as further input to the survey.

References

Beck, M., Dumpert, F. and Feuerhake, J. (2018) *Machine Learning in Official Statistics*. https: //arxiv.org/abs/1812.10422.

Yung, W., Karkimaa, J., Scannapieco, M., Barcaroli, G., Zardetto, D., Ruiz Sanches, J.A., Braaksma, B., Buelens, B. and Burger, J. (2014) *The Use of Machine Learning in Official Statistics*. UNECE.

4. Actors

One of the deliverables of WP4 is the actor analysis. The actor analysis identifies the roles (actors) that are important from the development of smart solutions until the smart solutions have reached the desired maturity[3]. The added value of this analysis is that by identifying at an early stage which roles are important, they can be involved in the ideas that exist with regard to smart solutions, but also to provide input in order to achieve the most successful possible development and implementation of smart solutions. They can be new roles, which were not necessary in other data collection modes, but they can also be existing roles, which you already needed in the other modes. This document contains actors that have already been described in other documents. It also contains new additions.

The following sources were used to analyse actors:

[1] essnet_smartsurveys_wp3_del_3.1_25022021.pdf (europa.eu)
[2] WPF_Deliverable_F1_BREAL_Big_Data_REference_Architecture_and_Layers_v.03012020
[3] wp3_deliverable_3.3_-enhanced_framework_08_04_2022.pdf (europa.eu)
[4] Essnet smartsurveys wp4 taxonomy

[1] contains a description aimed at developing a robust smart solution. It explores design requirements for (Trusted) Smart Surveys (TSSu) as opposed to traditional paper or online surveys. One of the issues addressed in the report relates to the required new personnel profiles.

[2] has been used as inspiration for determining which roles are needed at Smart Solutions. It describes, among other things, which resources are required for Big Data solutions.

[3] mentions a few actors that play a role in smart solutions. These actors are also important in other data collection modes.

Actors specifically relevant to smart solutions

The actors identified to play a specific role in smart solutions, and not or less so in other data collection modes, are:

Smart Data Methodologist: the Smart Data Methodologist is the one who generates new methodological ideas for surveys where smart solutions can be used. The Smart Data Methodologist understands how sensors work and sees application possibilities for these sensors in surveys. This actor has an eye for the representativeness and data quality of the observations via smart surveys and assesses the value of the data collected for the purpose of fulfilling the need of information.

Smart Data Engineer: A Smart Data Engineer *"transforms smart data into a useful format for analysis by building data pipelines from the points of data collection to the analytical environment or even to the distribution and publication environment. Thus, providing the tools needed for a processing smart data."* (quote from [1])

Data Scientist: A Data Scientist specializes in the analytical part of the statistical process, and machine learning in particular. The Data Scientist provides valuable insights, obtained from a large

[3] See also the Maturity Model

amount of structured or unstructured data. The Data Scientist looks for patterns in data and provides (automated) reports and visualizations of findings.

Developers and **testers** of:

- *Apps*: we distinguish four types of smart surveys [4], of which many of us often think of using the app with smart functionality. Research apps consist of pieces of software, created by an app developer, that are used for data entry and sensor data collection.
- *Sensors*: internal and/or external sensors can be used for smart surveys. There are internal sensors (on a smartphone), there are also external sensors (wearables e.g.). Some of the sensors have already been developed at some point. In addition, new sensors can be devised. Sensor developers are needed to create new sensors or improve existing sensors (for example, by attachments for the camera on the telephone that can be used to measure air quality). This does not necessarily have to be an actor present within an NSI. The development of sensors can also be outsourced.

UX designer: The user experience designer focuses on creating a particular product (e.g. an app) in such a way that it gives the user the best possible experience.

Actors, also relevant for other data collection modes

There are actors who play a role in smart solutions, but also in other forms of data collection. Some examples are:

(Smart) data provider, such as a respondent and/or third parties.

Interviewers: for surveys with interviewers supporting data collection. For example, interviewers could conduct a preliminary survey in advance, or recruit and motivate respondents during their participation in a smart survey.

Assistants: Actors that can provide direct or indirect assistance to respondents:

- Helpdesk (substantive): for substantive questions about the smart solution.
- Technical helpdesk: for technical questions about the use of the smart solution.
- Website administrator (content): e.g. publishing FAQ.

Data collection methodologist: for researching the most suitable method for the application of smart solutions. For example, approach strategy and questionnaire design (if a regular questionnaire is offered in addition to smart solutions). The research could be a literature review, conducting a pilot, or something else.

Production employees data collection: for example, employees who design letters, or design, build and test surveys.

Production employees processing and output: domain specialists, or actors responsible for e.g. the weighing or editing process.

Trainers: for instructing interviewers, but also trainers for the use of new tooling.

Analysts who monitor and analyse the smart solution process

Functional managers of tooling

Business and information analysts, who translate the requirements for official statistics in such a way that technical solutions can be realized by the developer.

Methodologists, specialized in statistical modelling. E.g. for determining the method effect as a result of using smart solution as an extra data collection mode

IT employees, such as infrastructure managers, security officer, (Enterprise) IT architect and application managers

Legal personnel, such as data protection officer (DPO), ethics committee and lawyer (also relevant when collaborating with other parties for the conclusion of processor agreements).

Government: that enacts the laws

Customers, like the government, media and citizens.

Chief data officer for the overall data management of an NSI.

In the next stage of this SSI project, the actors will be linked to the levels of the maturity model (see Chapter 5) and GSBPM. There will then also be more clarity about which roles will play a prominent role in, among other things, building an app for the pilot, production or maintenance of the smart survey. In addition, some roles that are mentioned quite generally in this chapter, such as IT employees, will be further broken down in the next stage of this project.

5. Maturity model

Introduction

Below is a first concept of the maturity model. In the end, the model has two goals:

- SSI develops smart solutions. SSI has to prove that these will be mature. The model can help with this. The assumption is that – to be mature – the solution should be (for a larger part and at least) at level 3 ('Production'). Not all of the criteria mentioned in the model will be applicable! That is because the model covers the maturity of an organization[1], not only a smart solution. Which criteria are applicable will be determined in the next SSI project phase.
- SSI (WP4) will define a benchmark as deliverable. The maturity model will form part of the benchmark. An NSI can use the benchmark to assess at which maturity level they are and can detect where improvements are possible. Furthermore, it can also be used to determine where you would like to be (ambition) and which steps are needed to get there.

Please see Appendix A for more background information and used terminology.

Some notes upfront:

- The model as described below should absolutely be seen as a **first concept**. In the next two SSI project phases the model will be elaborated and improved. This concept version lacks e.g. the balance between the different aspects, and e.g. level 5 ('Optimized') is barely described. Also, criteria need to be more specific and measurable defined. Also, the - what we call - level 0 (baseline) is lacking, that will describe the prerequisites to apply this model (e.g. NSI should be able to use mobile cawi).
- In the next SSI project phase, also the relationship between **actors** and the maturity levels will be described. That is because depending on the level other type of actors play a role.
- An organization will (almost) never be at one specific level for all criteria. Some criteria of a specific level will be met, others not. Sometimes even differs per survey. For example, the maturity of HBS and TUS can differ within an NSI. The centre of gravity is decisive to determine the level. But good to mention: it is not a goal in itself to determine the maturity level. It is just a tool for an organisation to see what steps have to be taken to get 'more mature'.

Model

We distinguish the following 5 maturity levels in our model: Awareness, Pilot, Production, Managed and Optimised. Each level is described by a general description. Furthermore, for each level, maturity criteria have been described. These criteria are described from 5 different aspects: Organisation, Methodology, IT, Logistics, and Legal.

The 5 levels, with some related terms:

- Awareness

  Pioneer, thinking, beginning, learning, literature

- Pilot

  Exploratory, investigating, starting, start-up, business case

- Production

  Implemented, used in production process, applied, separate solutions

- Managed

  Standardized, in control, integrated, sharable

- Optimized
  Continuous improvement, excellence, best in class, fluent

Maturity levels with maturity criteria

Awareness

There is still an **'Idea**'. This is about gaining insight that maybe you can use smart surveys, but not yet concrete what the added value is. It may be that you have a problem, it may be that you see opportunities for smart surveys. Just starting out, little to no experience. Initial vision and objectives as main guide. Little or no guidance. Much on gut feeling.

An exploratory study is being carried out for the idea to see what is known about the preconditions that apply: what is and what is not allowed/possible.

Organization

1. Methodologists are in the lead. It leans on individuals. Some methodologists are very enthusiastic. Others sceptical. The rest of the organization is rather unaware.

Methodology

1. The organization is still debating whether smart surveys have a future.
2. The organization is still thinking in which situations smart surveys would add value.
3. The organization studies literature and available use cases for orientation, to form a view, to form an opinion.
4. The methodology to be used is still unclear; to be devised. First ideas emerge.
5. Some basic standalone proof-of-concepts are devised.

Business process

No specific criteria yet for this level. To be added in next project phase.

IT

1. A rudimentary smart solution(s), like an app, with very limited functionality is available/developed.
2. A personal account is used to upload and manage smart solution (e.g. app) on app store(s).

Legal

1. The preconditions from a legal and ethical point-of-view can be explored. For example, a legal officer could be consulted for this.

Pilot

There are ideas for conducting **proof of concepts**. Pioneer. Informal collaboration, possible due to limited organisational size. Processes are ad hoc, scattered and unconnected. The idea of using smart surveys is put to practice. Searching for a business case, searching for 'good' methodology, IT possibilities and legal restrictions. The goal is not to make statistics yet.

Organization

1. Smart solution, as a subject, is mentioned in the vision of the NSI. The vision says that the subject should be investigated.
2. The innovation department (R&D) has the lead, drives the initiative. The focus is on methodology and IT, paving the routes for future projects.

Methodology

1. Alternative strategies are tested in field pilot(s), e.g. recruitment strategy, motivation strategy, etc.
2. It is examined what information can be fed back to a respondent.
3. Comparisons are made between smart and non-smart strategies, e.g. regarding response rate and representativity.
4. Human computer interaction is being investigated. Pilots are used to test UX/UI.
5. In-app respondent interaction is being evaluated and optimized through usability tests.
6. Machine learning routines are being developed and tested. The achievable quality is examined.
7. A goal is to prove that the methodology can eventually be raised to a level so that maturity level 3 or 4 can be achieved

Business process

1. Fieldwork is monitored via a minimum set of indicators based on the process data. Process data is collected; however not per se in a standardized way.

IT

1. Smart solution(s) with limited functionality is available on main devices and operating systems (like Android and iOS).
2. Company account is used to upload and manage smart solutions (e.g. apps) on app stores (like Google Play and Apple's App Store). Upload process is ad hoc.
3. Some mainstream (personal) devices are available for testing purposes.
4. The testing infrastructure to test apps and smart services is rudimentary.

Legal

1. Guidelines/recommendations have been accepted by (inter)national authorities
2. There are no ethical objections about what smart solutions are used for, how they are used and with whom. In case of doubt, the ethics committee is asked for advice.
3. A DPIA (including risk analysis) is available for each smart solution pilot and assessed by a legal officer. The pilot complies with the applicable rules and legislation. Thereby the privacy-by-design choices for the smart solution pilot are motivated, documented and approved.

4. Risks (e.g. security) in pilot(s) with smart solution have been assessed, evaluated and sufficiently mitigated according to the relevant officers (e.g. chief security officer).
5. Informed consent procedures for the smart solution pilot (like the use of sensors) are tested, optimized and compliant with legal requirements.

Production

The organisation is capable of producing results, but the process, IT, methodology and legal aspects are not standardized nor generic. Processes are **stovepipes**, lot is custom-made. However, the whole process (data collection, processing, analysis) is covered so that statistics can be made. Documentation is described and up to date. Some basic PDCA-cycle is in place, where analysis is done based on the survey results, improvements can be determined, and improvements can be made for the next cycle. Improvements are primarily focused on the survey at hand. The PDCA is primarily per survey, not across surveys.

Organization

1. The decision has been made to apply a smart solution for at least one survey. So, there is a positive business case. The organisation had determined that smart surveys have added value.
2. The production department has the lead. The focus is on: “the job has to be done”. Focus is not yet on the full chain, but on the individual departments.
3. The organization needs to plan how to implement the innovation into production processes and systems.
4. Relevant personnel have been trained and has the necessary knowledge for conducting the concerning surveys. However, knowledge is not yet widespread in the organization.

Methodology

1. The methodology used is effective.
2. For the smart solution(s) at least the following has been taken care of:
    a. used methodology is recorded and documented,
    b. the strategy (for e.g. recruitment, motivation) is done with a proven methodology,
    c. there are specific, defined, measurable acceptable goals for response rate and representativity,
    d. response rates are calculated according to international standards (e.g. AAPOR),
    e. representativity is assessed via R-indicators and/or coefficient of variation,
    f. registration and completion rates for relevant population subgroups are monitored,
    g. proven methodology is used to process collected data to statistical output, and
    h. plausibility checks are specified, and the smart data satisfies these plausibility checks.
3. It is clear what information the used smart solution is providing.
4. Machine learning routines follow literature best practices and result in sufficient quality for the concerning smart solution. The performance of these algorithms satisfies specified thresholds in out-of-sample use (e.g. through accuracy/F1 scores).
5. UI/UX design of the smart solution is usable and works for the specific solution. It is not yet optimized and cannot be shared between solutions. Usability includes:

    a. to attract and retain participants is checked against a) engagement, b) accessibility, c) instructions/information, d) time efficiency, e) feedback and error handling,
    b. willingness to share detailed personal information checked against a) trust and credibility, b) security and privacy, c) transparency, d) data collection efficiency, e) user control, and
    c. ability to complete complex tasks, a) intuitive, b) task flow and guidance, c) error prevention and handling, d) training and onboarding, e) feedback and support.
6. In multi-mode settings:
    a. smart nonresponse and selection differences can be evaluated,
    b. smart measurement differences can be evaluated, and
    c. smart measurement differences can be adjusted for, to warrant comparability in time and between relevant subgroups.

Business process

1. The process is not standardized. The process is rather ad hoc. The process is probably shaped as a stovepipe for a given survey.
2. There is a process for the data collection, but also for the processing and analyses sub processes.
3. Fieldwork is monitored via a minimum set of indicators based on the process data. Process data is collected; however not per se in a standardized way.
4. Interviewer feedback, if applicable, is evaluated and summarized
5. The contact centre / helpdesk has the capability to answer questions of respondents regarding the use of the smart solution.
6. App store analytics are performed (downloads, ratings, etc.).
7. App usage traffic measured by responses (data is received by NSI).

IT

1. Smart services have been demonstrated to follow input and output specifications as provided by the methodology level (quality metadata) and legal level (PET).
2. The IT architecture and all levels of the IT solution (smart solution, smart feature/service, machine learning algorithms, backend) are described.
3. Stress tests and technical tests for the smart solution(s) have been performed.
4. Process to incorporate changes (including improvements of user experience and usability) and bug-fixes into the smart solution is ad hoc.
5. Process to evaluate new operating systems and devices is ad hoc.
6. Process to update smart solutions for new version of libraries and operating systems is ad hoc.
7. Process to upload and manage smart solutions (e.g. apps) on app stores (like Google Play and Apple's App Store) using a company account is managed centrally.
8. The mainstream devices are available for testing purposes.
9. A testing infrastructure to test smart solutions is available and supported.
10. Deployment strategy has been described explicitly to host microservices and to connect to platforms
11. Pentest and LaP tests have been carried out and critical levels are addressed
12. Interaction with external databases is agreed on a functional, technical and legal level

Legal

1. A DPIA (including risk analysis) is available for each smart solution and assessed by a legal officer. The process complies with the applicable rules and legislation. Thereby the privacy-by-design choices are motivated, documented and approved.
2. The privacy-enhancing-techniques (PET) applied (privacy-by-design) are described.
3. In-house monitoring and handling of smart data errors have been motivated, documented and seconded by legal officers
4. Risks (e.g. security) in smart solution(s) have been assessed, evaluated and sufficiently mitigated according to the relevant officers (e.g. chief security officer).
5. Respondent data control has been tested and evaluated.
6. Informed consent procedures (like the use of sensors) are tested, optimized and compliant with legal requirements.
7. Internal monitoring and handling of smart data errors are motivated, documented and approved by legal personnel.
8. Respondent data control requests are evaluated and approved.
9. It has been proved that the smart solution does not lead to potential security breaches. For example, by conducting a penetration test.
10. Data collection reports (respondent requests, communications) are checked against PET and discussed with legal officers, but there is not yet a standard process in place

Managed

Smart solutions are applied in different surveys. The different aspects (Methodology, Process, IT, legal) are well-defined and managed. Uniform processes, procedures and systems. There is a **baseline platform**. Organisation is able to produce achievable results, which are well managed. The organisation is in control. Roles and responsibilities are well defined. The process is standardised and documented across organisational units. Processes are monitored to see if they are carried out accordingly. The PDCA-cycle is in place, where analysis is done based on the survey results, improvements can be determined, and improvements can be made for the next cycle. Improvements are focused across surveys. But probably still per main process stage (data collection, processing, analysis).

Organization

1. Smart solutions are applied for different surveys.
2. The production department has the lead. The focus is on efficiency and standardizing.
3. Knowledge of smart surveys / solutions is quite widespread. Mainly in the data collection, methodology and IT departments. More and more personnel are involved.

Methodology

1. Methodology is proven and effective.
2. Effective methodology is available regardless of the type of smart solution.
3. Quantitative monitoring and analysis of the methodology is done.
4. Relevant summaries of in-device paradata/audit trails for smart services have been defined.

Business process

1. Monitoring of app store analytics (downloads, ratings, etc.) is performed regularly.

2. Monitoring of app usage traffic measured by responses (data is received by NSI) has been setup.
3. There is a process in place where machine learning predictions with low classification probabilities (so, input that cannot be classified automatically), can be recognized and handled manual by personnel in the back-office.
4. There is a standardized process in place to update machine learning routines, to preserve the required performance.
5. Helpdesk and other interactions with respondents are evaluated and summarized

IT

1. Process to incorporate changes (including improvements of user experience and usability) and bug-fixes into smart solutions is well-defined and managed.
2. Process to evaluate new operating systems and devices is well-defined and managed.
3. Process to update apps for new version of libraries and operating systems is well-defined and managed. A backlog is maintained on a continuous basis.
4. Monitoring of new versions of libraries used (libraries update frequently) has been setup
5. Monitoring of new version of operating systems has been setup.

Legal

1. Risks (e.g. security) are regularly re-evaluated in terms of prevalence, likelihood and impact and discussed with security officers. Action is taken where necessary.
2. Data collection reports (respondent requests, communications) are checked against PET and discussed with legal officers, through a standardized process.

Optimized

The organisation is considered a leader in applying smart solutions. The organisation is a centre of knowledge and excellence. Quality and efficiency are beyond question and are seen as a key differentiator. The organisation has an **advanced platform** at its disposal. All the aspects (methodology, process, IT and legal) are continuously improved. Continuous improvement is embedded in all of the processes. It is in the veins of the organisation. Employees are empowered to take action and innovate. New innovations can be applied easily. The PDCA is focussed on still becoming better and more efficient and on the statistical business process as a whole.

Organization

No specific criteria yet for this level. To be added in next project phase.

Methodology

1. Smart method effects are evaluated at least every five years.

Others to be added in next project phase.

Business process

No specific criteria yet for this level. To be added in next project phase.

IT

1. In-app usage monitoring (on approval) for improving user experience and usability.

Others to be added in next project phase.

Legal

No specific criteria yet for this level. To be added in next project phase.

6. Benchmark

In WP4 one of the deliverables will be a benchmark. NSI's should be able to use it in practice. The SSI smart survey benchmark will be a mean to improve, to be able to start incorporating smart solutions in the business processes or to improve existing solutions getting more mature smart solutions. The benchmark will consist of the maturity model and guidelines how to implement smart solutions in more mature ways.

In the 'Review phase' of the SSI project, WP4 had a task to develop ideas about what a benchmark is and how a 'smart survey' benchmark may look like. The task was not to develop the benchmark itself! This chapter is the result of this task.

A benchmark in general

A **benchmark** is a standard or point of reference against which things may be compared.

In the context of smart surveys there is also another meaning for benchmark. Providing individual feedback to a respondent, e.g. on household expenditures, can be done with or without a benchmark. With a benchmark is meant that e.g. the expenditures are shown for a comparable household. This is *not* the type of benchmark used in the context of this document.

Benchmarking – in general – is the ongoing activity of evaluating an organization's processes, products and/or services or comparing the organization's performance with other organizations with the aim of improving performance and gain competitive advantage.

The process also involves looking beyond one's internal organization and industry for best practices. The best may be a non-competitor in another industry, but the focus still remains on **improving business processes**, understanding and effectively meeting customer's requirements.

In a dynamic business environment, businesses need to be **adaptable to change**. Processes that might have boosted sales and profitability a few years ago might no longer be appropriate for building success today.

There are various ways of benchmarking, for example, internal[1], functional[2], external[3] or generic[4]. Whichever approach is chosen depends on the needs of the organization.

For the process to be a success there has to be the willingness to adapt and learn from others.
By benchmarking your business processes, you can gain insights into industry best practices, identify areas of improvement, and set performance targets to enhance your organization's efficiency, effectiveness, and competitiveness. It allows you to learn from the success of others and drive continuous improvement within your own operations.

Example Imagine you run an e-commerce business and you want to evaluate and improve your customer service performance. One common benchmark used in this scenario is the average response time to customer inquiries.

You can start by researching industry standards or analysing the performance of leading competitors in your market. Let's say you find that the average response time for customer inquiries in your industry is around 24 hours.
Using this benchmark, you can compare your own customer service performance. If your average response time is 48 hours, you can identify that your response time is twice as long as the industry benchmark. This indicates a potential area for improvement.

With this benchmark in mind, you can set a target to reduce your response time to match or exceed the industry average. You might aim to respond to customer inquiries within 12 hours.

NSI's

Benchmarking within NSI's regarding smart surveys has a slightly different context. The above is primarily written from a perspective of commercial businesses. Government entities like NSI's have a different perspective. It is not about competing with other NSI's or being the best to beat others and being top of the class.

But still NSI's have a goal to improve their business and to be able to adapt to change. The world changes, new technologies arise, society changes, respondents' willingness to participate, changes.

- ⇨ SSI will provide in a benchmark where gained knowledge is incorporated. The SSI smart survey benchmark will be a mean to improve, to be able to start incorporating smart solutions in your business processes or to improve existing solutions getting more mature smart solutions.

Aspects of benchmarking

Aspects of a benchmark typically include the following components:

Performance Metrics: A benchmark involves specific performance metrics or indicators that are used to measure and evaluate the performance of a system, process, or entity. These metrics should be relevant, measurable, and align with the goals and objectives of the benchmark. These will be our **maturity criteria.**

Comparison Points: A benchmark requires a basis of comparison, which can be derived from historical data, industry standards, competitor performance, or best practices. It provides a **reference point** against which the performance being benchmarked can be evaluated.

Data Collection: In using a benchmark, data collection is essential. It involves gathering relevant and accurate data (i.c. within the NSI) related to the performance metrics being measured. The data can be collected through surveys, interviews, direct observations, or automated systems, depending on the context and availability.

Analysis and Evaluation: Once the data is collected, it needs to be analysed and evaluated to determine the performance (maturity) level and identify any gaps or areas for improvement. Statistical analysis, data visualization, and comparison techniques are often used to interpret the benchmark data effectively.

Target Setting: Based on the benchmark analysis, target setting involves establishing specific goals or performance targets that are desirable or attainable. These targets can be derived from the benchmark itself or from industry standards, best practices, or organizational objectives.

Improvement Initiatives: A benchmark often serves as a catalyst for improvement initiatives. It provides insights into areas where performance falls short of the benchmark and helps identify improvement opportunities. Organizations can implement changes, such as process improvements, technology adoption, training programs, or policy revisions, to bridge the performance gap.

Monitoring and Feedback: Continuous monitoring of performance against the benchmark is crucial to track progress and ensure that improvement initiatives are effective. Feedback loops and regular assessments help identify any deviations from the benchmark and allow for timely corrective actions.

By considering these aspects, benchmarks provide a structured and systematic approach to evaluating performance, identifying areas for improvement, and driving continuous improvement efforts in various domains.

SSI WP4 benchmark

The above is slightly modified to use in the context of the SSI project. 'Our' benchmark consists of two things:

- The maturity model (maturity levels and maturity criteria)
- A description of the statistical business process (GSBPM) focusing on the application of smart solutions. This is necessary to identify where in the process improvements can be made.

Besides that, it will also contain guidelines to use the benchmark.

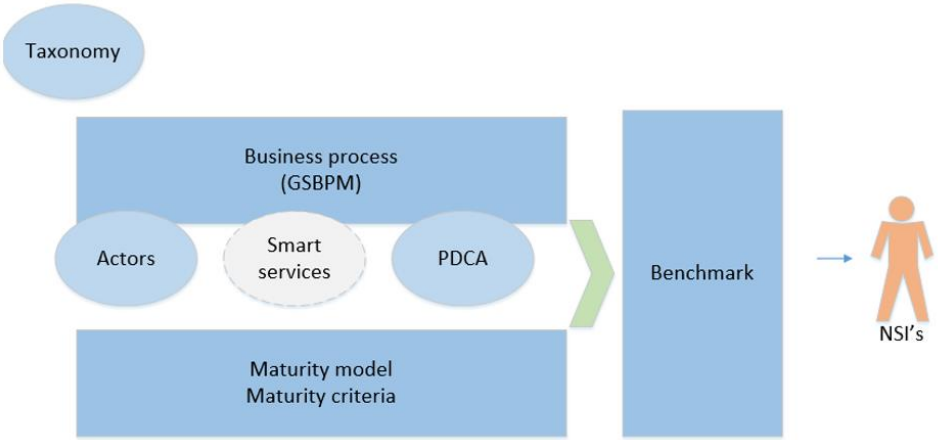


Illustration: coherence of the various topics

The benchmark will address different types of smart solutions. The various types will bring forth various aspects. So that is why we will make use of the *smart survey taxonomy (see chapter 2)*.

Because the benchmark is a deliverable of WP4 the focus is on the *business process*. It is about the process to conduct a smart survey. It will not address e.g., the software development process, the HR-process etc.

Process of benchmarking

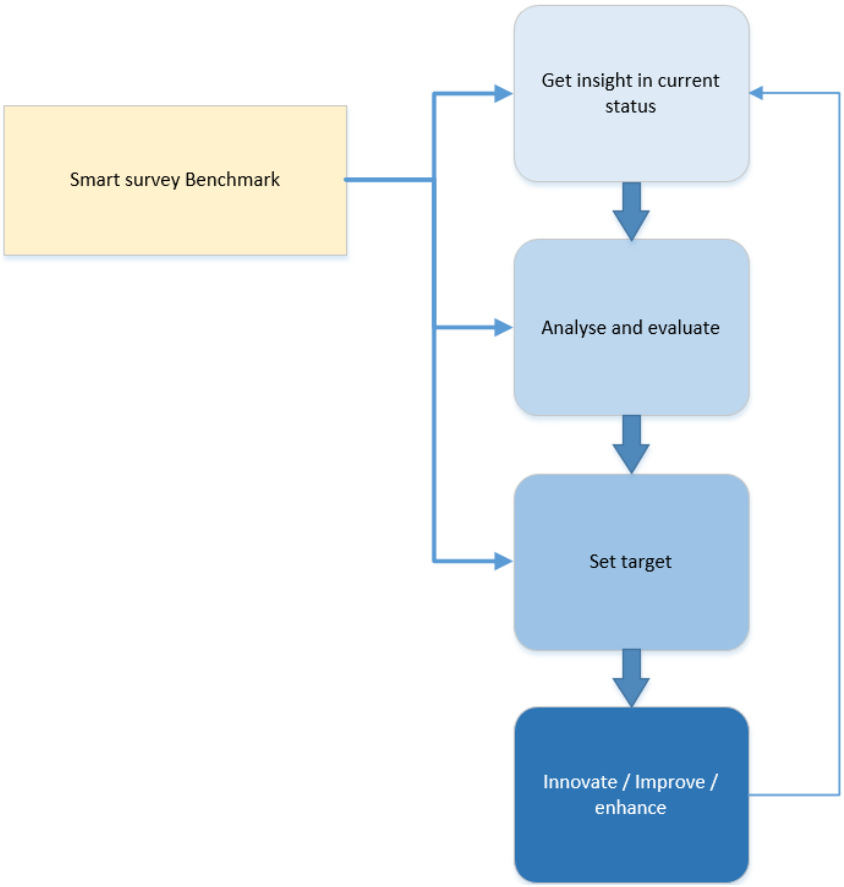
The **process of benchmarking** consists of 4 main steps.

First an NSI needs to get insight into the **current status** regarding smart solutions. The focus should be on the smart survey aspects. The benchmark will show you the aspects that you need to look at to address the current situation.

If the current status is clear then the NSI should compare itself with the maturity model. This is a process of **analysing and evaluating**.

After that **targets** can be set. The 'smart survey benchmark' will provide information. The benchmark gives insight in what next steps would be logical to conduct. So that a next level of maturity can be reached.

Using the 'smart survey' benchmark



When targets are set, **improvement** can be made. Improvement can be the improvement of an existing smart solution in the NSI to reach a 'next level', but it could also be an innovation activity where a new type of smart solution is introduced for a survey. But also, smaller enhancements in the current business process can be valuable.

Business process - GSBPM
The SSI smart survey benchmark will be linked to the GSBPM. This is necessary because when benchmarking you need to decide which process is in your field of attention. Obviously, in this case, that is the *statistical* business process.

This is also why we – in WP4 – first elaborated on the GSBPM, concerning smart surveys. We see that there is not an impact on all sub-processes when looking at smart surveys. So that also means that the focus of the smart survey benchmark is on the sub-processes that *are* affected in case of smart surveys.

That will give guidance in determining which part of the business process should be looked at to improve.

Iterative process
It should be clear that benchmarking is an **iterative** process. It is part of continuous improvement. Benchmarking is not a daily, weekly or monthly process. Think more of it as a yearly process. The process could give input to annual plans for improving your business processes (in a broader perspective). Smart survey benchmarking is supportive to the broader statistical business improvement process.

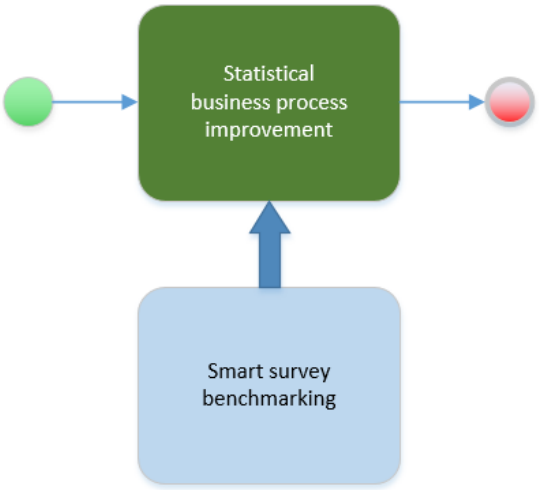


Illustration: benchmark can be used for improvement

WP4 activity
The goal for WP4 is to create a benchmark that is really useable by NSI's in practice. An NSI should be able to perform the benchmark process by using the benchmark. For WP4 to show/prove that the benchmark is really useable, in WP4 we will ask two NSI's to take the (draft of the) benchmark and to perform the benchmark process for one of their existing smart solutions. The results will then be used by WP4 to improve the benchmark to create the actual deliverable of WP4: a usable benchmark for smart surveys.

Benchmarking could take quite some time. For SSI there is only limited time available, and we don't want to be too much of a burden for the two NSI's. So, we need to look into this and decide together with the two NSI's how much time is available and how thoroughly the test-benchmarking can be done. From WP4 there is some time available to assist the two NSI's.

7. Looking ahead to the next project phase

In the next project stage (smart baseline stage) we will elaborate on different topics as mentioned above. The taxonomy, 'GSBPM' and maturity model all need to be supplemented and improved.

Different topics need to be linked more explicit to each other: e.g. the relationship between actors and the GSBPM and between actors and the maturity model.

We will start defining a first version of the benchmark and start describing a first version of the PDCA-cycle/process. Using the GSBPM as a framework we will describe the business process in more depth for relevant aspects and also indicate where microservices and machine learning fit into the business process.

This will all be written in more depth in the WP4 project stage plan.

Appendix A: Maturity framework

Goal of this topic

The goal of this topic is to give insight into the basis of our maturity model. This gives insight in the **maturity framework** that we are going to use in SSI. This framework is used to develop our **maturity model**.

First, we will see that SSI actually has two goals where maturity criteria are used.

Second, this appendix will define the different terms that are used concerning maturity. This is because a common glossary is necessary.

And third, this appendix will describe the maturity framework itself. The maturity model is described in chapter 5.

Goals of the maturity criteria

In SSI the development of maturity criteria serves two goals.

On the one hand, in WP4 one of the deliverables will be a **benchmark** (see chapter 6). The goal for WP4 is to create a benchmark that is useable by NSI's in practice. We want to help NSIs concretely with guidelines for introducing or further applying smart solutions. We think a benchmark can be helpful in this respect. With the benchmark, an NSI will be able to determine how mature they currently are regarding smart solutions, but more importantly, what steps they should take if they want to reach a higher level of maturity. A maturity model with maturity criteria will be an important part of the benchmark. These criteria will be used as performance indicators. These indicators are then used to measure and evaluate the performance (maturity) of an organization (NSI) concerning the application of smart solutions.

On the other hand, the SSI-project has the task to deliver an end-to-end solution for smart surveys. SSI will develop (the consisting parts of) two **end-to-end solutions**, as a proof-of-concept. These end-to-end solutions need to be mature. The two end-to-end solutions need to have a go on all four pillars: Methodology, IT, Logistics, Legal. The SSI-project will **demonstrate maturity** of these smart solutions (and will create strategies to maintain such solutions and to expand them with new features).

End-to-end solution

The SSI project assignment specifies that "the project must develop, implement and demonstrate the concept of Trusted Smart Surveys, realising a proof of concept for the complete, end-to-end, data collection process and demonstrating a solution combining:

1) involvement and engagement of citizens as active contributors

2) acquiring, processing and combining data collected from smart devices and other appliances

3) contributing to the trustworthiness by guarantying strong privacy safeguards."

So, for the first goal, the maturity criteria will be used exterior of the project. For the second goal, the criteria are used interior of the project, as a kind of quality indicator of the end-to-end solutions to be delivered. The set of criteria for both goals will have a large overlap. But it may also turn out that both sets – because of the different purpose – are different in some aspects: maybe other phrasing of the criteria, maybe a bigger or smaller set of criteria.

Terminology

Maturity framework and maturity model

When talking about maturity criteria different terms are used. The purpose of this appendix is to define these terms and give insight into the relation between the terms. Together, that provides a structure of how we can define a **maturity framework** that fits our purpose. This means that we should develop our own framework. It does not come out-of-the-box.

The framework consists of a focus area, focus aspects and maturity levels.

This is the maturity framework that we will use:

Focus area: Statistical business process regarding smart solutions					
Baseline / level 0: Prerequisites to use the model					
Maturity level ->	1 Awareness	2 Pilot	3 Production	4 Managed	5 Optimized
Focus aspects	*Maturity criteria*				
Methodology					
IT					
Business Process					
Legal					
Organization					

The SSI Maturity framework

A framework is just an 'empty' structure. The framework should lead to a model, a **maturity model**. So, the challenge is to develop 'our' maturity model. This model should define the maturity criteria to be used.

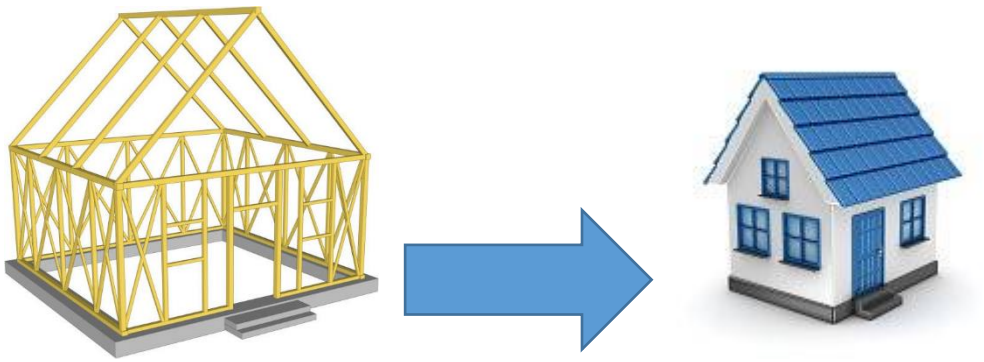


Illustration: Framework versus model

Focus area

First of all – when talking about maturity – the **focus area** needs to be determined. This could be e.g. building a house, procurement process, production process, human resource management process, etc. So, this is about defining the scope of your attention, for which you are going to define the framework.

In our case the focus area is defined as:

- The **statistical business process concerning smart surveys**,
- where the focus will be on the two **types** of smart solutions that cover the examples of HBS, TUS and energy, so ‘use of apps with device sensors’ and ‘data donation’.

Substantiation: We should look at the complete statistical business processes, so not just e.g. the data collection subprocess. As defined in the project plan, mature means that all GSBPM stages are considered and also a PDCA-cycle is described.

Focus aspects

Second, we need to determine the **focus aspects**. These are the aspects from which perspective we want to define the maturity criteria. The aspects should fit the focus area. The breakdown should be practical and usable: it should be possible to define maturity criteria from these perspectives. The aspects will give structure to the criteria.

As focus aspects we take the four pillars of the SSI project supplemented with the aspect “Organization”. So, the aspects are:

- Methodology
- IT
- Business process (Logistics)
- Legal
- Organization

Substantiation: The four pillars are the main pillars that should be considered when looking at smart surveys. The aspect ‘Organization’ is added because when talking about maturity it is not just about content and process. The organizational culture and the way an organization is structured and steered is also an aspect to be considered. E.g. mature methodology and a mature process, but without a mature organization, will lead to an imbalance and therefore will not work optimally.

So, from all these aspects, maturity criteria need to be defined.

Maturity criteria

Maturity is a measure of how ‘good’ a process or (end-to-end) solution is. 'Good’ means how well-defined, controlled, and optimized the business process or solution is to achieve its intended objectives.

The maturity criteria are the performance indicators that you would like to measure. The criteria should be relevant (in the context of the focus area), measurable, specific, objective.

When using the criteria, you are going to **measure,** thus giving a **value** to the indicator. Next this value needs **evaluation** and results in a **rating**; e.g. high/medium/low; green/orange/red, etc.

Criteria can have different rating categories. For criteria 1 the categories ‘yes/no’ are e.g. applicable, and for criteria 2 ‘very high/high/medium/low/very low’.

The total set of evaluated criteria gives insight in how mature the total process/solution is.

The criteria themselves are not part of the framework, but of our model.

Remark: in SSI the focus is on three smart solutions: HBS, TUS, energy. The question is whether all maturity criteria will be generic criteria. The aim should be to define generic criteria. However, maybe, when developing our model, we will see that at some point it will be necessary to include

some criteria specific for a type of smart solution. The source of these types is the **smart solutions taxonomy** (chapter 2).

Attributes of maturity criteria

What makes a maturity criteria a good, usable criteria? Here are some attributes:

- Measurability: Maturity criteria should be quantifiable or measurable to enable objective assessment and comparison.
- Objectivity: Maturity criteria should be objective, rather than subjective opinions or preferences. They should be applicable and relevant to the process being evaluated, and not influenced by personal biases or interpretations.
- Relevance: Maturity criteria should be directly related to the goals, focus area and focus aspects. They should reflect the critical factors that contribute to determining the maturity.
- Comprehensibility: Maturity criteria should be clear, specific, and unambiguous, leaving no room for misinterpretation. They should provide explicit guidelines.

Maturity levels

When maturity criteria are evaluated and rated the result does say something about how mature the process/solution is. However, it adds more value having a model with maturity levels. These levels provide in a 'growth model'. The rating of the criteria determines on what level you are. If you have insight into what level you are, you can determine what is necessary to get to a higher level. This makes it actionable. This is the goal of the benchmark (see chapter 6).

So, the maturity criteria can be structured into different levels or stages, allowing for a progressive assessment of maturity. This enables – general speaking – organizations to identify and prioritize areas for improvement and track their progress as they move from one level to another.

The levels should fit the purpose of the maturity framework. In WP4 we studied different existing models, such as Capability Maturity Model Integration (CMMI), Business Maturity Model (BMM), Integral Data Analysis Model (IDAM), Business intelligence (BI maturity model), Service Excellence Maturity Model (SEMM). These models focus mainly on business processes, data management, and IT development. We tried to map and translate these on the statistical business process and on the smart survey topic.

Almost all of these models distinguish 5 levels of maturity. We went with that, just because 4 levels seem to lack sufficient growth possibilities and more than 5 would be too complex for our purpose.

Because our focus area is the statistical business process, we were looking for levels that allow us to describe the innovative process of applying smart solutions. That process starts with generating ideas, then trying these ideas in pilots and deciding whether or not to apply the smart solution in production. When having a first process/solution in production, you could continue further developing your production process by managing and optimizing your methodology, process and solutions.

=> In our framework we define 5 maturity levels.

A good thing to mention upfront is that a precondition for using smart solutions – especially apps – is that an NSI is capable of using '**mobile cawi**'. If an NSI is not yet capable of providing cawi on a mobile device, it is a giant leap to use smart solutions. The focus of the maturity criteria and the maturity levels is the *smart* part. So, there are no criteria or levels for '*going cawi*'. However, the model should describe some prerequisites and requirements, that apply before smart surveys can be developed. This is what we will call the **baseline** or level 0. In fact, for many applications (e.g. HBS

and TUS) the smart survey is not a new survey but a ‘non-smart’ survey going smart(er). So, one needs to consider where we come from, i.e. the ‘non-smart’ version of the survey. The maturity level of a smart survey may be acceptable when the level of the non-smart scores weakly at one or more of the four design dimensions (meth, IT, logistic, legal).

These are the five maturity levels. These are elaborated in the chapter 5.

1. **Awareness**
2. **Pilot**
3. **Production**
4. **Managed**
5. **Optimized**

Each level has general properties and specific properties. A general property is descriptive, e.g. “*Processes are implemented 'ad hoc' and success depends on individual efforts*”.

A specific property defines what ‘rating’ of a criteria is necessary to be on a specific level. E.g. rating ‘good’ for criteria ‘X’ belongs to level ‘1’.

➔ The question is whether we need maturity levels for the goal ‘demonstrate maturity of end-to-end solution’. For the benchmark, we do need it. We dive deeper into that in the next project stage.