



Data Protection Impact Assessment (DPIA): smart perceptions survey 2023

STATISTICS NETHERLANDS (CBS)

DIRECTIE DATASERVICES, RESEARCH EN INNOVATIE (DRI)

The Hague, March 20, 2023

Maatregelen
nemen
Privacybewustwording
PIA
Doelbinding
Noodzaak
Effecten in kaart
bescherming van
persoonsgegevens
Risico's
minimaliseren
Richtinggevend
Rechtsgrond.
Met open vizier

Determination of controller responsible: 17 juli 2023

Naam: 5.1.2.e

Data Protection Officer Advice: 17 juli 2023

Naam: 5.1.2.e

Cio's advice: 17 juli 2023

Naam: Wim van Nunspeet

**CENTRAAL BUREAU VOOR DE STATISTIEK (CBS)
DIRECTIE DATASERVICES, RESEARCH EN INNOVATIE (DRI)**

Contact:

Centraal bureau voor de statistiek

5.1.2.e

Email: 5.1.2.e@cbs.nl

Telefoonnummer (Infoservice): (088) 570 70 70

Version: 2.0 17 juli 2023

Index

Index.....	4
Introduction.....	5
A. Description of data processing characteristics	6
1. proposal.....	6
2. Personal data.....	6
3. Data processing	6
4. Processing purposes	7
5. Parties involved.....	7
6. Interest in dataprocessing	8
7. Processing locations.....	8
8. Technique and method of data processing.....	8
9. Legal and policy framework	8
10. Retention periods	8
B. Assessment of lawfulness of data processing operations.....	9
11. Legal basis	9
12. Special personal data	9
13. Purpose limitation.....	9
14. Necessity and proportionality	9
15. Rights of the data subject.....	9
C. Description and assessment of risks for data subjects	10
16. Risks.....	10
D. Description of measures envisaged	11
17. Measures i	11
E. Appendices	13

Introduction

Centraal Bureau voor de Statistiek (CBS) processes data for the production of official statistics on the Dutch general population. For this purpose there is general CBS DPIA, baseline DPIA.

For divergent processes, separate DPIA's are composed. This is the case for innovative processes where new techniques and data collection methods are employed that are outside the scope of the company wide DPIA. This DPIA corresponds to a so-called smart sensor survey to be conducted in 2023 in IT, NL and SI. For this reason this DPIA is written in English.

This research is done within Eurostat-funded project Smart Survey Implementation (SSI) that runs from May 1st 2023 to April 30th 2025.

A. Description of data processing characteristics

1. proposal

The smart perceptions survey is a one-time-only survey conducted within Eurostat-funded project Smart Survey Implementation (SSI). In the following the survey is abbreviated to SSI-SPS. The Dutch name of the survey is 'Nieuwe manieren van meten'.

This DPIA concerns an additional process that is not covered by the general CBS DPIA. This process is employed in the context of the ESTAT-funded project SSI that aims to develop and evaluate effective push-to-smart respondent recruitment and motivation strategies. Push-to-smart refers to smart surveys that replace and/or supplement survey questions by sensor measurements and/or other forms of so-called smart data. These smart features are introduced to reduce respondent burden and to improve data quality. They are typically used for survey topics that are cognitively demanding, require detailed knowledge or recall, and/or concern concepts for which questions provide weak proxies. Examples are household expenditures, time use, media use, physical activity, indoor living conditions and working conditions.

In order to get to strategies that conform to respondent expectations but also hesitations, SSI conducts a series of pilots and field tests. The first in this sequence is the SSI-SPS that aims to collect general attitudes, motivations, hesitations and objections about smart surveys, in particular the use of sensors and other smart features. This survey is the prelude to more targeted surveys linked to existing surveys and the stepping stone to interviewer training, design of invitation letters, respondent assistance and respondent feedback. The SSI-SPS also forms the basis for privacy-by-design choices that are backed-up and understood by the general population.

The survey is conducted through regular CBS production systems and CBS Blaise servers from September 2023 to October 2023. Survey questionnaires are developed, tested and implemented by CBS Data Collection department. These processes are all covered by the general CBS PIA. See for details on the field test design: "Smart Survey Implementation: Annex B (5.1.2e et al 2022)". The new/additional process in SSI-SPS is the uploading of characteristics of photos of receipts and energy meters, the uploading of characteristics of digital receipts and location data to the Blaise server within the Blaise questionnaire. This DPIA describes the process, possible privacy issues and additional mitigation measures.

2. Personal data

The SSI-SPS collects ordinary and special personal data. The ordinary personal data are attributes in the form of a series of attitudes, a series of mobile/smart device usage variables, shopping receipts characteristics, energy meter data, step counts. Special data are location measurements (one measurement per respondent). See appendix: "Variables used in Field Test".

3. Data processing

The SSI-SPS survey consists of two parts: A paper questionnaire on general perceptions and digital literacy and an online questionnaire including four smart features. The paper questionnaire is introduced in order to also get opinions from those hesitant to do online surveys. The process involved constitutes of the regular data collection for a certain statistic, like household. But the questionnaires are changed as described and randomly distributed to a sample of only 4000 persons (Dutch version). When filling the questionnaires these respondents are asked to complete the additional smart blocks. The online survey follows the general survey and consists of four optional smart blocks. The order of the blocks is randomized in order to study order effects. All smart blocks are optional and respondents can decide to skip them. The smart blocks are: scan and/or upload of shopping receipts, manual data entry of step count from a personal activity tracker, one location measurement to determine travelled distance from home, and a photo of the energy meters in the respondent dwelling. The survey is conducted in Italy, Netherlands and Slovenia. ~~The Dutch version of the survey invites a random sample of 4000 persons.~~ See attachments: "Variables collected in SSI-SPS".

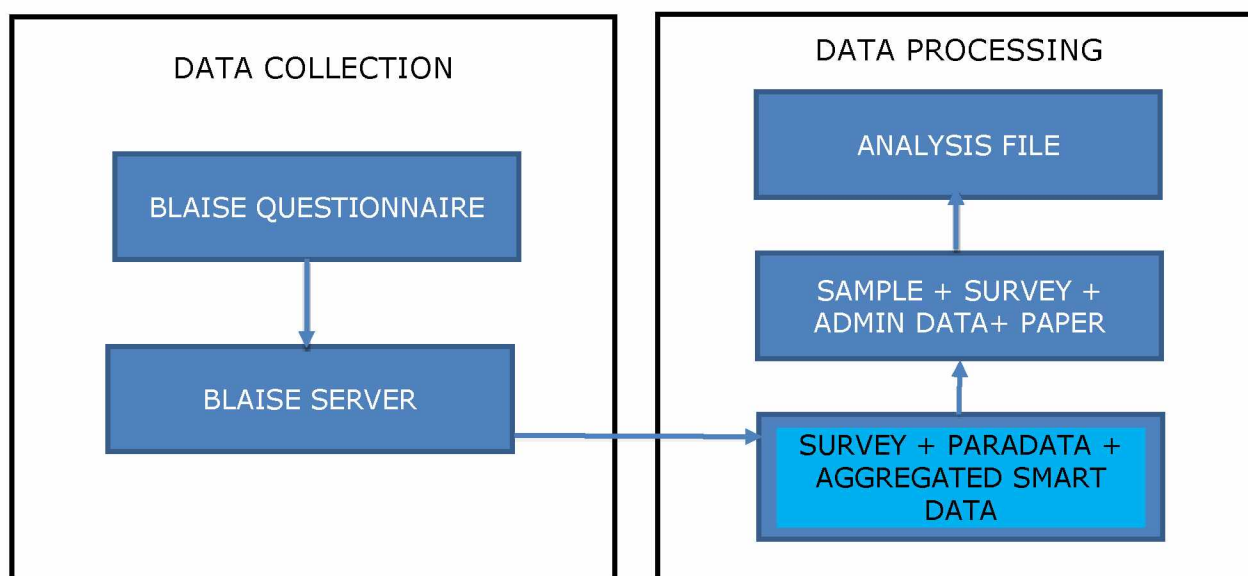
The samples with additional attributes are received through the regular process of data collection (assurance via CBS PIA) and are linked to a set of administrative variables. The resulting data are used by specialist to improve the data collection. The purpose is described in more detail in the following paragraph.

The SSI-SPS includes the following data processing (see also Figure 1, where the new parts are underlined):

- Respondents fill in the smart Blaise questionnaire and data is passed to the CBS Blaise server
- Online Blaise questionnaire data are combined with smart data

- Shopping receipts photos and files are prepared by respondents but NOT actually sent to CBS. Characteristics of the photos (type and size of file) are submitted.
- Energy meter photos are prepared by respondents but NOT actually sent to CBS. Characteristics of the photos (type and size of file) are submitted.
- The distance of the in-device location measurement to the respondent home address is estimated in-house.
- Paradata on the type of mobile device are derived and linked to the sample
- Sample data are linked to administrative data: age, gender, income quintiles, household type, urbanization degree of area of residence, socio-economic status/position
- Online questionnaire data are linked to the sample
- Paper questionnaires are digitized in-house and linked to the sample
- A single data set is created in which only the derived smart data (characteristics of photos/files of receipts and energy meters and distance to home address) are included. The original 'raw' smart data are not included and are deleted.

Figure 1: Overview of SSI-SPS processing. The light blue part is new for the SSI survey relative to regular data collection



4. Processing purposes

The data collected within SSI-SPS serves to understand motivations, hesitations and objections persons may have to smart surveys. The data collected consist of a set of questions and a set of smart data. A better understanding of such perceptions and opinions is used for the following purposes:

- Decide on privacy-by-design choices and data minimization principles for smart surveys, i.e. provide additional info for implementation of GDPR constraints
- Inform specific data protection impact assessments for specific survey themes
- Explore ethical boundaries of smart data collection
- Adapt/tailor content of invitation letters and instructions to the expectations respondents have
- Develop interviewer training acknowledging questions and issues candidate respondents may have
- Decide on the need for non-smart options as alternatives

In order to separate hypothetical willingness, known to be subject to social desirability, from actual willingness the respondents are asked to perform four smart tasks. The smart data corresponding to these tasks are processed as well in order to see whether the respondents did the tasks and how well they did the tasks.

5. Parties involved

This proces is conducted completely within the CBS, without any other external parties involved.

Parties involved:

- Responsible for processing: CBS methodology department (DRI – DRD)

CENTRAAL BUREAU VOOR DE STATISTIEK (CBS) - DIRECTIE DATASERVICES, RESEARCH EN INNOVATIE (DRI)

- Processor: CBS data collection department (DRI – DVZ)
- Provider:
 - CBS Blaise department advises and assists in sensor measurement through Blaise
 - CBS data collection department develops and implements questionnaires and collects data through its regular case management systems.
 - CBS methodology user lab reviews questionnaires and makes recommendations on content

Recipient:

- CBS methodology department together with CBS data collection team innovation analyze SSI-SPS data
- University Utrecht PhD 5.1.2.e and researchers 5.1.2.e and 5.1.2.e assist in the analysis. All three have appointments at CBS.
- NB1: Eurostat is the formal sponsor of the survey but data are not shared
- NB2: Stat Italy (ISTAT) and Stat Slovenia (SURs) conduct the SSI-SPS as well, but only use the CBS questionnaire and not the CBS backend/servers (i.e. they perform the survey themselves)

This DPIA does not relate to the production of statistics for policy, practice and science on the basis of Article 3 of the CBS Act. The data is only for internal use and is forwarded directly from the Blaise questionnaire to CBS without the intervention of another party. See also "Basisregels Informatiebeveiliging CBS".

6. Interest in dataprocessing

The SSI-SPS survey data and smart data are crucial in developing and implementing responsible, accountable and effective data collection recruitment strategies and informing legal officers and ethical committees.

7. Processing locations

The Netherlands (CBS in Den Haag and CBS in Heerlen).

8. Technique and method of data processing

The data for SSI-SPS are collected through a Blaise questionnaire with additional so-called sensor plugins developed by the CBS Blaise department. They have been tested technically and operationally prior to data collection. Submitted data are stored at a regular Blaise server. On the database queries are performed extracting aggregated summaries per respondent. All further processing is as usual and as described in the general CBS PIA.

There is no (semi-)automated decision-making or profiling and no micro survey or smart data are published.

9. Legal and policy framework

As stated in the standard CBS DPIA:

The Statistics Netherlands Act, the Statistical Law (European Regulation 223/2009), the European Statistics Code of Practice and the CBS Code of Conduct.

CBS may receive data under Section 33 of the SN Act. Section 33(4) of the Act states: "At the request of the Director General, the institutions, departments, bodies and independent government bodies referred to in Section 33(1), the legal entities referred to in Section 33(2) and the companies, independent professionals, institutions and legal entities referred to in Section 33(3) shall provide the data referred to in Section 33(1-3) free of charge within a period to be stipulated by order in council. In such cases no duty of confidentiality may be invoked, unless this duty is based on international regulations."

For data processing by CBS, which is responsible for producing Community statistics at national level, CBS' work programme is important:

The main lines of the work programme are laid down in a Multi-annual Programme (Section 14 of the SN Act). A work programme is adopted each year by the DG (Section 15 of the SN Act). The work programme also includes the Community statistics that CBS is required to compile under Section 4 of the SN Act. The data processing is therefore necessary for the performance of a task of general interest.

All processes other than the inclusion of new forms of (smart) data in the Blaise questionnaire are covered by the general CBS PIA. The processes related to the smart tasks and smart data in time will also be grouped in time under the same overarching task of CBS.

10. Retention periods

According to the CBS Selection List, the SSI-SPS falls under the Process: "Developing and drawing up methodologies for (performing) statistical research"

For this process, the agreement has been made that the data may be kept for 2.5 years after the final versions of the descriptions have been established or published.

B. Assessment of lawfulness of data processing operations

11. Legal basis

The statutory task and legal obligations form the legal grounds for processing personal data as stated in the SN Act (Sections 3, 4 and 5 of the SN Act). In words: The data collection and processing are 'necessary for the performance of a task carried out in the public interest'.

12. Special personal data

In SSI-SPS, one step count, as a proxy for physical activity, is submitted by respondents. These are manually entered by respondents as an answer to an SSI survey question and, thus, not imported through devices directly. A step count can be transformed directly to standard physical activity questions as asked in the CBS Health Survey (GEZO) and Lifestyle Survey (LSM). Since health data are special personal data, the original step counts are not kept but only a binary indicator whether the respondent conformed to the fit norm on the reference day.

13. Purpose limitation

The data collected in SSI-SPS match directly to the goals: Explore hesitations/motivations linked to willingness to share data through smart tasks, and measure hypothetical and actual willingness to do so. Four smart tasks have been included in order to evaluate the dependence on type of task and type of survey context. The four smart topics come from four real applications that are currently developed by CBS and other NSI's in SSI and together give a representative view on smart surveys.

14. Necessity and proportionality

The SSI-SPS is exceptional in that it is essentially a survey about how to do surveys. It thus does not by itself lead to population statistics or serve a particular official statistics' theme. The SSI-SPS has the aim to determine proportionality and subsidiarity of surveys yet to be fielded. Some of these smart surveys such as CBS Onderweg in Nederland (ODiN) and the Household Budget Survey (HBS or BO) have already been piloted by CBS. Here, SSI-SPS serves to finalize and perfect the respondent interaction.

In the SSI-SPS smart data are collected on four topics. Part of the smart data (household purchases, visited locations, energy usage) is currently collected through survey questions. These data are known to be of low data quality due to high respondent burden and/or high recall errors. Another part of the smart data (step counts) cannot be asked through questions and is currently derived through a series of questions. Also for these data it is known they are subject to high measurement error as respondents can only make educated guesses. Proportionality is ensured by allowing respondents to refuse smart data tasks and/or to provide data in traditional non-smart options. Furthermore, the outcome of SSI-SPS is in mutual interest of both CBS and respondents in order to reduce administrative burden through optimization of the data collection process. The data collected represents the minimal set of data to full fill the output needs.

Subsidiarity has been discussed with the CBS ethical committee (11-02-2023). The committee confirmed this SSI-SPS approach as the utility for fine tuning and perfecting smart surveys. There are no alternatives with less impact on privacy matters.

15. Rights of the data subject

Respondents always have the option to not do smart tasks or to choose alternative non-smart forms of providing data. For each smart option they are explicitly asked for checking and submitting the corresponding smart data.

Exception based on statistics law and disclosure of the data of this field test is not an issue.

²De selectielijst is organisatie-georiënteerd en primair bedoeld om de archivistische neerslag van het CBS als archiefvormer te kunnen waarderen. Onder archivistische neerslag wordt verstaan: alle documenten en andere gegevens (incl. data) die worden ontvangen, gegenereerd of opgesteld in het kader van de taakuitvoering, ongeacht hun vorm.

C. Description and assessment of risks for data subjects

16. Risks

Given that regular data collection systems are used with Blaise questionnaires, attention is limited to the smart data collected through the online questionnaire.

NB: It must be noted that while respondents take pictures of receipts and energy meters and select files of digital receipts, the corresponding files are not actually sent to the CBS backend/Blaise server.

Negative impact on the rights and freedoms of the respondent

- Smart data collection: Given that all smart data are collected after explicit checks by respondents, there is no inherent negative effect on rights and freedom. Respondents can always opt for not measuring and/or submitting.

A negative effect, i.e. third parties get access to the survey answers, would occur only if the temporary local in-device data storage and/or the transmission of the data would be breached.

- Smart data processing: Smart data are transformed before they are combined with other survey data and administrative data. Essentially, they are reduced to information that could also be asked directly but would require considerable effort and/or recall by respondents. Rights and freedoms would only be affected when third parties could enter the backend database and extract the underlying individual smart data.

What can cause negative consequences (origin risks)

- The sample person login details can be intercepted and used by a third party. However, this would not reveal any data of the intended respondent and render data useless. No individual data is sent to the respondent device.
- Third parties may breach the questionnaire and view the photos and selected files of e-receipts prepared by the respondent;

The chance of data being misused

- The chance of login credentials falling into wrong hands is extremely small as 1) samples are based on probability sampling with small inclusion probabilities, and 2) regular precautionary measures have been taken to deliver the credentials.
- The Blaise questionnaire is protected. Since files are stored only between taking photos/selecting e-receipts and acceptance by respondents, the chances of a breach are minimal.

The impact if smart data is misused

- The third party could learn the following: Households make purchases at certain shops, persons have/use a certain activity tracker, persons use certain types of energy, and the person was at one specific location on one time point which is labelled (home – work – shopping-etc). The impact of a breach is relatively minor as the SSI-SPS is a light version of smart surveys. It asks to only collect shopping receipts of one day, the photos of energy meters of one single time point and the location of a single time point. If these data would become available a third party would only get information on a very short time range and not on behaviour over time.

D. Description of measures envisaged

17. Measures



To determine the necessary measures, CBS uses the ISO 27001/2 standard and the BIR as a starting point. In this context, measures to protect privacy have already been included in the (proposed) processing processes. These measures are:

- If the letter of registration with login details falls into the wrong hands, logging in with that data does not lead to obtaining personal data about the respondent. No personal data were processed in the questionnaire.
- The collection of personal data can never be traced back to an identified person. Personal data are anonymised within the survey itself. The provisions of the general CBS PIA apply.
- Security of questionnaire access is arranged with a username and password.
- A penetration test was carried out on CBS backend and case management system. Annually CBS lets external auditors check the following: information security (ISO 27001), quality (ISO 9001) and privacy protection (NOREA's Privacy Control Framework). These audits and tests included Blaise servers. The certificates can be found on the website of CBS: [iso-en-privacycertificering](#)
- Encryption of data traffic between the respondent device and CBS.
- Incoming photos-scans-files are not actually sent to CBS
- No data are sent to the respondent device
- The data are only accessible to authenticated persons.

With these measures taken, the risks of data collection are covered and mitigated.

The measures regarding the processing of survey data from the SSI-SPS fall under the general CBS PIA, because the processing process of the field test does not differ from regular CBS processes for which a PIA already exists.

Maatregelen
nemen
Privacybewustwording
PIA
Doelbinding
Noodzaak
Effecten in kaart
Bescherming van
persoonsgegevens
Risico's
minimaliseren
Richtinggevend
Rechtsgrond.
Met open vizier

E. Appendices

Appendix 1: Variables in SSI-SPS smart questionnaire

Here, only the attributes in the smart online survey within SSI-SPS are described. The general questionnaire is administered by paper.

Smart data:

- Household expenditure
 - Willingness to submit photos or files of shopping receipts
 - Yes/no receipt submitted through scan
 - Yes/no receipt submitted as digital receipt
 - If willing: Characteristics of the photo of the receipt (type and size)
 - If willing: Characteristics of the e-receipt (type and size)
 - If not willing: Total amount spent on groceries in average week
 - Size of the household
- Energy usage
 - Type of dwelling
 - Number of residents
 - Availability of energy meters for gas, water and electricity
 - Yes/no willingness to take photos of energy meters
 - If willing: Per meter the characteristics of the photos (type and size)
- Physical activity
 - Rating of general health
 - Willingness to perform a step count through an activity tracker
 - If willing: Step count
 - If willing: Device details with which step count was made
 - If not willing: Number of minutes walking or running on average day
 - Yes or no doing sports other than hiking or running
- Travel
 - Purpose of stay at current location
 - Willingness to let location be measured
 - If willing: Location data
 - If not willing: Self-reported travelled distance
 - Means of transport modes that respondent uses in average week

Paradata:

- Type of device(s) (OS, brand, model year)
- Time stamps of the questionnaire modules
- Yes/no break-off
- Error logs of technical failures in smart tasks.