Smart Survey Implementation

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

Smart perceptions survey lessons learned

Document under revision

Version 1.0, 26-07-2024

Prepared by:

Ilaria Lunardelli, Janelle van den Heuvel, Barry Schouten (CBS, The Netherlands)

Alessandra Nuccitelli, Barbara D'Amen, Barbara Lorè, Monica Perez (ISTAT, Italy)

Mateja Zgonec (SURS, Slovenia)

Work package Leader:

Remco Paulussen (CBS, The Netherlands)
e-mail address : r.paulussen@cbs.nl
mobile phone   : +31 (0)6 2123 0392

Index

Summary

This note is addressed to privacy/legal officers and others with an interest in review and evaluation of smart features in official surveys. Based on the respondent answers and behaviour within the SSI smart perceptions survey, we give answers to questions that play a decisive role in the legal acceptance and implementation of smart surveys. A more detailed report is available covering a wider range of research questions (Annex A to Deliverable 1.2 of SSI).

1. Background

Between September 2023 and February 2024, the New Ways of Measuring (NWM) survey was conducted in Italy by ISTAT, the Netherlands by CBS and Slovenia by SURS. The NWM is part of the ESTAT-funded project Smart Survey Implementation (SSI). It consists of two parts: a general survey and a smart survey. The NWM-G(eneral) was conducted on paper and asked for the use and access to smart devices, the hypothetical willingness to perform smart tasks in a range of applications, for information and options respondents like to have, for options to control data being collected and for any reasons against participation. It contained questions after hypothetical willingness to seven smart features. The NWM-S(mart) survey is an online survey in which respondents are invited to perform four smart tasks (a subset of the seven asked in NWM-G): (i) share their location, (ii) scan or upload a shopping receipt, (iii) share their step count from an activity tracker, and (iv) take photos of their energy meters. Respondents could opt not to do a smart task and were then asked to answer a question.

The total sample size is 10000 with 4000 sample units in both the Netherlands and Italy and 2000 in Slovenia. Response rates on NWM-G vary from 25% in the Netherlands to 50% in Slovenia to 70% in Italy.

We answer the following questions relevant for legal-ethical review of smart features in surveys:
- What objections exist against smart features? And how uniform are these objections across different smart applications?
- Do respondents understand what they consent to?
- What (extra) assurances do respondents need or ask for?
- If and how do respondents like to control smart data?

We run by each of the four questions in the following sections. Follow-up questions can be mailed to
Barry Schouten (CBS) – jg.schouten@cbs.nl
Monica Perez (ISTAT) – perez@istat.it
Mateja Zgonec (SURS) – mateja.zgonec@gov.si

2. Smart feature hesitations

Let us first look at the hypothetical willingness rates. Table 1 shows the self-reported rates for the seven hypothetical tasks across the three countries. A few important conclusions can be drawn. The first is that there are sizeable groups that as a first response tend to refuse. In IT and SI, these groups even form majorities. The second is that groups that are hesitant ('maybe' or 'DK') are also fairly large. This implies that for them the logic of the request matters. The third is that there is a considerable variation across smart features, implying that context and type of feature may be decisive. The final and fourth is that countries differ quite strongly, with IT being more hesitant and NL being more willing, in general.

Table 1: Averages across all seven hypothetical smart tasks in (%) to the question: Would you participate in a(n) ISTAT/CBS/SURS survey which asks you to.

	Yes			Maybe			No			DK		
	IT	NL	SI	IT	NL	SI	IT	NL	SI	IT	NL	SI
Share location	9.2	24.9	20.9	15.5	24.9	22.5	65.5	37.8	50.9	9.7	4.1	5.6
Share pictures of house	7.1	11.8	7.4	9.3	17.9	13.3	76.5	58.9	76.0	7.1	3.1	3.4
Share energy data	24.2	40.9	17.5	17.5	24.8	22.5	49.0	28.0	54.1	9.2	3.8	5.8
Use air quality monitor	28.8	47.4	32.7	15.9	19.8	22.4	44.9	24.7	40.5	10.5	5.3	4.4
Give the step counts	21.0	39.0	29.7	14.3	22.8	21.8	56.1	32.4	45.0	8.6	3.1	3.5
Wear an activity tracker	12.1	20.2	19.2	14.0	20.0	18.5	64.7	48.3	57.7	9.2	3.0	4.5
Make/upload receipt pics	7.6	13.8	9.3	12.7	19.3	14.4	70.8	56.0	70.7	8.9	2.5	5.7

Can we find clues for these hesitations or refusals? From the analyses (Sections 5.3 and 6.2 in D1.2) three observations stand out:

- Persons reporting weak(er) digital skills less often go smart. It is, however, not true that persons with weaker skills never go smart.
- Persons reporting to be quite to very concerned about data being stolen or misused less often go smart. Again it is not a strict rule that they will never go smart; concerned persons are also present in smart respondents.
- Persons hesitant to go smart often report that they do not see the logic or point of providing the smart data. Given that the NWM survey was experimental in nature (respondents were required to express opinions on something that was not fully tangible and partially de-contextualized), this response is understandable. These respondents may still go smart once the option is offered next to non-smart data collection within a logical context.

We make the following recommendations:

- Assist persons that mention they have weaker digital skills and offer them alternative survey modes, in order to include them and facilitate their participation to the survey.
- Offer clear and easily accessible tools and material about what is collected and what happens to the data.
- Implement smart features only as options in settings where the added value is logical and easy to explain and understand.

We strongly believe that putting these recommendations into practice is much easier with a mix of written and personal communication. Written communication should address concerns as main points but leave further tailoring and elaboration to interactive/personal communication. Addressing all hesitations in detail simultaneously leads to long, discouraging texts. In case respondents are invited

to install an app, which is often true in smart surveys, we expect that applications should be more than one-time-only, one-way tools. This further strengthens the importance of tailored contact.

3. Respondent understanding

It is crucial that respondents understand what is asked of them, to begin with. Only then it is that a consent is really meaningful and we can believe that consequences are known and anticipated. We extract one table from Deliverable 1.2 that hints at the answer to the question whether respondents understand what is asked of them. Table 2 confronts actual willingness with hypothetical willingness. We now restrict to the four tasks that were implemented in the online questionnaire (NWM-S), as we believe they are sufficiently general.

Table 2 points at some tendencies. We observe shifts in all possible directions, e.g. a hypothetical 'yes' can turn into a real 'no' and vice versa. But clearly there is consistency; a 'yes' more often leads to a 'yes and a 'no' more often to a 'no'. This means that with the little information respondents got in the NWM-G, part of them was already able to determine a consistent standpoint. Next, we observe that the 'maybe' and 'DK' groups, that seem to want more information, make up their minds once offered the choice. They go in different directions. Those that do not do a task often give a meaningful reason why not (see Section 6.2 of D1.2). Furthermore, we observe that real willingness is lower than hypothetical willingness. Although disappointing from a response point of view, this is in the preferable direction. Once a respondent is actually confronted with a task, he/she still makes a decision in favour of 'protecting' him/herself. Finally, the variety in willingness rates across smart tasks points at a conscious thinking process. The different smart features are indeed very different in nature and this has been recognized by respondents.

Table 2: Hypothetical willingness in NWM-G against real willingness in NWM-S

NWM-G hypothetical	NWM-S observed willingness								
	Shares			Is not able to			Not share		
	IT	NL	SI	IT	NL	SI	IT	NL	SI
Share location									
Yes	63%	62%	49%	23%	30%	23%	14%	9%	28%
Maybe	39%	56%	43%	36%	19%	21%	26%	24%	36%
No	17%	28%	20%	63%	22%	12%	20%	51%	68%
Don't know	32%	47%	9%	46%	18%	27%	23%	35%	64%
Share step count									
Yes	47%	66%	84%	42%	33%	14%	11%	1%	2%
Maybe	42%	58%	85%	55%	40%	15%	3%	2%	0%
No	20%	24%	80%	68%	75%	4%	13%	1%	16%
Don't know	21%	29%	100%	67%	71%	0%	12%	0%	0%
Share receipt									
Yes	18%	48%	22%	63%	47%	66%	19%	5%	12%
Maybe	18%	32%	20%	66%	56%	67%	17%	12%	13%
No	7%	16%	13%	46%	48%	43%	47%	36%	44%
Don't know	9%	24%	24%	43%	59%	53%	47%	18%	24%
Share meter reading									
Yes	15%	63%	8%	16%	8%	42%	69%	29%	50%
Maybe	5%	42%	12%	15%	10%	35%	81%	48%	54%
No	5%	8%	4%	9%	12%	36%	87%	80%	60%
Don't know	2%	22%	0%	10%	17%	30%	88%	61%	70%

In Table 3, we give the answers respondents themselves gave to the open-ended question on suggestions to be well/better informed. The answers were clustered into four categories: 1) provide

online information through a webpage, tutorial, or other instructions, 2) provide offline information through a brochure or letter, 3) provide in-person information through interviewer or other staff, and 4) be explicit about what is collected and when (without reference to how). All options, except in-person contact, are suggested by part of the respondents. Hence, respondents do not necessarily think of interviewer, helpdesk or other staff as sources of information about the study. Differences between respondent that go smart and those that do not apply to the format (online vs offline) but also the data itself. Respondents going smart more often point at online information and the content/timing of data. Respondents not going smart more often mention offline information.

Table 3: Summary of Question 15: How could ISTAT/CBS/SURS help to explain what data will be collected through smart devices? split also for NWM-G respondents that did the NWM-S survey and those that did not.

	No NWM-S			NWM-S			All		
	IT	*NL*	*SI*	*IT*	*NL*	*SI*	*IT*	*NL*	*SI*
Online study information	28%	20%	19%	35%	29%	13%	32%	26%	16%
Offline study information	27%	22%	31%	18%	13%	15%	22%	16%	24%
In-person study information	7%	2%	4%	3%	1%	2%	5%	1%	3%
Content and timing of data	25%	25%	30%	32%	39%	50%	29%	35%	38%
Miscellaneous	13%	31%	17%	12%	18%	20%	12%	22%	18%

In-person communication seems to be a rather marginal channel (% is very low). However, it may favor those who cannot find alternative solutions on their own. The communication strategy of an NSI should be able to support all options, including in-person contact that helps respondents to understand where and how to find and use the information. Since in practice, this would be hard to predict, it seems best to have all options but let personal contact be accessible in where and how to find and use it.

4. Assurances

In the NWM-G, respondents were asked about facilities, information and/or tools that could help them to feel more secure in providing the data they share with CBS/ISTAT/SURS. Five options were suggested through closed-ended questions. These were followed by an open-ended question. We present the answers in Tables 4 and 5, respectively.

The suggestions listed in Table 4 are: 1) use dedicated apps, 2) go offline (i.e. paper), 3) let interviewers assist, 4) show/provide a security audit report by an external (trusted) party, and 5) provide (smart) devices by the statistical institute. Again a substantial part of the respondents gives a 'DK' answer, implying that they need more context. These proportions are largest for dedicated apps, audit reports and devices. Respondents that do go smart more often express that they prefer these three options. We speculate that these respondents may be more informed about the potential risks of cybercrime and/or the content of data being collected. The respondents not going smart more often value offline modes and interviewer assistance. The latter is not true for NL and here we most likely observe the impact of interviewer recruitment in IT and SI.

In Table 5 we give a summary of the answers to the open-ended question. The item-response rates in IT, NL and SI were, respectively, around 20%, 32% and 14%. Hence, in all countries a vast majority did not give suggestions. We cluster answers into four categories: 'Information by NSI' corresponds to what the statistical institute itself can do, such as providing information. 'Respondent sharing choice' corresponds to what the respondent can do themselves. 'Security measures' is about additional explicit security measures. 'Miscellaneous' is all else. Open-ended responses also highlight the importance of the NSI providing adequate information to make respondents feel comfortable about data security. Again the respondents that did go smart point at explicit information. Whilst for the Netherlands and Slovenia the possibility for the respondent to manage choices autonomously is also important, in Italy a substantial part of respondents claims that there is nothing else ISTAT has to do to make them feel comfortable about the security of the data they share with ISTAT.

Table 4: Summary of Question 12: What should CBS/ISTAT/SURS do in your opinion to ensure that the security of data you share with CBS/ISTAT/SURS feels secure? split also for NWM-G respondents that did the NWM-S survey and those that did not.

Dedicated apps	No NWM-S			NWM-S			All		
	IT	NL	SI	IT	NL	SI	IT	NL	SI
Yes	24.9%	23.9%	29.5%	42.1%	49.2%	44.9%	31.2%	39.0%	34.4%
No	24.6%	24.2%	22.9%	24.5%	14.7%	13.6%	24.5%	18.5%	19.9%
DK	50.6%	52.0%	47.6%	33.5%	36.2%	41.5%	44.3%	42.5%	45.7%

Offline modes	No NWM-S			NWM-S			All		
	IT	NL	SI	IT	NL	SI	IT	NL	SI
Yes	41.9%	62.5%	47.5%	43.4%	64.2%	48.1%	42.4%	63.5%	47.7%
No	20.9%	20.6%	25.3%	33.1%	21.1%	29.7%	25.3%	20.9%	26.7%
DK	36.9%	16.9%	27.2%	23.6%	14.8%	22.2%	32.1%	15.6%	25.6%

Interviewer assistance	No NWM-S			NWM-S			All		
	IT	NL	SI	IT	NL	SI	IT	NL	SI
Yes	32.3%	9.2%	41.2%	23.6%	13.8%	19.6%	29.1%	11.9%	34.4%
No	32.7%	71.8%	32.0%	55.4%	67.2%	56.3%	41.0%	69.1%	39.7%

| DK | 33.7% | 19.0% | 26.8% | 21.1% | 19.0% | 24.1% | 29.1% | 19.0% | 25.9% |

External audit report	No NWM-S			NWM-S			All		
	IT	NL	SI	IT	NL	SI	IT	NL	SI
Yes	29.6%	26.3%	37.6%	42.6%	39.0%	43.4%	34.4%	33.8%	39.4%
No	23.8%	34.6%	22.7%	26.8%	33.0%	16.1%	24.9%	33.7%	20.6%
DK	45.4%	39.1%	39.7%	30.1%	28.0%	40.5%	39.8%	32.5%	40.0%

Devices provided by NSI	No NWM-S			NWM-S			All		
	IT	NL	SI	IT	NL	SI	IT	NL	SI
Yes	21.5%	17.7%	32.2%	38.6%	36.8%	42.7%	27.7%	29.0%	35.5%
No	29.7%	38.5%	31.6%	27.6%	24.0%	22.8%	28.9%	29.0%	28.8%
DK	48.8%	43.8%	36.2%	33.8%	39.3%	34.5%	43.3%	41.1%	35.7%

Table 5: Summary of Question 13: Is there anything else CBS/ISTAT/SURS can do to make you feel comfortable about the security of data you share with CBS/ISTAT/SURS? split also for NWM-G respondents that did the NWM-S survey and those that did not.

	No NWM-S			NWM-S			All		
	IT	*NL*	*SI*	*IT*	*NL*	*SI*	*IT*	*NL*	*SI*
Information by NSI	38%	23%	34%	46%	48%	63%	42%	41%	46%
Respondent sharing choice	5%	38%	18%	5%	23%	4%	5%	29%	12%
Security measures	1%	1%	1%	1%	3%	4%	1%	2%	2%
Miscellaneous	7%	32%	48%	3%	26%	30%	5%	29%	40%
Nothing else	49			45			47		

Summarizing, NWM-G respondents agreed with suggestions given within the questionnaire. There is a clear distinction between respondents going smart and respondents not going smart. The options offered are underlined by those that go smart but do not necessarily convince that do not. Offline alternatives are stressed by those not going smart.

5. Smart data control

In NWM-G, respondents could react to and suggest facilities and tools that would give them more control over data and make them more autonomous in how and what they share. Three suggestions were given: 1) a personal landing page (include login credentials), 2) dedicated apps displaying data that has been collected and that offer options to (un)submit, and 3) a retention time between data collection and data submission. In addition, respondents could make suggestions themselves through an open-ended question. We show the results in Tables 6 and 7.

Table 6: Summary of Question 17: How could ISTAT/CBS/SURS assist you in controlling what data will be collected? split also for NWM-G respondents that did the NWM-S survey and those that did not.
a.   *Personal webpage/landing page*

	No NWM-S			NWM-S			All		
	IT	NL	SI	IT	NL	SI	IT	NL	SI
Yes	45.0%	52.8%	46.3%	76.8%	84.7%	74.4%	56.5%	71.1%	55.1%
No	21.0%	30.9%	22.0%	9.2%	7.5%	7.3%	16.7%	17.4%	17.3%
DK	34.0%	16.4%	31.7%	13.9%	7.8%	18.4%	26.7%	11.5%	27.5%

b.   *App that shows data being collected*

	No NWM-S			NWM-S			All		
	IT	NL	SI	IT	NL	SI	IT	NL	SI
Yes	42.0%	46.1%	46.6%	72.3%	78.0%	68.7%	53.0%	64.7%	53.5%
No	23.4%	32.2%	22.1%	13.3%	12.0%	7.6%	19.7%	20.4%	17.5%
DK	33.7%	21.7%	31.3%	15.3%	10.0%	23.7%	27.0%	14.9%	28.9%

c.   *Retention time before respondent data are really included*

	No NWM-S			NWM-S			All		
	IT	NL	SI	IT	NL	SI	IT	NL	SI
Yes	41.3%	44.3%	51.4%	55.0%	52.1%	63.0%	46.3%	48.8%	55.0%
No	20.2%	35.2%	21.5%	23.4%	28.1%	14.9%	21.4%	31.2%	19.4%
DK	37.6%	20.5%	27.1%	21.5%	19.7%	22.2%	31.8%	20.0%	25.5%

Table 7: Summary of Question 18: How could CBS help to explain what data will be collected? split also for NWM-G respondents that did the NWM-S survey and those that did not.

	No NWM-S			NWM-S			All		
	IT	NL	SI	IT	NL	SI	IT	NL	SI
Autonomy in sharing	1%	17%	7%	5%	41%	29%	3%	32%	16%
Authorize respondent	38%	6%	2%	34%	12%	2%	36%	10%	2%
Information on legal basis	35%	32%	9%	38%	38%	12%	36%	37%	10%
Nothing can be done by NSI	17%	45%	14%	19%	9%	7%	18%	22%	11%
Miscellaneous	8%		68%	5%		49%	7%		60%

Table 6 shows that for data control options respondents were more confident and provided fewer 'DK' answers. Apart from the retention time option, we see an overwhelming support among respondents that go smart for personalized landing pages and dedicated app. For respondents that did not go smart still a large proportion agrees with these same options. We conclude that explicit and transparent control options may be beneficial.

Table 7 closes our evaluation. To this open-ended question item-response rates were relatively low: 16% in IT, 16% in NL, and 10% in SI. So a large majority did not give any suggestions. The answers by those that did were clustered in four categories: 1) offer autonomy in what is shared and what not, 2) provide a tool that authorizes respondents what is collected and what not, 3) provide information on the legal basis, and 4) do not try as will not share. The last option means that respondents firmly state they not want to share the smart data. Especially, in NL this groups is relatively large among respondents not going smart. Proportions across the countries are very diverse, making it hard to draw strong conclusions. Given also the high item-nonresponse rate, we abstain from conclusions.

Summarizing, we see support for facilities and tools that help respondents see what data are (going to be) collected through smart devices and that allow them for some control.