Discussion Paper

From willingness to usable energy data: a feasibility study of energy insight data donation in a probability-based panel

Maaike Kompier (CBS)
Jeldrik Bakker (CBS)
Joris Mulder (Centerdata, LISS panel)
Natalia Kieruj (Centerdata, LISS panel)

January 2026

Abstract

Official statistics increasingly require timely and granular information on household energy use. While administrative smart meter infrastructures can provide high-frequency readings, a complementary approach is *data donation*: individuals actively share copies of their own digital data with researchers under informed consent. This paper reports a feasibility study in the Dutch probability-based LISS panel (N = 2,600; response rate 72.7%) that quantified (i) access to energy consumption *insights* (apps, websites, in-home displays, emails, and self-tracking), (ii) willingness to donate those insights under randomized incentives (€5/€10/€15), and (iii) completion and minimum usability of donated material.

Most respondents reported having access to some form of energy insight; only 19.7% reported having none. Among respondents with at least one insight method, 35.5% indicated willingness to donate and willingness increased with higher incentives. However, substantial attrition occurred between willingness, successful donation, and provision of data that met minimum requirements (at least seven days with at least daily frequency), yielding 270 minimally usable donations. Among these, 87.8% provided consumption in energy units (kWh for electricity; m^3 for gas), while the remainder provided values in euros, and two-thirds required extracting values from graphs. Finally, among smart-meter owners, 27.1% stated willingness to install a dongle (hypothetical scenario), with no evidence of an effect of the requested period or incentive.

The findings illustrate a central challenge in data donation for official statistics: participation is only the first hurdle; donation success and measurement usability require targeted design choices, respondent support, and quality control.

Keywords: data donation; smart surveys; energy insights; smart meters; incentives; LISS panel

Contents

1 Introduction

Official statistics are generally based on population registers and/or surveys, each with their own strengths and limitations. Registers are typically built for administrative rather than statistical purposes, which limits both their timeliness and the concepts they can capture. Surveys can be designed around statistical concepts directly, but they depend on respondents being able and willing to report accurately, which is not the case for all concepts. Energy consumption is an example of a concept that falls into this gap: it is not covered at the desired detail level by registers available to Statistics Netherlands, and reporting on energy consumption in a survey is incredibly burdensome and difficult as most respondents have little detailed knowledge of their own consumption patterns. As a result, neither registers nor surveys are well suited to capture energy consumption at the level of detail that policy and research increasingly demand.

At the same time, digitalisation has changed what is available outside these traditional sources. Smart meters, energy suppliers, and in-home displays now generate detailed behavioural traces of household energy consumption that respondents themselves can access, even if they cannot easily report on it from memory. This development has made an alternative route viable: *data donation*. Data donation is a consent-based approach in which individuals retrieve or access copies of their own data (e.g. via an app, web portal, or personal data export) and share them with researchers for a specified purpose (Boeschoten et al. 2022; Carrière et al. 2025). For official statistics, this is attractive precisely because it sidesteps the register/survey trade-off described above: it does not depend on administrative registers being built for statistical use, nor on respondents' ability to recall or estimate their own consumption. Additionally, data donation can (i) preserve participant autonomy, (ii) be aligned with privacy-by-design principles, and (iii) reduce reliance on proprietary, direct data sharing arrangements with commercial entities.

Data donation has been studied as a data collection method in survey methodology, for example in large-scale experiments on smartphone sensor sharing (Struminskaya et al. 2020; Struminskaya et al. 2021) and activity tracker data (Kraakman et al. 2023; Toepoel et al. 2021; Keusch et al. 2025; Kompier et al. 2024). In this paper, we examine this potential of data donation specifically for household energy consumption. Energy statistics on consumption are central to monitoring the energy transition, evaluating policies, and supporting national and European reporting obligations. Stakeholders request indicators that are more timely, more granular, and better aligned with end-use concepts (e.g. heating vs. other uses, electric vehicle charging, or the use of locally generated electricity). Currently, Statistics Netherlands uses monthly energy data from network operators for their energy statistics. Yet finer-grained measurements of household energy consumption are available, as a smart meter is installed in more than 90% of Dutch households. Many households can access this information through their energy supplier, in-home displays, solar monitoring apps, or web portals. Such information could support more granular energy statistics that better meet stakeholders' requests. In this feasibility study, we therefore investigate the willingness and ability of consumers to donate the insights they have, the selectivity of who does so, and the variety of the information these insights contain, in order to assess their suitability for use in official statistics.

Despite growing interest in data donation, the existing literature leaves two gaps that are decisive for official statistics. First, most studies stop at consent or stated willingness and do not follow the donated material through to whether it meets minimum analytical requirements; the few that assess donated data rarely do so in a probability-based sample that permits population-level statements about each stage (Boeschoten et al. 2022; Carrière et al. 2025). Second, selectivity and bias are themselves quality criteria for a candidate statistical source: if willingness to donate

is correlated with the target variable, the resulting estimates are biased in ways that demographic weighting cannot repair. This study addresses both gaps directly. We treat data donation as a multi-stage process (i.e., access, willingness, successful donation, and minimally usable data) and quantify attrition and selectivity at every stage in a probability-based panel. Although we develop this argument for the concrete and policy-relevant case of household energy insights, the staged, quality-oriented perspective is general: it applies to any digital-trace source that an official statistical institute might consider, and we return to this broader relevance in the discussion.

1.1 Theoretical Background

Data donation can reduce reliance on self-report, potentially lowering measurement error for behaviours that are hard to recall. It may also offer higher measurement frequency and new variables not typically collected in surveys. However, data donation designs introduce distinct error sources, including coverage and participation biases, donation failure, and heterogeneity in data formats (Boeschoten et al. 2022; Carrière et al. 2025). Boeschoten and colleagues propose a "total error" view for data donation, emphasising that error can be introduced at multiple stages: from platform availability and consent, to successful extraction and transfer, to the validity and comparability of the donated traces (Boeschoten et al. 2022). This multi-stage process can be interpreted as a Total Survey Error framework, where each stage introduces distinct sources of error. In addition, tooling and user experience matter: platforms that guide participants through exporting and donating their data can reduce friction and improve data quality (Araujo et al. 2022; Welbers et al. 2024).

For official statistics, these quality considerations are particularly salient. Challenges for data donation in official statistics are the response rate and bias of the respondents that are willing to donate data. Selectivity on demographic characteristics may be corrected for using weighting procedures, but when nonresponse is related to the target variable, response bias such as under- or overestimation may occur. Statistical production requires stability over time, documentation of processing steps, and quantification of selectivity and bias. Consequently, feasibility assessments need to go beyond willingness and focus on the entire chain from access to usable data. While prior studies often focus on consent or willingness, fewer studies assess whether donated data meet minimum analytical requirements.

These quality considerations take a concrete form in the energy domain, where the widespread adoption of smart meters in the Netherlands (approximately 90%) allows for more detail in energy statistics. Smart meters measure energy consumption with great detail (15 minute intervals for electricity and hourly intervals for gas), enabling distinction in prices throughout the day and additional services such as insight in one's own energy consumption for consumers. In the Netherlands, smart meter readings have already been explored as a statistical source for gas and electricity consumption (Statistics Netherlands (CBS) 2025), yet via the network operators only monthly readings were available for this exploration.

In a related mixed-methods feasibility study by members of the present author team at Statistics Netherlands, Bakker et al. (2026) tested the feasibility of collecting passively measured household energy data via a consumer-grade P1 dongle plugged directly into the smart meter and combining it with short diaries. That pilot highlighted that installation can be feasible and acceptable, but also that manual export workflows and limited measurement resolution constrain scalability and analytical potential. This dongle-based route is the second, more structured collection mode that we benchmark below (RQ4), alongside the manual donation of consumer-facing insights.

The present study complements the pilot described above by moving upstream in the donation chain: it assesses how many respondents have a smart meter, which would open up the possibility of donating energy insights. Furthermore, we investigate who is aware of these energy insights and what fraction can donate them through a self-service upload task. Additionally, respondent-facing insights differ from back-end metering data: they can be aggregated, delayed, presented in different units, or restricted by the device used. For example, some portals provide daily or monthly summaries, some show only costs, and some display graphs without numeric tables. Consumer-facing energy communications are not designed for statistical use, which can be a challenge for standardisation and comparability. Understanding the heterogeneity is necessary before energy insight donation can be used as a statistical source, and it motivates the empirical focus of the present study. Additionally, this study benchmarks stated willingness to install a dongle as a potential alternative that may yield more structured data.

1.2 Research questions

We organise the study around a staged view of data donation and address the following research questions:

RQ1 **Access:** What share of panel members have a smart meter (the technical precondition for high-frequency readings), what share have access to energy consumption insights, and through which methods (apps, websites, displays, emails, self-tracking)?

RQ2 **Willingness:** What share of respondents with insights is willing to donate them, to what extent can incentives increase willingness, and how does willingness depend on background characteristics?

RQ3 **Completion and usability:** Among those willing, what fraction successfully donates, how do these insights vary, and what share of the donated data meets minimum requirements (period and frequency)?

RQ4 **Alternative mode:** Among smart-meter owners, what share is willing to install a P1 dongle as an alternative to manual donation, and how does this depend on background characteristics?

2 Data and methods

2.1 Study setting and sample

The study was fielded in the Longitudinal Internet Studies for the Social sciences (LISS) panel, a probability-based household panel in the Netherlands administered by Centerdata (Scherpenzeel 2011; Centerdata 2024). The panel is based on a true probability sample drawn from the Dutch population register by Statistics Netherlands (CBS). Self-registration is not possible. Households are sampled and subsequently all household members aged 16 years and over are invited to participate as panel members. Panel members participate in monthly web surveys and receive monetary incentives proportional to the survey length (€15 per hour).

For this study, a total of 3,574 panel members were invited in April 2024. Of these, 2,653 logged in and 2,600 completed the questionnaire (response rate 72.7%).

Table 2.1 summarises key background characteristics of the respondents. Whereas gender was balanced in the sample, elderly and people with higher education were overrepresented.

Table 2.1 Sample characteristics (N = 2,600).

Gender	N	Age	N	Education	N
Men	1,274	18-34	379	Low	519
Women	1,322	35-49	497	Medium	914
Other	4	50-64	670	High	1,161
		65+	1,054	Unknown	6

Table 2.2 provides housing characteristics reported in the study by respondents. The largest share of respondents live in a row / terraced / corner dwelling type and are home owner.

Table 2.2 Housing characteristics (N = 2,600).

Dwelling type	N	Tenure	N
Detached	370	Owner-occupied	1,739
Semi-detached	382	Rented	844
Row / terraced / corner	1,076	Free housing	16
Ground-floor dwelling	84	Unknown	1
Apartment	629		
Other	59		

2.2 Procedure

Respondents entered the web questionnaire via an email invitation. This web questionnaire contained all questions and instructions to the data donation task. Respondents would donate their data by means of a file upload within the questionnaire. After closing the questionnaire, they were done with the study.

2.3 Experimental conditions

Incentives were varied between respondents to assess the height of the financial compensation needed to offset the cognitive and technical burden and privacy concerns. These experimental incentives were paid on top of the regular incentive for completing the questionnaire (€15 per hour) and did not replace it. LISS panellists are often invited for experimental and innovative studies and receive additional incentives for tasks such as data donation or installing equipment at their home, so an incentive of zero was not used: it would have been inconsistent with panellists' expectations and would not have allowed us to study the effect of incentive height, which was our aim. In this study these additional incentives were varied for both donating the data and willingness to install the dongle, and the requested installation period for the dongle was randomized as well to assess how willingness depends on the required installation period. The two experimental incentives concerned two distinct tasks: a given respondent could in principle be offered both the donation incentive (for uploading insights) and, if they owned a smart meter, the dongle incentive (for the hypothetical dongle), but the two were randomized independently and never applied to the same task.

- **Insight donation:** Respondents who have and check their insights were randomized to an incentive of €5, €10, or €15 for donating their insights.
- **Install dongle:** Smart-meter owners were randomized to a requested installation period (one week, one month, or one year) and, if they indicated to doubt or were not willing to install a dongle, to an incentive of €10, €15, or €20.

2.4 Measures

In the questionnaire, respondents were first introduced to the topic by answering several contextual questions about energy-related characteristics. Subsequently, respondents were asked whether their dwelling had a smart meter and whether they had any insight into their household's energy consumption. Furthermore, they were asked how often they checked their insights and whether they adopted their behavior based on the insights. Those reporting to have and check their insights, were then asked whether they were willing to donate the available insights. Respondents who indicated willingness were asked to download or export their data and upload it within the questionnaire. The donation request and task required respondents to locate, extract, and upload files independently. No file type was specified, as the aim of the study was to investigate the diversity in file types. If they could not upload a file, respondents were instructed to upload a screenshot or photo of the displayed insight. Some evaluation questions were posed to those who had successfully donated data, respondents who failed were asked to explain what went wrong and respondents who were not willing were asked for their reasons for not donating. Finally, respondents with a smart meter were asked a hypothetical question about willingness to install a P1 dongle that would enable direct data sharing without manual downloads or screenshots. Again the respondents who were not willing were asked for their reasons. The entire questionnaire can be found in Appendix A (in Dutch).

The procedure above defines a sequence of stages that a respondent passes through: from having access to insights, to being willing to donate, to a successful upload, to material that is actually usable. We refer to this sequence as the donation pipeline and use it as the organising structure for the results; the attrition across its stages is shown later in Figure 3.3. Based on the responses to the questionnaire and data uploads, we defined one key indicator for each stage:

- **Access indicators:** self-reported smart meter (yes/no/don't know) and whether respondents had at least one insight method and checked this. Insight methods included apps, websites, in-home displays, emails, and "other" (including self-tracking and home automation).
- **Willingness to donate insights:** binary indicator among those who have and check insights.
- **Donation success:** whether the respondent donated a file or screenshot among those willing.
- **Minimum usability:** whether the donated material met a minimum requirement of at least seven days of data and at least daily frequency. These requirements were a trade-off between inclusiveness and analytical usability. Additional quality attributes included unit (kWh/m^3 vs. euros), whether values were explicit or only visible as a graph, and whether electricity and/or gas were included.
- **Willingness to install a dongle:** binary indicator among smart-meter owners for the hypothetical scenario.

2.5 Analytical approach

Descriptive analysis was used to determine whether respondents have a smart meter and whether they have insight into their energy consumption. Descriptive analysis was likewise used to assess through which methods respondents have insights, how often they check these insights, and whether they use the insights to adapt their energy behaviour.

Binary logistic regression models were used to test the dependence of willingness to donate insights or install a dongle on the background characteristics gender, age and education level. The general significance per variable was determined using likelihood ratio tests. For willingness to donate insights the randomized incentive was also added in the model. For willingness to install the dongle, only chi-squared tests were done for the randomized incentive and period conditions, as these were presented sequentially depending on the provided answers (the incentive was only shown when respondents were initially unwilling or in doubt).

Descriptive statistics were used again to describe the donated files and the reasons for being unwilling to donate data or install a dongle.

3 Results

3.1 Access to smart meters and energy insights (RQ1)

The majority of respondents reported having a smart meter (76.7%), while 13.5% did not know. Figure 3.1 shows that the dominant insight method is via app (45.7%) followed by mail (26.8%). Furthermore, 14.8% of respondents reported having insights via multiple methods, which is indicated in the figure by the colouring of the bars.

Interestingly, insights were not limited to smart-meter owners: 47.4% of respondents without a smart meter (or unsure) still reported having some insight into energy consumption, again predominantly via app or mail. Only 19.7% reported having no insights, implying that more than 80% of respondents would in principle be able to donate data.

When asked what kind of apps, respondents reported energy suppliers or solar panels. Other methods that provided respondents with insights of their energy consumption were in-home

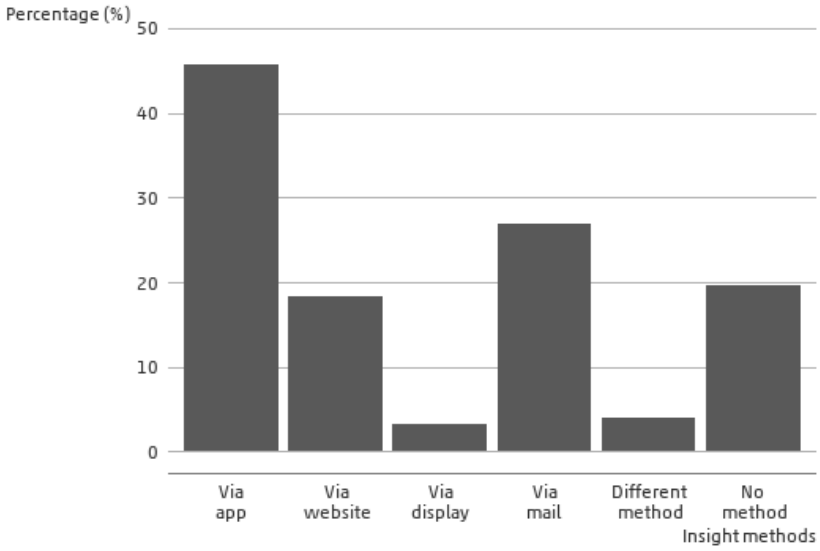

Figure 3.1 Methods through which respondents have insight into their energy consumption.

displays (often supplier-linked) and self-tracking (spreadsheets, paper logs, and home automation tools). Figure 3.2 shows that most respondents reported checking their insights at least monthly. Those who checked their insights more often were also more likely to report using the information to adjust their behaviour.

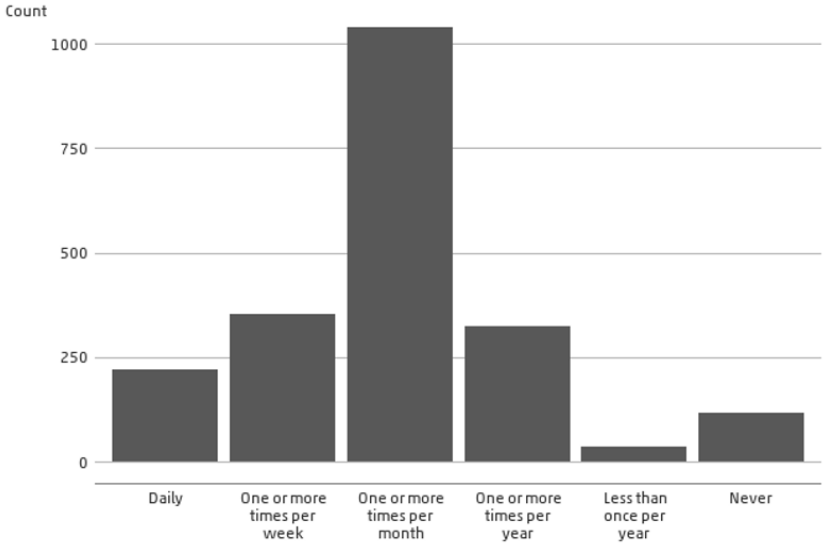

Figure 3.2 How often respondents check the insights they have.

3.2 Willingness to donate insights and incentive effects (RQ2)

Among respondents who have and check their insights (N = 1,971), 35.5% indicated willingness to donate. Willingness per incentive level is shown in Table 3.1, and the corresponding logistic regression in Table 3.2. The association between incentive and willingness was statistically significant ($\chi^2(2) = 8.1, p = .02$): offering €10 or €15 both increased the likelihood of donating relative to €5. Notably, the estimated effects of €10 (OR = 1.33) and €15 (OR = 1.32) were almost identical, with strongly overlapping confidence intervals, so raising the incentive from €10 to €15

Table 3.1 Willingness to donate energy insights by incentive amount.

Incentive	N in condition	Willing to donate
€5	655	31.1%
€10	656	37.2%
€15	660	38.0%

Note. Only respondents who have and check their insights were asked to donate; $N = 1,971$.

does not appear to buy additional willingness. If this holds under a direct test of the €10 versus €15 contrast, an NSI could realise the same participation at the lower amount. Willingness was also higher among men, younger respondents, and those with higher education (gender: $\chi^2(1) = 7.0, p = .01$; age: $\chi^2(3) = 102.9, p < .001$; education: $\chi^2(2) = 11.9, p = .003$; see Table 3.2).

Table 3.2 Binary logistic regression predicting willingness to donate data.

Variable	Category	B	SE	OR	$95\%CI$	p
Condition	(ref. = €5)					
	€10	0.29	0.12	1.33	[1.05,1.68]	.02
	€15	0.27	0.12	1.32	[1.04,1.67]	.02
Gender	(ref. = man)					
	Woman	-0.35	0.10	0.71	[0.58,0.86]	<.001
Age	(ref. = 18-34)					
	35-49	-0.24	0.17	0.79	[0.56,1.10]	.16
	50-64	-0.80	0.16	0.45	[0.33,0.62]	<.001
	65+	-1.21	0.16	0.30	[0.22,0.41]	<.001
Education	(ref. = Low)					
	Middle	0.23	0.15	1.26	[0.94,1.68]	.13
	High	0.46	0.14	1.59	[1.20,2.10]	.001

Note. $N = 1,967$. Respondents identifying as "Other" gender were excluded due to small cell size. SE = standard error, OR = odds ratio, CI = confidence interval. Reference categories are marked as ref.

3.3 From willingness to usable donated data (RQ3)

Figure 3.3 shows the attrition along the full data donation pipeline. The gap between willingness and usable data is substantial: of the 699 respondents willing to donate, 578 (82.7%) uploaded a file or screenshot, but only 270 of those uploads met the minimum usability threshold. In other words, fewer than four in ten willing respondents produced data that was analytically usable, and this group amounts to less than 10% of the full sample.

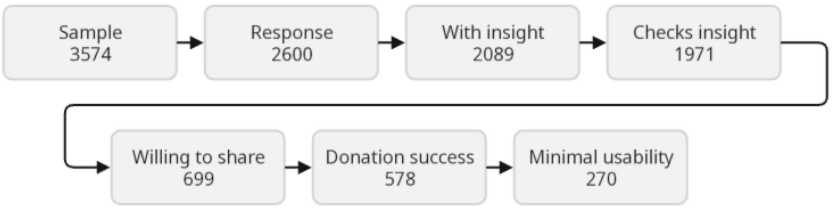

Figure 3.3 Response flow through the data donation pipeline, from the invited sample to the minimally usable donations.

Donation success also varied by the type of insight respondents had. Table 3.3 relates the number of successful donations to the number of respondents who reported having each insight

method. The highest donation rate was observed among respondents who accessed insights through an app (27.7%), compared with 15.2%–19.2% for the other methods. This is consistent with the expectation that app-based information is easier to retrieve and donate than information shown on in-home displays or ad-hoc tracking systems.

Table 3.3 Donation success by insight method.

Method	Donation success	Insight method	Donation rate
App	329	1,187	27.7%
Website	86	475	18.1%
Display	13	85	15.2%
Email	134	698	19.2%
Other	16	105	15.2%

Each successfully donated file was assessed on the period it spanned, its measurement interval, the type of consumption it described, the unit used, and how the numbers could be extracted; Table 3.4 reports these attributes for the 570 files that contained data. The period requirement was rarely the bottleneck: 545 of 570 files (95.6%) covered at least seven days. The binding constraint was measurement frequency, where only 294 files (about half) contained at least one measurement per day. Applying both requirements jointly leaves 270 files that meet the minimum usability threshold.

Table 3.4 Characteristics of donated data (N = 570).

	Yes	No
Period *(at least 7 days)*	545	25
Interval *(one measurement per day or more)*	294	276
Type *(electricity)*	543	27
Unit *(kWh)*	518	52
Derivation method *(numeric)*	309	261

Note. $N = 570$. 8 donated files did not contain any data.

Among the 270 minimally usable cases, some covered only electricity (45.2%) or only gas (5.6%), with 47.8% covering both. 87.8% provided consumption in energy units (kWh for electricity; m^3 for gas), while the remainder provided values in euros. Only 30.7% provided explicit numeric values; for 67.8%, values would need to be derived from a graph.

Although most respondents who managed to donate a file reported that doing so was (very) easy (see Figure 3.4), respondents who attempted to donate but failed explained this by the lack of technical skills and device-related constraints (e.g. the insight not being available on the device used for the questionnaire). Among those who were not willing to donate, the most frequently mentioned reasons were that donating would take too much time, privacy concerns, and perceived difficulty (see Figure 3.5).

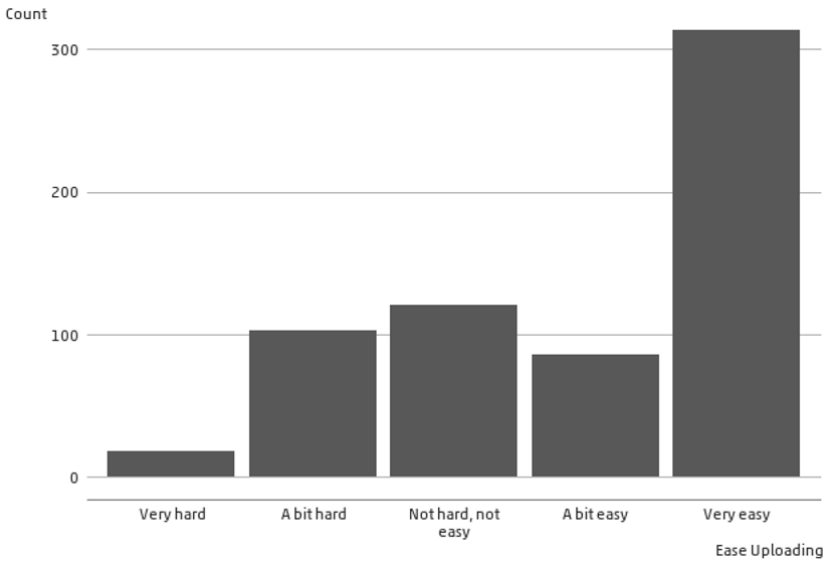

Figure 3.4 Ease of donating amongst respondents who were successful.

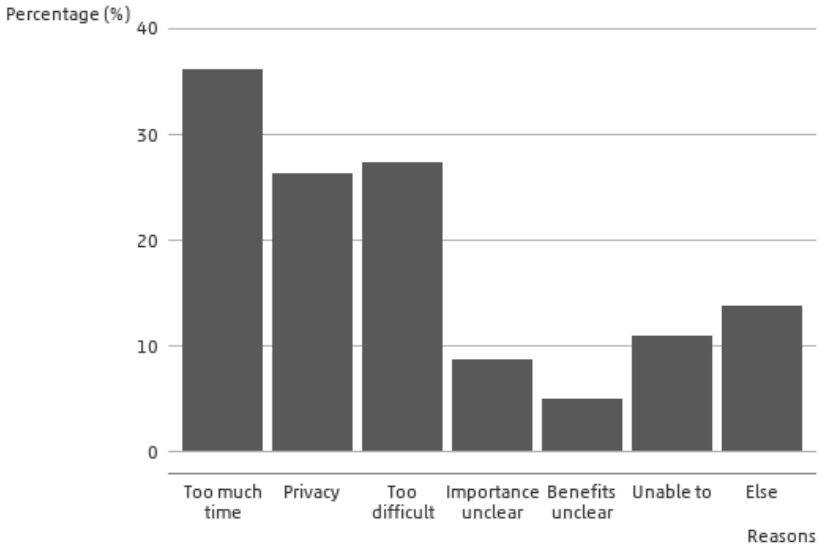

Figure 3.5 Reasons for not donating data.

3.4 Willingness to install a smart meter dongle (RQ4)

Among smart-meter owners (N = 1,995), 19.6% indicated willingness to install a dongle before any incentive was specified. As Table 3.5 shows, willingness was essentially flat across the randomized requested periods (18.9% for a week, 20.0% for a month, 20.1% for a year); these differences are not significant ($\chi^2(2) = 0.4, p = .83$). The requested installation period thus had no detectable effect on willingness.

Table 3.5 Willingness to install a dongle by period (all respondents with smart meter; N = 1,995).

Period	N in condition	Willing to install dongle
week	688	18.9%
month	660	20.0%
year	647	20.1%

The 1,603 respondents who were initially unwilling or in doubt were then told about an incentive. Offering this incentive raised overall willingness from 19.6% to 27.1%, but, as Table 3.6 shows, willingness within this group was not associated with the height of the incentive (7.3%, 10.8%, and 9.8% for €10, €15, and €20 respectively; $\chi^2(2) = 4.1$, $p = .13$).

Table 3.6 Willingness to install a dongle by incentive (respondents with smart meter who initially did not want to install the dongle; N = 1,603).

Incentive	N in condition	Willing to install dongle
€10	523	7.3%
€15	509	10.8%
€20	571	9.8%

This overall willingness to install a dongle (27.1%) depended on gender and age, but not on education (gender: $\chi^2(1) = 36.9$, $p < .001$; age: $\chi^2(3) = 34.2$, $p < .001$; education: $\chi^2(2) = 4.7$, $p = .10$). As the logistic regression in Table 3.7 shows, men and younger respondents were more willing to install the dongle, mirroring the selectivity found for insight donation.

Table 3.7 Binary logistic regression predicting willingness to install dongle.

Variable	Category	B	SE	OR	$95\%CI$	p
Gender	(ref. = man)					
	Woman	-0.68	0.11	0.51	[0.41,0.62]	<.001
Age	(ref. = 18-34)					
	35-49	0.00	0.19	1.00	[0.69,1.46]	>.999
	50-64	-0.22	0.18	0.81	[0.56,1.16]	.24
	65+	-0.64	0.18	0.53	[0.37,0.76]	<.001
Education	(ref. = Low)					
	Middle	0.24	0.16	1.27	[0.94,1.73]	.12
	High	0.32	0.15	1.38	[1.03,1.86]	.03

Note. $N = 1,991$. Respondents identifying as "Other" gender were excluded due to small cell size. SE = standard error, OR = odds ratio, CI = confidence interval. Reference categories are marked as ref.

Respondents who did not want to install a dongle most often mentioned privacy concerns (solid bars in Figure 3.6). Those who were undecided more often cited uncertainty about the effort and unclear personal benefit (open bars in Figure 3.6). A practical barrier that emerged in open answers was that the P1 port was already in use (e.g. by an existing device).

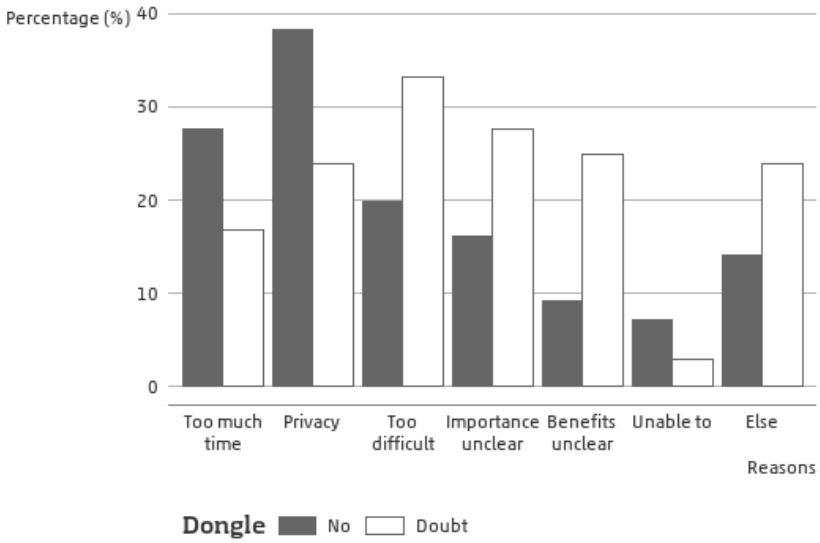

Figure 3.6 Reasons for not being willing to install a dongle. Solid bars: respondents who were not willing; open bars: respondents who were in doubt.

4 Discussion

This feasibility study provides a benchmark for energy insight data donation in a probability-based panel. The results suggest that *coverage of insights* is relatively high: most respondents report having insight into their energy consumption, often through energy supplier apps (RQ1). However, only about one-third of respondents with insights report willingness to donate (RQ2), and an even smaller fraction provides data that is minimally usable for statistical analysis (RQ3). Willingness to install a smart meter dongle, at 27.1% after incentives, was of a similar order as willingness to donate insights (RQ4). This staged attrition aligns with the broader data donation literature, which stresses that consent and willingness are only part of the quality equation: extraction, transfer, and usability matter equally (Boeschoten et al. 2022; Carrière et al. 2025). Together, these results make a concrete contribution to assessing the feasibility of using data donation of household energy insights in official statistics.

4.1 Implications for official statistics

Although smart meters are near-universal in the Netherlands, the methods through which respondents obtain insight vary a great deal, resulting in donated data that is too heterogeneous in content and format to be integrated directly. Donated material was sometimes presented in euros, only visible as a graph, aggregated at monthly intervals, or fragmented across electricity and gas. Preferably, data is presented by every provider in a highly structured format. Without additional standardisation, such heterogeneity can undermine comparability and increase processing costs. The findings therefore support two strategic directions:

- **Design for usability.** Provide clearer guidance, specify acceptable export formats, and embed real-time validation (e.g. in-survey checks that the file contains at least daily values covering the requested period). Best-practice work recommends piloting and user-centred improvements to reduce friction and quality loss (Carrière et al. 2025; Welbers et al. 2024). Probability-based panels such as LISS can facilitate development hereof in piloting and test phases before scaling to broader populations.

– **Consider assisted or automated collection.** While willingness to install a dongle is lower than willingness to donate insights, a dongle-based approach can yield more structured, high-frequency, machine-readable data and may reduce respondent burden after installation (Bakker et al. 2026). Potentially, hybrid approaches are possible: insight donation can serve as a low-barrier entry mode, while a subset of respondents can be recruited for more structured measurement.

These strategic directions can be generalized to other focus areas. Data donation for use in official statistics is most feasible for data sources that are widely adopted by the entire population and for which data is stored in a highly structured format.

4.2 Selectivity and bias

As argued in the introduction, selectivity and bias are not a side issue but a core quality criterion for any candidate statistical source: a source that is easy to donate but systematically skewed relative to the target population is of limited value for official statistics. Our results speak directly to this criterion. Both insight donation and dongle willingness were higher among men and younger respondents, and insight donation willingness was higher among those with higher education. This suggests potential nonparticipation bias in donated datasets, consistent with earlier findings on sensor and trace data sharing (Struminskaya et al. 2020; Struminskaya et al. 2021). The response bias found in this study based on demographic variables can be corrected for using weighting procedures. Problems occur when the selectivity of the respondents results in selection bias, because of a correlation between the response rate and the target variable. In the case of household energy consumption, it is likely that willingness to donate is related to technical skills and interest in energy consumption. These factors might also be correlated with actual energy use, therefore resulting in a bias in the target variable that cannot be corrected for using demographic variables. Future studies should further investigate the dependency of response in energy related studies on actual energy consumption to quantify this response bias in more detail.

4.3 Limitations

First, the study was designed as a feasibility assessment, with intentionally limited instructions to accommodate the wide variety of insight sources. This design choice likely reduced the fraction of machine-readable, standardized donations. Additionally, limited instructions may have resulted in fewer successful data donations: with clear instructions also the less technical respondents may have been able to donate data. Second, the dongle question was hypothetical and asked at the end of the questionnaire; respondent fatigue may have lowered stated willingness.

4.4 Conclusion

This study assessed willingness to donate household energy insights and the heterogeneity of the donated files. It contributes to the emerging literature on data donation by conceptualising and empirically assessing donation as a multi-stage process, distinguishing between data access, willingness to donate, successful donation, and the provision of minimally usable data. Using a probability-based panel, we provide rare population-based evidence on each stage of this process in the context of energy usage data. Our findings show that, although data access and willingness to donate are relatively high, significant attrition occurs at later stages, leaving a small

usable dataset. Whereas many studies examine only willingness and response bias, this study also assessed the donated data itself.

While our empirical case is energy, the central lesson is not specific to it. The same staged, quality-oriented lens (i.e., access, willingness, successful donation, minimally usable data, and selectivity at each step) applies to any digital-trace source that a statistical institute might consider, and it clarifies when donation is a realistic option: it is most feasible for sources that are widely held across the population and stored in a structured, machine-readable format. For official statistics, then, the main challenge is not participation itself, but ensuring that donated data is usable, comparable, and scalable.

Acknowledgments

The views expressed in this paper are those of the authors and do not necessarily reflect the policies of Statistics Netherlands. We thank Centerdata and the LISS panel participants for enabling this study.

Funding

This work was carried out as part of the working agreement between Statistics Netherlands (CBS) and Centerdata. Both provided in-kind support in the form of staff time.

Data availability

The underlying microdata originate from the LISS panel and contain donated files and screenshots. Access is restricted due to privacy and confidentiality requirements. Aggregated outputs and analysis code may be made available upon reasonable request to the authors, subject to approval by the data owner and Statistics Netherlands.

References

Araujo, T., J. Ausloos, W. van Atteveldt, F. Loecherbach, J. Moeller, J. Ohme, D. Trilling, B. Van der Velde, C. de Vreese, and K. Welbers (2022). "OSD2F: An open-source data donation framework". In: *Computational Communication Research* 4.2, pp. 372–387. DOI: 10.5117/CCR2022.2.001.ARAU.

Bakker, J., M. Kompier, and V. Toepoel (2026). "What can we ask respondents in terms of data donation? The case of energy usage". In: *Statistical Journal of the IAOS*. Under review.

Boeschoten, L., T. Araujo, J. Ausloos, J. Möller, and D. L. Oberski (2022). "A framework for privacy preserving digital trace data collection through data donation". In: *Computational Communication Research* 4.2, pp. 388–423. DOI: 10.5117/CCR2022.2.002.BOES.

Carrière, T. C., L. Boeschoten, B. Struminskaya, H. L. Janssen, N. C. de Schipper, and T. Araujo (2025). “Best practices for studies using digital data donation”. In: *Quality & Quantity* 59.1, pp. 389–412. DOI: 10.1007/s11135-024-01983-x.

Centerdata (2024). *Referencing and citing the LISS panel*. Reference document (version 6.1). URL: https://www.lissdata.nl/app/uploads/sites/4/2024/08/Reference_LISS_v6.1_20240828.pdf.

Keusch, F., B. Struminskaya, J. Mulder, and S. Jongerius (2025). “Data donation as a method to measure physical activity in older adults: Cross-sectional web survey assessing consent rates, donation success, and bias”. In: *Journal of Medical Internet Research* 27, e69799.

Kompier, M., A. Elevelt, A. Luiten, J. Mulder, and V. Toepoel (2024). “Data donation of personal physical activity trackers”. In: *Survey Practice* 17.

Kraakman, R., M. Kompier, A. Luiten, V. Toepoel, et al. (2023). “The feasibility of using consumer-level activity trackers in population monitoring of physical activity”. In: *Survey Research Methods* 17.4, pp. 429–445.

Scherpenzeel, A. C. (2011). “Data collection in a probability-based Internet panel: How the LISS panel was built and how it can be used”. In: *Bulletin de Méthodologie Sociologique* 109.1, pp. 56–61. DOI: 10.1177/0759106310387713.

Statistics Netherlands (CBS) (2025). *Slimme-meterdata: nieuwe bron voor gas- en elektriciteitsverbruik*. News release, 25 April 2025. URL: https://www.cbs.nl/nl-nl/corporate/2025/17/slimme-meterdata-nieuwe-bron-voor-gas-en-elektriciteitsverbruik.

Struminskaya, B., P. Lugtig, V. Toepoel, B. Schouten, D. Giesen, and R. Dolmans (2021). “Sharing data collected with smartphone sensors: willingness, participation, and nonparticipation bias”. In: *Public Opinion Quarterly* 85.S1, pp. 423–462. DOI: 10.1093/poq/nfab025.

Struminskaya, B., V. Toepoel, P. Lugtig, M. Haan, A. Luiten, and B. Schouten (2020). “Understanding willingness to share smartphone-sensor data”. In: *Public Opinion Quarterly* 84.3, pp. 725–759. DOI: 10.1093/poq/nfaa044.

Toepoel, V., A. Luiten, and R. Zandvliet (2021). “Response, willingness, and data donation in a study on accelerometer possession in the general population”. In: *Survey Practice* 14.1.

Welbers, K., F. Loecherbach, Z. Lin, and D. Trilling (2024). “Anything you would like to share: Evaluating a data donation application in a survey and field study”. In: *Computational Communication Research* 6.2, pp. 1–28. DOI: 10.5117/CCR2024.2.5.WELB.

Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, Grafimedia

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information