# ESSnet Smart Surveys

**Grant Agreement Number: 899365 - 2019-DE-SmartStat**

Link to our CROS website

## Workpackage 2

## Smart Survey Pilots

# Deliverable 2.8: Functional and technical descriptions of tools - Living conditions (WP2.4)

**Version 1.0, 8-12-2021**

**Prepared by:**

Annemieke Luiten (CBS, Netherlands)
a.luiten@cbs.nl

Workpackage Leader:

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

SUMMARY: WP2 performs four diverse pilots to inform WP3 on the specifications of a smart survey platform in the European Statistical System (ESS). This deliverable describes how the outcomes of WP2.3 (health) are fed back to WP3. The Inventory of Innovative Tools and Sources is described, using CSPA recommendations. A CSPA compliance template is filled in and added as appendix. The tools and sources inventory is supplemented by a methodological level, a logistics level and a legal-ethical-policy level. On top of these levels, a mapping of the WP2.3 pilots to the WP3 POCS and smart survey features is described. <u>It is stressed that empirical motivation for specifications in this deliverable are given in deliverable 2.1 d.</u>

## 1. Introduction

In WP2.4 we investigate possibilities to use sensor systems that inform on indoor environment quality (IEQ). Indoor environment has a large impact on health. As such the findings can be integrated in health surveys, but can also be used for housing surveys, or to monitor energy transition effects, for example. The measurements are not so much a replacement of present survey questions, but offer additional and richer insights and quality.

WP2.4 was designed as a series of small scale pilots with a relatively large focus on Phase 0 (tool selection), and Phases 1 (conceptualisation and data collection instruments) and 2 (methodology). Phase 2' (IT and infrastructure) are touched upon, but are no main focus of the project. A description is given of the IT and architectural environment as needed for the chosen device. Other solutions are possible though, with additional or other requirements.

Phase 0 ended with the selection of the uHoo air as the tool of choice for all pilots, although it was clear from the onset that this was not the ideal tool. It would however, only be possible to do a number of pilots within the time span of the ESSnet Smart Surveys if we used an OTS product. There was not time to develop a device, plus the dedicated software. This also means that the description of tools in this deliverable is but one of possible solutions. In the ideal solution, both the tool and the software environment would be developed in house, or by a trusted third party.

Deliverable 2.7 (Schouten, 2020) describes the elements of a smart survey tool, being the IT frontend, IT backend, Methodology and Logistics. Not all of the elements described in the Deliverable 2.7 are relevant for a smart survey using the uHoo, and there are on the other hand additional elements: the logistics of preparing, sending out, receiving back and cleaning of the uHoos.

In the remainder of this deliverable, these four elements will be described for surveys using the uHoo. In Section 5 the legal-ethical-policy features of measuring IEQ in this way are described. Sector 6 shortly mentions how country specific application could influence comparability of outcomes. Finally, in section 7, the mapping of these findings on the WP3 POCs is summarized.

For the IT frontend and backend, the existing Eurostat initiatives to create and maintain an inventory following CSPA (Common Statistical Production Architecture) are mostly sufficient. The inventory, as described at https://webgate.ec.europa.eu/fpfis/wikis/display/ISTLCS/INVENTORY is filled in for the IT front and backend for the uHoo tool. See Annex A for the inventory questionnaire.

## 2. IT frontend

Requesting participants to place a uHoo in their home will most probably take place within the context of a survey. At an appropriate position in the survey, the respondents will be asked if they are willing to undertake the additional task of placing the uHoo in their home for a number of weeks. Hence, case management is primarily a task of the NSI's standard survey administration. For respondents consenting to the additional task, a new branch has to be developed towards the management of the uHoo's.

### 2.1 The uHoo

The uHoo sensor system measures CO2, CO, NO2, TVOC, CO, temperature, humidity and Ozone. The device is placed in a designated room in the house of the participant. It needs to stay connected to electricity and Wi-Fi at all times. Every minute, measurements are taken and transmitted to the uHoo cloud environment. The participants needs to connect the device to internet by downloading the uHoo app from playstore or apple store. The participant registers his device with a user name and password. During this process, the app demands access to location. Without consent, the app does not continue the download. The respondent can at the end of the process withdraw this access, but at that time the location is obviously already known. Location is not transmitted with the raw data, however. See figure 1 for a picture of the uHoo.

**Figure 1. the uHoo air.**

Because the project group needs to access the data, the research group communicated the password and user name to the respondent. Each device had its own user name and password. For this pilot, every uHoo was used in three pilots and the passwords and user names for each of the devices were the same for the three pilots. Once more structural fieldwork would be performed, an other solution obviously needs to be found for this procedure. The most elegant solution from the point of view of data security would be to let the participant choose his own name and password. The down side of that solution would be that the research organization or NSI cannot monitor progress, cannot download the data, and cannot generate the feedback.  The participant would have to generate the data, and transmit these to the NSI, either direct or via data donation.

The uHoo and the WiFi connection make use of a 128-bit AES end-to-end encryption to secure the data.

Table 1 gives an overview of the sensors used in the uHoo, the detectable range, resolution and estimated accuracy of the measurements.

Table 1. Detectable range, resolution, and estimated accuracy of the tested low-cost multi-channel monitors.

| Sensor | Type | Model | Detectable range | Resolution | Estimated accuracy |
|---|---|---|---|---|---|
| PM2.5 | Optical scattering | Shinyei Kaisha PPD42-60 | 0–200 μg/m3 | 0.1 μg/m3 | ±20 μg/m3 or 10% of reading |
| CO2 | NDIR | ELT Sensor T-110-3V | 400 to 10,000 ppm | 1 ppm | ±50 ppm or 3% of reading |
| CO | Electrochemical | Figaro Engineering TGS5342 | 0–1000 ppb | 0.1 ppb | ±10 ppm |
| TVOC | Metal oxide | Cambridge CMOS CC881B | 0 to 1000 ppb | 1 ppb | ±10 ppb or 5% of reading, based on the types of VOC |
| O3 | Metal oxide | SGX Sensortech MICS-2714 | 0–1000 ppb | 1 ppb | ±10 ppb or 5% of reading |
| NO2 | Metal oxide | SGX Sensortech MICS-2714 | 0–1000 ppb | 1 ppb | ±10 ppb or 5% of reading |

Note: table copied from Baldelli, 2021.

## 2.2. Setting up the uHoo for use

The app that enables the use of the uHoo is available from playstore and apple store:
https://apps.apple.com/us/app/uhoo/id1084953997
https://play.google.com/store/apps/details?id=com.uhooair&hl=en&pcampaignid=pcampaignidMKT-Other-global-all-co-prtnr-py-PartBadge-Mar2515-1

Once downloaded, there is an in-app instruction that helps respondents through the process. This instruction is sufficient for most participants, but a quite substantial part of participants ran into trouble while installing the app and we needed to get into contact with uHoo's technical helpdesk quite frequently to try to help them. We did not succeed in all cases. Some participants never succeeded to install the app. We have tried developing an instruction for the participants, but this ran into the problem that circumstances could be quite different for different platforms and different telephones, even from the same type.

In addition, the uHoo needs to be signed up to the monitoring platform. CBS has received passwords for each uHoo, different ones than the ones that are used by respondents. CBS has an administrator account, where we can monitor all uHoos in use for any pilot. For practical reasons, CBS has performed this task also for the pilots in Germany and Belgium.

## 2.3. Calibration

Sensors need to be calibrated at regular intervals to align the measurements (of indoor substances in this case) to known values, e.g., as measured with expensive dedicated sensors of research institutes like the Dutch RIVM or TNO, the Belgian Sciensano or VITO, and the German Robert Koch

Institute or Fraunhofer. The following is an excerpt from Baldelli (2021): "*The device generates readings each minute and auto-calibrates the $CO_2$ sensor every 168 h (7 days) and the TVOC, O3 and NO2 sensors every 24 h. This auto-calibration estimates exposure levels that the sensor is subject to during operation. then the software updates and becomes more accurate to the specific operating conditions. Due to this auto-calibration, a 48-h "warm up" period is needed for the uHoo monitor to provide better measurements… Auto-calibration is important because chemical sensors age faster in polluted environments and their baselines need to be recalibrated frequently* ".

### 2.5 Helpdesk support

Technical support was necessary in the pilots, and will need to be available if we roll out this system to a large(r) population. Some technical assistance is available online, but generally, the assistance offered there was not sufficient for the problems that we ran into.

## 3. IT backend

IT backend describes case management, data processing, the web server, data models, the user database, linkage to existing data, etc.

### 3.1. Data processing

The output of the uHoo data consists of a .csv file, with the values of the 8 measurements, taken every ten minutes. See table 2 for an example.

**Table 2. Output of the uHoo**

| Time | Device | CO2 | Temperature | Humidiy | PM2.5 | VOC | NO2 | Ozon | CO |
|---|---|---|---|---|---|---|---|---|---|
| 11-1-2021 17:50 | CBS-24 | 646 | 19.7 | 41.6 | 9 | 0 | 19.4 | 8 | 0 |
| 11-1-2021 18:00 | CBS-24 | 831 | 21.1 | 39.9 | 3 | 11 | 19.4 | 8 | 0 |
| 11-1-2021 18:10 | CBS-24 | 869 | 21.9 | 35.9 | 11 | 0 | 19.4 | 8 | 0 |
| 11-1-2021 18:20 | CBS-24 | 843 | 22.3 | 34.7 | 10 | 0 | 19.4 | 8 | 0 |
| 11-1-2021 18:30 | CBS-24 | 818 | 22.5 | 33.8 | 6 | 0 | 19.1 | 8 | 0 |
| 11-1-2021 18:40 | CBS-24 | 813 | 22.6 | 33.4 | 6 | 0 | 18.8 | 8 | 0 |
| 11-1-2021 18:50 | CBS-24 | 814 | 22.7 | 33 | 5 | 0 | 19.4 | 8 | 0 |
| 11-1-2021 19:00 | CBS-24 | 806 | 22.9 | 32.7 | 12 | 0 | 19.4 | 8 | 0 |
| 11-1-2021 19:10 | CBS-24 | 833 | 22.9 | 33.1 | 9 | 0 | 18.4 | 7.9 | 0 |
| 11-1-2021 19:20 | CBS-24 | 836 | 23 | 32.8 | 4 | 0 | 19.4 | 8 | 0 |
| 11-1-2021 19:30 | CBS-24 | 846 | 23.1 | 32.6 | 7 | 0 | 17.7 | 7.9 | 0 |
| 11-1-2021 19:40 | CBS-24 | 843 | 23.1 | 32.5 | 6 | 0 | 19.4 | 8 | 0 |
| 11-1-2021 19:50 | CBS-24 | 851 | 23.3 | 32.3 | 2 | 0 | 17.6 | 7.9 | 0 |
| 11-1-2021 20:00 | CBS-24 | 850 | 23.4 | 32.2 | 8 | 0 | 17.6 | 7.9 | 0 |

### 3.2. Plausibility checks and validation rules

The data need to be checked for non-use, originating from for example unplugging the device , either deliberately or (mostly) undeliberately. Also change in the WiFi setup can disrupt measurement.

### 3.3. Algorithms to subtract the substantive information

A series of algorithms have been developed by uHoo to transform the raw data from each sensor into user-friendly data showing measurements of each indoor pollutant. A limited amount of data analysis is performed on the platform. The participant sees his IEQ values in colours of, green, orange and red, indicating if certain thresholds are passed. The thresholds are initially defined according to WHO norms, but participants may change these thresholds to their personal needs. A warning signal is sent to the participant's cell phone is a threshold is passed. This signal can be suppressed by the participant.

For data analysis by the NSI, additional algorithms are necessary to distill the data needed by the NSI. In Phase 2' a number of studies were performed on how to use this data. These will be outlined in Deliverable 2.1. During measurement, no machine learning takes place.

# 4. Methodology

The methodology level, as described in deliverable 2.7, consists of:

- User interface of frontend
- Data collection strategy, with the additional components of recruitment materials and incentives
- Plausibility checks and validation rules
- Machine Learning development

The user interface is determined by the uHoo manufacturer. If ever we decide to have a similar device developed according to our specifications, substantial attention needs to be given to this aspect. Now, too many people were not able to set up the device.

## 4. 1. Data collection strategy

The data collection strategy exists of the following elements: the choice whom among the population to measure with IEQ meters (every sample person, a random sample, unhealthy people, those with potentially an unhealthy IEQ, etc.). Other topics in this section are the choice of recruitment vehicle, recruitment materials and incentives.

### *4.1.1. Will everyone be measured with a device?*

In the documentation for WP2.3 is described how this choice may impact measurement, as those who will not wear the sensor will most probably fill in a questionnaire to the same goal. For WP2.4, the situation is a bit different: there are not questionnaires measuring this information. However, the choices made by NSI's may be different, and lead to different outcomes.

### *4.1.2. How to offer the device to the respondent*

In the three pilots conducted with the uHoos, the uHoos were sent to the participants, or in some cases were handed out personally. An alternative offered by the vendor is to place and install the uHoo with the participant. It is evident that in this case, unlike the situation in WP2.3, we cannot send interviewers out with a number of uHoos.

The various scenarios of getting the uHoo to the respondent and various scenarios of getting them back again have implications for the willingness of respondents to participate in this kind of survey research.

1. Sample units are approached with the express purpose of gaining participation in a research project on indoor air quality, to be measured by the uHoo (or an alternative device). Those who are willing receive the measurement system and are also asked to provide additional information on their living condition, aspects of their housing and health.

2. The respondent in the (health or living conditions) survey is asked, e.g., at the end of the survey, if they would be willing to place the measurement device in their house for one month (or another period). The device is sent in the mail, with an instruction how to install it. A return box and label are provided, and the respondent sends the device back.
3. Alternatively, an appointment is made for a representative of the vendor (or a suitable alternative) to come to the participant's home to place and install the device.

#### *4.1.3. Recruitment materials*

- Depending on the format chosen, see 4.1.2, the recruitment takes place via an invitation letter, or a question in the questionnaire. Regardless, the text of the recruitment warrants careful consideration. People are asked to do something that is alien to them. Therefore, the invitation needs to be clear, and to address worries that people have expressed in pilots. How this invitation can best be formulated is a topic of ongoing research. This topic will be detailed in deliverable 2.1.
- Additionally, people need to receive a careful instruction on how to handle the device. The in-app description and the short description that is sent to the participant with the uHoo is sufficient for most, but certainly not for all participants.

#### *4.1.4. Incentives*

There has been but one incentive in these pilots, the feedback that people receive, both from the app on a constant / daily basis, and the once monthly feedback that they received from the research group. It is unclear whether people will need additional monetary incentives to participate in this kind of research. In a pilot that is presently on the agenda, we will ascertain that. It is my feeling however that the insights are sufficient incentive for participants, and that willingness will he high, even without additional monetary incentive.

### 4.2. Machine Learning development

Our thinking on how to use this kind of data is not yet strongly developed. The analyses so far have used regular linear analysis techniques. Some machine learning algorithms may be necessary, e.g., to determine if and when participants moved their uHoos to another room in the house, as requested in these pilots. We have also used ML to determine how long we should measure before a pattern is established in participants' behaviour and hence IEQ. Another algorithm envisaged is understanding if respondents change their behaviour over the duration of measurement period, and if they keep up that behaviour.

However, a suitable or satisfying algorithm has not yet been developed. Machine learning has been used to understand the relation between patterns of IEQ and participants' behavior. In the pilot, we have used a questionnaire to understand the influence of certain aspects of people's living conditions (age of the house, insulation, kind of distractor hood, cooking habits etc.) on IEQ. If we want to understand, for example, what the influence is of the increasing amount of insulation on IEQ, we have to be able to subtract other relevant influences.

## 5. Legal-ethical-policy

Deliverable 2.7 describes that smart surveys need a new framework of legal-ethical-and policy decisions. The bullet points below come from deliverable 2.7. For each, the relevance for WP2.4 is described.

- The data collection may use the computing and storage options of personal devices. In WP2.4 storage partly takes place on devices that are given to the respondent, not on personal devices. However, the participants' smartphone is necessary to connect to WiFi. Potentially, security issues may arise from this use.
- Part of the data collection may be passive. In WP2.4 all data collection is passive, although obviously, the respondent takes the active role of accepting the device and connecting it to internet.
- Part of the data collection processing may be shifted to the personal devices rather than in-house at the institute. This is not relevant for WP2.4. Some participants may have their own indoor measurement systems, but these will be so limited to be unusable. There exist citizen science initiatives to measure air quality, for which participants use their own devices, but these are targeted at outdoor air quality. It is very important to know the measurement quality of the sensor; differences between participants' devices would be most unwelcome.
- Part of the collected data may be of specialist content of which part of the respondents has no knowledge, even under informed consent. For WP2.4, this has two aspects: part of the collected data is indeed of specialist content, to the extent that most of us do not really understand what some of the measurements mean and how they should be interpreted. The other aspect is that the analysis may show aspects of the living arrangements that the respondent is not aware that we can gauge based on $CO_2$ levels, e.g., the number of persons living in the house and at what time they leave, when visitors arrive, activity in the sleeping room, etc. This is however not something we would analyse, so the question is whether we would need to make the respondent aware of this aspect.
- In order to be able to participate, one may have to possess a specific device or to accept that one needs to use a device that is provided. For WP2.4, or more generally the measurement of indoor air quality, the participant needs to accept a device in their home. No alternative is available.

## 6. Comparability over countries

Comparability of (smart) statistics over countries is crucial. Deliverable 2.7 makes a distinction between generic features, that are shared by all NSI's and that are considered to be MUST haves. For measuring IEQ, two generic aspects are relevant: the device and the algorithms used. Comparability can potentially also be threatened if countries deploy the device on varying subgroups within the population.

There are also features that a majority of countries prefer, but that not strongly effect measurement, the SHOULD haves. For example, differences in fieldwork strategy. This will most likely apply to measurement with IEQ devices as well. Countries may differ in their fieldwork strategy (the use of interviewers for example, or the use and amount of incentives). In the case of measurement with IEQ measurement devices some of these aspects will impact measurement comparability of statistics over countries if differing groups will be measured.

Finally, country-specific minority features concern those that are preferred by a few NSI's and do not affect comparability. They are considered COULD haves. For this particular case, no features like these have been described (yet).

## 7. WP3 mapping

Table 1 describes how the information described in this document maps to the areas of specific interest where WP3 will develop PoCs for modular prototype elements: Active-passive trade-offs, machine learning methodology, privacy-preservation, metadata, incentives and front-end configuration

TABLE 1: WP3 POC elements and pilots

| | Indoor Air Quality |
|---|---|
| Active-passive | Passive |
| Machine learning | Transformation of sensor data |
| Privacy preservation | Indoor climate data |
| Metadata | Sensor data and derivations |
| Incentives | Personal statistics |
| Frontend configuration | Not in pilot |

In Deliverable 2.7, an overview is given of the (smart) features employed by smart surveys:

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

Table 2 describes how these elements work out for WP2.3.

TABLE 2: Smart features and pilots

| | Indoor Air Quality |
|---|---|
| Device intelligence | Not in this pilot |
| Internal sensors | YES |
| External sensors | Possible, but not used in this pilot |
| Public online data | YES |
| Personal online data | NO |
| Big data linkage | Possible (data on dwellings), but not in this pilot |

## 8. References

Baldelli, A. (2021). Evaluation of a low-cost multi-channel monitor for indoor air quality through a novel, low-cost, and reproducible platform. *Measurement: Sensors,* 17, doi.org/10.1016/j.measen.2021.100059.

Schouten, B. (2020). Deliverable 2.7: Template for functional and technical descriptions of tools. Available at https://webgate.ec.europa.eu/fpfis/wikis/pages/viewpage.action?spaceKey=EstatBigData&title=ESSnet+Smart+Surveys+2020-2021