



STATIS
Statistisches Bundesamt



IN
Instituto
Nacional de
Estadística

STATEC



Statistics Poland

**Office for
National Statistics**



STATBEL
België in cijfers



Statistisk sentralbyrå
Statistics Norway

eurostat 

ESSnet Smart Surveys

Grant Agreement Number: 899365 - 2019-DE-SmartStat

[Link to our Confluence website](#)

Workpackage 2 **Smart Survey Pilots**

Deliverable 2.9: HBS app documentation

Version 1.0, 15-02-2021

Prepared by:

Jack Janssen (CBS, Netherlands)
Ole Mussmann (CBS, Netherlands)
Barry Schouten (CBS, Netherlands)
Nick de Wolf (CBS, Netherlands)

Workpackage Leader:

Barry Schouten (CBS, Netherlands)

Jg.schouten@cbs.nl

telephone : +31 70 3374905

SUMMARY: This WP2 deliverable is dedicated to the Household Budget Survey application in WP2.1 Consumption. It provides links to the Gitlab repositories for frontend and backend source code. Furthermore, it gives updated descriptions of the CSPA level and preliminary descriptions for the other levels distinguished in deliverable 2.7. In deliverable 2.8 all levels are described. The mappings to WP3 PoC's and general smart features are also done in deliverable 2.8.

1. INTRODUCTION

Deliverable 2.7 proposes four levels of description of smart survey pilots as well as mappings to the WP3 proof-of-concept's and to the six smart survey features. In 2020, functional tests have been performed by six WP2.1 countries. In 2021 also phase 3 field tests will be conducted by three WP2.1 countries. Deliverable 2.8 will provide all details mentioned in deliverable 2.7. This deliverable describes the HBS at the end of 2020 based on the functional tests only and is a stepping stone to deliverable 2.8. Furthermore, links to source code are provided.

The four levels are:

1. CSPA conceptual, logical and physical levels
2. Methodology level
3. Logistics level
4. Legal-ethical-policy level.

The levels are crossed against frontend and backend. Generic specifications and country-specific features are sketched as well, but provided in detail in deliverable 2.5 on the shareability of tools.

2. DESCRIPTION OF THE HBS APPLICATION

We first note that the HBS application is under development. This means the descriptions will be subject to change in the coming years. We discuss each of the deliverable 2.7 levels.

2.1 CSPA conceptual-logical-physical levels

For the IT frontend and backend, we conform to the inventory following CSPA (Common Statistical Production Architecture) by Eurostat. The inventory can be found at <https://webgate.ec.europa.eu/fpfis/wikis/display/ISTLCS/INVENTORY> Annex A gives the completed questionnaire for HBS.

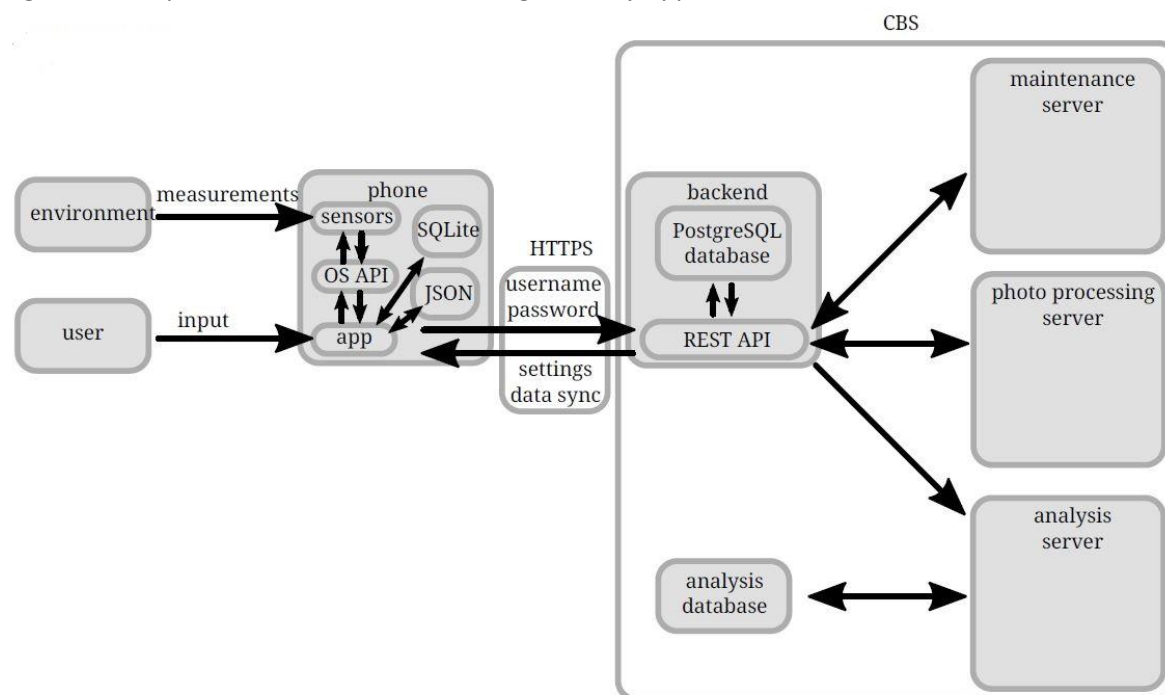
The app is listed as Household Budget Survey in Google and iOS app stores and the source code can be found at <https://gitlab.com/tabi/projects/budget>

The Household Budget Survey app is at version 3 at the time of writing and two new versions are planned for 2021:

- Version 0: Version prepared to evaluate technical problems across countries and devices (June/July 2019)
- Version 1: Version with revisions of technical issues for test round 1 (August 2019)
- Version 2: Version with revisions based on @HBS test round 1 (December 2019)
- Version 3: Version including receipt scanning pipeline (January 2021)
- Version 4: Version with revisions based on @HBS test round 2 and ESSnet usability tests (May 2021)
- Version 5: Version including introduction questionnaire and multi-device data entry (November 2021)

The app components and data flows are presented in Figure 1. The app components are the frontend, backend, maintenance server, photo processing server and the analysis server.

Figure 1; Components in the household budget survey application



The components are described one by one.

2.1.1 App frontend

The app is available in ten country versions: BE, DE, ES, FI, NL, NO, LU, PL, SI, UK. Description of the user interface and detailed English screenshots of the app frontend can be found in the former report on the action of project @HBS (Schouten, Bulman, Järvensivu, Plate and Vrabič-Kek 2020). Annex B includes screenshots of the main screens.

Generic features of the app frontend are:

- Diary functionality
- Manual data entry and receipt scan data entry
- The use of separate product search lists linked to COICOP
- The use of separate store lists to prepare receipt processing and classification of products
- Four main screens: calendar, list of submitted expenditures, expenditure statistics and settings
- Check of receipt scan data quality
- Feedback of classified receipt

Implemented configurable features of the app frontend are:

- Length of data collection period
- App language
- Content of product search lists

- Content of store search lists (names and types)
- Helpdesk functionality (helpdesk phone number and email)
- Timing of statistics feedback to respondent (instantaneous, delayed, not)

Not yet implemented configurable features are:

- App colour scheme and NSI logo

Features that some countries prefer to be included (will be detailed in deliverable 2.5):

- Quantities and number of units of purchased products
- Within household filtering of expenditures
- Within household anonymization
- Distinction of biological/ecological products
- Inclusion of homegrown products

2.1.2 App backend

A preliminary description of the HBS backend was provided at the end of project @HBS (Janssen 2020). Since then, the backend has been implemented and expanded. The backend communicates via a REST API interface with the apps on the respondent's phones as well as the maintenance-, photo-processing- and analysis-servers. Communications between the REST API and the apps is encrypted via HTTPS and authenticated with usernames and passwords.

See Annex C for a detailed description of the backend.

2.1.3 Maintenance server

The maintenance server has three crucial tasks:

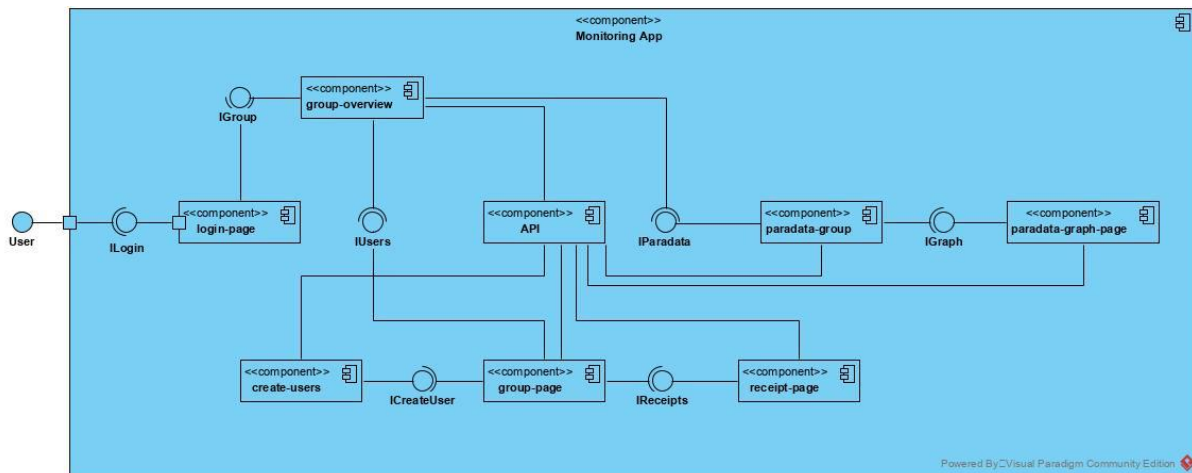
- Case management:
 - Assignment of app protocols
 - Generation of users and passwords
 - Initiation of contact, reminder and motivation strategies
- Monitoring of data collection:
 - Linkage to sampling frame and administrative data
 - Queries on recruitment and participation rates
 - Queries on in-app answer and navigation behaviour
- Helpdesk support:
 - Background on technical errors or other data entry issues
 - Checks of conditional incentive requirements

The maintenance server is accessed through a dashboard web application. Access to the various tasks is dependent on roles and permissions assigned to a person.

The maintenance server is implemented in Angular. The use-case diagram in Figure 2 and the component model diagram in Figure 3 show the main functionalities of the maintenance server:

- Login
- Management of survey groups

Figure 3: Component diagram of maintenance server



2.1.4 Photo processing server

The photo processing server handles the processing of receipts scanned by respondents. Receipt scanning has the following steps:

1. In-app scanning
 - a. Taking a picture
 - b. Check of scan quality to interact with respondent
 - c. In-app OCR to interact with respondent
2. In-house OCR and language processing
3. In-house classification to COICOP
4. In-app feedback of classified receipt

The in-house codebase can be found in the following gitlab repository:

<https://gitlab.com/Nickdewolf/hbs.wp4-shopping-receipt-recognition>

The in-house codebase uses Tensorflow and OpenCV for some of the image pre-processing and Tesseract for the OCR. For COICOP classification we used a library called FastText. All pre-trained models are included in the gitlab repository.

The code has been divided into four categories: *mob*, *ocr*, *coicop*, and *api*.

1. **mob**: Handles the image processing steps, such as background removal, skew correction, and brightness/contrast/color improving.
2. **ocr**: Handles the steps involved with OCR and language processing. The functions in these files call Tesseract on the pre-processed images and return raw text, after which this is further processed to retrieve informations such as purchased products and their corresponding prices.
3. **coicop**: Handles the training classifiers and contains the functions to classify product descriptions into their corresponding COICOP code. The current pre-trained model can handle 112 different COICOP codes.
4. **api**: Contains all the code for retrieving and sending data between the backend. This files uses the data as contained in `requests_config.json` to find the correct urls.

Some example code has been included in the form of jupyter notebooks:

- **00_Prepare_CBS_data:** shows example code on how we processed the internal data, which for privacy reasons cannot be shared on the gitlab repository.
- **01_Train_Models:** shows example code on how to train various models for the COICOP classification. The input .csv files requires just 2 columns: the first containing product descriptions and the second containing the corresponding COICOP code.
- **02_Test_Pipeline:** Shows code for each of the steps included in the pipeline, such as: retrieving an image from the server, apply image processing, apply OCR and language processing, COICOP classification and preparing data to be sent back to the backend database through the rest API.

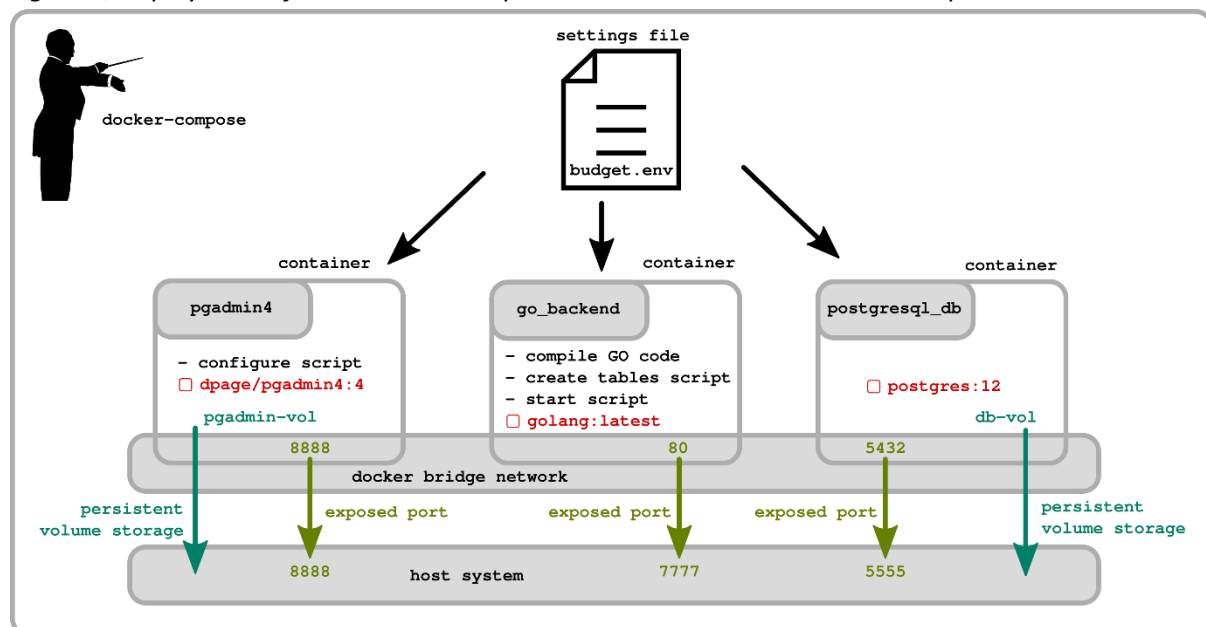
To run the code with the previously mentioned backend, the **main.py** can be used to query the server for changes every 5 minutes. This file essentially runs code similar to the code found in 02_Test_Pipeline.ipynb with additional error handling.

2.1.5 Analysis server

The analysis server is the link to the statistical analysis. It is the starting point for further data processing and production of statistics. It contains no intermediate survey and sensor data and only data of completed data collections.

The Analysis server and database are not implemented yet. While this functionality is not available the input database can be queried directly with the pgAdmin_4 tool for PostgreSQL databases.

Figure 4; Deployment of the backend components with Docker and Docker-compose



2.2 Backend deployment

The backend can be deployed in micro-services. Every logical component is put into a Docker container. These containers are then deployed – and tied together – in a set. A single settings file contains the configuration options for all containers. The orchestration happens via docker-compose.

The docker-compose.yml file contains all the glue that holds the containers together and makes them operate as a team. It takes care of building the containers (if necessary), setting up an isolated network for inter-container communication, exposing container ports to the host system (so that one can still interact with the micro-services individually) as well as creating persistent storage that remains intact even when the ephemeral containers are destroyed.

So far the orchestration includes a PostgreSQL database, the database administration tool PGAdmin4 as well as the Golang REST API. It can easily be expanded to include the maintenance server or the photo processing server as well. See Figure 4.

2.3 Methodology level (preliminary)

The methodology level consists of:

- User interface of frontend: The most influential design choices in the user interface.
- Data collection strategy: This concerns the use of contact modes, contact and reminder strategies, incentive strategies, recruitment materials, the use of “non-smart” modes.
- Data quality checks: Soft and hard checks of plausibility of entered data and notifications of missing data.
- Machine learning models in processing or mixing/fusing sensor data

We discuss each of the levels for the household survey application. These discussions will be updated in later deliverables.

User interface influential choices:

- Search algorithm: Manual data entry of products can range from fully open to fully closed. Under a fully open data entry, respondents type product names and these are classified later. Under a fully closed data entry, respondents can only choose from a list. There are various hybrid options. The current HBS lets respondents type products and then shows matches based on the Jarro-Winkler distance,
- Product lists: In a closed or semi-open manual data entry, product lists language and diversity determine the ease with which respondents can find products. It can be decided to include frequently occurring typos, even though they are grammatically incorrect. The quality and richness of lists varies greatly between countries.
- Shop lists: To aid receipt processing, a list of shops can be added. Shops have different receipt formats. Shop lists are not common in ESS countries.
- Discount/rebate: Products that have a discount or are rebated complicate data entry both manually and through scans. Choices how to present discount and rebate turned out influential. Currently, there is not yet a fully satisfactory way of doing this across different settings.

Data collection strategies:

- Interviewer assistance: Interviewers can be involved in recruiting respondent and keeping them motivated. Interviewers may be involved right from the start or only after nonresponse. In the ESSnet two strategies will be tried; one without interviewers and one

where interviewers visit household to recruit them. This will be used to construct effective strategies for the future.

- Recruitment material: Since app data collection requires downloading and registering an unknown app, recruitment material includes instructions, basic overview of screens, possibly a landing page and possibly a brochure explaining what data is collected and for what purpose. For the HBS recruitment without interviewers, recruitment material consists of a manual with overview of screens and a brochure displaying what kind of statistics respondents will get about their own expenditures. In-app per screen there is a brief tutorial on how to navigate.
- Multi-mode HBS: HBS can be a mix on app data collection, online data collection through a website, interviewer questionnaires and paper diaries. Modes/devices may be offered concurrently or sequentially. In the HBS field test, only the app itself will be used, but accessible on smartphones, tablets and laptops/desktops. In other field tests, more advanced multi-mode strategies are tested.
- Household reference person: HBS is a household survey. On one extreme one household member completes all questionnaires and diary and, thus, also provides info on expenditures by household members. On the other extreme, all members submit expenditures individually. An important design decision is whether household members can view each other's expenditures. In the current HBS app, all household members can submit expenditures but there is no anonymization. The data collection strategy will be to assign one contact or reference person who does most of the data entry and makes sure other members submit as well.
- Incentives: HBS is a burdensome survey and many countries include unconditional and/or conditional incentives. There is, however, great variety in the amount of incentive. In the HBS field test this is mostly left to countries. As a general incentive the respondents get in-app feedback. This feedback consists of two graphics, a circle chart and a time series chart. The feedback can be instant, delayed to the end of data collection or omitted completely.

Data quality checks:

- In-app scan quality: Scans of receipts may vary greatly in quality depending on light conditions, camera settings and type and respondent behaviour. Currently, the HBS app informs the respondent on scan quality while taking a picture. It also displays a bounding box that the respondent can adjust. Furthermore, it provides in-app OCR so that the respondent can decide whether the scan is good enough or needs to be retaken.
- Checks of amounts and process: Respondents may produce errors in number of products and or prices when entering data manually. Currently, there is no check on outliers, but the total amount is calculated and shown to the respondent. This can be seen as a soft check. Also in the individual statistics, outliers may be noticed.
- Plausibility of expenditures: Expenditures of households are unknown and may take any form from diverse and many to monotone and few. It is unlikely that a respondent will not buy food for many weeks, but not impossible. The current HBS, therefore, only asks respondents to validate each day. It asks whether all expenditures are included. Respondents can ignore these reminders and validate a day without expenses.
- Respondent validation: Respondents get classified receipts from the backend. When being online, the feedback is available within 10 seconds. If the respondent works offline, then processing is done at the first instance the respondent goes online. The classified receipts can be edited, but it is still an open question what should be editable and what not.

ML models:

- OCR and Natural Language Processing: See earlier descriptions. The HBS applies in-app OCR as a form of pre-scan of data quality and in-house OCR for processing. In time the in-house OCR may be included in the app itself. The ML model which is sizeable can be compressed.
- Classification: See earlier descriptions. Classification of products can be in-house and in-app as well depending on how the ML models are trained and. There are three options. The first is to train ML models on annotated receipts. This could be done through active and online learning, making sure that models are re-trained. The second is to use EAN/GTIN product descriptions as typically available in scanner data and to train models to classify to these. The descriptions are linked to COICOP by CPI departments on a regular basis. The third is where receipt texts are directly obtained from shops. The last option is ideal but leans heavily on cooperation of shops. The second option is used for NL and some other countries. Countries without scanner data or limited scanner data have to resort to the first option. In the second and third option, ML models have to be applied in-house as shops demand that these data are not exported.

2.4 Logistics level (preliminary)

The logistics level distinguishes:

- Recruitment: Many of the ESS surveys have an interviewers-assisted data collection such as doing the starting questionnaire, recruiting and assisting respondents in a diary, motivating respondents during data collection, and/or picking up closing questionnaires.
- Monitoring: Sample units may not participate, drop-out or deliver insufficient data quality. Monitoring dashboards may be inspected on a frequent basis in order to determine whether interventions are needed at overall level.
- Assistance: Respondents may be assisted in starting and using smart survey tools passively through a helpdesk and website or actively through interviewers and or technical experts.
- Human-in-the-loop machine learning: In sensor data applications, models seldom reach 100% accuracy. Certain population subgroups or certain survey statistics may require manual inspection. In ESSnet Smart Surveys where feedback of statistics to respondents is deemed important, such human-in-the-loop processes may even occur during data collection.

We give a first exposition of how the different logistics are implemented for the household budget survey application.

Recruitment:

- Interviewer involvement training: Most countries employ interviewers to recruit respondents. This interviewer role, implies that interviewers are themselves experienced app users and should be able to answer at least the most basic FAQ. It also means that they should periodically be instructed when the HBS app changes.
- Interviewer guidance: Some countries let interviewers keep track of respondents during data collection. The role of monitoring respondents implies that app data collection monitoring is accessible by interviewers through frequent reports or through direct access to the backoffice. It also implies they must be even more knowledgeable about how the

app works and what issues may occur. Currently, in the HBS there is no access of interviewers to the backoffice. For now it is planned that they get weekly updates.

Monitoring:

- Recruitment and activity: Monitoring of registration and submission of data are fairly standard. It means that on a continuous basis queries can be run. In the maintenance server this is indeed possible.
- Answer and navigation behaviour: Monitoring of in-app behaviour is not at all standard. This concerns consultation of certain app screens and help options, time lags when filling in expenditures, problems when scanning receipts, and so on. Currently, the HBS app has a paradata plug-in that logs all behaviour, but this needs to be made more sophisticated and targeted at certain behaviour.
- Technical errors: Respondents may experience problems that are specific to their own devices or to general system failure such as the backend server being inaccessible. These errors can partially be logged. Currently, type of device and operating system are derived and can be used in detecting problems.

Assistance:

- Survey helpdesk: Respondents may call the NIS (as usual) to ask questions about the purpose of the survey and about the task that they are supposed to (such as definitions etc). This type of assistance is not very different from the traditional HBS, except for user interface questions.
- Technical helpdesk: When respondents experience problems, then these are generally too complicated for the regular helpdesk. Currently, the respondent can send messages in-app to the NSI. Obviously, this only is available when the app is installed. The technical helpdesk email is also available in recruitment letters but cannot be contacted by phone. The survey helpdesk can forward questions.
- Landing page/FAQ: It is recommendable to have a dedicated webpage for the survey including explanations, instructions and FAQ. In previous experiments, it was experienced that respondents do not make much use of the webpage and tend to use in-app options or contact the helpdesk.

Human-in-the-loop ML:

- Inspection of classified receipts: It is yet an open decision to what extent OCR and classification of receipts contains a human check. Both the OCR and classification return indicators of accuracy. Based on lower thresholds to these indicators, it can be decided to flag processed receipts for inspection. This may be country-dependent as ML models for classification depend heavily on the quality of training and retraining of models. For now, it is foreseen that a manual check of performance is included until more experiences are available.

2.5 Legal-ethical-policy level (preliminary)

The legal-ethical-policy level is not discussed here. It is the topic of the ESSnet working group legal-ethical. We mention only that the CBS data protection impact assessment (DPIA) is currently supplemented for the HBS through a dedicated risk analysis. The DPIA is translated to English. The DPIA will be made available by the end of February 2021.

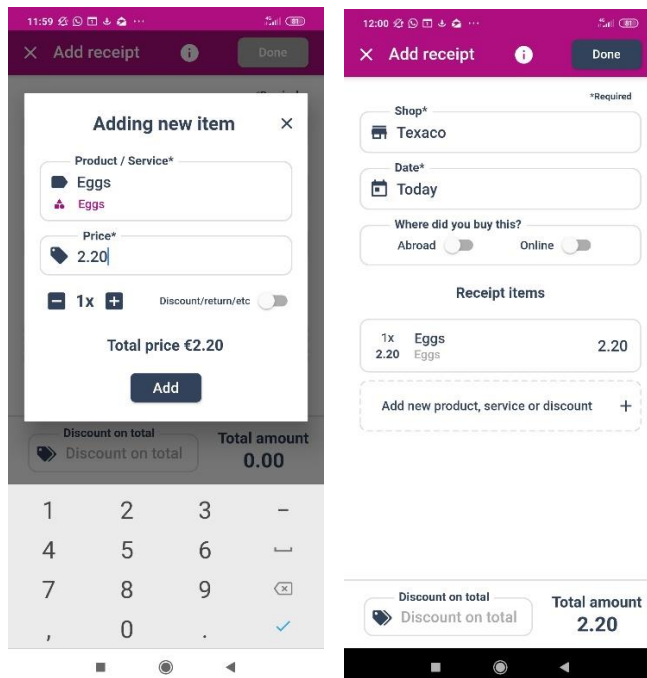
ANNEX A – EUROSTAT CSPA INVENTORY QUESTIONNAIRE
See attachment

ANNEX B – SCREENSHOTS OF MAIN APP SCREENS

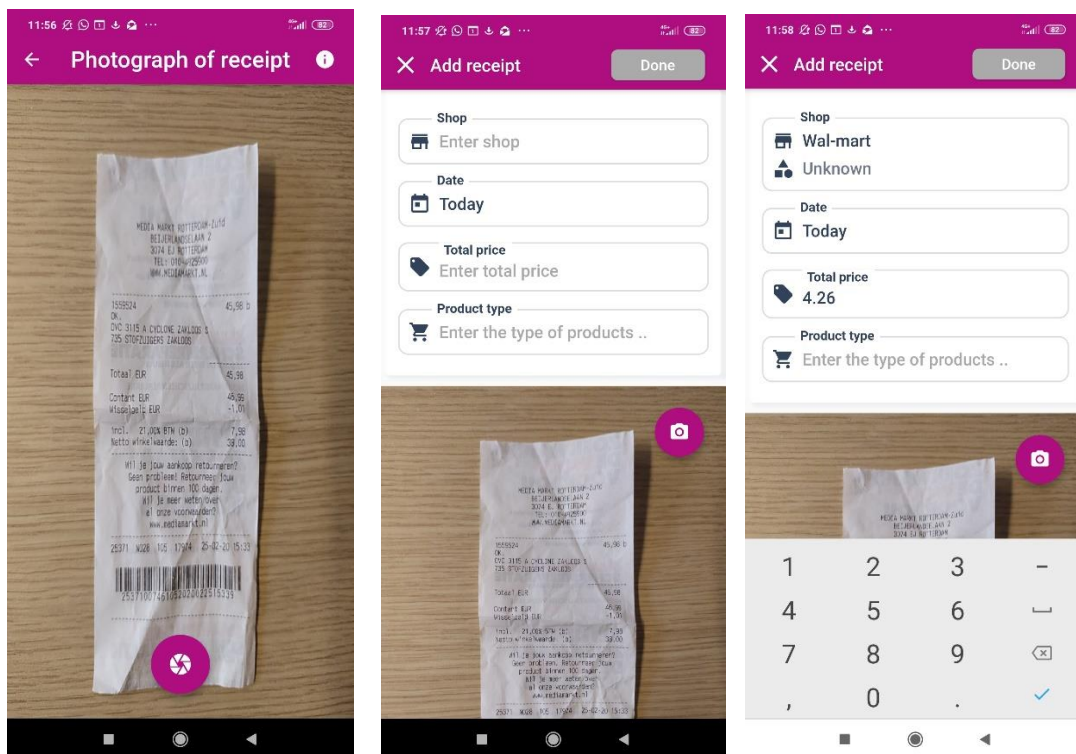
Calendar:



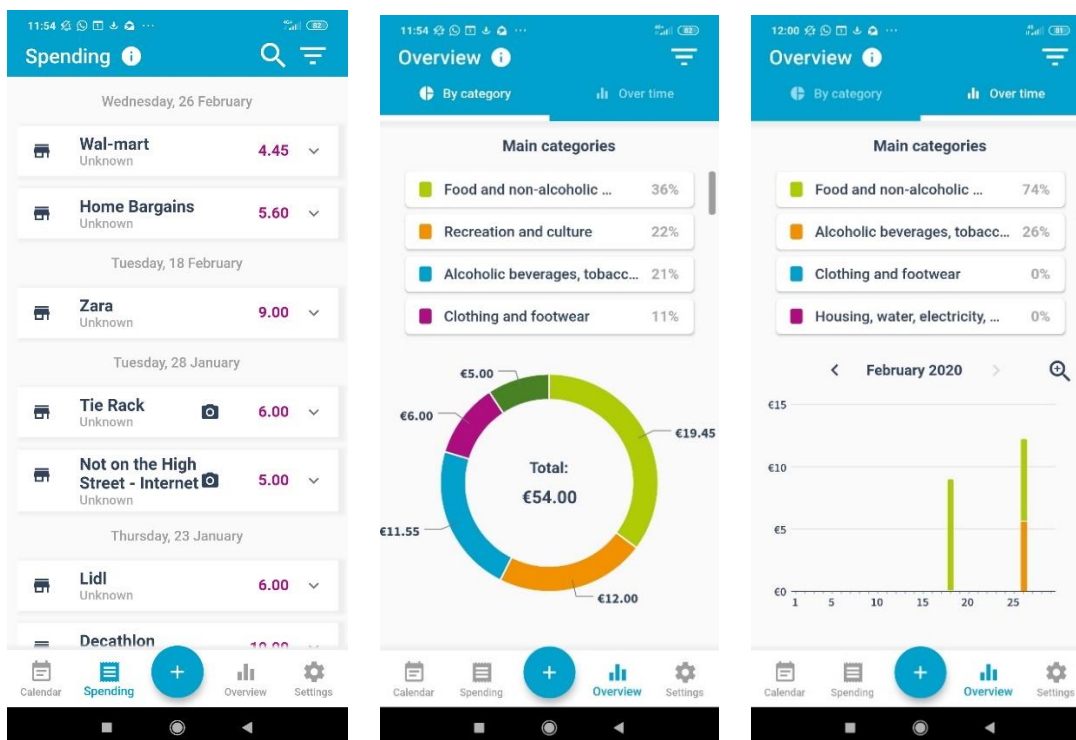
Manual data entry:



Scan data entry:

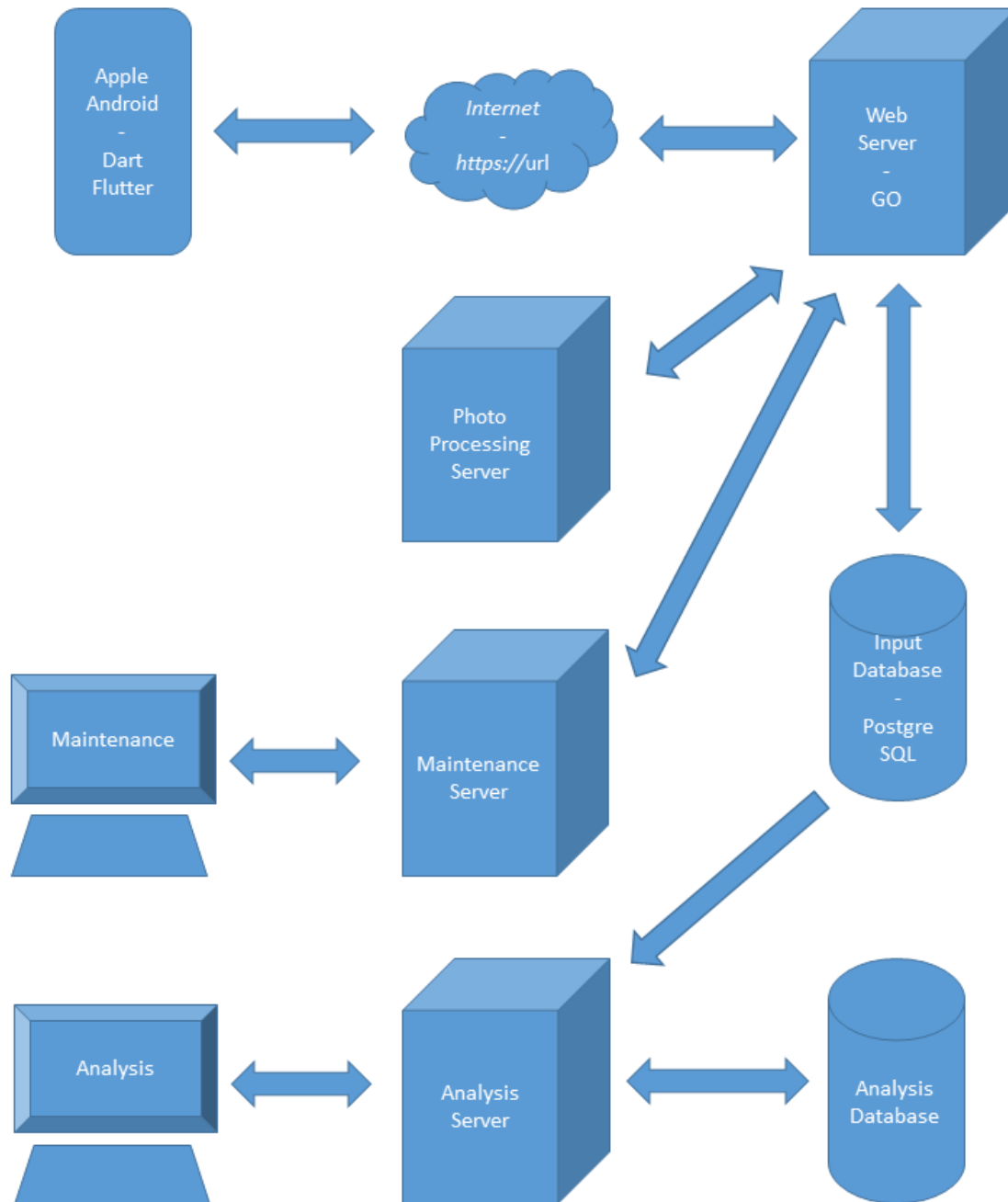


Expenditure overview and statistics overview:



ANNEX C – DETAILED DESCRIPTION OF BACKEND INPUT DATABASE

Household Budget Survey - Mobile App - System overview



System parts (implemented):

- Users enter/manage data on Apple or Android smartphones with HBS app. HBS app is developed in Dart/Flutter.
- Communication over internet via secure HTTPS URL.
- The Web Server monitors HTTPS URL and synchronizes data on smartphones with data in input database. Users can manage data on multiple smartphones. Data is timestamped on smartphone, data with latest timestamp is valid. The web server is developed in GO language.
- The Input Database stores and synchronizes data from smartphones. The database is also used for management of usernames and passwords (hashed). The input database is developed in PostgreSQL.
- The Photo Processing Server analyses photos of receipts. Optical character recognition is used to detect descriptions and prices of products. Descriptions in turn are linked to the Coicop classification. The resulting list of products is returned to the Web Server and Input Database and is synchronized with the phones of the user. At this moment the Photo Processing Server initiates the processing of receipts: it asks the Web Server for new photos to process. In the future the processing of receipts should be initiated by the Web Server: the Web Server asks the Photo Processing Server to process a new photo.
- The Maintenance and monitoring of user response, management of passwords is accessible through a web browser.
- The Maintenance Server is a web server developed in Angular. It provides an user interface for the management of groups of users (passwords) and the monitoring of the usage of the HBS app by the users.

System parts vision (not yet implemented):

- Analysis of data by statisticians.
- Analysis Server for analysis of data in analysis database, and data import from input database to analysis database.
- Analysis Database for final data storage and data analysis.

Functionality

The main functionality of the HBS backend is:

- User authentication: check whether the user password matches the (hashed) password in the database
- Store receipt data: the data gathered on the smartphone is also stored in the database.
- Synchronize receipt data: keep the data synchronized between smartphones if the user uses multiple smartphones.
- Classification of receipt photos: find list of products on receipt photo (descriptions, prices and Coicop categories).
- Managing HBS app users: create groups of users and passwords.
- Monitoring: respondent activity and app usage is monitored and may be input to follow-up actions

The HBS-App maintains the receipt data locally on a phone. When the HBS-App is started, or when a receipt is entered or changed the app tries to connect to the GO server to synchronize the locally stored data with the data in the database. The HBS-App does not need a continuous connection with the GO server, when there is a connection all data will be synchronized.

Status

The HBS backend is operational but is still being developed further. Several countries tested a part of the system on the CBS infrastructure:

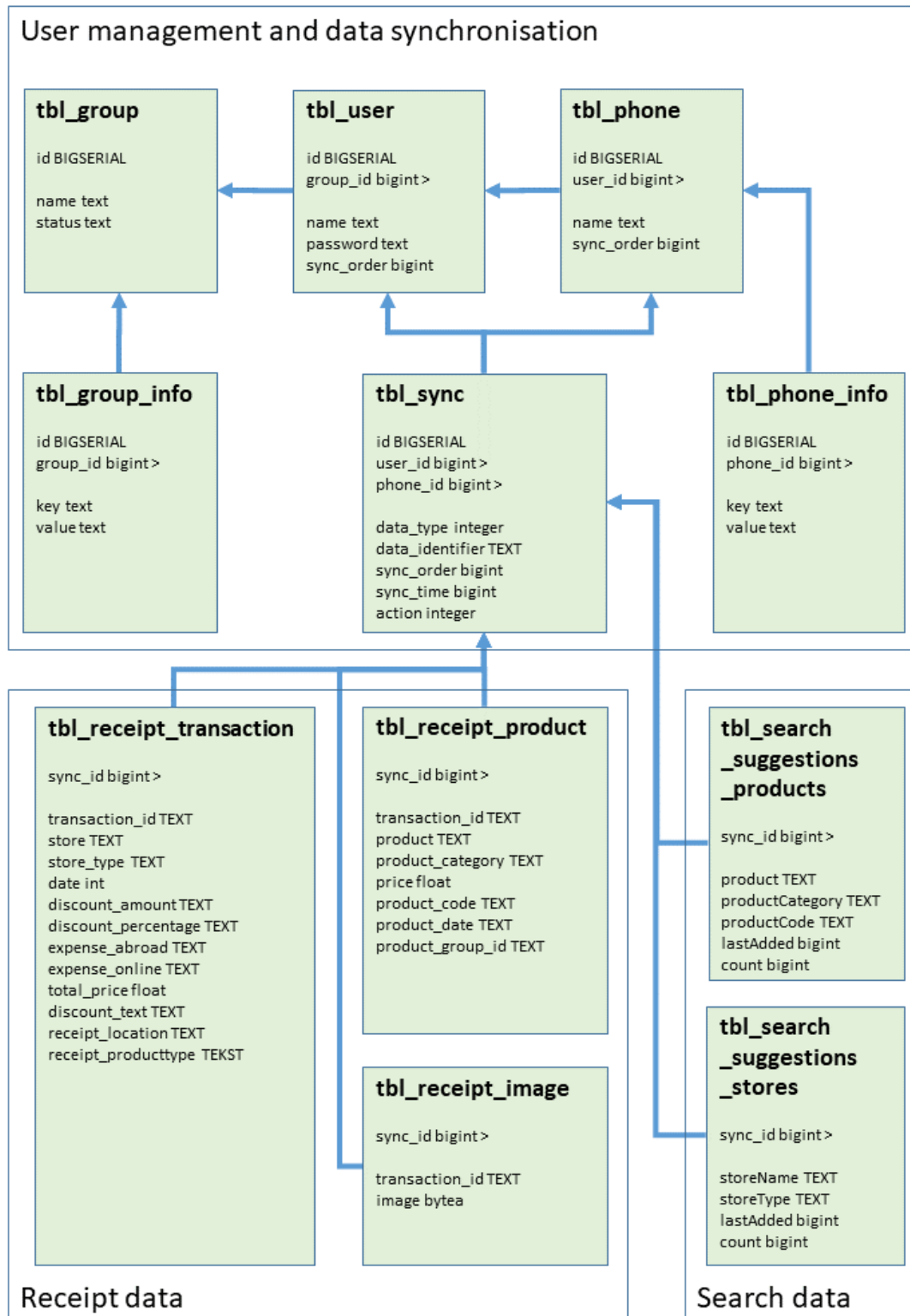
- HBS app
- HTTPS URL
- Web Server
- Input Database

At this moment we are preparing a large stress test for this part of the system. A small scale experiment showed that it can handle 100 receipt photos (250 KB per photo) in 15 seconds. This is without the Photo Processing functionality.

On a small scale we are also testing:

- Photo Processing Server (+/- 10 seconds processing time per photo)
- Maintenance Web Server (basic functionality)

Household Budget Survey - Data Model



PostgreSQL Input Database

The Input Database is implemented in PostgreSQL. It contains 11 tables (see data model above):

- **tbl_group**: survey group
- **tbl_group_info**: properties of survey group (controls functionality of HBS app)

- `tbl_user`: users in survey groups (hashed passwords)
- `tbl_phone`: phones of users (a user can have multiple phones)
- `tbl_phone_info`: properties of phones (information gathered on phones)
- `tbl_sync`: data synchronization mechanism
- `tbl_receipt_transaction`: receipt transaction (synchronized data)
- `tbl_receipts_product`: receipt products (synchronized data)
- `tbl_receipt_image`: receipt photo (synchronized data)
- `tbl_search_suggestions_products`: searched products (synchronized data)
- `tbl_search_suggestions_stores`: searched stores (synchronized data)

There is no further functionality in the database. The contents of these tables is managed by the Web Server.

The required size of the database depends on the number of users, the length of the survey period and the number of receipts a user enters per day. Per receipt the required database size is: +/- 5 MB, this is roughly the size of a photo + 1KB of textual data. Example: If 100 users take 2 photos per day for a period of one month, the required database size is: $100 * 2 * 5 \text{ MB} * 30 = 30.000 \text{ MB} = 30 \text{ GB}$. After a receipt photo is processed it could be removed from the database thus freeing up a lot of space (not implemented yet).

GO-Language Web Server

The Web Server is implemented in GO-Language. The file structure of the code is:

backend

- database
 - `init.go` initialize database connection
 - `tbl_<...>.go` insert, delete, update, select per table
 - `data_<...>.go` data synchronization mechanism
 - `dashboard_<...>.go` monitor user activity
 - `create_tables.sql` sql script for database creation
- global
 - `constants.go` global constants
 - `functions.go` global functions
 - `dashboard.go` types for dashboard
 - `data_<...>.go` types for data synchronization
 - `type_<...>.go` types other
- restapi
 - `test.go` test if web server is up and running
 - `group.go` manage groups
 - `users.go` manage users
 - `phone.go` register phones of users
 - `process.go` photo processing
 - `general.go` general methods
 - `data_<...>.go` data synchronization
 - `dashboard_<...>.go` monitor user activity
 - `manual.txt` API manual

- backend.go main entry point https url

The API of the Web Server can be used to manage groups and users, to monitor user activities, to synchronize data over multiple phones and to access receipt images for photo processing.

The Web Server is configured via environment settings:

- PORT http port
- DB_HOST database host
- DB_NAME database name
- DB_PORT database port
- DB_USER database user
- DB_PASSWORD database password

Angular Maintenance Server

The Maintenance Server is implemented in Angular. It provides a user friendly interface to the API of the Web Server. At the moment we are testing a first basic version of this Maintenance Server, it can:

- manage groups and users
- generate a set of usernames and passwords for a specific survey group
- monitor survey progress (number of received receipts per user)
- show phone types used by users
- show receipt photos of individual user

The Maintenance Server is still under development. We also want to use it to monitor the usage of the HBS app (dwell times on pages, how often does a button get clicked), but for this purpose the HBS app has to be adapted (not implemented yet) to gather the required information.

Future steps

The not yet implemented part of the system overview is:

- Data analysis by substantive researchers and methodology
- Analysis server for analysis of data in analysis database, and data import from input database in analysis database.
- Analysis database for final data storage and data analysis.

While this functionality is not available the input database can be managed/queried directly with the pgAdmin_4 tool for PostgreSQL databases.

The HBS backend as it is, is equipped to authenticate users, store receipt data in an input database, manage groups of users, monitor survey progress and process receipt images. All parts of the HBS backend system will require further development/maintenance. Most interesting will be the Photo Processing Service. We are just starting tests to find out how accurate it can classify a photographed receipt (when the photographer is not holding a thumb over the receipt while taking a photo).



Innovative Tools for Time Use Survey or Household Budget Survey

Fields marked with * are mandatory.

INTRODUCTION

Dear colleague,

Welcome to the survey on 'Innovative Tools for HBS and TUS'.

This survey has been developed by the Task Forces on Innovative Tools and Sources for Time Use Survey (TUS) and Household Budget Survey (HBS).

It aims in first instance at collecting information on the existence, expertise and usage of tools supporting the modernisation of data collection for completing the Time Use Survey and/or the Household Budget Survey.

Secondly, it aims at maintaining up-to-date the information on existing tools which have been already identified and listed in the Inventory of Innovative Tools and Sources for HBS and TUS.

This inventory has been designed and is maintained by the two Task Forces. All information collected via this survey will be valuable input for the inventory. It can refer to tools either in development or already in use.

The inventory provides an overview on ESS level and will be used as a basis for the future modernisation of data collection in TUS and HBS.

The answers will be regarded as coming from your organisation. Therefore, please report the type of organisation (e.g. NSI, University, Ministry, private company,...) you represent. No reference will be made to the actual respondents.

To support the understanding of terminology used in the questionnaire the TF developed a Glossary with a list of concepts. The Glossary file is accessible in the questionnaire under the header "Background Documents", click on the link "Glossary - List of concepts".

In the name of the Task Force members I would like to thank you in advance for your cooperation!

Hubertus Cloudt

1 AN OVERVIEW OF YOUR ORGANISATION

We would like to know more about your organization and YOU - the person responsible for THIS questionnaire within your organisation

(even though more than one person will contribute to the answers).

* 1.1 Full name

Barry Schouten

* 1.2 Function

Senior methodologist and program coordinator data collection innovation

* 1.3 Email address

jg.schouten@cbs.nl

* 1.4 Organisation

Official name of your organisation and English translation (if applies)

Centraal Bureau voor de Statistiek (CBS)/ Statistics Netherlands

1.5 Department

Research and Development

1.6 Unit

Methodology

* 1.7 Country

The Netherlands

1.8 Head of department/unit responsible for the survey (HBS/TUS)

Name of the person

Jos Schiepers (Social Statistics Division)

1.9 Which of the surveys is the main focus of the tool?



Household Budget Survey



Time Use Survey



Both

* 1.10 Has your organisation ever organised a Household Budget Survey in the past?



Yes



No

*** 1.11 How many Household Budget Surveys has your organisation organised?**

Since 2015 once every 5 years, before 2015 every
year

*** 1.12 Is your organisation currently involved in a HBS data collection, or is it planning to carry out one in the next years?**

Choose one answer

- ☐ No, we are neither involved nor planning
- ☐ Yes, we are currently involved in the data collection
- ☒ Yes, we are planning the data collection in 2020
- ☐ Yes, we are planning the data collection in 2021
- ☐ Yes, we are planning the data collection in 2022
- ☐ Yes, we are planning the data collection in 2023
- ☐ Yes, we are planning the data collection in 2024
- ☒ Yes, we are planning the data collection in 2025
- ☐ Yes, we are planning the data collection in 2026-2030
- ☐ Don't know
- ☐ No answer

*** 1.13 Has your organisation ever organised a Time Use Survey in the past?**

- ☒ Yes
- ☐ No

*** 1.14 How many Time Use Surveys has your organisation organised in the past?**

2011, 2016

*** 1.15 Is your organisation currently involved in a TUS data collection, or is it planning to carry out one in the next years?**

Choose one answer

- ☐ No, we are neither involved nor planning
- ☐ Yes, we are currently involved in the data collection
- ☐ Yes, we are planning the data collection in 2020
- ☒ Yes, we are planning the data collection in 2021
- ☐ Yes, we are planning the data collection in 2022
- ☐ Yes, we are planning the data collection in 2023
- ☐ Yes, we are planning the data collection in 2024
- ☐ Yes, we are planning the data collection in 2025
- ☐ Yes, we are planning the data collection in 2026-2030
- ☐ Don't know
- ☐ No answer

2 IDENTIFICATION OF THE TOOL

This part of the questionnaire deals with tools that your organisation is already using or developing on its own, or in partnership with another organisation. If you are using a ready-made purchase tool and you are not able to answer to some of the questions, please let us know who developed the tool and they will be invited to fill out the questionnaire. If the tool is in Conceptual/Design phase you might opt for a shorter version of this questionnaire focusing on early stage projects. Let us know if you wish so.

By **tool** we mean any software platform that combines both front-end and back-end applications and their functionalities.

2.1 Name, ownership, development phase

* 2.1.1 Is your organisation using or developing (on its own or in partnership with another entity) a tool able to collect household budget/time use data online?

- ☒ We are using our own tool (we are service owners of the tool)
- ☒ We are developing our own tool (we are in-house developers of the tool)
- ☐ We are using a tool developed by - or in partnership with - others (public organisation, government, university, private company, ...)
- ☐ We are developing a tool in partnership with others (public organisation, government, university, private company, ...)
- ☐ Other (please specify below)

* 2.1.2 If Other – please specify

* 2.1.3 What is the name of this tool?

Household Budget Survey

* 2.1.4 What is the aim of this tool?

What was the original business need that triggered the tool's development? What is the reason for using the tool?

Primary data collection with Apps, increasing response, lowering burden

* 2.1.5 Can you provide a short description of the main function(s) of this tool?

Primary data collection on mobile devices for the household budget survey

* 2.1.6 Which part(s) of the HBS are addressed by this tool?

Diary (whole or parts of it) and/or Questionnaire (whole or parts of it) A

- ☐ part of the Diary
- ☒ The whole Diary
- ☒ A part of the Questionnaire
- ☐ The whole Questionnaire

* 2.1.7 Briefly describe the outcomes of the tool

- ☐ Data collection
- ☐ Exchange of data - API
- ☐ Statistical production
- ☐ Research
- ☐ Visualisation
- ☐ Other - please specify below
- ☐ Don't know
- ☐ No answer

*** 2.1.8 If Other – please specify**

*** 2.1.9 Are there any dependencies with other tools and sources?**

e.g. Matching and Data Linking Service

- ☐ Yes, the tool receives data from other existing sources
- ☐ Yes, the tool depends on the results of other tools
- ☐ Yes, the results of the tool are used by another tool
- ☐ Yes, results of the tool create a new data source used for further processing
- ☐ Other - please specify below
- ☐ Don't know
- ☐ No answer

*** 2.1.10 If the options "...receives data from other existing sources" or "Other" are selected please provide a brief explanation**

*** 2.1.11 Is there any documentation available about this tool?**

- ☐ Yes, online guidelines
- ☐ Yes, online articles
- ☐ Yes, in the app store
- ☐ Yes, other documentation
- ☐ No
- ☐ Don't know
- ☐ No answer

2.1.12 If Yes please provide the link(s) below. links to online guidelines, and/or online articles, and/or to the app stores, etc.

<https://play.google.com/store/apps/details?id=nl.cbs.householdbudgetsurvey>

2.1.13 If it is possible to share the documentation, would you please upload the file

The maximum file size is 1 MB

*** 2.1.14 Please indicate the stage of development of this tool.**

- ☐ Development phase
- ☒ Test phase
- ☒ Pilot phase
- ☐ Data collection phase - release maintenance
- ☐ Don't know
- ☐ No answer

*** 2.1.15 Who has the Intellectual Property ownership of the tool?**

Intellectual property refers to creations of the mind and is divided into two categories: Industrial Property (includes patents for inventions, trademarks, industrial designs and geographical indications) and Copyright for artistic work.

Statistics Netherlands (CBS)

2.1.16 Is this tool patented or protected by other property protection rights (if applicable)?

Patent is the exclusive right granted by a government to an inventor to manufacture, use, or sell an invention for a certain number of years.

Other protection rights: for example i-depot in Benelux

Open Source, no copyright

*** 2.1.17 Which of the following elements of the data collection design are part of the tool?**

at least 1 choice(s)

Automated communication: providing automated feedback, instructions and alerts to the respondents in a form of SMS, e-mail etc.

Fully prepared database: download of a database ready for statistical analysis (in .xlsx, .csv, .sav, ... formats)

Online calibration procedure: an online module in order to define weights based on defined parameters (population numbers on age, gender, education; numbers of days completed, dispersion of the year, ...)

Online data analysis: a statistical software package that makes it possible to analyse the data from within the tool (e.g. R)

- ☐ Online questionnaire
 - ☐ Online diary
 - ☒ Smartphone diary app with online or offline data collection
 - ☐ Automatic communication
 - ☐ Online invitation procedure
 - ☐ Automatic data collection flow
 - ☐ Online follow-up/overview of fieldwork
 - ☐ Fully prepared database
 - ☐ Online calibration procedure
 - ☐ Complete metadata information
 - ☐ Online data analysis
 - ☐ Don't know
 - ☐ No answer
-

2.2 Parameters in an online time use diary

NOT APPLICABLE

2.3 Parameters in an online household budget diary

*** 2.3.1 Which of the parameters listed below are included in the tool in relation to the online household budget diary? These parameters can be changed/modified depending on the data collection goals.**

at least 1 choice(s)

- ☒ Length of the fieldwork period (e.g. a year, a month ...)
- ☐ Number of consecutive diary days (e.g. 1 week, 2 weeks, 3 weeks ...)
- ☐ The registration method (e.g. Retrospective, Continuous, Self-administered, Interview, ...)
- ☒ Input method activities (e.g. Own wordings, category selection, suggestion via external sources ...)
- ☒ Definition of an expenditure list (e.g. expenditure determination, level structure ...)
- ☐ Definition of item specific questions (e.g. please specify unit of measure: kg, g, litres...?) Validation
- ☒ check on input (e.g. reporting in the future, no quantity, no price, no place/country of purchase...)
- ☐ Start of the diary - date and hour (e.g. specific date; at entry, at midnight,...)
- ☐ Learning period – period before the actual registration (e.g. 5 hours, 2 hours ...)
- ☒ Quality check during diary
- ☐ Quality checks at the end of the diary
- ☐ Overview and correction/completion modus for respondent
- ☐ Don't know
- ☐ No answer
- ☐

*** 2.3.2 What registration method(s) are included in the online household budget diary? If you choose multiple methods, it means that a combination of these methods can be provided to the respondent. *at least 1 choice(s)***

- ☐ Retrospective
- ☒ Continuous
- ☐ Other – please specify below
- ☐ Don't know
- ☐ No answer

*** 2.3.3 If Other – please specify**

*** 2.3.4 Via which (combination of) method(s) can the respondent fill in the activity in the online diary?**

at least 1 choice(s)

Multiple answers are possible **Input**

- ☐ through typing out
- ☐ Input via speech recognition
- ☐ Input based on selection from a pre-defined list
- ☐ Input based on a search via key words to alter a pre-defined list
- ☐ Input through the use of other data sources (scanning of receipts, use of bank account logs) Other –
- ☐ please specify below
- ☐ Don't know
- ☐ No answer

* 2.3.5 If Other – please specify

3 NON-TECHNICAL FEATURES OF THE TOOL

In this Section of the questionnaire we will cover the following topics:

Business goal(s) of the tool; Validation; Accessibility and Usability; Assistance and Feedback to the respondent; Fieldwork monitoring.

3.1 Business goals

* 3.1.1 What is the focus of the tool? *at least 1 choice(s)*

- ☐ General population data collection
- ☐ Integration broader statistical network
- ☐ Government policy research (unpaid work, gender equality, transportation, leisure, sport ...) Multi-disciplinary data collection/research
- ☐ Target specific data collection
- ☐ Experimental data collection – Test environment
- ☐ Other – please specify below
- ☐ Don't know
- ☐ No answer

* 3.1.2 If Other – please specify

* 3.1.3 What are the business goals of the tool? *at least 1 choice(s)*

Scalability is the capability of a system, network, or process to handle a growing amount of work, or its potential to be enlarged to accommodate that growth.

- ☐ In-house data collection
- ☐ Scalability – governance tool
- ☐ Product to others
- ☐ Service to others (Software as a Service - SaaS)
- ☐ Don't know
- ☐ No answer

*** 3.1.4 Who are the stakeholders? at least 1 choice(s)**

A **stakeholder** is an organization or a person with an (in) direct (economic, policy, research, etc.) benefit to the output of a business process/function.

- ☐ My own organisation
- ☐ (Other) NSIs
- ☐ Other governmental structures (international, regional, city, community level)
- ☐ NGOs
- ☐ Academic
- ☐ Commercial
- ☐ Other – please specify below
- ☐ Don't know
- ☐ No answer

*** 3.1.5 If Other – please specify**

3.2 Validation

*** 3.2.1 What validation checks are covered in the tool? at least 1 choice(s)**

Data validation is the process of monitoring the results of data compilation and ensuring the quality of the statistical results. Data validation specifies methods and processes for assessing statistical data, and how the results of the assessments are monitored and made available to improve statistical processes.

Administrative validation: a check to ensure that general profile, contact information, etc. has been provided.

Input validation - questionnaire: a check to validate that user input and responses are in right format (e.g. numbers, dates, URLs, etc.).

Input validation - diary: a check that validates whether a diary is filled in.

Process validation: a critical part of quality assurance procedures to confirm whether the process is effectively controlling the quality of the data collected.

Database validation: automatic check to ensure that the database structures are not corrupt and the data entered is sensible and feasible.

Security validation: decreasing the likelihood of fraud, e.g. CAPTCHA, SMS verification, requiring login, etc.

- ☐ Administrative validation (profile, contact information, ...)
- ☐ Input validation – questionnaire
- ☐ Input validation – diary
- ☐ Process validation
- ☐ Database validation
- ☐ Security validation
- ☐ Don't know
- ☐ No answer

*** 3.2.2 How does the tool take into account validity aspects?** *at least 1 choice(s)*

- ☐ There is no quality control (e.g validation rules management, execution of validation rules, reports on processed data, ...)
- ☐ Via a dashboard (response rates, period overview, state overview, validation reports, validation rules editor, ...)
- ☐ Via paradata (start date, end date, registration time, device, ...)
- ☐ Via quality parameters (TUS: nr. of registered activities, sleep, eat, travel, unspecified time, ...; HBS: nr. of registered expenses, durable goods, food, leisure...)
- ☐ Via validation procedures automatically executed during the time-use/expense registration Other –
- ☐ please specify below
- ☐ Don't know
- ☐ No answer

*** 3.2.3 If Other – please specify**

3.3 Accessibility and Usability

Front-end (or front-office) is an User Interface or respondent interface that facilitates the respondent to complete a survey or diary (or whatever task). It provides functionality (business logic – CRM) and data necessary to complete the demanded tasks from the respondent.

Business logic is the programming that manages communication between an end user interface and a database. The main components of business logic are business rules and workflows. A business rule describes a specific procedure; a workflow consists of the tasks, procedural steps, required input and output information, and tools needed for each step of that procedure. Business logic describes the sequence of operations associated with data in a database to carry out the business rule.

Back-end (or back-office) is a data collector/researcher interface that facilitates the data collector/researcher to build a data collection/research or fieldwork. The back-end is an evolving computer system that not only designs the data collection/research. It also includes decision models on how the fieldwork is organized and administered. Data collected through the front-end are stored in databases of which the criteria are defined through the back-end. The back-end can also be able to communicate with other devices and sources. This way, data coming from the respondent can be fused with data captured via connected devices or sensors (also called Internet of Things). To do this an Application Programming Interface (API) needs to be defined.

Usability is the extent to which the tool can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

*** 3.3.1 Which functionalities does the front-end provide?**

- ☐ Information pages about the data collection
- ☐ Task overview (e.g. monitor the fieldwork of one or more surveys - according to user permissions - or showing a progress bar of the data collection period, or other monitoring aspects)
- ☐ Language selection
- ☐ Instructions to respondents
- ☐ Business logic to complete data collection
- ☐ Responsive design (cross device and browser usage)
- ☐ Usable for people with disabilities (e.g screen readers)
- ☐ Mode switching (e.g. mix-mode of interfaces used for data entry, data capturing: partly via web application online, partly via App on smartphone or tablet,..)
- ☐ Manage respondent reminders
- ☐ Other functionalities - please specify below
- ☐ Don't know
- ☐ No answer

*** 3.3.2 If Other – please specify**

Providing feedback to respondents on their own expenditure statistics (can be instant or delayed depending on protocol)

*** 3.3.3 Is it a multipurpose tool (it can be used by different users for different data collection/research purposes)?**

- ☒ Yes
- ☐ No
- ☐ Don't know
- ☐ No answer

3.3.4 On a scale from 1 to 5, how do you rate its effectiveness...

1 = "very poor"; 2 = "poor"; 3 = "moderate"; 4 = "good" ; 5 = "very good"

	1	2	3	4	5	Don't know	No answer
* ...in the defined phases of the collection?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
* ...for the underlying business logic of the collection flow?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

3. 3.5 On a scale from 1 to 5, how supportive is the tool in passing from ...

1 = "not at all supportive"; 2 = "not enough supportive"; 3 = "somehow supportive"; 4 = "supportive enough" ; 5 = "very supportive"

	1	2	3	4	5	Don't know	No answer
* ...the definition of data needs to the setup of the tool?	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
* ...the collection of data to their availability?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
* ...the availability of data to their valorisation?	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

3.4 Assistance and Feedback to the respondent

* 3.4.1 What modes of assistance are foreseen to explain the use of this tool?

at least 1 choice(s)

- ☐ Download documentation/instructions website
- ☒ In app instructions
- ☒ Instructions provided in the app store, or app site
- ☐ Instruction video
- ☒ Dedicated website/page
- ☐ Real time helpdesk – chat function
- ☐ Support team
- ☒ FAQ
- ☐ Chatbot (chatting with a robot)
- ☐ Other – please specify below
- ☐ Don't know
- ☐ No answer

* 3.4.2 If Other – please specify

* 3.4.3 Is it possible to give feedback to the respondent via this tool?

at least 1 choice(s)

For example: overview answers, time expenditure, graphics...

- ☐ No
- ☒ Yes, during the completion of the data collection
- ☐ Yes, after the completion of the data collection
- ☐ Yes, after the fieldwork
- ☐ Other – please specify below
- ☐ Don't know
- ☐ No answer

* 3.4.4 If Other – please specify

3.5 Fieldwork monitoring module

This module (feature) gives the possibility to monitor or control how the online tool supports the fieldwork organiser – the statistical institute. It is a kind of surveillance system to monitor whether a given respondent opened and/or filled in the questionnaire and/or diary.

3.5.1 Does the tool include a fieldwork monitoring dashboard either as built-in or plug-in module?

- ☐ Yes, as a built-in module
- ☒ Yes, as a plug-in module
- ☐ No, but this feature will be developed in the future
- ☐ No
- ☐ Don't know
- ☐ No answer

3.5.2 Does the tool allow a connection to the selected sample: monitoring of respondent's actions like opening / neglecting, filling, finishing of the questionnaire during the data collection period?

- ☒ Yes
- ☐ No
- ☐ Don't know
- ☐ No answer

3.5.3 What kind of information is possible to collect via the module?

Registration/in-app answering behavior/type device and OS/technical errors/linkage to sampling frame data

3.5.4 Does the tool allow a connection to cost calculation?

- ☐ Yes
- ☒ No
- ☐ Don't know
- ☐ No answer

4 OTHER DATA SOURCES CONNECTED TO THE TOOL

In this context **sources** are organized streams of data inflow, via a multifaceted approach. These streams can be active or passive.

* 4.1 Is the tool able to request and receive data from other data sources (internet data, scanner data, log files, administrative data, etc.)?

at least 1 choice(s)

- ☐ Yes, electronic data via file upload and using predefined fields
- ☐ Yes, electronic data via Application Programming Interface (API)
- ☐ Yes, other - please specify below
- ☒ Not yet, but this feature is in development
- ☐ No, and not in development
- ☐ Other – please specify below
- ☐ Don't know
- ☐ No answer

* 4.2 If Other – please specify

*** 4.3 What sources can be connected to the tool to obtain external data?** *at least 1 choice(s)*

Data capture is the process by which collected data are put in a machine-readable form. Elementary edit checks are often performed in sub-modules of the software that does data capture.

Scanner data are detailed data on sales of consumer goods obtained by 'scanning' the bar codes for individual products at electronic points of sale in retail outlets. The data can provide detailed information about quantities, characteristics and values of goods sold as well as their prices.

- ☐ Administrative data
- ☐ Proxy registration (e.g. parent brings children to school, when the activity is registered it is also shown in children's diary)
- ☐ Smartphone (GPS, accelerometer, gyroscope, Near Field Communication (NFC), Bluetooth, Noise, Camera, Heart Rate, Blood Pressure ...)
- ☐ External GPS
- ☐ Wearable – activity tracker
- ☐ External sensors (CO2, Temperature, Velocity, ...)
- ☐ Loyalty cards data
- ☐ Cash register/scanner data
- ☐ Credit/debit cards data
- ☐ Other – please specify below
- ☐ None
- ☐ Don't know
- ☐ No answer

*** 4.4 If Other – please specify**

Receipts scanned by respondents are processed and classified through machine learning models and then returned. Linkage to bank transactions data is under research

*** 4.5 What sort of information is captured?** *at least 1 choice(s) Multiple answers are possible*

- ☐ Profile data (e.g. name, gender)
- ☐ Survey data (e.g. professional status)
- ☐ Activity data (e.g. sleeping) or Expenditure data
- ☐ Context data (e.g. location, with whom?)
- ☐ Don't know
- ☐ No answer

*** 4.6 Where is the data coming from these devices and sensors stored?**

A **Data source** is a location or service from where data or metadata can be obtained.

- ☐ Externally – (a copy of) the data source is provided afterwards
- ☐ Externally – via an API-key the data source can be consulted
- ☐ Externally – via an API-key the data is stored on a proprietary server
- ☐ Internally – the data is collected on a proprietary server
- ☐ Other – please specify below
- ☐ Don't know
- ☐ No answer

*** 4.7 If Other – please specify**

*** 4.8 Is the data collected through external sources used to provide suggestions or ask additional questions to the respondent?**

- ☒ No
- ☐ Yes, based on input from connected devices and sensors and a developed algorithm the respondent receives suggestions on their past activities (e.g. based on frequently visited locations "WORK" as an activity is suggested – “Are you working?”; “Did you stop working?”)
- ☐ Yes, based on input from connected devices and sensors extra questions are asked to the respondent (e. g. someone is in a shopping centre and based on this information extra questions are asked – “are you shopping?” or reminders - "please do not forget to register the purchases in the diary") Other –
- ☐ please specify below
- ☐ Don't know
- ☐ No answer

*** 4.9 If Other – please specify**

Scanned receipts are submitted to the backend where they are processed and classified. Respondents get a copy of the results which they can edit

*** 4.10 How are these data used? at least 1 choice(s)**

Multiple answers are possible

- ☒ Paradata (data are stored as extra variables)
- ☐ Direct input (the data are automatically used as input in the survey)
- ☐ Controlled input (the respondent validates the input first)
- ☐ As input for decision models (an algorithm interprets the data before input is presented to the respondent)
- ☐ Don't know
- ☐ No answer

*** 4.11 Are there any databases, wearables, sensors or connected devices linked to the tool so far?**

- ☒ Yes
- ☐ No
- ☐ Don't know
- ☐ No answer

4.12 Which databases, wearables, sensors or connected devices are linked so far?

Input database accessible through a REST-API that connects to receipt processing server, case management/maintenance server and analysis server

* 4.13 Is the tool able to deliver and provide data to another data source? *at least 1 choice(s)*

- ☐ Yes, electronic data via file upload and using predefined fields
- ☐ Yes, electronic data via Application Programming Interface - API
- ☐ Yes, other - please specify below
- ☐ Not yet, but this feature is in development
- ☐ No, and not in development
- ☐ Other – please specify below
- ☐ Don't know
- ☐ No answer

* 4.14 If Other – please specify

5 TECHNICAL FEATURES OF THE TOOL

This part deals with the technical features of the **tool**. In order to fill it in you might need the assistance of a colleague from the development team.

5.1 Version and last update

5.1.1 Can you indicate the version of the development of the tool?

Format vX.Y.Z where X is the major version, Y is the minor version, and Z is the patch version, eg. v2.3.0

2.0.6

5.1.2 Can you indicate when the tool was last updated?

Please indicate the last update, even if minor.

November 16, 2020

* 5.1.3 Does the tool have any connected or built-in modules? *at least 1 choice(s)*

- ☐ Yes, modules based on open source software
- ☐ Yes, it makes use of external software modules (e.g. Google, Amazon ...)
- ☐ Yes, modules developed inside the organization
- ☐ No
- ☐ Don't know
- ☐ No answer

* 5.1.4 What functions of the tool are covered by the built-in or connected modules? e.g. file format conversion, validation, etc.

Modules are listed as dependencies in: <https://gitlab.com/tabi/projects/budget/-/raw/master/app/pubspec.yaml>

5.2 Software architecture

* 5.2.1 Where is the functionality of the tool performed?

- ☒ Server accessed by the application
- ☒ Within the application itself
- ☐ Don't know
- ☐ No answer

* 5.2.2 Which type of application is this tool?

at least 1 choice(s)

A **web application** is a software application that runs on a remote server. It can be reached via a web browser of a computer, tablet or smartphone.

A **mobile application** is installed from an app store on a tablet, smartphone or watch. A mobile application can be **native** or **hybrid**. A **desktop application** is an application that runs stand-alone in a desktop or laptop computer.

- ☐ Web application
- ☐ Mobile native application
- ☒ Mobile hybrid application
- ☐ Desktop application
- ☐ Don't know
- ☐ No answer

* 5.2.3 Does the tool have a management website (Content Management System - CMS) to program data collection (e.g. prepare the collection tools, execute the fieldwork, define the database, etc.)?

A **Content Management System** is a computer application that supports the creation and modification of digital content.

- ☒ Yes
- ☐ No
- ☐ Don't know
- ☐ No answer

5.2.4 For the web application: which programming language and framework(s) are used?

	Programming language	Framework(s)
* Front-end		
* Back-end		

5.2.5 For the mobile native application: which programming language and framework(s) are used?

	Programming language	Framework(s)
* Front-end		
* Back-end		

5.2.6 For the mobile hybrid application: which programming language and which framework(s) are used?

Front-end Dart Flutter
Back-end Go

	Programming language	Framework(s)
* Front-end		
* Back-end		

5.2.7 For the desktop application: which programming language and framework(s) are used?

	Programming language	Framework(s)
* Front-end		
* Back-end		

5.2.8 For the management website: which programming language and framework(s) are used?

	Programming language	Framework(s)
* Front-end		
* Back-end		

*** 5.2.9 For which operating systems is the desktop application functional? *at least 1 choice(s)***

Multiple answers are possible

- ☐ Linux
- ☐ iOS
- ☐ Windows
- ☐ Other OS – please specify below
- ☐ Don't know
- ☐ No answer

*** 5.2.10 If Other – please specify**

5.2.11 Considering the data storage organization, can you provide information about what database management system is used to design the database.

PGAdmin4 as well as scripts

5.2.12 Considering the data storage organization, can you provide information about possible other elements which play a role here like security, interfaces to access the database (front-end) and how the back-end (Administration, Maintenance, Back up procedures,...) is organised.

- Front-end (app) and back-end communicate via a REST API
- The connection is encrypted with HTTPS
- The different functional elements of the backend are Docker containers
- Orchestration of the containers happens with docker-compose

5.3 Security and Privacy

* 5.3.1 Concerning password security, which precautions are applied? *at least 1 choice(s)*

Password **composition** policy includes e.g. the minimum number of characters from the set of lowercase letters, uppercase letters, special characters, and numbers. **Protection/encryption**

- ☐ Password composition
- ☐ Reuse password
- ☐ Reset password
- ☐ Password security protocol
- ☐ Other – please specify below
- ☐ Don't know
- ☐ No answer

* 5.3.2 If Other – please specify

* 5.3.3 Concerning communication security: which precautions are applied? *at least 1 choice(s)*

A **communication protocol** is a defined set of rules and regulations that determine how data is transmitted in telecommunications and computer networking.

- ☐ Automatic communication (no one reads/sends out the emails personally; left aside the emails send to the help-desk)
- ☐ Token based communication (no address but a token is used)
- ☐ Communication protocol
- ☐ Other – please specify
- ☐ Don't know
- ☐ No answer

* 5.3.4 If Other – please specify

* 5.3.5 Concerning server/data storage security: which precautions are applied? *at least 1 choice(s)*

A **Virtual Private Server** is a virtual machine sold as a service by an Internet hosting service. A VPS runs its own copy of an operating system (OS), and customers may have superuser-level access to that operating system instance, so they can install almost any software that runs on that OS.

A **client/server protocol** is a communications protocol that provides a structure for requests between client and server in a network. **Virtual**

- ☐ Private Server (VPS)
- ☐ Back-up strategy

Protection/encryption

Software on server in data collection countries

- ☐ Database on server in data collection countries
- ☐ Data transmission protocol
- ☐ Server protocol
- ☐ Data storage protocol
- ☐ Other – please specify
- ☐ Don't know
- ☐ No answer

*** 5.3.6 If Other – please specify**

*** 5.3.7 Concerning privacy: which precautions are applied? *at least 1 choice(s)***

Informed consent is a permission granted in full knowledge of the possible consequences, the risks involved and the alternatives. An **anonymization protocol** allows anonymizing personal data within the data transmission from data holders to a data collector without privacy breaches.

- ☐ Informed consent
- ☐ Split-up personal information from collection data
- ☐ Anonymization protocol
- ☐ Software on server
- ☐ Software on client
- ☐ Other – please specify
- ☐ Don't know
- ☐ No answer

*** 5.3.8 If Other – please specify**

*** 5.3.9 Is the setup of the tool in conformity with ... ?**

at least 1 choice(s)

Data treatment **confidentiality** refers to rules applied for treating the data set to ensure that private information from individual units cannot be accessed and to prevent unauthorised disclosure.

- ☐ ESOMAR/ISO regulations on data protection and confidentiality
- ☐ National privacy law – please specify below
- ☐ EU privacy law - GDPR
- ☐ None of above
- ☐ Don't know
- ☐ No answer

*** 5.3.10 Which country/countries privacy law? – please specify**



5.4 Functionalities of the applications

* 5.4.1 For which browsers is the web application compatible/tested? *at least 1 choice(s)*

Multiple answers are possible Internet

- ☐ Explorer
- ☐ Microsoft Edge
- ☐ Mozilla Firefox
- ☐ Google Chrome
- ☐ Safari
- ☐ Opera
- ☐ Other – please specify below
- ☐ Don't know
- ☐ No answer

* 5.4.2 If Other – please specify

* 5.4.3 For which screen sizes is the web application employable? *at least 1 choice(s)*

Multiple answers are possible

- ☐ Less than 5 inches (smaller Smartphones)
- ☐ Between 5 and 6,8 inches (bigger Smartphones)
- ☐ Between 6,9 and 9,0 inches (smaller Tablets)
- ☐ Between 9,1 and 10,2 inches (bigger Tablets)
- ☐ Between 10,3 and 12,9 inches (large Tablets and small Laptops)
- ☐ Between 13 and 15,6 inches (middle-sized Laptops and computer screens)
- ☐ Higher than 15,6 inches (large Laptops and computer screens)
- ☐ Don't know
- ☐ No answer

* 5.4.4 Is the web application accessible via a public link?

- ☐ Yes
- ☐ No
- ☐ Don't know
- ☐ No answer

* 5.4.5 Please provide the link here:

* 5.4.6 For which operating systems is the mobile application (native or hybrid) functional?

at least 1 choice(s)

Multiple answers are possible Android

iOS

- ☐ Windows Phone
- ☐ Symbian, Java ME, BlackBerry, Kindle
- ☐ Other OS – please specify below
- ☐ Don't know
- ☐ No answer

* 5.4.7 If Other – please specify

* 5.4.8 Is the mobile application accessible via a link and/or via the app store?

between 1 and 2 choices

- ☒ Yes, via a link provided for the survey
- ☒ Yes, via the app store
- ☐ No
- ☐ Don't know
- ☐ No answer

5.4.9 Please provide the link(s) below if possible:

* 5.4.10 Which of the following devices can be used in a survey using this tool?

at least 1 choice(s)

Multiple answers are possible

- ☐ Computer
- ☒ Tablet
- ☒ Smartphone
- ☐ Smartwatch
- ☐ Other – please specify below
- ☐ Don't know
- ☐ No answer

* 5.4.11 If Other – please specify

* 5.4.12 Which of the following features (if any) apply to the tool.

at least 1 choice(s)

Multiple answers are possible

- ☒ Respondents can use different devices through the study and the content on these devices is automatically synchronized.
- ☐ Respondents can use different devices through the study, but the content on these devices is not automatically synchronized.
- ☐ Respondents can log online and offline
- ☐ The User Interface (UI) of the web and mobile application is consistent
- ☐

- ☐ Don't know
- ☐ No answer

*** 5.4.13 Please indicate which functionalities are present in the web and/or mobile application(s)?**

at least 1 choice(s)

Application Programming Interface is a way in which computer applications communicate and cooperate with external applications or sources of information. An API key is used to track and control how the API is being used. An API key makes use of an Universally Unique Identifier

(UUID) to ensures his uniqueness. **Login screen: user name & password**

- ☐ **Login validation & communication**
- ☐ Personal settings/menu (name, contact detail, change password, ...)
- ☐ **Data collection overview (tasks to be completed: e.g. questionnaire – diary – questionnaire)**
- ☐ Communication/notification overview
- ☐ **Data collection info (details, privacy, FAQs, ...)**
- ☐ Inclusion of external API-key(s) from wearables, sensors and connected devices
- ☐ Other functionalities – please specify below
- ☐ Don't know
- ☐ No answer

*** 5.4.14 If Other – please specify**

5.4.15 Can multiple languages be offered during a survey? Which of the following options apply?

Yes

No

	Yes	No	Don't know	No answer
* For every data collection multiple languages can be offered	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
* Respondents can switch between the offered languages	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

*** 5.4.16 In which language(s) is the survey offered?**

at least 1 choice(s)

- ☐ Bulgarian
- ☐ Croatian
- ☐ Czech
- ☐ Danish
- ☐ Dutch
- ☐ English
- ☐ Estonian
- ☐ Finnish
- ☐ French
- ☐ Gaelic
- ☐ German
- ☐ Greek
- ☐ Hungarian
- ☐ Italian
- ☐ Latvian
- ☐ Lithuanian
- ☐ Maltese
- ☐ Polish
- ☐ Portuguese
- ☐ Romanian
- ☐ Slovak
- ☐ Slovenian
- ☐ Spanish
- ☐ Swedish
- ☐ Other (please specify below)

* 5.4.17 If Other – please specify

Norwegian

5.5 Functionalities of the management website

* 5.5.1 How is the login process managed? at least 1 choice(s)

Multiple answers are possible

- ☐ Login screen: user name & password
- ☐ Login validation & communication
- ☐ Different user levels and roles
- ☐ A management level to define the user level and roles
- ☐ Other – please specify below
- ☐ Don't know
- ☐ No answer

* 5.5.2 If Other – please specify

- ☐
- ☐

*** 5.5.3 Which phases in the statistical production are covered by the tool?**

at least 1 choice(s)

- ☐ Selection of the application(s) and devices to be used (web app, native app, hybrid app; computer, laptop, tablet, Smartphone, Smartwatch)
- ☐ Development of questionnaires
- ☐ Development of diaries
- ☐ Use/inclusion of other data sources through use of API
- ☐ Definition of communication (paper, on screen, email, notification, ...)
- ☐ Definition of respondents
- ☐ Set up of data collection flow
- ☐ Execution of fieldwork/data collection
- ☐

Calibration of the data (method to weigh the collected data based on population representation, and dispersion over the days)

- ☐ Download/Export of database
- ☐ Download metadata
- ☐ Other – please specify below
- ☐ Don't know
- ☐ No answer

* 5.5.4 If Other – please specify

* 5.5.5 Which kind of information can be downloaded? *at least 1 choice(s)*

The **paradata** of a survey are data about the process by which the survey data were collected. Codes of

- ☐ variables, activities
- ☐ Time points (begin & end time)
- ☐ Variables names & labels (questionnaires, context questions)
- ☐ Text/category of variables names & labels (questionnaires, context questions)
- ☐ Paradata (actual logging information)
- ☐ Other – please specify below
- ☐ Don't know
- ☐ No answer

* 5.5.6 If Other – please specify

* 5.5.7 To which formats can the database be exported?

- ☐ .csv (Comma Separate Values)
- ☐ .xlsx (Excel)
- ☐ .sav (SPSS)
- ☐ .por (R)
- ☐ .xpt (SAS)
- ☐ Other – please specify below
- ☐ Don't know
- ☐ No answer

* 5.5.8 If Other – please specify

PostgreSQL

5.5.9 To which level does the metadata relate?

- ☐ Individual/household level (e.g. age, profession, family composition on the respondents/cluster level)
- ☐ Statistical production level (e.g. having a multiple choice question with a number of answer categories)
- ☐ Calibration level (e.g. having a sample of males and females, in different age categories)

- ☐ Other – please specify below
- ☐ Don't know
- ☐ No answer

5.5.10 If Other – please specify

END OF THE QUESTIONNAIRE
THANK YOU FOR YOUR CONTRIBUTION