Personalised Health Monitoring and Decision Support Based on Artificial Intelligence and Holistic Health Records

D3.6 – Secondary data capture and interoperability II

WP3 Personalized Holistic Health Records

Dissemination Level: Public
Document type: Report
Version: 1.0
Date: October 31, 2022

Document Details

Project Number	101017441
Project Title	iHelp - Personalised Health Monitoring and Decision Support Based on Artificial Intelligence and Holistic Health Records
Title of deliverable	Secondary data capture and interoperability II
Work package	WP3 Personalized Holistic Health Records
Due Date	October 31, 2022
Submission Date	October 31, 2022
Start Date of Project	January 1, 2021
Duration of project	36 months
Main Responsible Partner	Innovation Sprint
Deliverable nature	Report
Author name(s)	Aristodemos Pnevmatikakis, Spyros Spanos, Harm op den Akker (iSPRINT), Pavlos Kranas (LXS), Jake Griffiths (ICE)
Reviewer name(s)	Pavlos Kranas (LXS), Fabio Melillo (ENG)

Document Revision History

Version History			
Version	**Date**	**Author(s)**	**Changes made**
0.1	2022-09-30	Aristodemos Pnevmatikakis (iSPRINT)	Restructure previous version
0.2	2022-10-01	Aristodemos Pnevmatikakis (iSPRINT)	Executive summary and introduction (section 1)
0.3	2022-10-02	Aristodemos Pnevmatikakis (iSPRINT)	Section 5.1.4
0.4	2022-10-05	Aristodemos Pnevmatikakis, Spyros Spanos, Harm op den Akker (iSPRINT)	Section 4
0.5	2022-10-07	Aristodemos Pnevmatikakis (iSPRINT)	Sections 5.1.2 and 5.1.3
0.6	2022-10-09	Aristodemos Pnevmatikakis (iSPRINT)	Section 5.2
0.7	2022-10-11	Jake Griffiths (ICE)	Section 6.3
0.8	2022-10-12	Aristodemos Pnevmatikakis (iSPRINT)	Conclusions (section 7)
0.9	2022-10-21	Fabio Melillo (ENG)	1st Internal review
0.10	2022-10-25	Pavlos Kranas (LXS)	2nd Internal review

0.11	2022-10-25	Aristodemos Pnevmatikakis (iSPRINT)	Reviewer comments addressed
0.12	2022-10-27	Pavlos Kranas (LXS)	Quality review
1.0	2022-10-31	Dimosthenis Kyriazis (UPRC)	Final version

Table of Contents

Table of Figures

Executive summary

The purpose of this iHelp deliverable, D3.6 – "Secondary data capture and interoperability II", a public report, is to complement the first version, D3.5 – "Secondary data capture and interoperability I", as an inventory of secondary data for the iHelp clinical studies, i.e., the information extracted from their end-users, individuals that are mostly patients, but in cases of preventive studies, also the general public. No matter their status (*patient*, or *individual*), end-users in iHelp are study participants, and this is how they are referred to in this document. The deliverable thus documents the choices made on what to measure and what to simply ask the study participants to provide, trying to strike a balance between loading people with devices (and cost) but also asking too much of them in terms of data entry. In that respect it is of interest outside the iHelp community, serving as a data collection manual outside of the strict clinical setting. It also considers how the data can be collected and how they can be represented in the iHelp big data platform.

This document is delivered in M22 (October 2022), after the compliment of the requirement collection as reported in D2.3 – "State of the art and requirements analysis III" in M20. This document presents the final secondary data inventory, as discussed between the partners responsible for secondary data collection and the clinical partners.

1 Introduction

We define "Secondary data" as the data collected from the study participants in a non-clinical setting. Such data are collected, ingested into iHelp and complement the primary, clinical data in the HHR.

1.1 Deliverable purpose

The purpose of this iHelp deliverable, D3.6 – "Secondary data capture and interoperability II", a public report, is to act as a manual for secondary data collection, both within and beyond iHelp. It answers the questions on what to collect, how to do so, and how to represent the collected information. Secondary data are introduced, discussing what can be collected, what is actually collected for the clinical studies of iHelp, and how this is facilitated, using both automated measurements and reports. Actually, a successful secondary data collection system is the one that achieves a balance between usage of devices for measurements (which incur cost and demand the attention of the study participants) and the request for reports (which incur manual data entry from the study participants).

1.2 Target audience

The target audience for this deliverable consists of individuals both within and outside of the iHelp project. The following iHelp partners will be using this deliverable:

- Internal to the secondary data capture task, software engineers will use the information for implementing the necessary widgets the study participants will be reporting data with. Also, the questionnaires involved will be identified and implemented using the designated computer-readable representation. Finally, the secondary data inventory will be used by the partners mapping the data attributes to the HHR.
- The data scientists building the HHR, as well as the ML engineers building the iHelp models, will be informed on the secondary data inventory.
- The clinical partners will be informed of the options available for collecting information from their study participants, in order to guide them through the use of the data collection system at the recruitment phase.

As to the clinical study community beyond iHelp, this deliverable is documenting the project's attempt for striking a balance between the measurement burden (hardware that needs participants' attention and its cost) and the reporting burden (repeated requests for manual entry of data).

1.3 Deliverable context

The requirements from the clinical partners are the starting point for determining what to capture. Their final version is found in D2.3 – "State of the art and requirements analysis III", delivered in M20. These requirements have been analyzed in bilateral meetings between the clinical partners and those responsible for the secondary data extraction to finalize the needs for secondary data in iHelp.

The system collecting the secondary data is Healthentia, an e-clinical platform, developed by Innovation Sprint, the iHelp partner leading the secondary data collection efforts of iHelp. This is a mature product, and hence only the new modules needed are specified by iHelp, both in this deliverable and more formally in D2.7 – "Functional and Non-Functional Specifications II," delivered in M20 (August 2022).

Healthentia is used as an edge processing platform in iHelp, playing the role of a composite sensor. The role of Healthentia as a source is described in D2.5 – "Conceptual model and reference architecture II" delivered in M18 (June 2022) and the details on how this is achieved are presented here.

The primary and secondary data utilize the same ingestion pipeline in iHelp. Hence data ingestion in general, and not only primary data, is discussed in D3.4 – "Primary data capture and ingestion II" also to be delivered in M22. Pointers to that document are only included here.

Secondary data mapping into the HHR is also part of the secondary data efforts of iHelp. Preliminary information on the HHR is to be found in D3.1 – "Data Modelling and Integrated Health Records: Design and open specification I" delivered in M8. The mapping itself is discussed here, and the information presented should affect the final HHR specification in D3.2 – "Data Modelling and Integrated Health Records: Design and open specification II" to be delivered in M32.

1.4 Deliverable structure

The secondary data are defined in contrast to the primary in the beginning of Section 2. They are split into the four different categories characterising the life of people outside the clinical setting. A discussion then begins from D2.1 – "State of the art and requirements analysis I", towards the analysis and presentation of today's clinical partners' wish-list.

The way to collect secondary data is discussed in Section 3. The discussion is originally generic, as befits a public deliverable, but then drills down to the particulars of the iHelp secondary data. The wish-list of Section 3 is broken down in attributes to be measured by devices and attributes to be reported by the study participants. The section concludes with the processing the raw data into more usable quantities, like habits and trends.

In introduction of the Healthentia platform is carried out in Section 4, to be followed by the details on how the iHelp studies are setup, to facilitate secondary data capture.

Section 5 details the sources of secondary data actually collected in the studies presented in Section 4, leading to the detailed secondary data catalogue of attributes.

All these secondary data need to be ingested into the iHelp platform and mapped into HHR. Section 6 follows this process, indicating the use of the Healthentia e-clinical platform as a composite sensor in iHelp, outlining the secondary data ingestion and detailing their mapping into HHR.

Finally, the deliverable is concluded in Section 7.

1.5 Changes in this version

There are several changes in this second version of the deliverable, mostly spanning the how to collect secondary data and how to represent it. The question of what to collect has been covered in the first version. The detailed list of changes is as follows:

- Executive summary: Updated.
- Section 1.1: Deliverable purpose extended.
- Section 1.3: Deliverable context updated.
- Section 1.4: Deliverable structure updated.

- Section 3.3: Moved here the processing of raw secondary data.
- Section 4: New section expanding from the existing introduction to Healthentia, into actually how Healthentia is used to facilitate (configure and run) the iHelp studies.
- Section 5.1: Consolidated the secondary data sources. In detail, sections 5.1.2 and 5.1.3 are updated to reflect the final pilot choices in terms of questionnaires, while in section 5.1.4 the description of the nutrition widget has been added.
- Section 5.2: New section detailing the secondary data attributes.
- Section 6.3: Secondary data mapping into HHR.
- Section 7: The new conclusions are drawn.

2 Secondary data: What to collect

Data is the new medicine [1]. Data (when shared, reused in a privacy-respecting way and maintaining the control to the people providing it) can improve patient outcomes, foster research and accelerate deployment of novel health services [2]. In the health domain, and for the purpose of this deliverable, data comes in two flavours: inside and outside of the clinical setting.

2.1 Primary vs. secondary data

Health data are readily associated with clinical tests performed invasively on samples taken from our bodies, or non-invasively using modern depicting techniques. Such data, obtained in a clinical setting, is of paramount importance, is termed primary in the iHelp project, but certainly does not form the complete spectrum of health data.

Hippocrates (approx. 460-370 BC), the father of medicine believed that disease was not a punishment inflicted by the gods but rather the product of environmental factors, diet, and living habits, a fact that is well-established today. Our living habits can be enumerated using data attributes about our lifestyle, obtained in our natural environment, outside of the clinical setting. This data is termed Real-World Data (RWD) and in iHelp is distinguished from the clinical, primary data by being termed secondary data, more to indicate the different collection setting than to attempt to compare the importance of the two (Figure 1).

Figure 1: The two sources of iHelp data – primary (clinical) and secondary (real-world data).

RWD are formally defined by the FDA as "*data related to patient health status and/or the delivery of health care routinely collected from EHRs, claims and billing data, data from product and disease registries, patient-generated data including home-use settings, and data gathered from other sources that can inform on health status, such as mobile devices.*" [3].

While clinical data is generated sparingly, during testing or hospitalisation sessions, RWD is generated continuously, throughout the days, weeks and months of our lifetime. It would thus be good practice to actually collect RWD in a continuous mode, rather than the sporadic mode in which clinical data are collected. Section 3 discusses how this can happen, and the practices followed in iHelp.

RWD comprises attributes that enumerate different important aspects of the way we live our lives. The attributes are grouped in the physiological, psychological, social and environmental categories discussed next.

2.1.1   Physiological attributes

The physiological attributes of RWD have to do with the human body, its activities and adverse events. They are mostly measured using activity trackers and/or smartphones, but are also reported.

Activity related attributes are *steps walked*, *distance walked*, *elevation* (or *floors climbed*), energy dissipation, time spent in different activity intensity zones (e.g., mild, moderate and high intensity physical activity, as it is formally defined as a function of age) and exercise activities (walking, running, cycling, etc.), as well as their distribution in the day. Presence indoors or outdoors can also be of importance. Specialized physical activity can also be measured via composite tests like the six-minute walk test (6MWT), the frailty test or games specifically designed to measure muscular responses (tapping on a mobile phone screen for Parkinson's disease or performing other exercises while monitored and analysed by depth cameras to measure features important in stroke or accident rehabilitation). All these tests are scripted and hence can be measured using sensors and audio-visual instructions to the people on their smartphone.

Attributes related to the functioning of the heart include the continuous measurements of the heart rate variability and the time spent in different heart rate zones, as well as the daily resting heart rate measurement.

Sleep related attributes include continuous measurements on the time spent in the different sleep stages (awake in bed, light, REM, deep sleep).

Other physiological attributes can be self-reported by the participant. Symptoms of interest in general or to specific therapeutic areas (e.g., headache, body temperature, blood pressure, pains, diarrhoea, fatigue, nausea), including their intensity can be collected. Weight should be regularly reported, and less so height (especially at younger or older ages). Nutrition is paramount, starting at a higher level with the consumption of food categories of interest, but more detailed analysis can also be used when available. Water, coffee, tea, refreshments and alcohol intake can be reported. Finally, the menstrual cycle can also be of importance.

2.1.2   Psychological attributes

The psychological RWD attributes refer to the emotions of the study participants. They are mostly reported and include at a high-level simple emotional state self-assessment, but when deemed necessary the collected information goes deeper using standardized reports from professional therapists monitoring the patients.

Measurements can also be used to indirectly capture psychological aspects. Emotion can be recognized from video recordings of the face, the voice audio or the social media text posts. Places visited (which, how diverse they are) are also an indication of the psychological state. Aspects like the weather or spending unusual time in commuting can have some importance.

2.1.3   Social attributes

The social RWD attributes enumerate study participants' social life. Such information can be measured indirectly using attributes about the usage of the phone (diversity, duration, frequency of calls) and social media (diversity, number, frequency of interactions).

More direct information can be reported using questionnaires on activities with friends, family or co-workers, or can be obtained in conversation with a digital virtual coach.

2.1.4 Environmental attributes

The environmental RWD attributes attempt to enumerate the environment the study participants live in. They include reported environmental indicators for the assessment of the quality of life, like the 11 attributes of the OECD better life index [4].

Precise measurements of living or working environment quality can be obtained by integrating relevant commercial devices (e.g., for air quality analysis), or by integrating with data services that report e.g., Air Quality Index at specific locations.

2.2 Discussing clinical requirements

The RWD attributes presented in Section 2.1 are generic, and should be specialized for every therapeutic domain. Nevertheless, they had been the starting point for the possibilities available to clinical partners in terms of secondary data collection. The initial processing resulted to the secondary data requirements in D2.1 "State of the art and requirements analysis I", Section 4.3. Further analysis after the submission of D2.2 "State of the art and requirements analysis II" in M12 of the project has resulted to the consolidated requirements across all clinical partners shown in Table 1 (physiological) and Table 2 (psychological, social and quality of life).

Table 1: Physiological RWD consolidated requirements across the five clinical partners.

Category	Attributes/parameters	UOM	FPG	HDM	MUP	TMU
Physical activity	Steps	x	x	x		x
	Floors		x			x
	Energy burned	x	x			x
	Time in different activity zones		x			x
Exercise sessions	Start time	x	x			
	Duration	x	x			
	Type	x	x			
	Energy burned	x	x			
Heart	Resting heart rate	x	x	x		
	Max heart rate	x	x	x		
	Time in different heart rate zones	x				
Sleep	Start time	x	x	x	x	x
	Duration	x	x		x	x
	Time in different sleep stages	x	x		x	
	Pittsburgh Sleep Quality Index [5]					x
Nutrition	Consumption of food categories (meat, veg, grain etc)	x	x	x	x	x

	Liquids	x	x		x	x
	Food Frequency Questionnaire [6]	x				
	Food diary with rations [7]	x				
Symptoms	Symptom assessment questionnaire [8]					x

The physical activity attributes and exercise sessions can be collected using an activity tracker, or a smartphone (as long as it is kept with the study participant most of the time). Physical activities cannot be entered manually, while exercise sessions (at least start time and duration) can.

The heart can only be monitored continuously with an activity tracker. Sporadic measurements can be added manually.

The sleep attributes can be collected by an activity tracker. Start and duration can be entered manually. The questionnaire is a self-assessment.

All nutrition, symptom, psychological and health/quality of life attributes are manually reported.

Table 2: Psychological, social and quality of life RWD consolidated requirements across the five clinical partners.

Category	Attributes/parameters	UOM	FPG	HDM	MUP	TMU
Psychological	Mood self-assessment via 5-level smileys	x	x	x	x	x
	The Warwick-Edinburgh Mental Wellbeing Scales [9]	x	x	x	x	
	Rosenberg Self-Esteem Scale [10]	x	x	x	x	
	Body Awareness Questionnaire [11]	x	x	x		
	Subjective Happiness Scale [12]	x	x	x	x	
	Motivation, risk averse, big 5	x	x	x		
	Locus of Control Questionnaire [13]	x	x	x		
	Health Behaviour and Stages of Change Questionnaire [14]	x	x	x	x	
	Symptom assessment questionnaire [8]					x
Social	Types and frequent use of social platform		x	x	x	x
	Supportive messaging with moderation		x	x		
	Opportunities for forming new health habits	x	x	x	x	
	Use of mobile apps		x	x		x
Health/Quality of life	Overall health rating by participants [15]	x	x	x	x	x
	Health literacy (HLS-EU_Q16) [16]	x		x		
	Diseases: diabetes, hypertension, pancreatitis, cirrhosis, hepatitis	x				x
	Family history of pancreatic cancer	x		x		x
	Process of change [17]	x	x	x	x	
	Smoking					x

Height & weight	x				x
Quality of life (EQ-5D) [18]	x		x	x	
Quality of life for patients (EORTC QLQ-C30) [19]		x	x		x
Quality of life for pancreatic cancer patients (EORTC QLQ - PAN26) [20]		x	x		
Activities of Daily Living: hygiene, continence mgmt., dressing, feeding, ambulating			x		
Frailty measure			x		

Regarding the social attributes, the supportive message is an opportunity to send a message from one study participant to all others, after being approved by a moderator. The collection of the message is a free text report. The moderation can happen at the portal app. The delivery of the approved message can be done by the virtual coach. The opportunities for forming new health habits is a questionnaire. The use of mobile apps and social platforms can be reported but also can in principle be measured. There are authorization implications though that make such measurements difficult to implement but also to enable by the study participants.

The RWD in Table 1 and Table 2 constitute the iHelp clinical partners' wish-list. In the following section this will be analysed in terms of feasibility to derive the secondary data inventory.

3 Secondary data: How to collect

RWD are collected outside the clinical setting, directly from the study participants using ubiquitous, easy to operate devices, or simply by asking people about the necessary information.

3.1 Measuring from devices

RWD measurements involve devices that are commonly used by people. An important source of physiological information is the activity tracker. These are consumer devices gaining popularity amongst health- and wellness- aware people. Study participants can already own one, easily get a cheap one, or the studies can distribute them for free as a participation incentive, especially for longer studies. Other devices can be scales and ubiquitous medical devices (e.g., thermometers, blood pressure monitors, SPO2 monitors). Even more specialised medical devices can be included here, of the type used at home not by the general population, but by certain patient categories.

No matter the devices, there are two modes of measurement collection: the automatic and the manual. Automatic measurement collection refers to having the measurement device integrated with the data collection system, the measurements flowing from the device into the RWD collection system in an unattended manner. Manual measurement collection refers to having the study participant reading out the measurement from the device and manually reporting the measurement to the RWD collection system.

The automatic is clearly the preferred option, the only one when the measurements have high volume and/or frequency. It minimises study participants' burden and data entry errors. It is of this option only that "measurement from devices" refers to. The methods to do so are outlined in Section 3.1.1, while devices involved in the iHelp measurements are discussed in Section 3.1.2. The manual measurement option is actually a case of study participant report, covered in Section 3.2.

3.1.1 Methods

Devices can be integrated with RWD capturing systems in two ways: employing the devices manufactures' provided API or SDK. To understand the difference, it is important to understand the flow of information from the device.

All devices have a short-range communication capability, which is almost always Bluetooth Low Energy (BLE). This necessitates the use of another device that collects the information, called a controller, base station, edge node or bridge depending on the manufacturer. In case of domotics devices being fixed in some location, this controller is a fixed device, located somewhere in the home. In case of wearables, the most natural choice is to use the mobile phone that is mostly around the study participant as a controller, in which case the controller is some software installed on the phone. The controller receives the information from the device(s) it is paired with using the BLE protocol of its communication module. It then transmits this information to the servers of the device manufacturers using WLAN. The study participants control the devices using SW on their mobile phones and/or on the web. They also view (and control) their data using the same apps. The process is depicted in the upper section of Figure 2, the 3rd-party device section.

Since the study participants need to be more in control of their data and the advent of GDPR, device manufactures typically offer means to the study participants to get their data, not just view it. Data exports in files have long been available, but recently there are more automated options, allowing the study participants to get their data in online ways. APIs and/or SDKs are being offered so that requests are forwarded by third-party SW systems to get the data. The study participants only need to authorise these third-party SW systems to collect the data on their behalf. This is exactly the method used in iHelp, with the details being discussed in Section 6.1.

When a device manufacturer offers an API to get data, then the data is captured from the cloud platform of the integrated device. The API offers endpoints to be used by an authenticated entity to get data they are authorized for. When the entity is a data collection system, the API endpoints are called by the data collection cloud platform. Depending on the manufacturer, the API can passively wait to be utilised to offer the requested data, or can notify the data collection system of the availability of new data to be collected. The data loop back to the study participant is closed by one of the tasks of the mobile app of the data collection system: It utilizes another API, that of the data collection system, to get the integrated device data and visualise it alongside the rest of the secondary data it collects. While the usage of a 3rd-party API is very easy to support by the data collection system, the downside of the API approach is that many SW entities are involved (the two apps and the two cloud platforms) and the round trip of the data that travels from the device to the mobile phone, to two different cloud platforms back to the mobile phone, is very long. The device integration via the provided API is presented in Figure 2.
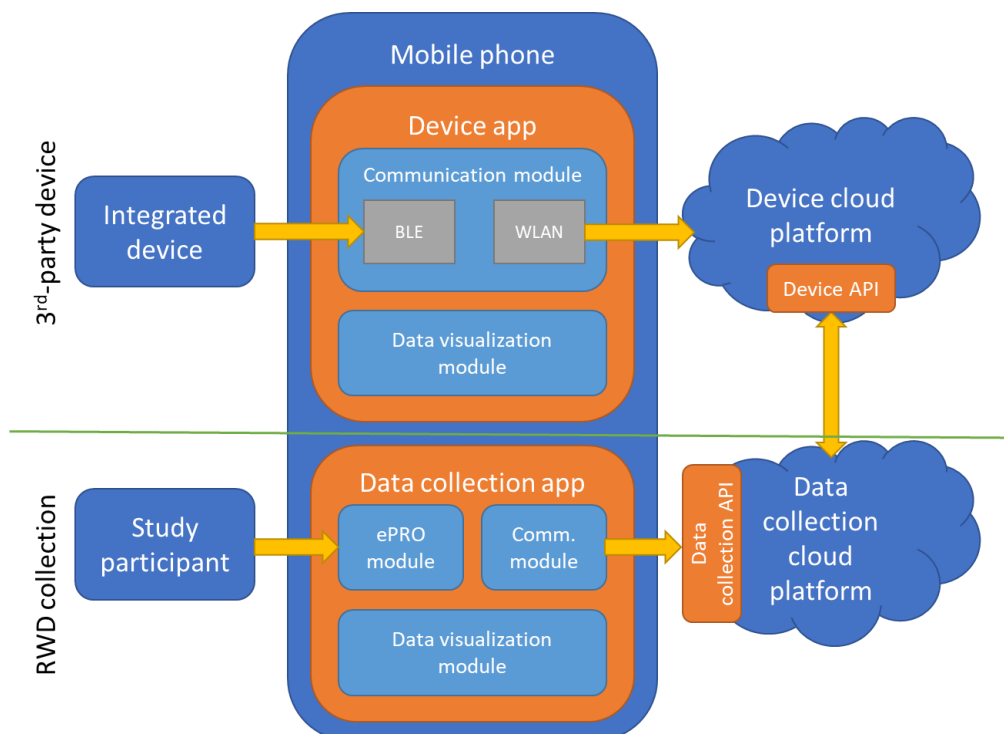
Figure 2: Generic RWD collection mechanism where measurements utilize device API.

When a device manufacturer offers an SDK, then the data collection app receives the data from a local source, without involving the device cloud platform. The SDK can offer two different things. One option is direct access to the device. In this case the device is programmable and a part of the data collection system needs to be programmed in it. This will send the collected data to the data collection app of the

mobile phone. Measuring via direct access to a programmable device via its SDK is depicted in Figure 3. The other option is indirect access to the device, via the device app running on the same mobile phone, as shown in Figure 4. This is easier to implement, but demands both the device app and the data collection app to be running on the mobile phone.
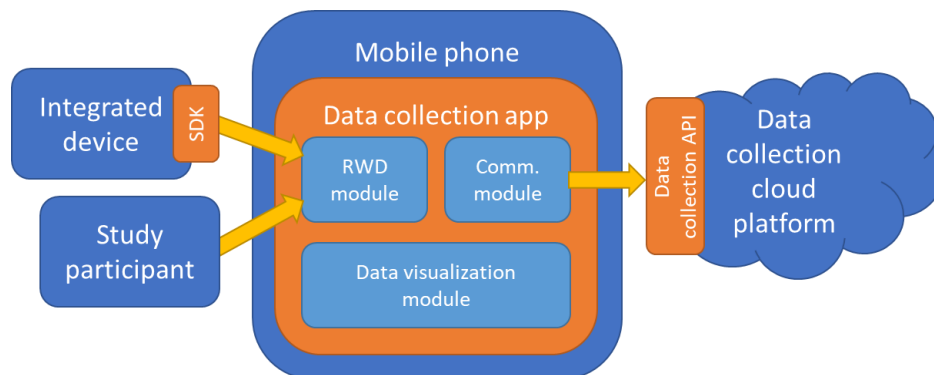
Figure 3: Generic RWD collection mechanism where measurements utilize direct access to programmable device via its SDK.
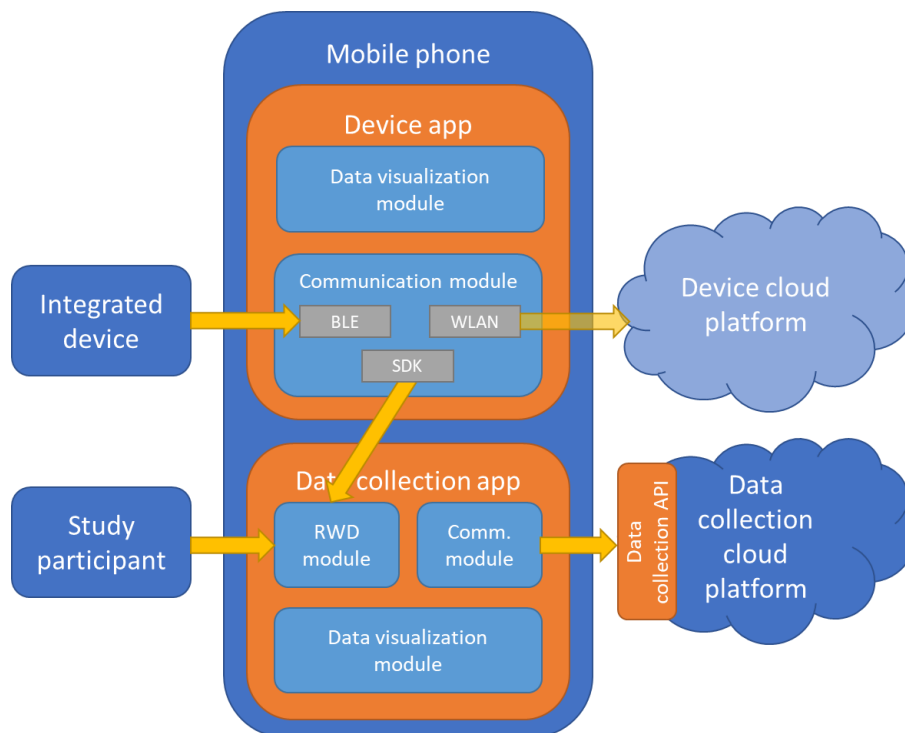
Figure 4: Generic RWD collection mechanism where measurements utilize indirect access to device via its SDK on the device app.

A variant of the SDK access is when the devices being integrated are the actual sensors on the mobile phone. In this case, the SDK used is the phone's SDK, which is always well-known and documented. On Android devices a data collection service is programmed within the data collection app. On iPhone devices, the Apple Health Kit is used to access the data from Apple Health, the app that collects all health and wellness sensor data of the phone or other integrated devices. Similar solutions to Apple Health exist from all the big Android players (e.g., Samsung Health) but the Android manufacturer segmentation does not allow a data collection system to interface with one such app and get data from a significant portion of the possible study participants. Apart from Android manufacturer segmentation, another usability issue

arises. For the mobile phone sensors to be used, the mobile phone must be used as a wearable, always in the pocket of the study participant, which is clearly not feasible for most people. Additionally, the location of measuring (i.e., the pocket) is of influence in the accelerometer-based estimation of a person's energy expenditure throughout the day. When using the mobile phone as a measurement device, this location of wearing is not always known, and not always the same. Wrist-worn devices – especially when it is known whether the user is wearing it on the dominant or non-dominant hand, and therefore preferred.

Obviously, the easiest integration method is the API, but the most open to the needs of the data collection system is the SDK that allows direct access to a programmable device. Unfortunately, the device measurement method is not for the data collection system to select. Almost all device manufacturers offer API access to the collected data (the manufacturer maintains ownership of the data), while only a handful offer an SDK, and those that do usually represent more experimental and less commercial devices. Also, API usage is usually free, while SDK access is reserved for very large clients of the device manufacturers, or comes with a very high cost. Finally, device manufacturers can apply usage constraints in both modes of access, limiting the data attributes offered, the temporal granularity, the number of times data can be requested, or the volume of transferred data. All these aspects need to be taken into account when selecting the device to integrate.

3.1.2 Measuring physical activity in iHelp

The iHelp clinical partners have specified directly the use of a device for physical measurements, most of them agreeing that an activity tracker is much preferred over mobile phone sensing. They also indirectly specified the use of scales for weight collection. Only the physical measurement device (activity trackers or mobile phones) is reporting measurements automatically. The scales give the information to the study participant, to be entered manually into the system. It is thus the devices for physical activity measurements that are discussed in this section.

Currently, the iHelp physical measurement options in place are summarised in Table 3. Two well-known brands of activity trackers are directly covered: Fitbit and Garmin, both using API. Fitbit allows a certain number of API calls per day, and offers all information on the day up to that point, albeit without any information on when this data has been last updated. Garmin notifies whenever new information is available, and does provide some time series data, allowing intra-day information, though this is of no importance to iHelp, where no intra-day information is required. More activity trackers can be covered indirectly if devices are integrated with Apple Health. Integration with Apple Health utilised the Apple Health Kit SDK. Finally, the Android SDK is used in a module of the mobile app that captures the Android phone sensor signals.

Table 3: iHelp physical measurement options – integration methods and RWD collected.

Integrated devices	Integration method	RWD collected
Fitbit activity trackers	Fitbit API	Physical activity, exercise sessions, heart & sleep
Garmin activity trackers	Garmin API	Physical activity, exercise sessions, heart & sleep
iPhone sensors and devices integrated with Apple Health	Apple Health Kit (SDK)	Physical activity (with just the phone sensors)

Android sensors	Android SDK	Physical activity & exercise sessions

There is one difference between the Fitbit and Garmin measurements integration. Currently the integration with Garmin does not transfer any heart zone information requested by one of the clinical partners. A work-around this problem is currently sought, to get this information indirectly.

There is nothing to be done for the desired continuous heart monitoring without dedicated activity trackers. For sleep though there is the solution of start and end time reporting by the study participant, as discussed in Section 3.2.3.

Although automated measurements are secure in terms of data objectivity (a device measures and reports data without the intervention of the study participant) and quite low in terms of compliance burden onto the study participant, they do have some shortcomings:

- Devices do have an acquisition cost that has to be paid, and a study participant might need a few to fulfil the study protocol.
- Devices operate on battery; the study participant needs to make sure they are always charged, especially those meant for almost 24/7 usage.
- Not all devices can be integrated. It is easier for low volume and frequency measurements to be simply copied from the device's screen into the data collection system by the study participant.
- Some attributes cannot be measured; only the study participant can inform the system about certain aspects of their lives.

For these reasons, patient reports are still very important in studies. These reports are covered in Section 3.2.

3.2 Reporting from study participants

Study participants need to report on outcomes and their experiences, quantifying measures about them: the PROMs and PREMs. Traditionally the collection of the reports had been a manual process involving pen and paper, but now ePRO systems are facilitating this process in three important ways:

- Scheduled questionnaires can be addressed to study participants and the answers can be collected and processed much easier with the online ePRO tools.
- Spontaneous answers can be included with study participants reporting on outcomes and experiences at the moment they happen, without needing to collect them and report them at predefined times. This is very important for event-type outcomes like a symptom, affliction or discomfort. It is also important for manual measurement entry, like weight logging.
- Continuous editing for incremental data collection can be facilitated for attributes that need to be accumulated throughout the day (e.g., the daily water consumption is easier to report when the study participant just adds this information regularly in the day, or retrospectively for the previous one), or automated measurements that need some correction (e.g., the sleep start and end times). Both can be achieved using widgets, i.e., UI elements at the disposal of the study participant for continuous data entry.

Scheduled and ad hoc questionnaires, as well as widgets are discussed in the next sections.

3.2.1 Scheduled questionnaires

Scheduled questionnaires are addressed to study participants at regular intervals (daily, weekly, etc.). Although they can be custom, non-standard questionnaires, they usually are formal, well-established ones, used throughout the scientific community to collect a specific type of information. Often each question needs to be considered on its own, but it is not uncommon that an aggregated score is calculate over groups of questions.

The ePRO system facilitates scheduled questionnaires by presenting them to the study participant in a timely way (even in a clever one, if the participant is known to be otherwise pre-occupied at the exact moment). It utilises a variety of UI elements to facilitate data entry (e.g., multiple choice lists, tick-boxes slide-bars, numerical entries, selection on images). It also offers advanced routing enabling or disabling groups of questions based on the study participant's status or previous answers.

3.2.2 Ad hoc questionnaires

Ad hoc questionnaires are at the disposal of the study participant for some spontaneous report. In many cases they are single-question, prompt-type ones (e.g., "please enter your weight in kilos," or "Indicate the severity of your migraine." All ad hoc questionnaires can be accessed using some "add event" UI element.

Ad hoc questionnaires should also be retrospective, allowing the study participant to change the time the report correspond to from the default "now" value to anything in the past. It is not easy to report a migraine the moment one suffers from it, but it is quite easy to remember about it in the next morning.

3.2.3 Widgets

Widgets are UI elements at the disposal of the study participant for continuous data monitoring, entry and correction. Attributes like liquid and food consumption can be collected more easily incrementally throughout the day instead of a one-off questionnaire at the end of it or the beginning of the next. Omissions and errors can be corrected retrospectively by revisiting previous days. Data can be visualised in natural ways, giving information back to the study participant, thus facilitating their compliance to the study protocol.

Also, widgets can be used to edit automatic measurements that are error-prone. For instance, the detected sleep start and end time, or the physical exercise type are automatically measured attributes that are usually error prone. The interested study participant can edit these attributes using the provided widgets.

3.3 Processing secondary data

Section 3 details the collection of raw secondary data, in the sense that it is data from measurements and reports as is, without any attempt to extract any higher-level meaning from it. This section is about processing raw secondary data, trying to establish higher-level information.

Currently the only higher-level information of interest to the clinical partners is the temporal evolution of the secondary data. For this habits and trends around them are established.

3.3.1 Habits

Behaviour is about habits and deviation from them. It is not so much about the current status. Hence any reasoning on behavioural data should regard the temporal evolution of secondary data. The temporal information for any quantity can be included by just using all *d* past samples of the quantity, but this leads to an unmanageable amount of input data. Another option is to use models of the past values: temporal evolution can be represented by assuming the normal distribution of the quantity, using its average and standard deviation.

Averages updated at every time step n with memory $a_n^{(d)}$ can approximate expectations of the *k*-th power of any RWD sequence x_n:

$$\overline{x_n^k}^{(d)} = a_n^{(d)} \cdot \overline{x_{n-1}^k}^{(d)} + \left(1 - a_n^{(d)}\right) x_n^k \tag{1}$$

where the memory is given by:

$$a_n^{(d)} = \begin{cases} 1 - 1/n & n < d \\ 1 - 1/d & n \geq d \end{cases} \tag{2}$$

The parameter *d* can be considered as the length of the memory, i.e., the approximate days that influence the average.

Habits are established as the long-term averages of RWD. The memory in (2) can be short- or long-term depending on the choice of the memory length parameter *d*. E.g., a choice can be 7 for short-term averages (approximately, the last week is influencing short-term averages), and a value of 84 for long-term averages (approximately 12 weeks or 3 months influencing long-term averages).

Valid choices of *d* for getting habits and short-term averages are actually dictated by the duration of the studies.

3.3.2 Trends

There are deviations from some habit that can carry significant information about a study participant. The first order ($k = 1$) and second order ($k = 2$) averages yield the standard deviation estimates:

$$\sigma_n^{(d)} = \sqrt{\overline{x_n^2}^{(d)} - \left(\overline{x_n}^{(d)}\right)^2} \tag{3}$$

The trend of some secondary data is defined as the variation in the short-term average from its long-term counterpart (the habit), normalized by the long-term standard deviation. The trend is then given by:

$$T_n = \frac{\overline{x_n}^{(d_{long})} - \overline{x_n}^{(d_{short})}}{\sigma_n^{(d_{long})}} \tag{4}$$

This definition of the trend makes it independent of the range of the secondary data in question. Small values around zero always designate a quantity that is currently at the level of the habit, not changing significantly. Large positive trends indicate a deviation from a habit towards larger values, while large negative trends indicate deviating from a habit towards smaller values.

3.3.3 Temporal information

Temporal information about all collected RWD is added by using the short- and long-term averages, as well as their trends. While missing automated measurements are very unusual and an indication of some failure, missing values in the reports are normal. Nobody expects study participants to be entering normal body temperatures and lack of symptoms, thus in the above calculations one should consider missing reports in a day as indicating normality, and use the equivalent normal values in the updates.

4 Healthentia e-clinical platform

The secondary data collection is facilitated by Healthentia, an e-clinical platform by Innovation Sprint. The platform provides secure, persistent data storage and role-based, GDPR-compliant access. It collects the data from the mobile applications of all study participants, facilitating smart services such as risk assessment, and providing both original and processed information to the mobile and portal applications for visualization. The high-level architecture of the platform is shown in Figure 5. This is a layered architecture, comprising of the API, data management, core functionalities, study management and services layers.
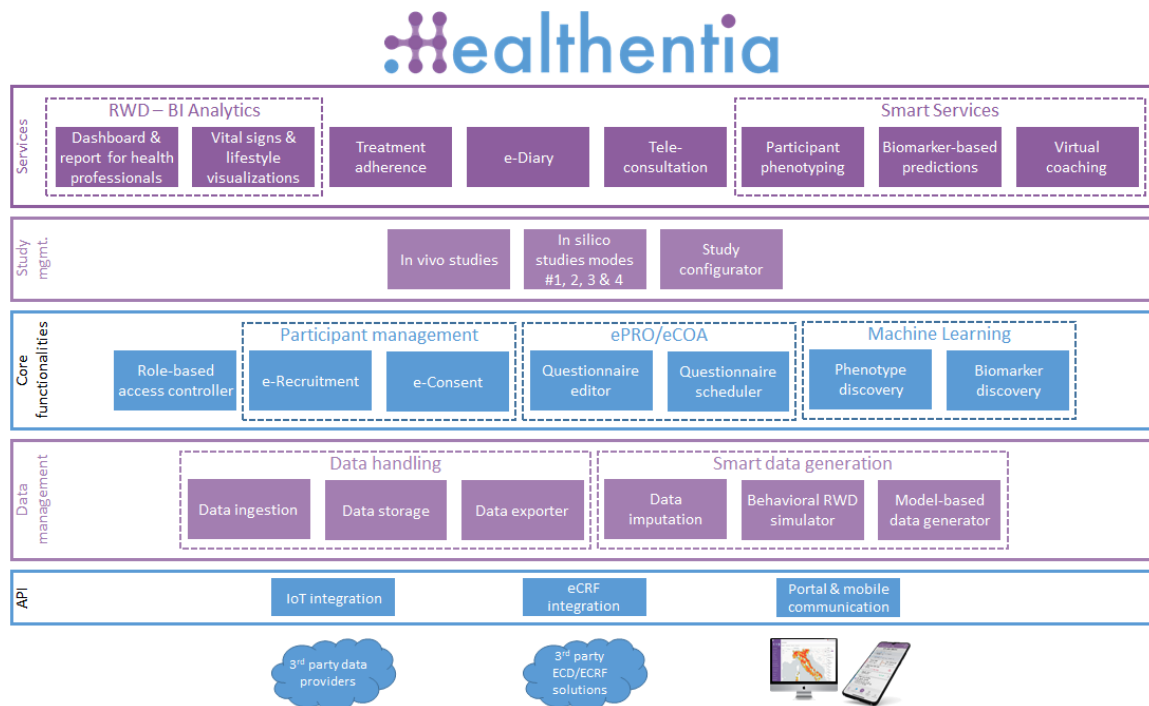
Figure 5: Healthentia high-level architecture.

The Healthentia API layer provides the means to connect Healthentia with the outside world. Data importing (IoT integration) and exporting (eCRF integration), as well as internal communication with the portal and mobile apps are facilitated through it.

The low-level operations on the data are hosted in the data management layer. The data handling functionalities are utilised by the API I/O endpoints. The smart data generation functionalities drive synthetic data generation.

The Healthentia core layer comprises of high-order functionalities on top of the data, like role-based control, participant management, participants' reports management and ML functionalities. Please note that the reporting and the ML functionalities of Healthentia are not used in iHelp. These are provided by the iHelp Big Data Platform.

The study layer allows managing of the studies, the entities in which HCP, participants and their data are organized. They can be formal clinical studies, or informal ones managed by pharmaceutical companies, hospitals, or research centres.

Finally, the services layer implements the necessary functionalities of the web portal and the mobile application. These include dashboard services in both portal and mobile apps, e-diary for reporting by study participants and other services like treatment adherence, teleconsultation and smart services.

The two applications providing the user interface for study participants (mobile app) and the HCP (web portal) are detailed in the following subsections.

4.1 Mobile app for study participants

The Healthentia mobile application enables data collection at the study participant end. Screenshots are shown in Figure 6, including the main activity with all the widgets, the info drawer where the only reference to Healthentia is visible, the sleep widget details and the nutrition widget details to name a few components.
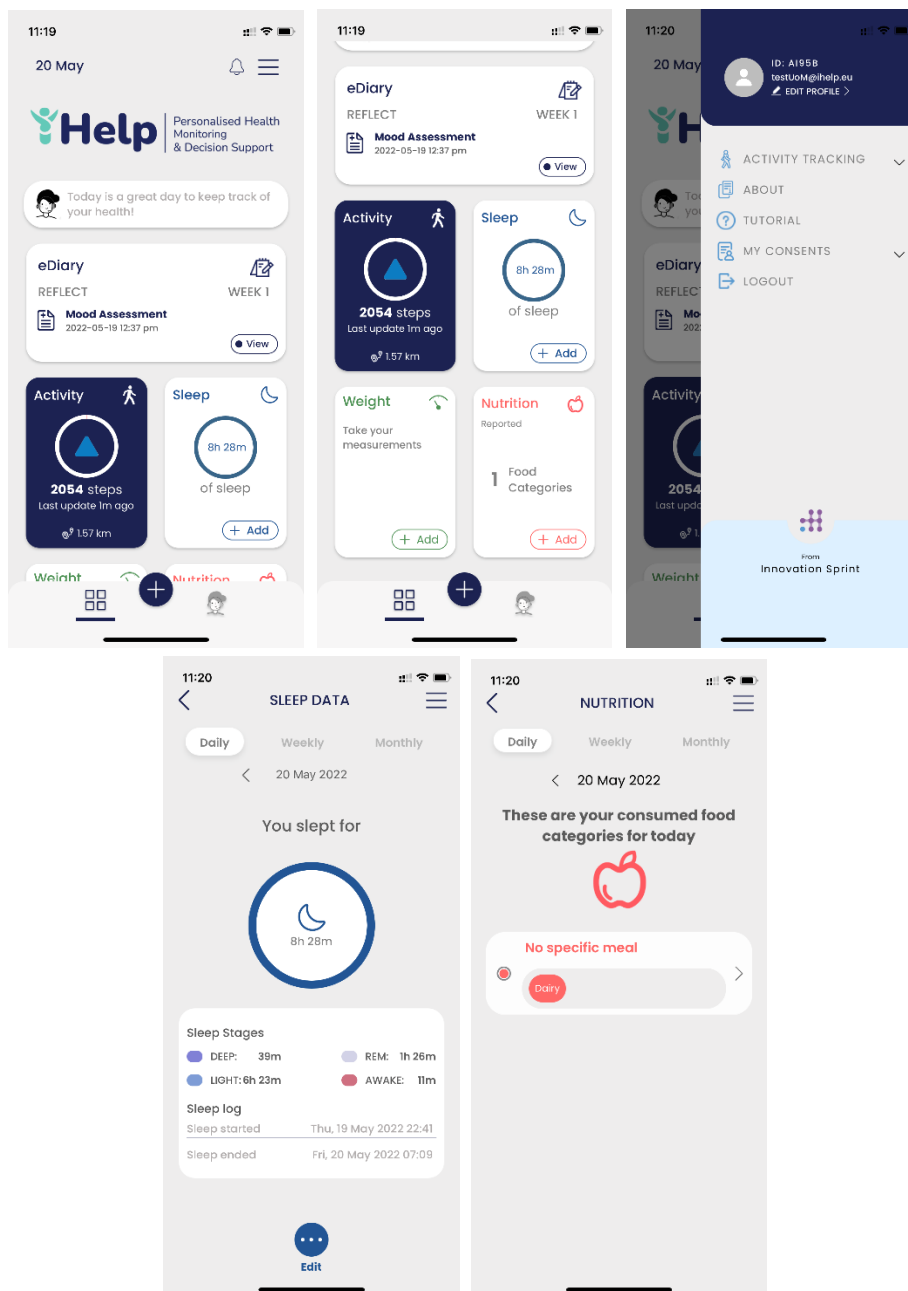
Figure 6: Healthentia mobile application.

Measurements are obtained from IoT devices, third-party mobile services or a proprietary sensing service. Study participants' reports are obtained via answering questionnaires that are either regularly pushed to the study participants' phones or are accessed on demand by the study participants themselves. Both the measured and reported data are displayed to the study participants, together with any insights offered by the smart services of the platform.

The Healthentia mobile app can be obtained from the Google and Apple app stores, but once an iHelp study participant is logged in, they are offered a version branded with iHelp logos and colours, offering an iHelp look and feel, unified across all five iHelp studies.

4.2 Web portal for study administrators

The Healthentia web portal application targets the study administrators. After logging in using two-factor authentication, and depending on their role, study administrators will have access to one, or all the iHelp studies, in an overview shown in Figure 7.
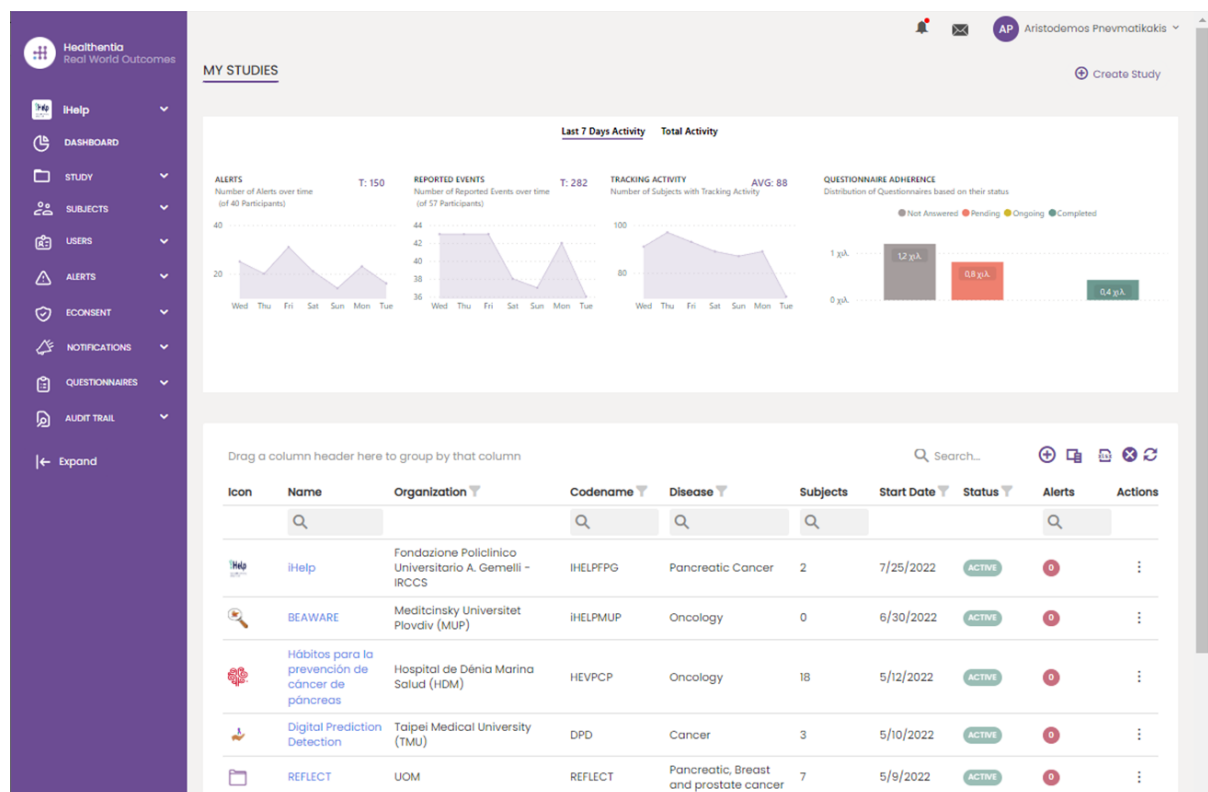
Figure 7: Healthentia portal application – viewing iHelp studies.

Selecting their study, the study administrators land on the study configuration pages. General configuration (see Figure 8) is about the study name and purpose, including the use of icons to give a study-specific look both at the portal, but mainly at the mobile app. This is followed by the per-study mobile application configuration. The configuration of the mobile app is about consents and widgets, since the questionnaires themselves are handled by a different portal mechanism.

The study participant registration section (Figure 9) covers items of the general terms, as well as different consents that can be asked of the participant. The widgets' section follows (Figure 9 and Figure 10), where they can be enabled and some of their features can be configured.
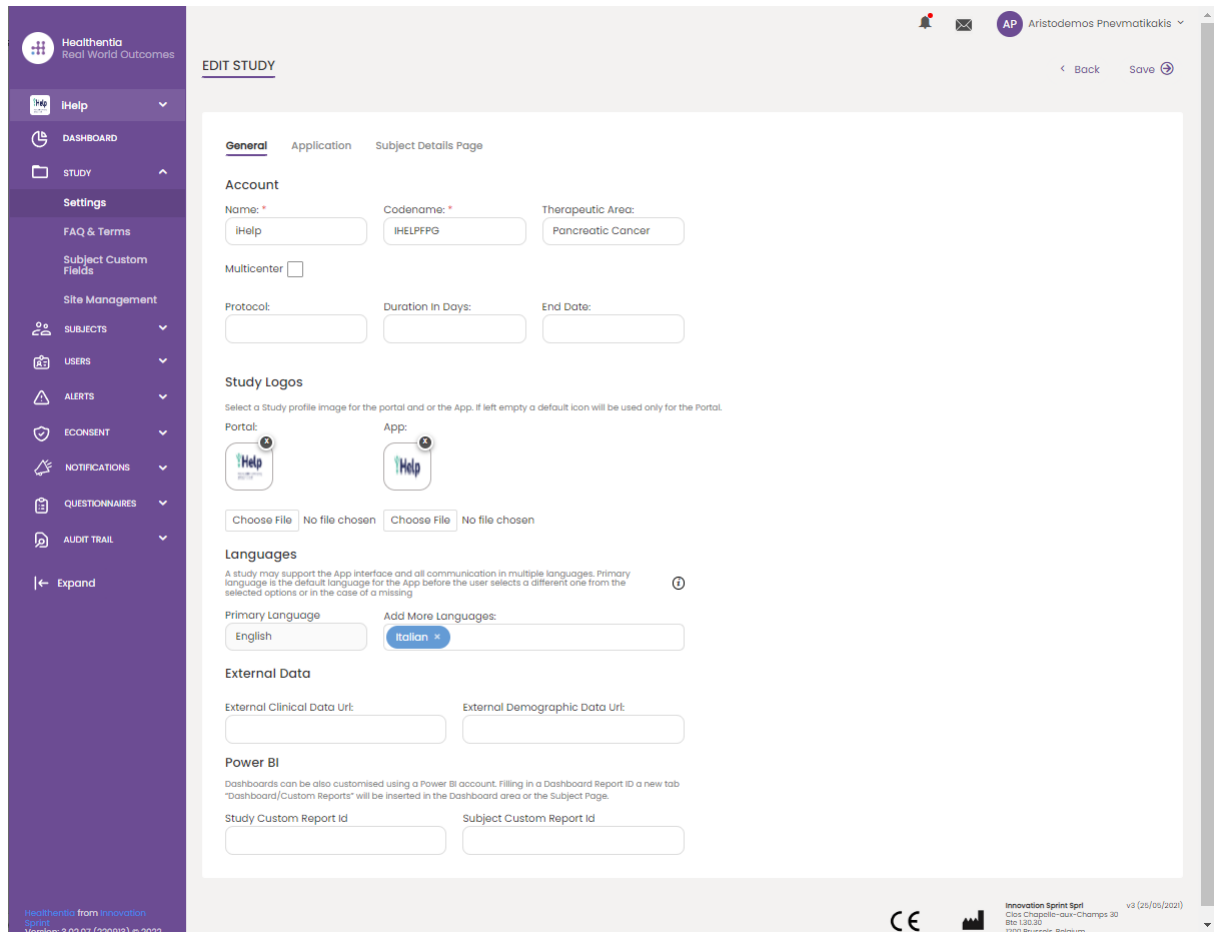
Figure 8: Study general configuration.
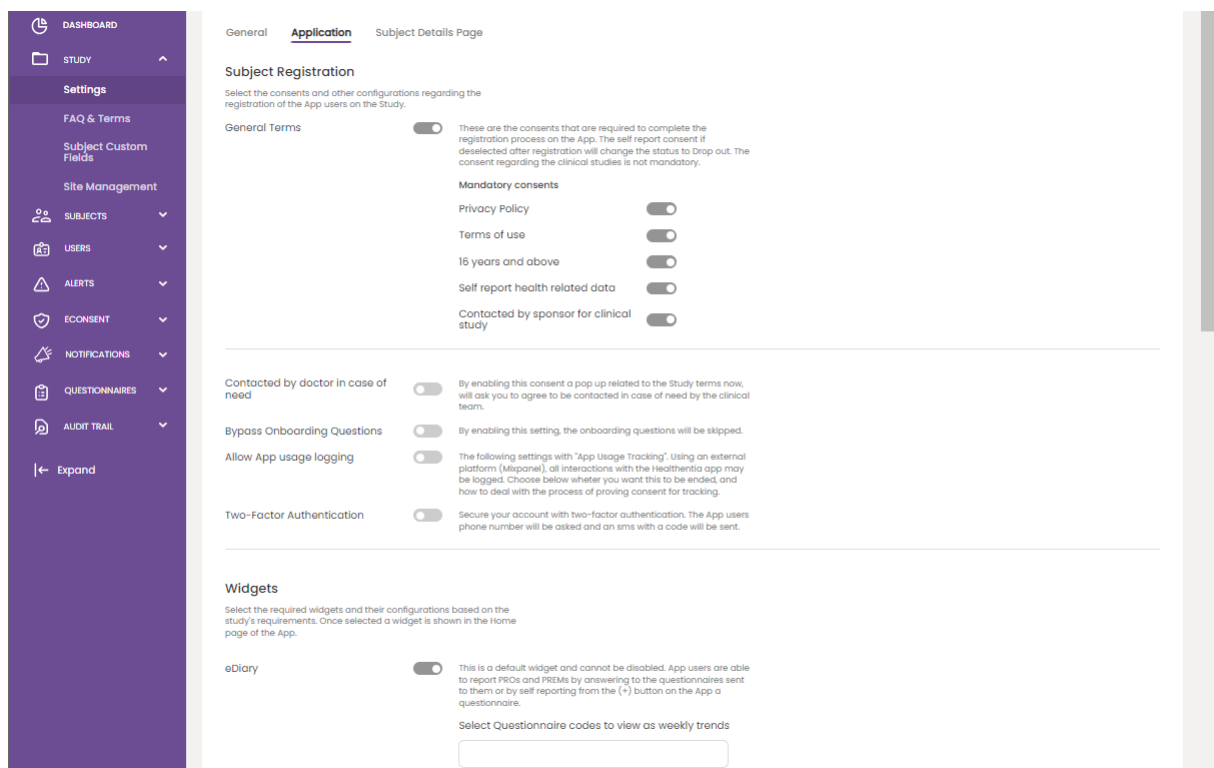
Figure 9: Per-study mobile app configuration – Registration terms and e-Diary.
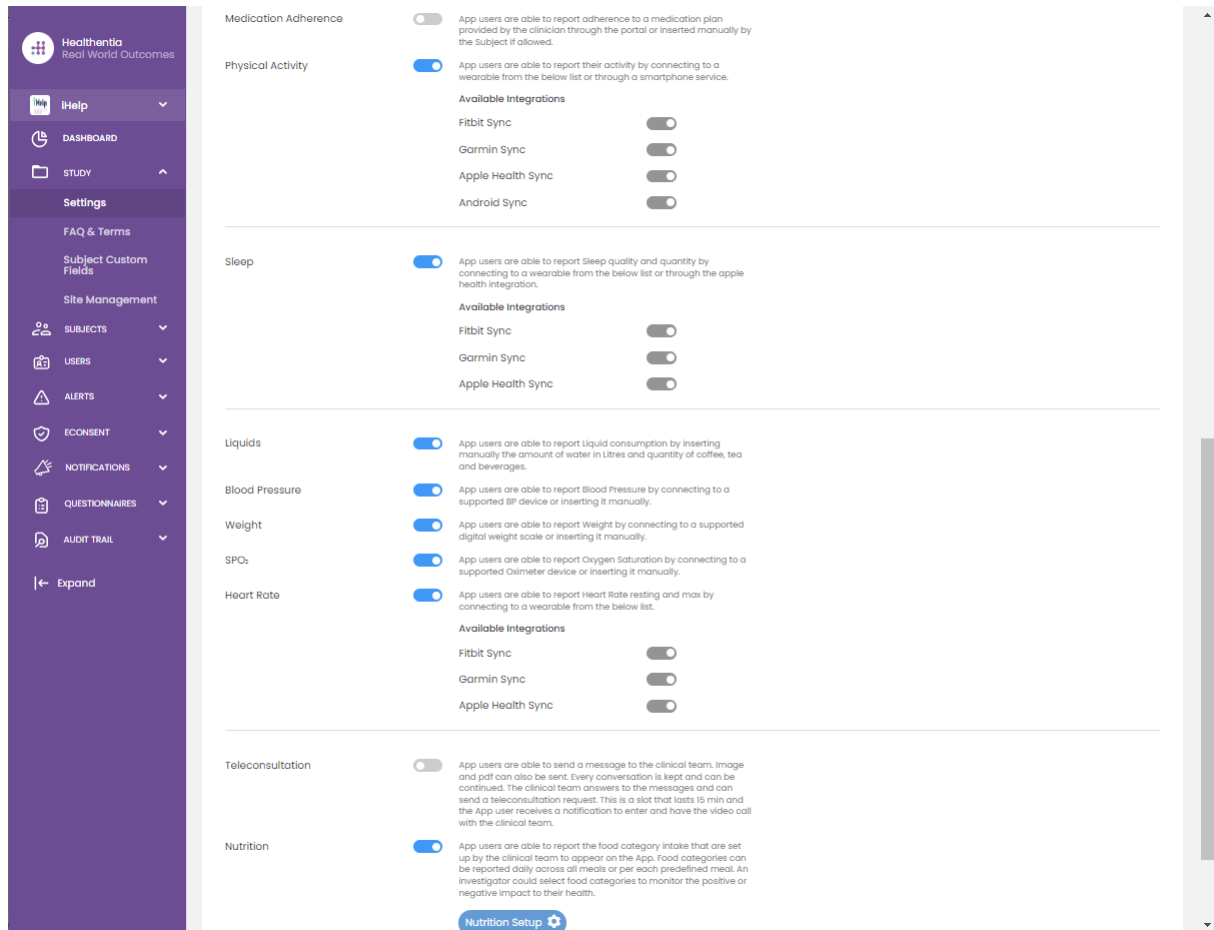
Figure 10: Per-study mobile app configuration – Enabling and configuring widgets.

In particular for the configuration of the nutrition widget a modal dialog is used for the HCP to select the nutrition items of interest from a hierarchical list (Figure 11). Only those selected are then shown to the study participant as options to select from.
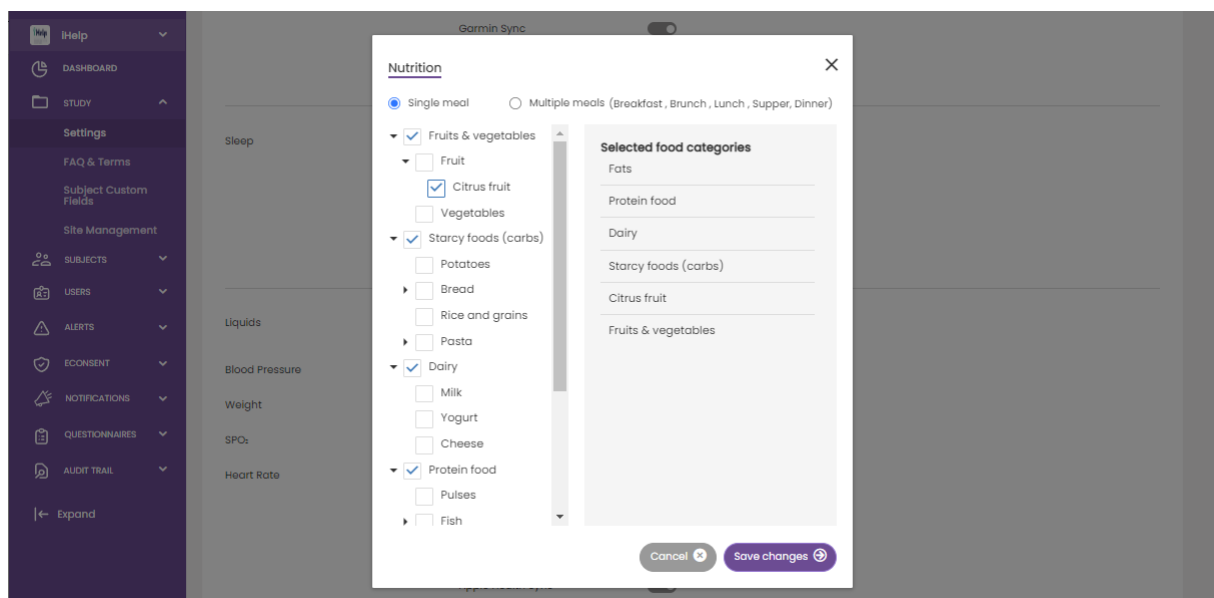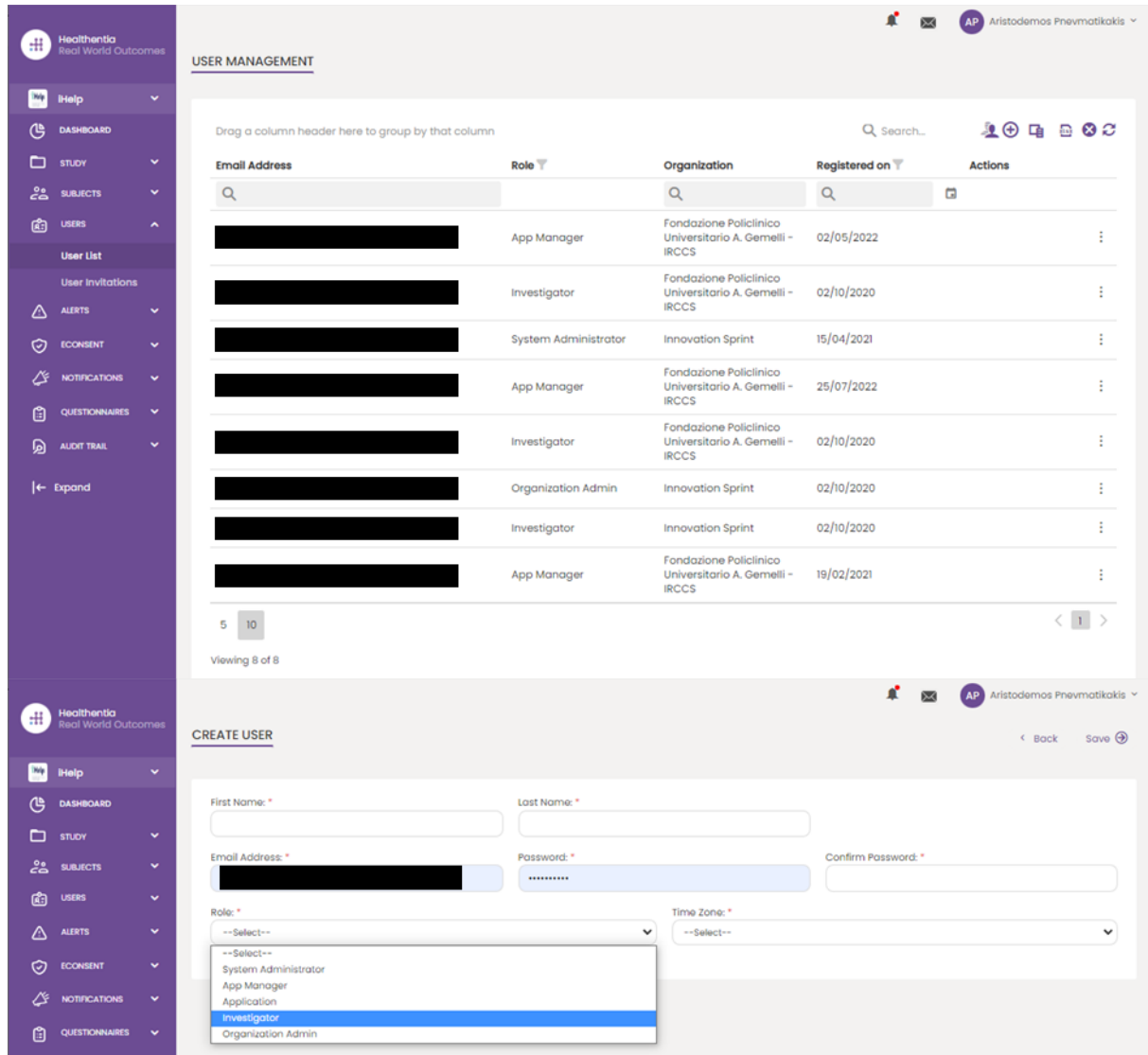
Figure 11: Per-study mobile app configuration – Configuring the nutrition widget.

More than one study administrators can be associated with each study. This is facilitated by the study user management shown in Figure 12. The current study administrators associated with the study are listed, together with their roles and organizations. More study administrators can be invited to the study, if they are existing users of Healthentia, or can be created as new users.

Figure 12: Study HCP user management – list of users (top) and creation of a new one (bottom).

The bulk of the secondary data that is reported to iHelp, is collected via questionnaires. The questionnaires of each study are listed, indicating the languages they are offered in and the number of questions they comprise of. The list is shown in Figure 13.

Every questionnaire can be edited. Initially, some info about the questionnaire, including possible graphics, is followed by the list of questions (and page separators), as shown in Figure 14. Each question can then be edited, to configure it and its possible answers (Figure 15). Finally, the routing between questions is arranged (Figure 16), with routing to the next question being the default. Routing can be fixed across all answers, but also for an individual answer.
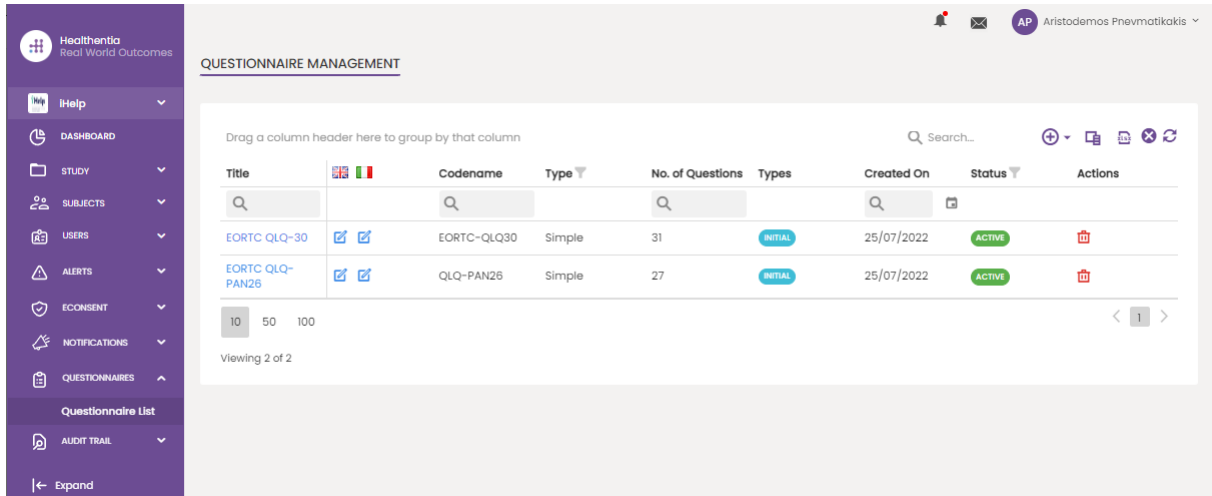
Figure 13: List of questionnaires.
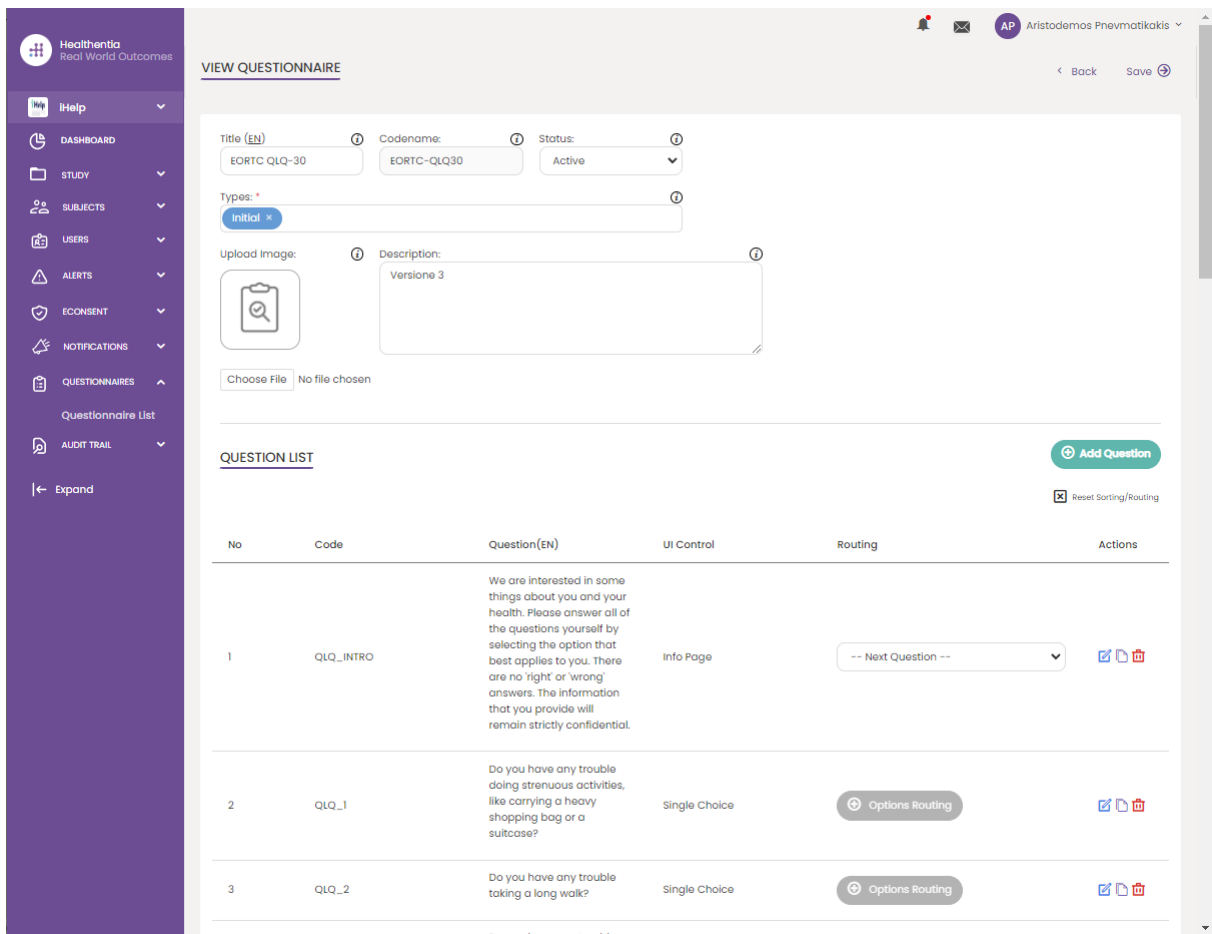
Figure 14: Editing a questionnaire.
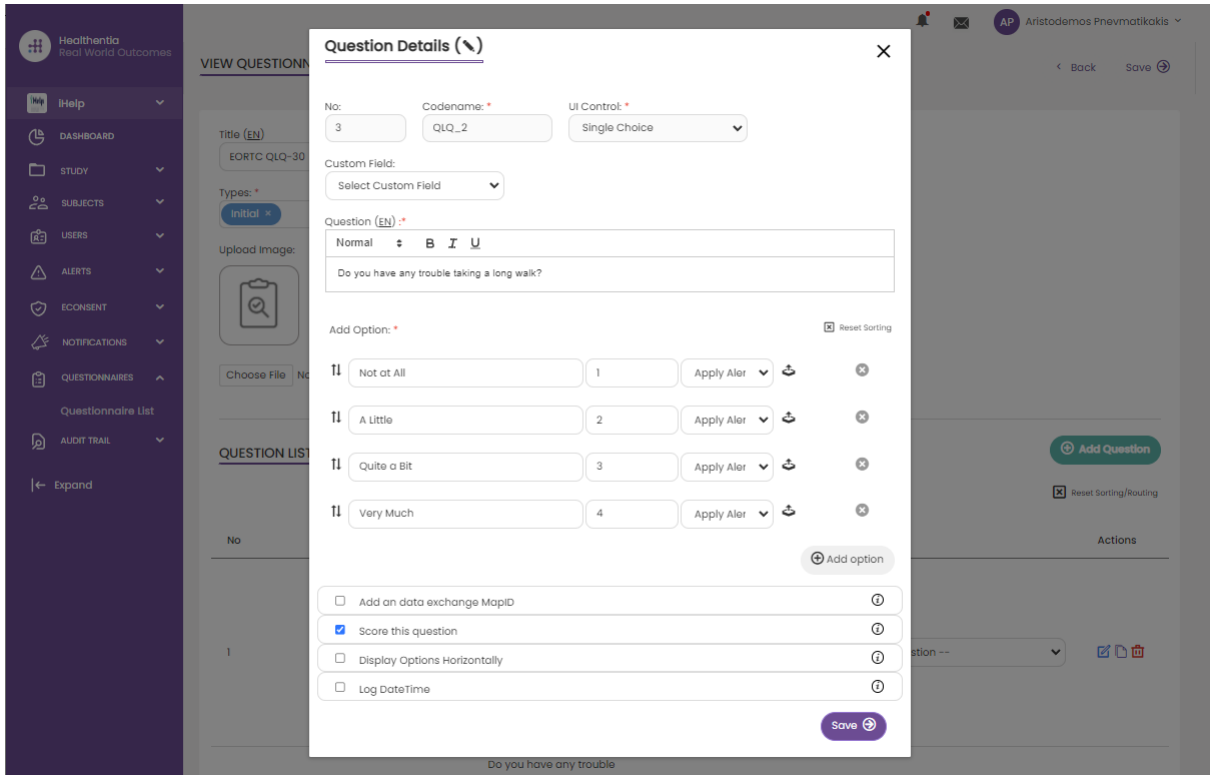
Figure 15: Configuring a question and its answers.
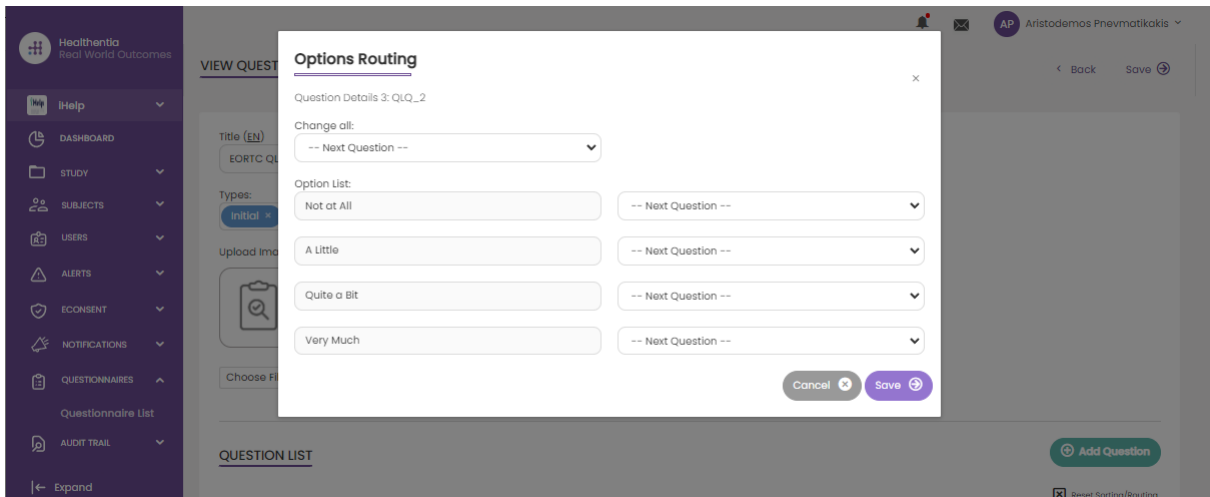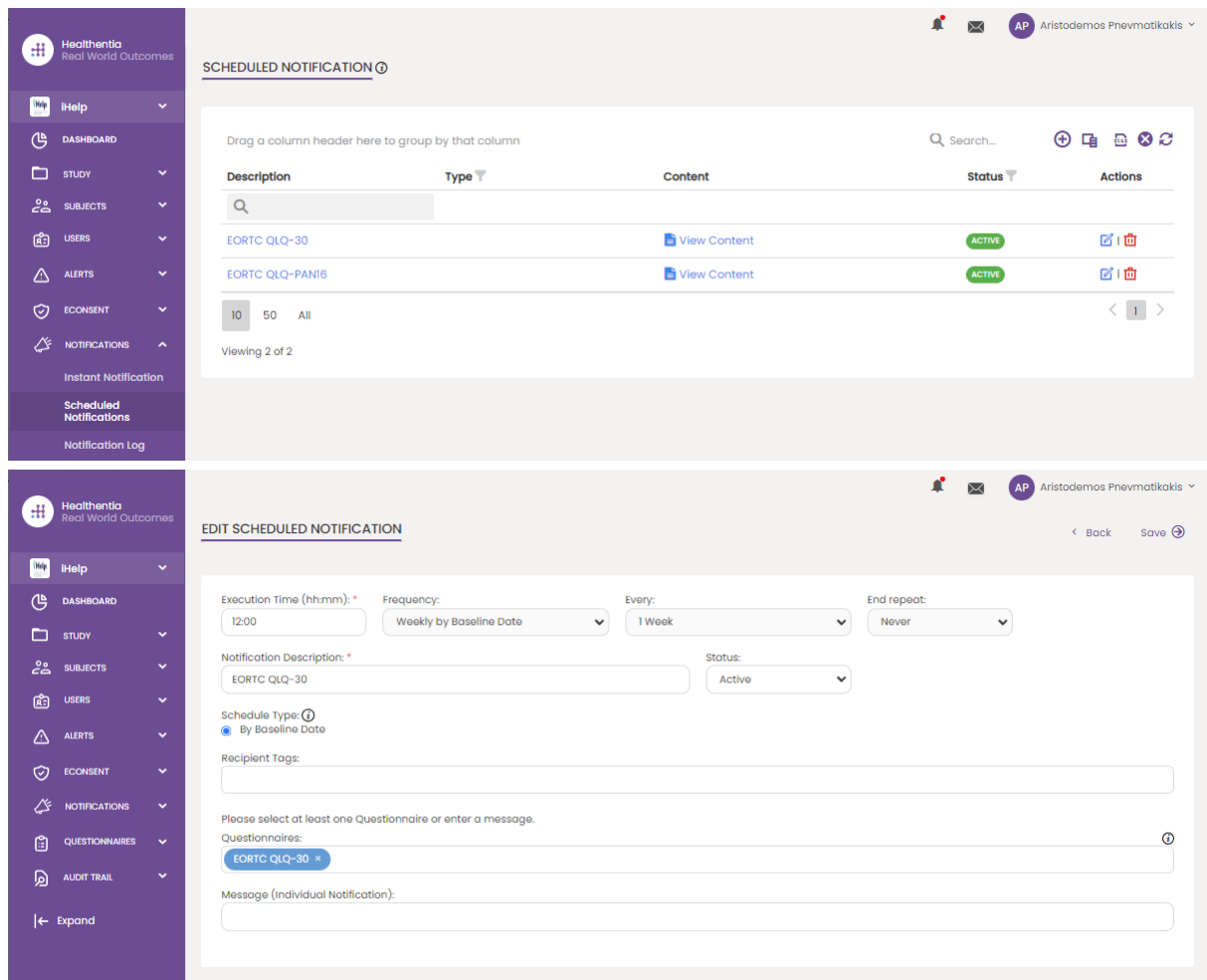
Figure 16: Routing between questions – generic and response-specific.

The questionnaires are pushed to the study participants for answering via the scheduled notifications mechanism. All configured notifications are listed as in Figure 17. Each notification can be configured regarding its timing, and the questionnaire it will forward. It is also possible to address a questionnaire to study participants based on their condition, as it is expressed by tags associated with them.

Figure 17: Pushing questionnaires to study participants via scheduled notifications – Notifications' list (top) and details (bottom).

Apart from configuring the studies, the Healthentia portal web application facilitates study participants' monitoring by the study administrators. This can be done at a study level with a registration, adherence, demographics and symptoms' overview, as well as with a physical activity overview. For an example see Figure 18, but note that the data come from an informal study currently running for some time, not from any of the iHelp ones that have just been configured.
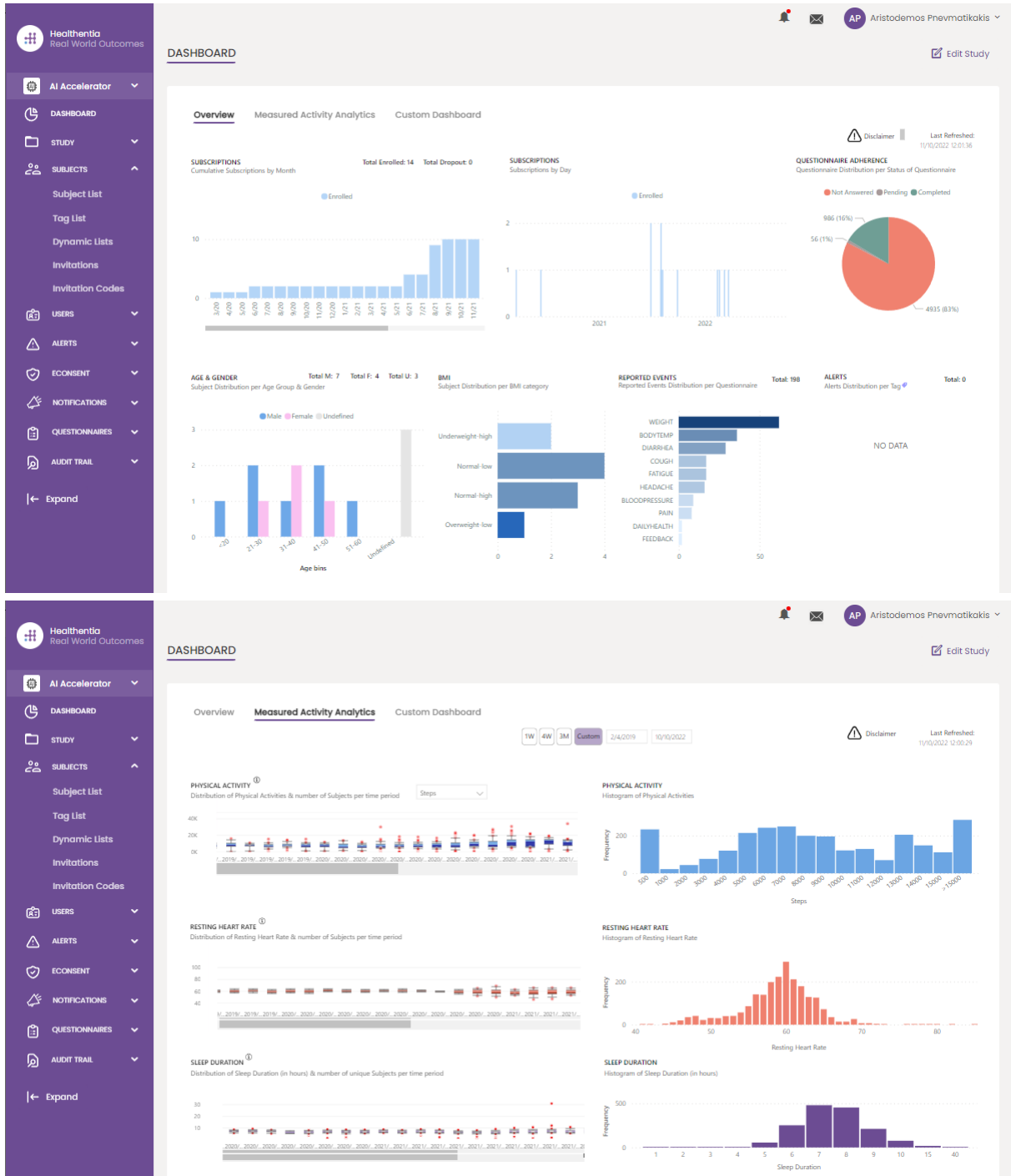
Figure 18: Study-level overview – Registration, adherence, demographics and symptoms (top) and physical activity (bottom).
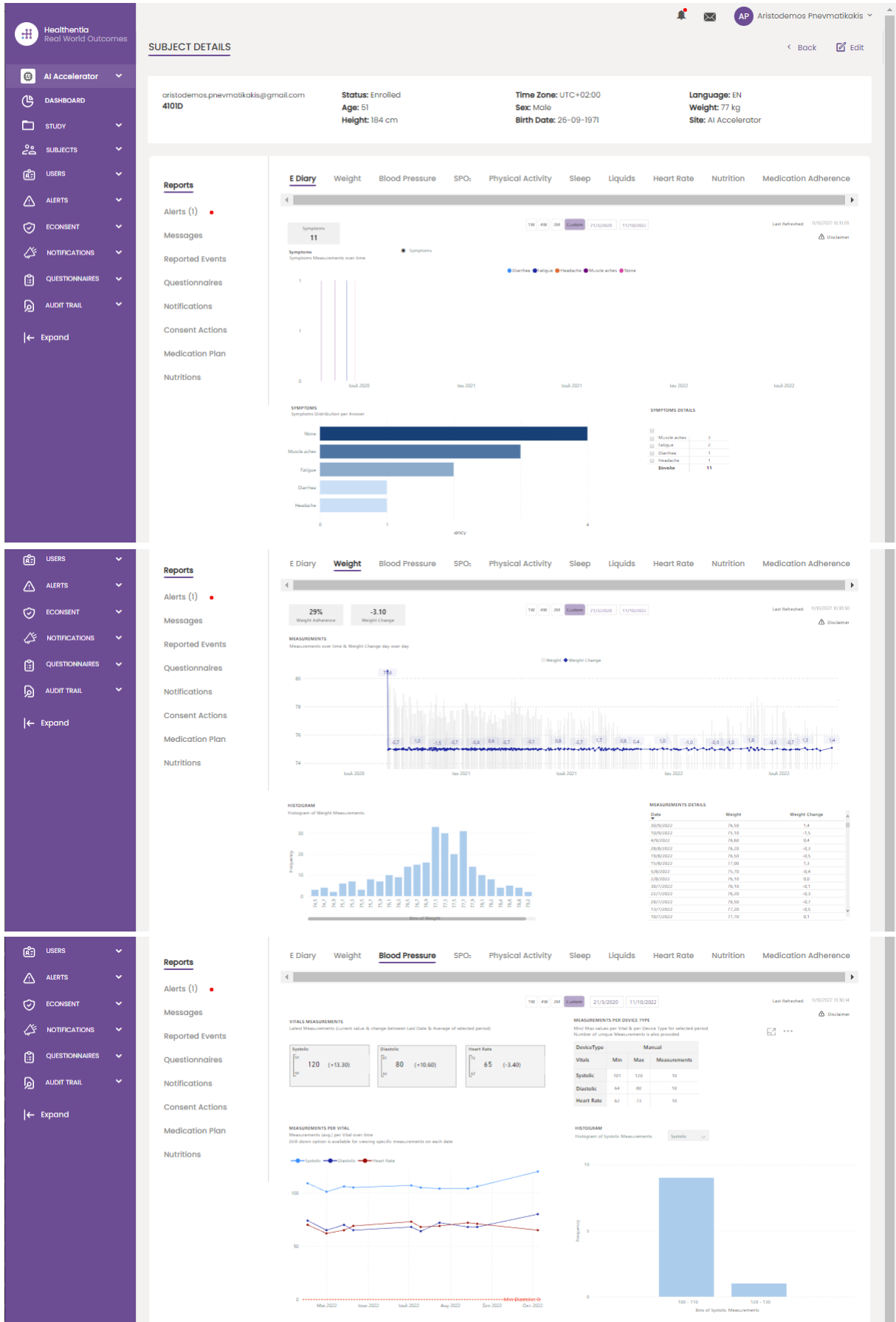
Figure 19: Participant-level overview – e-Diary (top), weight (middle) and blood pressure (bottom).
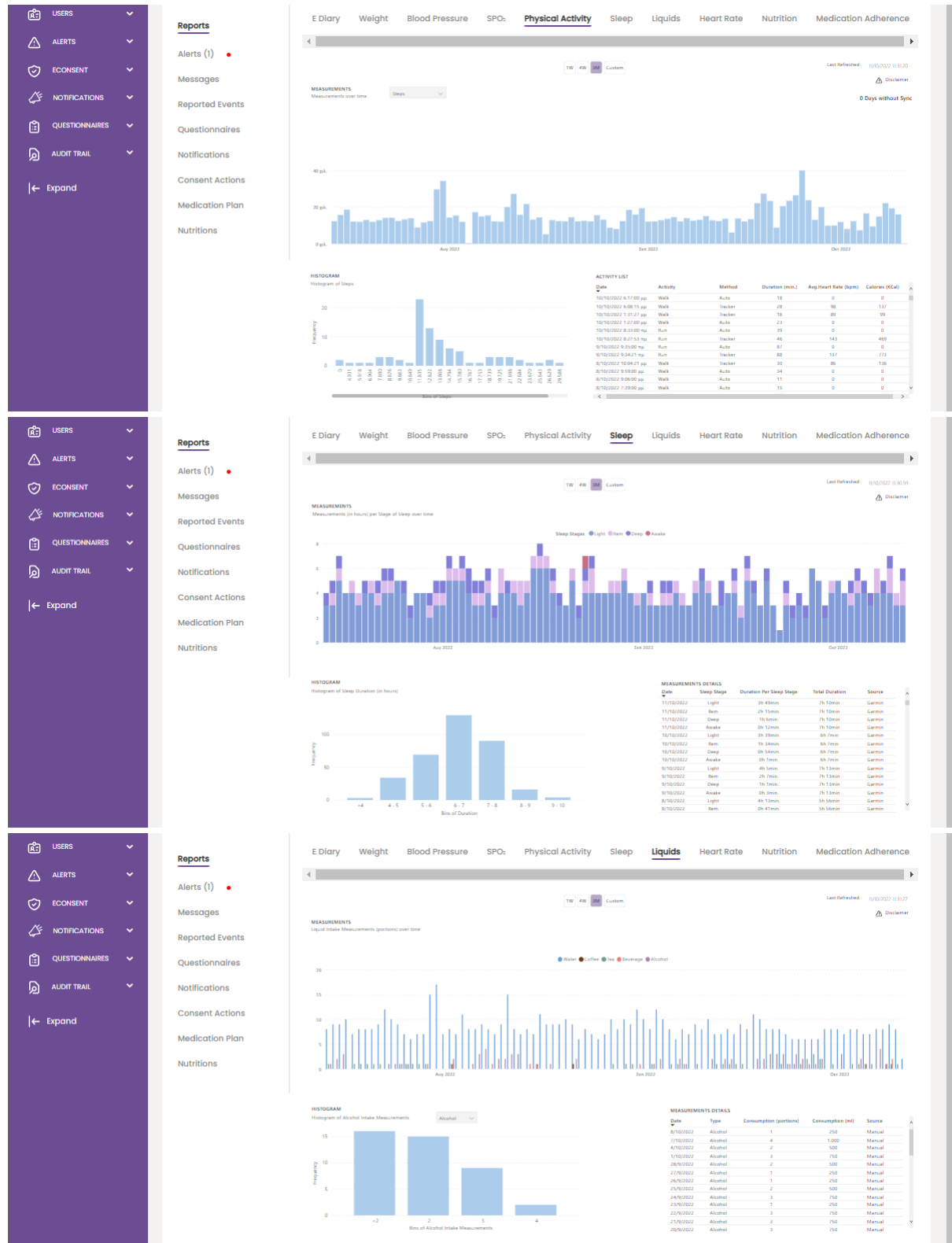
Figure 20: Participant-level overview – physical activity (top), sleep (middle) and liquid consumption (bottom).

The monitoring can also be done for the individual study participant. In the following Figure 19, Figure 20, Figure 21 and Figure 22 secondary data about one of the deliverable authors is shown, and he explicitly consents on doing so. The data span the e-Diary, weight, blood pressure, physical activity, sleep, liquid consumption, heart rate, nutrition, medication adherence and questionnaires.
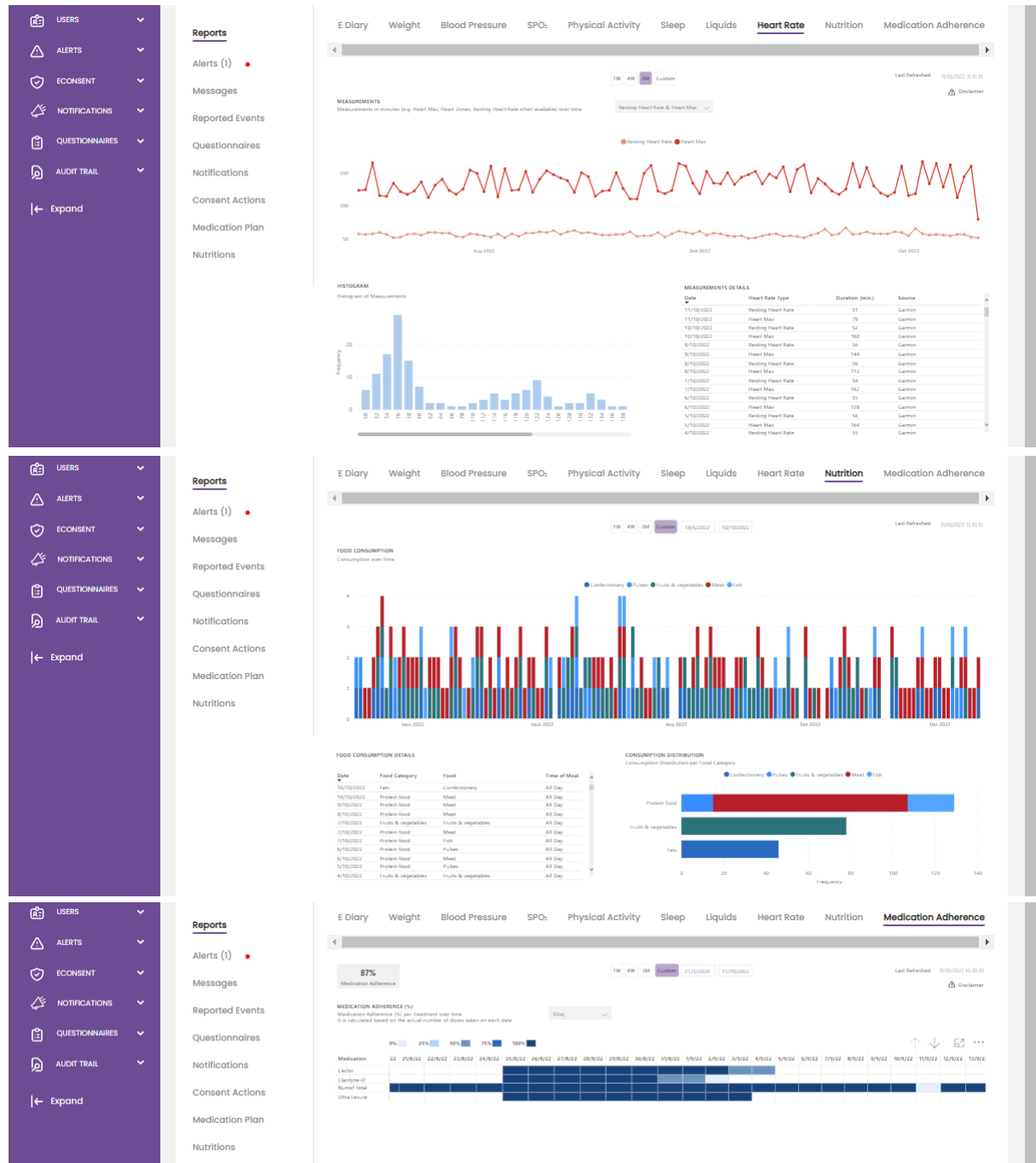
Figure 21: Participant-level overview – heart rate (top), nutrition (middle) and medication adherence (bottom).

Figure 22: Participant-level overview – questionnaires and their status.

5 iHelp secondary data catalogue

Section 2 concluded with the iHelp clinical partner's wish-list for secondary data to be collected, while section 3 gave an overview of the ways secondary data can be collected and section 4 detailed how the collection is performed using Healthentia. In this section, all this information is brought together, building the iHelp secondary data inventory. In the following subsections, first the secondary data sources are detailed, and then the secondary data inventory is built.

5.1 Secondary data sources

The automated measurements, questionnaires and widgets used for the iHelp secondary data collection are detailed in this subsection.

5.1.1 iHelp automated measurements

The iHelp automated measurements are shown in Table 4. For many attributes alternative measuring options and their shortcomings are given.

Table 4: iHelp automated measurements.

Category	Attributes	Measuring options
Physical activity	Steps	Activity trackers via API (Garmin, Fitbit) Mobile phone via SDK (Android, Apple Health Kit)
	Floors Climbed	
	Energy burned	
	Time in different activity zones	Only with activity trackers
Exercise sessions	Start time	Activity trackers via API (Garmin, Fitbit) Mobile phone via SDK (Android, Apple Health Kit)
	Duration	
	Energy burned	
	Type	Multiple with API, only walk & run with Android SDK
Heart	Resting heart rate	Activity trackers via API (Garmin, Fitbit)
	Max heart rate	
	Time in different heart rate zones	Fitbit only, Garmin under investigation
Sleep	Start time	Activity trackers via API (Garmin, Fitbit) Mobile phone via SDK (Android, Apple Health Kit)
	Duration	
	Time in different sleep stages	Only with activity trackers

5.1.2 iHelp scheduled questionnaires

The iHelp scheduled questionnaires are detailed in Table 5. For every questionnaire an analysis of the implementation is also given.

Table 5: iHelp scheduled formal questionnaires.

Questionnaire	Frequency	Scoring	Questions	Implementation
Pittsburgh Sleep Quality Index [5]	Monthly	7 component scores summarised into a single global score	Two numerical and two time of day questions	Numerical and time input
			20 questions to choose single out of four options	Single choice UI
The Warwick-Edinburgh Mental Wellbeing Scales [9]	Start, middle and end of intervention	N/A	14 statements about feelings and thoughts, select one out of five options about their frequency in the past two weeks	Single choice UI
Rosenberg Self-Esteem Scale [10]	Start, middle and end of intervention	Each statement is given 1 (strongly disagree) to 4 (strongly agree) points, negative statements 2, 5, 6, 8, 9 are reverse scored	10 statements about self-esteem, select one out of four options (strongly agree, agree, disagree, strongly disagree)	Single choice UI matrix with 4 columns
Body Awareness Questionnaire [11]	Start, middle and end of intervention	Total scale score as a sum of the items, some being reverse-scored	18 statements about body awareness, scored on a 1-7 scale	Single choice UI matrix with 7 columns
AUDIT questionnaire	Start, middle and end of intervention	N/A	10 statements about drinking	Single choice UI
Locus of Control Questionnaire [13]	Start, middle and end of intervention	Total score as a sum of the items, some being reverse-scored	10 binary questions, each scoring 2 or 0	Single choice UI matrix with 2 columns
Health Behaviour and Stages of Change Questionnaire [14]	Start, middle and end of intervention	N/A	5 groups of phrases related to different behaviours. 5th group only applies to women	Single choice UI
Health literacy (HLS-EU_Q16) [16]	Start and end of intervention	Sum of the items (does not handle the missing items)	16 questions on how easy it is to follow different aspects of health (easy: 1, hard: 0, do not know: missing)	Single choice UI matrix with 3 columns
Fagerstrom Test for Nicotine	Biweekly	N/A	6 questions about smoking	Single choice UI

Dependence				
Family history – Pancreatic cancer	Start of intervention	N/A	19 questions about family history, related to pancreatic cancer	Single choice UI
Comorbidities questionnaire	Start of intervention	N/A	45 questions about comorbidities	Single choice UI
Quality of Life (EQ-5D-5L) [18]	Start, middle and end of intervention	N/A	5 dimensions, 3 or 5 levels	Single choice UI matrix with 3 or 5 columns
Quality of life for patients (EORTC QLQ-C30) [19]	Bi-monthly	N/A	28 questions with 5 levels	Single choice UI matrix with 4 columns
			2 questions with 7 levels	Single choice UI matrix with 7 columns
Quality of life for pancreatic cancer patients (EORTC QLQ - PAN26) [20]	Bi-monthly	N/A	A continuation of the above questionnaire with 26 more questions with 4 levels	Single choice UI matrix with 4 columns
Edmonton Symptom Assessment System [8]	Daily	N/A	Intensity of nine common cancer patients' symptoms (pain, tiredness, nausea, depression, anxiety, drowsiness, appetite, well-being and shortness of breath)	0-10 horizontal slide-bars in single page
			Pain location	If pain, branch to body map
Perceived stress scale	Weekly	N/A	14 questions about stress	Single choice UI

Please note that:

▪ The Edmonton Symptom Assessment System [5] has a final question about a custom patient reported symptom that cannot be automatically handled by the system. All symptoms of interest must be pre-defined. Other or the same symptoms can be collected also via the symptom ad hoc questionnaires shown in Table 6.

▪ The Health Literacy questionnaire (HLS-EU_Q16) [16] allows a third "do not know" answers that have no scoring info. The scoring system needs to be considered.

5.1.3 iHelp ad hoc questionnaires & reported measurements

The the ad hoc questionnaires and manually reported measurements used in iHelp are shown in Table 6. Their implementation with UI elements is also discussed. The symptoms are from [8].

Table 6: iHelp ad hoc questionnaires and manually reported measurements.

Attribute	Implementation
Height	Numeric entry of manually measurement using non-integrated scales
Mood	Self-assessment using 5-level smileys

Pain symptom [8]	0-10 horizontal slide-bar, in not 0 also pain location body map
Tiredness, nausea, depression, anxiety, drowsiness, appetite, well-being and shortness of breath symptoms [8]	0-10 horizontal slide-bar
Supportive messaging	Free-form text input, capturing a message from a study participant towards all others. This message is pushed to the other study participants only after moderation from an HCP.
Overall health rating by participants [15]	Single question with six answers using the single choice UI
Pictorial questionnaire for self-assessment (Rockwood Frailty index) [22]	Single question with nine answers using the single choice UI

Please note that the frailty measure can be the specified ad hoc questionnaire, but it can also be the result of a composite measurement, having the study participants perform a sequence of scripted physical exercises, monitored with the mobile phone sensors.

5.1.4 iHelp widgets

Four widgets are currently envisioned for the iHelp secondary data collection: Sleep, liquids, nutrition and e-diary. All these widgets can be enabled or disabled by the study administrators.

The role of the sleep widget is to view sleep duration and stages, as well as edit sleep duration. Its appearance on the main screen and its details once tapped are shown in Figure 23.
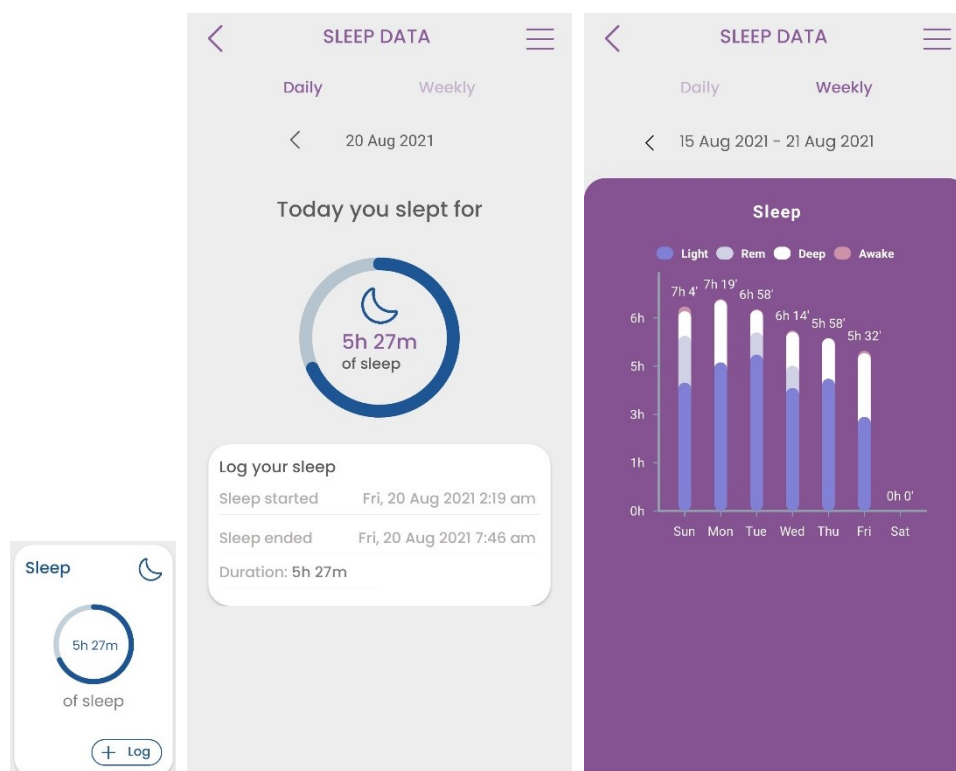
Figure 23: Sleep widget in main screen (left). Daily and weekly views (middle and right). The daily view facilitates edit.

The role of the liquids' widget is to view and progressively or retrospectively add liquid consumption. Its appearance on the main screen and its details once tapped are shown in Figure 24. The widget has water consumption as the main attribute to be collected, but allows the study administrators to enable any of the coffee, tea, beverages and alcohol consumption attributes.

The nutrition widget is similar to the liquids one, only now there are different food categories to select from, that are enabled or disabled by the study administrator. The widget's appearance on the main screen and its details once tapped are shown in Figure 25.

The e-diary also displays reported information and allows access to the ad hoc questionnaires and manually reported measurements as shown in its daily and weekly details in Figure 26.

Finally, there are widgets that have only a data viewing and not a data collection role. These are the physical activity and the virtual coach widgets.
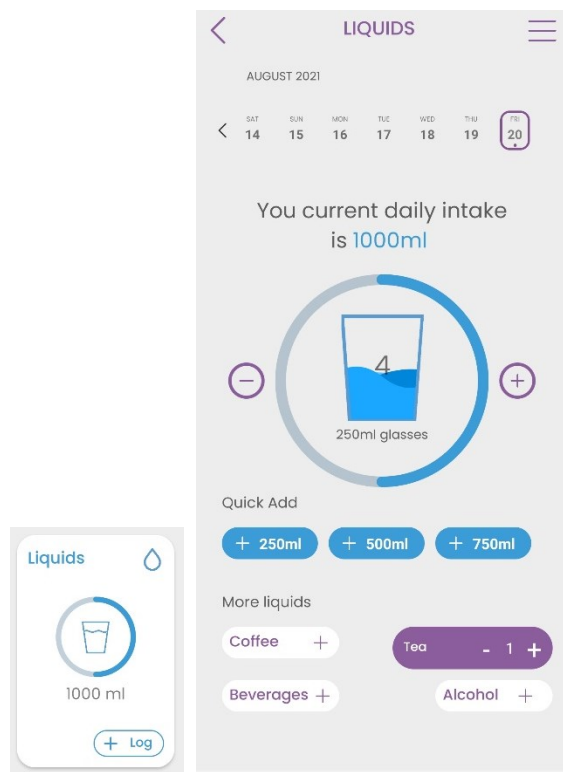
Figure 24: Liquids' consumption widget in main screen (left) and daily details for data entry (right).
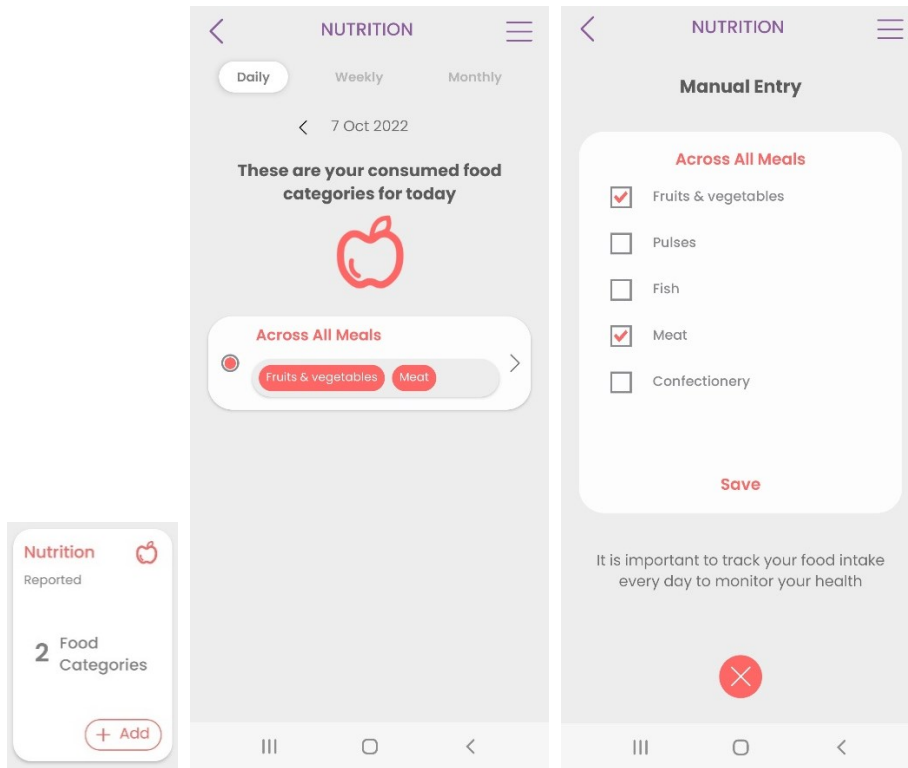
Figure 25: Nutrition widget in main screen (left) and details for information display & data entry (right)
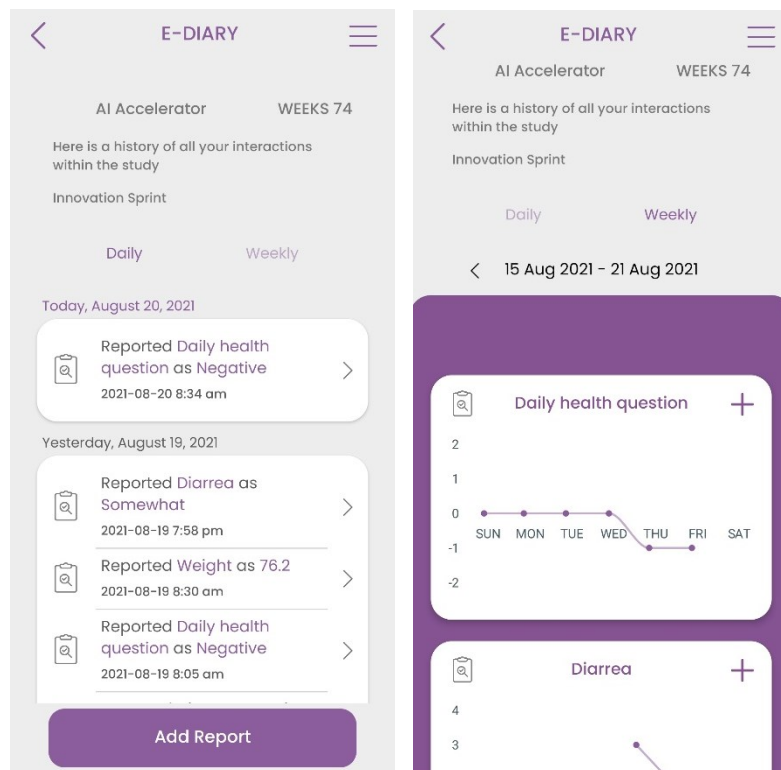
Figure 26: e-diary widget daily (left) and weekly (right) details. Data can be entered via the "Add report" button and the "+" UI elements.

5.2 Secondary data attributes

Subsection 5.1 detailed the secondary data sources for all iHelp pilots. The exact data attributes are shown in this subsection. The data attributes are first summarised in Table 7. In total there are 435 secondary data attributes.

Then, the secondary data attributes are detailed in Table 8, which is the secondary data catalogue for iHelp.

Table 7: Summary of data attributes across all iHelp studies.

Category	Type	Attributes	Total
Objective measurements	Daily activity	10	47
	Daily heart	8	
	Daily sleep	8	
	Activity sessions	21	
(Manual) measurements - widgets	Weight	2	12
	Blood pressure	5	
	Oxygen saturation	5	
Subjective - widgets	Liquids	6	16
	Treatment adherence	6	
	Nutrition	4	
Ad-hoc symptom questionnaires	Body temperature	2	12
	Cough	3	
	Diarrhea	2	
	Fatigue	2	
	Pain	3	
Subjective - questionnaires	AUDIT	11	360
	Body awareness	19	
	Comorbidity	46	
	Edmonton Symptom Assessment	14	
	EORTC QLQ-30	31	
	EORTC QLQ - PAN26	27	
	Fagerstrom Test for Nicotine Dependence	7	
	Family History - Pancreatic Cancer	20	
	General Health	2	
	Health Behaviour Stage of Change	6	
	Health Habits	18	
	Health Literacy Scale	33	
	Height	2	
	Locus of Control	11	
	Mood Assessment	2	

	Perceived Stress Scale	15	
	Pittsburgh Sleep Quality Index	27	
	Quality of Life	7	
	Rosenberg Self-Esteem Scale	21	
	Warwick Edinburgh Mental Wellbeing Scale	29	

Table 8: Secondary data catalogue.

Category	Subcategory	Attribute	Type	ID
Objective	Daily activity	Date	String	Date
		Steps walked	Int	Steps
		Distance travelled (meters)	Int	Distance
		Calories burned	Int	Calories
		Floors climbed	Int	Floors
		Lightly active minutes	Int	LightlyActive
		Moderately active minutes	Int	FairlyActive
		Highly active minutes	Int	VeryActive
		Sedentary minutes	Int	Sedentary
		Total active minutes	Int	TotalActive
	Daily heart	Date	String	Date
		Resting heart rate	Int	RestingHeartRate
		Min heart rate	Int	HeartMin
		Max heart rate	Int	HeartMax
		Out of range minutes	Int	HeartOutOfRangeMinutes
		Fat burn minutes	Int	HeartFatBurnMinutes
		Cardio minutes	Int	HeartCardioMinutes
		Peak minutes	Int	HeartPeakMinutes
	Daily sleep	Date	String	Date
		Sleep start (hours relative to midnight)	Float	SleepStart
		Sleep end (hours relative to midnight)	Float	SleepEnd
		REM minutes	Int	RemSleep
		Light minutes	Int	LightSleep
		Deep minutes	Int	DeepSleep
		Awake minutes	Int	AwakeTime
		Total minutes	Int	TotalSleep
	Activity sessions	Activity type	String	Title
		Start date/time	String	StartTime
		Active duration in miliseconds	Int	ActiveDuration
		Duration in miliseconds	Int	Duration
		Calories	Int	Calories
		Speed	Float	Speed
		Pace	Float	Pace
		Steps	Int	Steps
		Distance	Float	Distance
		Distance unit	String	DistanceUnit
		Elevation	Int	ElevationGain
		Average heart rate	Int	AverageHeartRate
		Out of range minutes	Int	OutOfRangeMinutes
		Fat burn minutes	Int	FatBurnMinutes
		Cardio minutes	Int	CardioMinutes
		Peak minutes	Int	PeakMinutes
		Heart rate zone definitions	String	HeartRateZones
		Lightly active minutes	Int	LightlyMinutes
		Moderately active minutes	Int	FairlyMinutes
		Highly active minutes	Int	VeryMinutes
		Sedentary minutes	Int	SedentaryMinutes
(Manual) measurements	Weight	Weight (kg)	Float	Weight
		Date	String	Date
	Blood pressure	Diastolic (mmHg)	Float	Diastolic
		Systolic (mmHg)	Float	Systolic
		Heart rate (bpm)	Float	HeartRate
		Arythmia	Boolean	Arythmia
		Date time	String	DateTime
	Oxygen saturation	Date time	String	DateTime
		Oxygen (%)	Int	Oxygen
		Pulse (bpm)	Int	Pulse
		PI	Int	PI
		Notes	String	Notes
Subjective	Liquid consumption	Date	String	Date
		Water (glass - 250ml)	Int	Water
		Coffee (cup)	Int	Coffee
		Tea (cup)	Int	Tea
		Refreshment (glass - 250ml)	Int	Refreshment
		Alcohol (serving, depending on type)	Int	Drink
	Treatment reminder (adherence)	Date	String	Date
		JSON description of progress	String	DosesTaken
		Number of doses	Int	DailyDoses
		Unit	String	Unit
		Quantity per dose	String	Dose

		Treatment name	String	TreatmentName
	Nutrition	Date	String	Date
		Name of food category	String	Name
		ID of food category	Int	Id
		Meal of the day	String	Meal
	AUDIT Questionnaire	Date time	String	DateTime
		How often do you have a drink containing alcohol?	Text	AUDIT_1
		How many drinks containing alcohol do you have on a typical day when you are drinking?	Text	AUDIT_2
		How often do you have six or more drinks on one occasion?	Text	AUDIT_3
		How often during the last year have you found that you were not able to stop drinking once you had started?	Text	AUDIT_4
		How often during the last year have you failed to do what was normally expected from you because of drinking?	Text	AUDIT_5
		How often during the last year have you needed a first drink in the morning to get yourself going after a heavy drinking session?	Text	AUDIT_6
		How often during the last year have you had a feeling of guilt or remorse after drinking?	Text	AUDIT_7
		Have you or someone else been injured as a result of your drinking?	Text	AUDIT_9
		How often during the last year have you been unable to remember what happened the night before because you had been drinking?	Text	AUDIT_8
		Has a relative or friend or a doctor or another health worker been concerned about your drinking or suggested you cut down?	Text	AUDIT_10
	Body Awareness Questionnaire	Date time	String	DateTime
		I notice differences in the way my body reacts to various foods.	Numerical	BAQ_1
		I can always tell when I bump myself whether or not it will become a bruise.	Numerical	BAQ_2
		I always know when I've exerted myself to the point where I'll be sore the next day.	Numerical	BAQ_3
		I am always aware of changes in my energy level when I eat certain foods.	Numerical	BAQ_4
		I know in advance when I'm getting the flu.	Numerical	BAQ_5
		I know I'm running a fever without taking my temperature.	Numerical	BAQ_6
		I can distinguish between tiredness because of hunger and tiredness because of lack of sleep.	Numerical	BAQ_7
		I can accurately predict what time of day lack of sleep will catch up with me.	Numerical	BAQ_8
		I am aware of a cycle in my activity level throughout the day.	Numerical	BAQ_9
		I don't notice seasonal rhythms and cycles in the way my body functions.	Numerical	BAQ_10
		As soon as I wake up in the morning, I know how much energy I'll have during the day.	Numerical	BAQ_11
		I can tell when I go to bed how well I will sleep that night.	Numerical	BAQ_12
		I notice distinct body reactions when I am fatigued.	Numerical	BAQ_13
		I notice specific body responses to changes in the weather.	Numerical	BAQ_14
		I can predict how much sleep I will need at night in order to wake up refreshed.	Numerical	BAQ_15
		When my exercise habits change, I can predict very accurately how that will affect my energy level.	Numerical	BAQ_16
		There seems to be a best time for me to go to sleep at night.	Numerical	BAQ_17
		I notice specific bodily reactions to being overhungry.	Numerical	BAQ_18
	Body Temperature	Date time	String	DateTime
		What is your body temperature?	Numerical	FEVER
	Comorbidity Questionnaire	Date time	String	DateTime
		At what age have you been diagnosed with Gastric/Duodenal ulcer?	Numerical	ULCER_AGE
		Please tell the name of the other condition(s) that you have been told by the doctor.	Text	OTHER_SPECIFY
		At what age have you been diagnosed with Periodontal disease?	Numerical	PD_AGE
		At what age have you been diagnosed with Acute pancreatitis?	Numerical	AP_AGE
		At what age have you been diagnosed with Choledocholithiasis (bile duct stones or gallstones)?	Numerical	CHOLEDOC_AGE
		At what age have you been diagnosed with Acute gastroduodenitis?	Numerical	AG_AGE
		At what age have you been diagnosed with Chronic pancreatitis?	Numerical	CP_AGE_4
		At what age have you been diagnosed with Cysts of the pancreas?	Numerical	CP_AGE_38
		At what age have you been diagnosed with Hypertension?	Numerical	HYPERTENSION_AGE
		At what age have you been diagnosed with Operations of the pancreas?	Numerical	OP_AGE
		At what age have you been diagnosed with Type 1 diabetes mellitus?	Numerical	T1DM_AGE
		At what age have you been diagnosed with Chronic gastroduodenitis?	Numerical	CG_AGE
		At what age have you been diagnosed with Type 2 diabetes mellitus?	Numerical	T2DM_AGE
		At what age have you been diagnosed with Gastric surgery?	Numerical	GS_AGE
		At what age have you been diagnosed with Gallbladder removal surgery (Cholecystectomy)?	Numerical	CHOLECYS_AGE
		At what age have you been diagnosed with Liver cirrhosis?	Numerical	LC_AGE
		At what age have you been diagnosed with Hepatitis by other reasons?	Numerical	HEPOTHER_AGE
		At what age have you been diagnosed with Helicobacter Pylori infection?	Numerical	HPI_AGE
		At what age have you been diagnosed with Hepatitis type B?	Numerical	HEPB_AGE
		At what age have you been diagnosed with Hepatitis type C?	Numerical	HEPC_AGE
		At what age have you been diagnosed with Cholecystitis?	Numerical	CHOLEC_AGE
		Have you ever been told by doctors that you have/had any of the following conditions? Gastric/Duodenal ulcer	Text	ULCER
		Have you ever been told by doctors that you have/had any of the following conditions? Systemic Lupus Erythematosus (SLE)	Text	SLE
		Have you ever been told by doctors that you have/had any of the following conditions? Acute pancreatitis	Text	ACUTEPANCREATIS
		Have you ever been told by doctors that you have/had any of the following conditions? Chronic pancreatitis	Text	CHRONICPANCREATIS
		Have you ever been told by doctors that you have/had any of the following conditions? Hypertension	Text	HYPERTENSION
		Have you ever been told by doctors that you have/had any of the following conditions? Type 1 diabetes mellitus	Text	T1DM
		Have you ever been told by doctors that you have/had any of the following conditions? Type 2 diabetes mellitus	Text	T2DM
		Have you ever been told by doctors that you have/had any of the following conditions? Liver cirrhosis	Text	LIVERCYRRHOSIS
		Have you ever been told by doctors that you have/had any of the following conditions? Hepatitis type B	Text	HEPATITISTYPEB
		Have you ever been told by doctors that you have/had any of the following conditions? Hepatitis type C	Text	HEPATITISTYPEC
		Have you ever been told by doctors that you have/had any of the following conditions? Hepatitis by other reasons	Text	HEPATITISOTHER
		Have you been told by doctors that you have/had any other condition?	Text	OTHER
		Have you ever been told by doctors that you have/had any of the following conditions? Cholecystitis	Text	CHOLECYSTITIS
		Have you ever been told by doctors that you have/had any of the following conditions? Gallbladder removal surgery (Cholecystectomy)	Text	CHOLECYSTECTOMY
		Have you ever been told by doctors that you have/had any of the following conditions? Acute gastroduodenitis	Text	ACUTEGASTRODUODENITIS
		Have you ever been told by doctors that you have/had any of the following conditions? Chronic gastroduodenitis	Text	CHRONICGASTRODUODENITIS
		At what age have you been diagnosed with Systemic Lupus Erythematosus (SLE)?	Numerical	SLE_AGE
		Have you ever been told by doctors that you have/had any of the following conditions? Helicobacter Pylori infection	Text	HELICOBACTERPYLORIINFECTION
		Have you ever been told by doctors that you have/had any of the following conditions? Gastric surgery	Text	GASTRICSURGERY
		Have you ever been told by doctors that you have/had any of the following conditions? Operations of the pancreas	Text	PANCREASOPERATIONS
		Have you ever been told by doctors that you have/had any of the following conditions? Cysts of the pancreas	Text	PANCREASCYSTS
		Have you ever been told by doctors that you have/had any of the following conditions? Periodontal disease	Text	PERIODONTALDISEASES

		Have you ever been told by doctors that you have/had any of the following conditions? Choledocholithiasis (bile duct stones or gallstones)	Text	CHOLEDOCHOLITHIASIS
		At what age were you told you had this other condition(s)?	Numerical	OTHER_AGE
	Cough	Date time	String	DateTime
		Do you have a cough?	Text	COUGH
		Define your cough type:	Text	COUGH_TYPE
	Diarrhea	Date time	String	DateTime
		Do you have diarrhea?	Text	DIARRHEA
	Edmonton Symptom Assessment System	Date time	String	DateTime
		Please select the number that best describes how you feel NOW:	Numerical	ESAS_1
		Please select the number that best describes how you feel NOW:	Numerical	ESAS_2
		Please select the number that best describes how you feel NOW:	Numerical	ESAS_3
		Please select the number that best describes how you feel NOW:	Numerical	ESAS_4
		Please select the number that best describes how you feel NOW:	Numerical	ESAS_5
		Please select the number that best describes how you feel NOW:	Numerical	ESAS_6
		Please select the number that best describes how you feel NOW:	Numerical	ESAS_7
		Please select the number that best describes how you feel NOW:	Numerical	ESAS_8
		Please select the number that best describes how you feel NOW:	Numerical	ESAS_9
		Please mark on the picture where it is that you hurt:	Text	ESAS_11
		Please select the number that best describes how you feel NOW this other symptom:	Numerical	ESAS_10
		Would you like to report an additional symptom?	Text	ESAS_10intro
		What is the symptom that you would like to report about? (e.g., constipation)	Text	ESAS_10opt
	EORTC QLQ-PAN26	Date time	String	DateTime
		During the past week...Has the information given about your physical condition and treatment been adequate?	Text	PAN26_24
		During the past week...Have you received adequate support from your health care professionals?	Text	PAN26_23
		During the past week...Were you limited in planning activities in advance (e.g. meeting friends)?	Text	PAN26_22
		During the past week...Were you worried about your health in the future?	Text	PAN26_21
		During the past week...To what extent have you been troubled with side-effects from your treatment?	Text	PAN26_20
		During the past week...Have you been dissatisfied with your body?	Text	PAN26_19
		During the past week...Have you felt physically less attractive as a result of your disease and treatment?	Text	PAN26_18
		During the past week...Did you feel the urge to move your bowels quickly?	Text	PAN26_17
		During the past week...Did you have frequent bowel movements?	Text	PAN26_16
		During the past week...To what extent was your skin yellow?	Text	PAN26_15
		During the past week...Have you had itching?	Text	PAN26_14
		During the past week...Have you felt less interest in sex?	Text	PAN26_25
		During the past week...Did you have a dry mouth?	Text	PAN26_13
		During the past week...Have you worried about your weight being too low?	Text	PAN26_11
		During the past week...Were you bothered by gas (flatulence)?	Text	PAN26_10
		During the past week...Have you had indigestion?	Text	PAN26_9
		During the past week...Did food and drink taste different from usual?	Text	PAN26_8
		During the past week...Were you restricted in the amounts of food you could eat as a result of your disease or treatment?	Text	PAN26_7
		During the past week...Were you restricted in the types of food you can eat as a result of your disease or treatment?	Text	PAN26_6
		During the past week...Did you find it uncomfortable in certain positions (e.g. lying down)?	Text	PAN26_5
		During the past week...Did you have pain during the night?	Text	PAN26_4
		During the past week...Have you had back pain?	Text	PAN26_3
		During the past week...Did you have a bloated feeling in your abdomen?	Text	PAN26_2
		During the past week...Have you had abdominal discomfort?	Text	PAN26_1
		During the past week...Did you feel weak in your arms and legs?	Text	PAN26_12
		During the past week...Have you felt less sexual enjoyment?	Text	PAN26_26
	EORTC QLQ-30	Date time	String	DateTime
		During the past week...Has your physical condition or medical treatment caused you financial difficulties?	Text	QLQ_28
		During the past week...Has your physical condition or medical treatment interfered with your <u>social</u> activities?	Text	QLQ_27
		During the past week...Has your physical condition or medical treatment interfered with your <u>family</u> life?	Text	QLQ_26
		During the past week...Have you had difficulty remembering things?	Text	QLQ_25
		During the past week...Did you feel depressed?	Text	QLQ_24
		During the past week...Did you feel irritable?	Text	QLQ_23
		During the past week...Did you worry?	Text	QLQ_22
		During the past week...Did you feel tense?	Text	QLQ_21
		During the past week...Have you had difficulty in concentrating on things, like reading a newspaper or watching television?	Text	QLQ_20
		During the past week...Did pain interfere with your daily activities?	Text	QLQ_19
		During the past week...Were you tired?	Text	QLQ_18
		During the past week...Have you had diarrhea?	Text	QLQ_17
		During the past week...Have you been constipated?	Text	QLQ_16
		How would you rate your overall <u>health</u> during the past week?	Numerical	PSQI_29
		During the past week...Have you vomited?	Text	QLQ_15
		During the past week...Have you lacked appetite?	Text	QLQ_13
		During the past week...Have you felt weak?	Text	QLQ_12
		During the past week...Have you had trouble sleeping?	Text	QLQ_11
		During the past week...Did you need to rest?	Text	QLQ_10
		During the past week...Have you had pain?	Text	QLQ_9
		During the past week...Were you short of breath?	Text	QLQ_8
		During the past week...Were you limited in pursuing your hobbies or other leisure time activities?	Text	QLQ_7
		During the past week...Were you limited in doing either your work or other daily activities?	Text	QLQ_6
		Do you need help with eating, dressing, washing yourself or using the toilet?	Text	QLQ_5
		Do you need to stay in bed or a chair during the day?	Text	QLQ_4
		Do you have any trouble taking a short walk outside of the house?	Text	QLQ_3
		Do you have any trouble taking a long walk?	Text	QLQ_2
		Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?	Text	QLQ_1
		During the past week...Have you felt nauseated?	Text	QLQ_14
		How would you rate your overall <u>quality of life</u> during the past week?	Numerical	PSQI_30
	Fagerstrom Test for Nicotine Dependence	Date time	String	DateTime
		How soon after you wake up do you smoke your first cigarette?	Text	FAGERSTROM_1
		Do you find it difficult to refrain from smoking in places where it is forbidden?	Text	FAGERSTROM_2
		Which cigarette would you hate most to give up?	Text	FAGERSTROM_3
		How many cigarettes do you smoke per day?	Text	FAGERSTROM_4

			Do you smoke more frequently during the first hours after waking, than during the rest of the day?	Text	FAGERSTROM_5
		Do you smoke even if you are so ill that you are in bed most of the day?	Text	FAGERSTROM_6
	Family History Pancreatic Cancer	Date time	String	DateTime
		At what age was this person diagnosed with cancer?	Numerical	PER2_AGE
		What is the relationship to you of the person who have/had cancer?	Text	PER5_REL
		Is there any other first-degree relative (parents, siblings, or children) who has/had cancer?	Text	PER5_QUE
		At what age was this person diagnosed with cancer?	Numerical	PER4_AGE
		What type of cancer does/did this person have/had?	Text	PER4_TYPE
		What is the relationship to you of the person who have/had cancer?	Text	PER4_REL
		At what age was this person diagnosed with cancer?	Numerical	PER1_AGE
		What type of cancer does/did this person have/had?	Text	PER1_TYPE
		What is the relationship to you of the person who have/had cancer?	Text	PER2_REL
		Is there any other first-degree relative (parents, siblings, or children) who has/had cancer?	Text	PER4_QUE
		At what age was this person diagnosed with cancer?	Numerical	PER3_AGE
		What type of cancer does/did this person have/had?	Text	PER3_TYPE
		What is the relationship to you of the person who have/had cancer?	Text	PER3_REL
		Is there any other first-degree relative (parents, siblings, or children) who has/had cancer?	Text	PER2_QUE
		What type of cancer does/did this person have/had?	Text	PER5_TYPE
		Is there any other first-degree relative (parents, siblings, or children) who has/had cancer?	Text	PER3_QUE
		What type of cancer does/did this person have/had?	Text	PER2_TYPE
		What is the relationship to you of the person who have/had cancer?	Text	PER1_REL
		At what age was this person diagnosed with cancer?	Numerical	PER5_AGE
	Fatigue	Date time	String	DateTime
		Do you experience fatigue?	Text	FATIGUE
	General Health	Date time	String	DateTime
		Which of these descriptions best fit you?	Text	ADL
	Health Behaviour Stage of Change Questionnaire	Date time	String	DateTime
		Smoking	Text	HBSCQ_SMOKING
		Nutritional Consultation	Text	HBSCQ_NUTRITION
		Physical Activity	Text	HBSCQ_PHYACTIVITY
		Alcohol Consumption	Text	HBSCQ_ALCOHOL
		Use of mammography and/or ultrasound	Text	HBSCQ_HEALTHUSAGE
	Health Habits Questionnaire	Date time	String	DateTime
		How many servings of fruits or vegetables do you have a day? (One serving is most easily identified by the size of your fist)	Numerical	HHQ_1
		Based on your answers, is there ONE thing you would be interested in changing now?	Text	HHQ_10
		How many 8-ounce servings of Nonfat (skim), low-fat (1%), or reduced-fat (2%) milk do you drink a day?	Numerical	HHQ_96
		How many 8-ounce servings of Soda or punch do you drink a day?	Numerical	HHQ_94
		How many 8-ounce servings of Fruit or sports drinks do you drink a day?	Numerical	HHQ_93
		How many 8-ounce servings of Water do you drink a day?	Numerical	HHQ_92
		How many 8-ounce servings of 100% juice do you drink a day?	Numerical	HHQ_91
		Is there a television set or Internet-connected device in your bedroom?	Text	HHQ_8
		How many times a week do you eat fast food or takeout?	Numerical	HHQ_7
		How much time a day do you spend at faster breathing/heart rate or sweating?	Numerical	HHQ_61
		How much time a day do you spend being active?	Numerical	HHQ_6
		How many times a week do you eat breakfast?	Numerical	HHQ_5
		How many times a week do you eat dinner at the table together with your family?	Numerical	HHQ_4
		How much recreational (outside of school work) screen time do you have daily?	Numerical	HHQ_3
		How many hours do you sleep each night?	Numerical	HHQ_2
		How many 8-ounce servings of Whole milk do you drink a day?	Numerical	HHQ_95
		On a scale of 1 to 10, how confident or sure do you feel about carrying out your plan?	Numerical	HHQ_11
	Health Literacy Scale "How easy would you say it is to…"	Date time	String	DateTime
		understand information in the media on how to get healthier?	Text	HLS_15
		understand advice on health from family members or friends?	Text	HLS_14
		find information on how to manage mental health problems like stress or depression?	Text	HLS_8
		decide how you can protect yourself from illness based on information in the media?	Text	HLS_12
		judge if the information on health risks in the media is reliable?	Text	HLS_11
		understand why you need health screenings?	Text	HLS_10
		understand health warnings about behaviour such as smoking, low physical activity and drinking too much?	Text	HLS_9
		follow instructions from your doctor or pharmacist?	Text	HLS_7
		use information the doctor gives you to make decisions about your illness?	Text	HLS_6
		judge when you may need to get a second opinion from another doctor?	Text	HLS_5
		understand your doctor's or pharmacist's instruction on how to take a prescribed medicine?	Text	HLS_4
		understand what your doctor says to you?	Text	HLS_3
		find out where to get professional help when you are ill?	Text	HLS_2
		find information on treatments of illnesses that concern you?	Text	HLS_1
		find out about activities that are good for your mental well-being?	Text	HLS_13
		judge which everyday behaviour is related to your health?	Text	HLS_16
		find information on treatments of illnesses that concern you?	Text	HLS_1
		find out where to get professional help when you are ill?	Text	HLS_2
		understand what your doctor says to you?	Text	HLS_3
		understand your doctor's or pharmacist's instruction on how to take a prescribed medicine?	Text	HLS_4
		judge when you may need to get a second opinion from another doctor?	Text	HLS_5
		use information the doctor gives you to make decisions about your illness?	Text	HLS_6
		follow instructions from your doctor or pharmacist?	Text	HLS_7
		find information on how to manage mental health problems like stress or depression?	Text	HLS_8
		understand health warnings about behaviour such as smoking, low physical activity and drinking too much?	Text	HLS_9
		understand why you need health screenings?	Text	HLS_10
		judge if the information on health risks in the media is reliable?	Text	HLS_11
		decide how you can protect yourself from illness based on information in the media?	Text	HLS_12
		find out about activities that are good for your mental well-being?	Text	HLS_13
		understand advice on health from family members or friends?	Text	HLS_14
		understand information in the media on how to get healthier?	Text	HLS_15
		judge which everyday behaviour is related to your health?	Text	HLS_16
	Height	Date time	String	DateTime
		Please indicate your height in cm	Numerical	HEIGHT_QUESTION
	Locus of Control Questionnaire	Date time	String	DateTime
		Do you make your own decisions, regardless of what other people say?	Text	LCQ_4
		Do you usually manage to resist being persuaded by other people's arguments?	Text	LCQ_9

		Question	Type	Variable
		Do you often feel you are the victim of outside forces you cannot control?	Text	LCQ_8
		Are most of the things you do designed to please other people?	Text	LCQ_7
		Do you take steps, such as exercise and diet to control your weight and fitness?	Text	LCQ_2
		If something goes wrong, do usually reckon it's your own fault rather than just bad luck?	Text	LCQ_6
		Do you find it a waste of time to plan ahead because something always causes you to change direction?	Text	LCQ_5
		Do you believe that your personality was firmly laid down in childhood so there is little you can do to change it?	Text	LCQ_3
		Is there some bad habit, such as smoking, that you would like to break but can't?	Text	LCQ_1
		Are you sceptical about the extent to which your horoscope can tell you what you should do and what's going to happen to you?	Text	LCQ_10
Mood Assessment		Date time	String	DateTime
		How do you feel at this moment?	Text	MOOD_QUEST
Pain		Date time	String	DateTime
		Do you experience pain?	Text	PAIN
		Please indicate where you experience pain	Text	BODYMAP
Perceived Stress Scale		Date time	String	DateTime
		In the last month, how often have you been upset because of something that happened unexpectedly?	Text	PSS1
		In the last month, how often have you felt that you were unable to control the important things in your life?	Text	PSS2
		In the last month, how often have you felt nervous and stressed?	Text	PSS3
		In the last month, how often have you felt confident about your ability to handle your personal problems?	Text	PSS6
		In the last month, how often have you felt that things were going your way?	Text	PSS7
		In the last month, how often have you found that you could not cope with all the things that you had to do?	Text	PSS8
		In the last month, how often have you been able to control irritations in your life?	Text	PSS9
		In the last month, how often have you felt that you were on top of things?	Text	PSS10
		In the last month, how often have you been angered because of things that happened that were outside of your control?	Text	PSS11
		In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?	Text	PSS14
		In the last month, how often have you dealt successfully with irritating life hassles?	Text	PSS4
		In the last month, how often have you felt that you were effectively coping with important changes that were occurring in your life?	Text	PSS5
		In the last month, how often have you been able to control the way you spend your time?	Text	PSS13
		In the last month, how often have you found yourself thinking about things that you have to accomplish?	Text	PSS12
Pittsburgh Sleep Quality Index		Date time	String	DateTime
		If you have a room mate or bed partner, ask him/her how often in the past month you have had . . .Episodes of disorientation or confusion during sleep	Text	PSQI_10d
		If you have a room mate or bed partner, ask him/her how often in the past month you have had . . .Legs twitching or jerking while you sleep	Text	PSQI_10c
		If you have a room mate or bed partner, ask him/her how often in the past month you have had . . .Long pauses between breaths while asleep	Text	PSQI_10b
		If you have a room mate or bed partner, ask him/her how often in the past month you have had . . .Loud snoring	Text	PSQI_10a
		Do you have a bed partner or room mate?	Text	PSQI_10
		During the past month, how much of a problem has it been for you to keep up enough enthusiasm to get things done?	Text	PSQI_9
		During the past month, how often have you had trouble staying awake while driving, eating meals, or engaging in social activity?	Text	PSQI_8
		During the past month, how often have you taken medicine to help you sleep (prescribed or 'over the counter')?	Text	PSQI_7
		During the past month, how would you rate your sleep quality overall?	Text	PSQI_6
		Please describe this other reason that kept you awake.	Text	PSQI_5jextra
		During the past month, how often have you had trouble sleeping because of. . .Any other reason	Text	PSQI_5j
		During the past month, how often have you had trouble sleeping because you . . .Have pain	Text	PSQI_5i
		During the past month, how often have you had trouble sleeping because you . . .Had bad dreams	Text	PSQI_5h
		During the past month, how often have you had trouble sleeping because you . . .Feel too hot	Text	PSQI_5g
		During the past month, how often have you had trouble sleeping because you . . .Feel too cold	Text	PSQI_5f
		During the past month, how often have you had trouble sleeping because you . . .Cough or snore loudly	Text	PSQI_5e
		During the past month, how often have you had trouble sleeping because you . . .Cannot breathe comfortably	Text	PSQI_5d
		During the past month, how often have you had trouble sleeping because you . . .Have to get up to use the bathroom	Text	PSQI_5c
		During the past month, how often have you had trouble sleeping because you . . .Wake up in the middle of the night or early morning	Text	PSQI_5b
		During the past month, how often have you had trouble sleeping because you . . .Cannot get to sleep within 30 minutes	Text	PSQI_5a
		During the past month, how many hours of <u>actual sleep</u> did you get at night? (This may be different than the number of hours you spent in bed.)	Numerical	PSQI_4
		During the past month, what time have you usually gotten up in the morning?	Text	PSQI_3
		During the past month, how long (in minutes) has it usually taken you to fall asleep each night?	Numerical	PSQI_2
		During the past month, what time have you usually gone to bed at night?	Text	PSQI_1
		If you have a room mate or bed partner, ask him/her how often in the past month you have had . . .Other restlessness while you sleep	Text	PSQI_10e
		Please describe the other restlessness while you sleep.	Text	PSQI_10eextra
Quality of Life Questionnaire		Date time	String	DateTime
		Anxiety / Depression TODAY	Text	ANXDEPSC
		Your usual pain/discomfort TODAY	Text	PAINSC
		Please tap on the scale to indicate how your health is TODAY.	Numerical	EQVASSC
		Your self-care TODAY	Text	CARESC
		Your usual activities TODAY (e.g. work, study, housework, family or leisure activities)	Text	USACTSC
		Your mobility TODAY	Text	MOBSC
Rosenberg Self-Esteem Scale		Date time	String	DateTime
		All in all, I am inclined to feel that I am a failure.	Text	RSES_9
		I am able to do things as well as most other people.	Text	RSES_4
		At times I think I am no good at all.	Text	RSES_2
		I wish I could have more respect for myself.	Text	RSES_8
		I feel that I'm a person of worth, at least on an equal plane with others.	Text	RSES_7
		I certainly feel useless at times.	Text	RSES_6
		I feel that I have a number of good qualities.	Text	RSES_3
		I feel I do not have much to be proud of.	Text	RSES_5
		On the whole, I am satisfied with myself.	Text	RSES_1
		I take a positive attitude toward myself.	Text	RSES_10
		On the whole, I am satisfied with myself.	Text	RSES_1
		At times I think I am no good at all.	Text	RSES_2
		I feel that I have a number of good qualities.	Text	RSES_3
		I am able to do things as well as most other people.	Text	RSES_4

		I feel I do not have much to be proud of.	Text	RSES_5
		I certainly feel useless at times.	Text	RSES_6
		I feel that I'm a person of worth, at least on an equal plane with others.	Text	RSES_7
		I wish I could have more respect for myself.	Text	RSES_8
		All in all, I am inclined to feel that I am a failure.	Text	RSES_9
		I take a positive attitude toward myself.	Text	RSES_10
	Warwick Edinburgh Mental Wellbeing Scale	Date time	String	DateTime
		I've been feeling optimistic about the future	Text	WEMWBS_1
		I've been feeling useful	Text	WEMWBS_2
		I've been feeling relaxed	Text	WEMWBS_3
		I've been feeling interested in other people	Text	WEMWBS_4
		I've had energy to spare	Text	WEMWBS_5
		I've been dealing with problems well	Text	WEMWBS_6
		I've been thinking clearly	Text	WEMWBS_7
		I've been feeling good about myself	Text	WEMWBS_8
		I've been feeling close to other people	Text	WEMWBS_9
		I've been feeling confident	Text	WEMWBS_10
		I've been able to make up my own mind about things	Text	WEMWBS_11
		I've been feeling loved	Text	WEMWBS_12
		I've been interested in new things	Text	WEMWBS_13
		I've been feeling cheerful	Text	WEMWBS_14
		I've been interested in new things	Text	WEMWBS_13
		I've been feeling optimistic about the future	Text	WEMWBS_1
		I've been dealing with problems well	Text	WEMWBS_6
		I've been feeling relaxed	Text	WEMWBS_3
		I've been feeling useful	Text	WEMWBS_2
		I've been able to make up my own mind about things	Text	WEMWBS_11
		I've been feeling confident	Text	WEMWBS_10
		I've been feeling close to other people	Text	WEMWBS_9
		I've been feeling interested in other people	Text	WEMWBS_4
		I've been feeling good about myself	Text	WEMWBS_8
		I've had energy to spare	Text	WEMWBS_5
		I've been thinking clearly	Text	WEMWBS_7
		I've been feeling loved	Text	WEMWBS_12
		I've been feeling cheerful	Text	WEMWBS_14

6 Secondary data in the iHelp pipelines

Up to this section, the description of the system facilitating secondary data capture has been abstract, focusing on the data collection options and not the specifics. This section dives into the specifics, addressing how Healthentia, a 3rd-party system is used "as a sensor" to gather secondary data and how this data finds its way into the iHelp platform and the HHR.

6.1 Integrating Healthentia with iHelp

In this section we first introduce the Healthentia platform, the measurement platform used in the iHelp project that is brought in as background and modified for the needs of the project by Innovation Sprint. Then, in Section 6.2 and Section 6.3, we discuss how Healthentia is integrated with the rest of the iHelp platform.

6.1.1 Healthentia as a sensor

Other platforms employing Healthentia "as a sensor" utilize the eCRF integration API endpoints to get data. These endpoints facilitate individual or bulk data exporting towards authorized 3rd-party systems. In order for such systems to utilize the endpoints, they need to be authenticated. To this extend, Healthentia supports the "application" user role, a role intended for software systems and not humans. Such a system is authorized to access the data of certain studies and is given a pair of credentials. It then starts interacting with Healthentia by logging-in to obtain a token. Bearer authentication is then used in each call to authenticate the system and authorize access to certain RWD.

The data handling functionalities of the Healthentia data management layer is then employed under the hood to fetch the requested data.

The role-based control of the Healthentia core layer then handles authorization of data access. Other functionalities of this layer are also silently used to facilitate RWD collection, for managing participants and their reports.

Certainly, for RWD collection by different healthcare organizations, the functionalities to create and configure in-vivo studies of the Healthentia study layer are utilised at the setup phase, but also throughout the studies to monitor and manage them.

Finally, the functionalities of the Healthentia services layer allow the population of the mobile app dashboards and the collection of reports (e-diary). Both are important for the user experience of the study participants, to keep them interested in using the app by providing the sense that some information is fed back to them, allowing them to understand their conditions better.

6.2 Ingesting secondary data

Secondary data in iHelp are collected and processed with the help of a modified version of Healthentia capable of covering the needs of the clinical partners of iHelp. The Healthentia-powered iHelp secondary data system is positioned over the overall iHelp architecture in an adaptation from section 6.1 of D2.5 – "Conceptual model and reference architecture II" shown in in Figure 27.
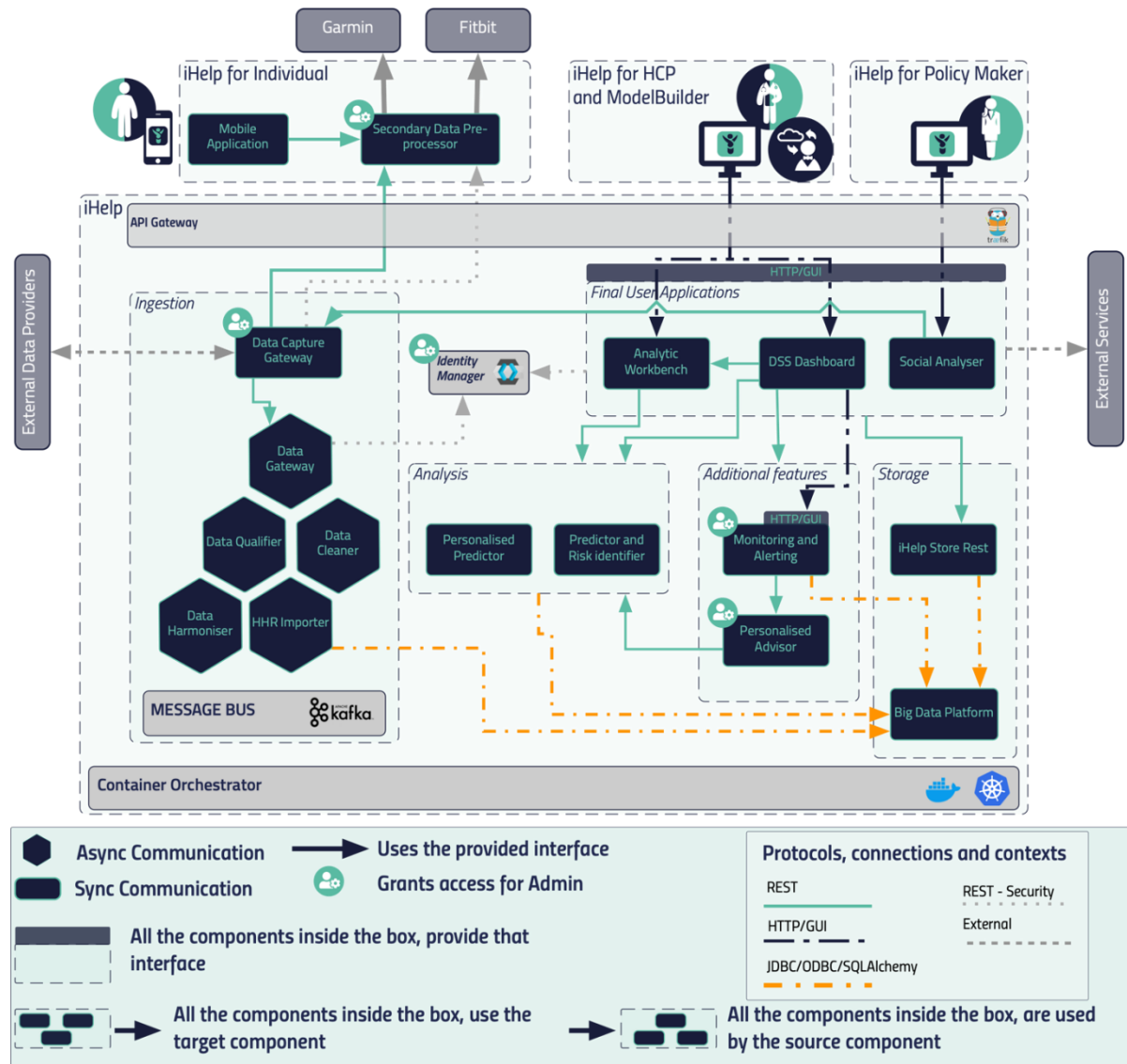
Figure 27: iHelp architecture and the role of Healthentia in it.

Secondary data collection spans two iHelp components, as detailed in D2.4 – "Conceptual model and reference architecture I":

- The mobile application handles reports from study participants. It is implemented by the Healthentia mobile application, suitable modified and branded for iHelp. It communicates with the secondary data pre-processor.

- The secondary data pre-processor handles automated measurement collection from 3rd-party devices and processes all raw secondary data into higher-level information. It is implemented by the Healthentia platform functioning as a sensor for iHelp. It communicates with the mobile application to receive reports and the Data Connectors component to offer the secondary data.

Identifiers of the study participants are registered in iHelp; as a result, both the iHelp platform and the secondary data collection system powered by Healthentia are aware of their users. The registration process is outlined in Section 6.2.1. Then, data ingestion can start, as outlined in Section 6.2.2.

6.2.1 Study participant registration

iHelp maintains the HHRs in a way so that it is possible to trace back all data to a particular study participant resource. The HHRs comprise of the primary and the secondary data. To do so, the study participants are registered with iHelp. The secondary data are gathered by a data collection system powered by Healthentia, thus Healthentia needs to know about the different studies involved and the participants in each one, i.e., the study participants are registered to studies of Healthentia.

There are five clinical partners in iHelp, each running their own study with different target populations and goals. Although there is a core of secondary data that is common across most studies, it is clear from Table 1 and Table 2 that there are many differences as well, especially in the scheduled and ad hoc questionnaires. To be able to support these changes, the secondary data collection system is based on five different Healthentia studies, one per clinical partner. This allows participants' segregation both for privacy and for management reasons. The process is as follows:

- A DPA is mutually agreed between each clinical partner and Innovation Sprint, where the clinical organization is designated as the study controller and Innovation Sprint the study processor.
- A Healthentia administrator (a user role of Healthentia and not used in iHelp at all) creates the organizations for the clinical partners that are not already using Healthentia, alongside with at least one organization administrator from each clinical partner.
- The organization administrator creates the iHelp study for their healthcare organization using the Healthentia portal application. If the organization administrator needs help with the subsequent steps, they invite some Healthentia administrator into their study.
- Any of the administrators involved can now proceed configuring the study. This is again done at the Healthentia portal application. The configuration involves creation of invitations for the study participants, creation of the questionnaires, selection of the widgets to be activated and creation of the investigators (the healthcare professionals monitoring the study). This latter role can be covered by the organization admin, or not be covered at all, since the use of Healthentia in iHelp stops at that point. Healthentia data of the studies are ingested into the iHelp big data platform, and any monitoring should be carried out with the tools therein.

For the study to begin, first the participants need to be registered to their respective studies in Healthentia. Healthentia does not need to know about the primary data and the complete HHR of the study participants. iHelp on the other hand needs to have the complete picture, so iHelp is naturally selected as the master in this process: new study participants are registered there, and this iHelp registration triggers the Healthentia registration. The registration is undertaken by an iHelp user with the HCP role and its details are beyond the scope of this deliverable. The iHelp HCP and the Healthentia investigator can of course be the same person.

Healthentia study registration is done using an invitation code created during study configuration. This code is communicated to the iHelp study participant upon their registration, together with instructions on obtaining the branded version of the Healthentia mobile app for the particular study. The participants download the app and register themselves, adding their invitation code. This way the study participants also become Healthentia users with the study participant role. The email does not need to be real, but it does need to be unique. The registration is completed with every Healthentia study participant being assigned an ID.

After the registration is complete, the study participants are solely represented by their IDs. Any secondary ingestion will request individual data using an ID, or will receive bulk data for these IDs. For this reason, the IDs are made known to iHelp employing an API call that gets all registered participants for a study.

The registration process described uncovers the involvement of certain people via their use of SW components, in which case they have become users with different roles. During their involvement, they perform certain actions. The people, their user roles, their user actions and the SW components they use to perform them are listed in Table 9.

Table 9: The people involved in the iHelp studies, their user roles, their user actions and the SW components they use to perform them.

Person	User role	User actions	SW component
Healthentia administrator	Healthentia administrator	Creates new Healthentia organizations	Healthentia portal app
		Assists with study configuration	
HCP	iHelp HCP	iHelp study participants' registration, study monitoring	iHelp DSS Dashboard
	Healthentia organization administrator	Creates new study in their organization, and optionally configures it or invites Healthentia administrators to help out with the task	Healthentia portal app
	Healthentia investigator	Healthentia Study configuration, study participants invitation from iHelp to Healthentia, and optionally study monitoring	Healthentia portal app
Study participant	Healthentia study participant	User of the Healthentia mobile app to enter data and receive advice	Healthentia mobile app
	iHelp study participant	Passive role, having their account created at iHelp, always interacting with iHelp via the Healthentia mobile app	None

6.2.2 Data ingestion

After secondary data has been collected and pre-processed, they are periodically sent to the integrated iHelp platform so that they can be persistently stored and be accessible by the platform's analytical tools. For that, a data ingestion pipeline is periodically established that takes the responsibility of acquiring the data from Healthentia, applying all data processing related functionalities, and finally storing the data to the Big Data Platform of iHelp.

The data ingestion process involves two separate processes: the capture of the data, and the ingestion of the data. The data gateway implements the data capturing functionality. It provides different means for data connectivity in order to access the data that might reside outside of the deployment of the integrated platform of iHelp. This is exactly the case with the secondary data, as they live inside the Healthentia platform that is considered an external source. Therefore, it makes use of REST clients to be able to connect and receive the related information periodically. The data provider, which in this case is

the processor of the iHelp studies running with Healthentia, configures the gateway by invoking its interface and providing:

- The URI of the Healthentia exposed web services,
- The frequency that the data gateway needs to connect, and
- The definition of the schema of the provided data.

Then, a scheduler is being deployed that will periodically send REST requests to the provided URI, grab the data, transform them to the internal abstract entities and put them to a Kafka queue, along with data related meta-information (i.e., the definition of the schema).

After data has been initially captured by the data gateway, a specific data pipeline is established that consists of various analytical functions, whose purpose is to ensure the level of quality assurance, transform them to the common data model of HHR and finally persistently store them to the Big Data Platform of iHelp. These analytics functions related with the data ingestion might be domain/schema agnostic, and therefore they will rely on the definition of the schema of the secondary data, so that they can be in a position to properly deserialize and process them. They listen to the Kafka topic where the data gateway has put the captured data, and they apply their algorithms. Eventually, secondary data is received by the secondary *data mapper* that transforms them to HHR entities. As soon as they are transformed to HHR entities, the *data importer* receives them from a specific Kafka topic, establishes connection with the Big Data Platform and persistently stored them to the relational data model of the storage.

The deployment of the data pipeline related with the secondary data ingestion can be static or dynamic, thus can be automatically deployed each time such ingestion takes place. This happens in order not to waste valuable resources of the underlying infrastructure of the time that this data pipeline is inactive. However, this is the responsibility of the task T3.2 ("Primary Data Capture and Ingestion") and it happens transparently for what concerns the activities described in this report. More information regarding the data ingestion can be also found in the corresponding deliverable D3.3 "Primary data capture and ingestion I".

6.3 Mapping secondary data into HHR

The Secondary Mapper sub-component will enable the mapping of secondary data (e.g., from mobile, wearable and social-media platforms) into the common data model used by the iHelp platform (HHR). The component will provide necessary transformation functions that are required to map the secondary data from heterogeneous sources to the holistic health records that will be eventually stored in the iHelp platform – a conceptual schema of the HHR model, at the time that this report was written, is shown in Figure 28. The Secondary Data Mapper will be an integral part of the iHelp platform as it supports the enrichment of typical health records or EHRs individuals with the personalised secondary data (e.g., lifestyle, behavioural, social etc) coming from heterogeneous sources, including mobile and wearable devices.

Based on the microservice platform architecture model adopted in the iHelp project, the Secondary Data Mapper component will also be developed as a microservice. This microservice will expose API(s) to perform the conversions from a native data format or data model to a standard data modal and format, while applying any remapping and unit conversion required for different types of data fields or

parameters. In this respect, the secondary data mapping functionality will comprise of the following (one or more) steps:

- Conversion to the standardised HHR format adopted in the iHelp platform
- Unit conversions to transform individual data parameters according to the predefined format
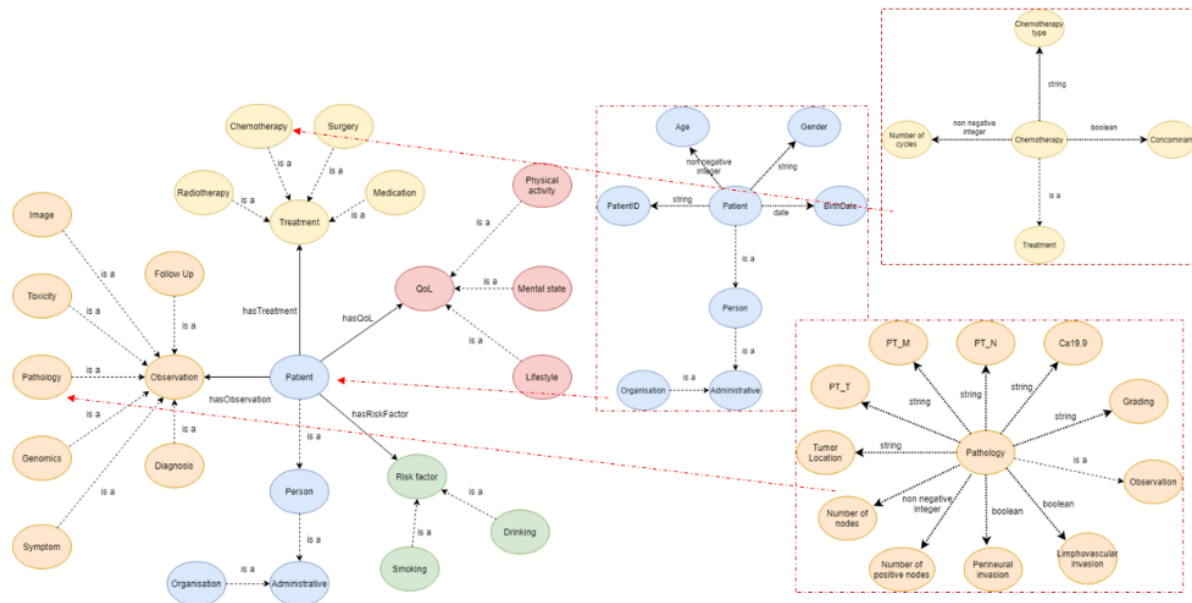- Mapping specific concepts or data parameters to standardised HHR structure

Figure 28: Conceptual schema of the HHR model used in iHelp

In terms of implementation of the Secondary Data Mapper component, a schema library and data driven ingestion pipeline can be created to avoid heavily scenario specific code being developed.

The concept is to further breakdown the micro services approach to a chain of converters, mappers and meta data format tools that are built form a palette of common processes in each case being data driven on both the individual operation and the overall conversion process definition. The components of the Secondary Data Mapper microservice will be created as Docker containers drawn from a palette of images and controlled by environment variables and configuration files. In terms of functioning or operations, the Secondary Data Mapper will be invoked by its exposed API by the data harmonizer function. The latter subscribes to the to the necessary Kafka message bus topics in the iHelp platform, do the needed mapping using the API of the Secondary Data Mapper, and then publish the mapped data back to the message bus for the data importer to receive and store in the iHelp big data platform.

Since the previous version of this document, we have achieved a conceptual mapping of the full data dictionary to FHIR, to do this we have primarily used the Observation and QuestionnaireResponse classes. These have been assigned to subcategories within the data to best represent the information provided.

Answers which can be physically measured in an objective manner have been mapped to the Observation class as this allows association of multiple values using the "Observation.component.value" mappings, this is an enhancement of the basic "Observation.value" which only allows one value to be stored for an observation object. This class also allows us to use a mapping called "Observation.effective" which means we can store a time at which the observation is made from either the user entering the data, or the

information being gathered from a wearable device. This class does require a field "Observation.status" which stores a value explaining how finalized the observation is. In most cases this will be final as the data will be gathered and future corrections or amendments will not be needed.

The other data which cannot be measured in an objective manner are subjective pieces of data and include information on mental well-being and family health history or addiction to substances such as nicotine. For these mappings we used the QuestionnaireResponse class. Similarly, to the Observation class this has a requirement for a status which would be set in the same way as the Observation class. This class allows entry of multiple answers in the "QuestionnaireResponse.Item.Answer" mapping, these answers can be restricted by assigning a Questionnaire to this class which has defined questions and potentially answer options which the user can choose from to minimize error. However, in this mapping this is not used as the questionnaire data we collect is mostly text based and will not be applicable for answer options. It also allows storage of a date in the mapping "Questionnaire.Authored" as required for regularly filling in the questionnaires at different time increments.

Table 10 shows the subclasses within the full data dictionary provided by iSprint and the class which has been used to represent that data in FHIR.

Table 10: Subcategories and classes

Subcategory	Class
Daily activity	Observation
Daily heart	Observation
Daily sleep	Observation
Activity sessions	Observation
Weight	Observation
Blood pressure	Observation
Oxygen saturation	Observation
Liquid Consumption	Observation
Treatment reminder	Observation
Nutrition	Observation
AUDIT Questionnaire	QuestionnaireResponse
Body Awareness Questionnaire	QuestionnaireResponse
Body Temperature	Observation
Comorbidity Questionnaire	QuestionnaireResponse
Cough	Observation
Diarrhea	Observation
Edmonton Symptom Assessment System	QuestionnaireResponse
EORTC QLQ-PAN26	QuestionnaireResponse
EORTC QLQ-30	QuestionnaireResponse
Fagerstrom Test for Nicotine Dependence	QuestionnaireResponse
Family History Pancreatic Cancer	QuestionnaireResponse
Fatigue	QuestionnaireResponse
General Health	QuestionnaireResponse

Health Behaviour Stage of Change Questionnaire	QuestionnaireResponse
Health Habits Questionnaire	QuestionnaireResponse
Health Literacy Scale	QuestionnaireResponse
Height	Observation
Locus of Control Questionnaire	QuestionnaireResponse
Mood Assessment	QuestionnaireResponse
Pain	QuestionnaireResponse
Perceived Stress Scale	QuestionnaireResponse
Pittsburgh Sleep Quality Index	QuestionnaireResponse
Quality of Life Questionnaire	QuestionnaireResponse
Rosenberg Self-Esteem Scale	QuestionnaireResponse
Warwick Edinburgh Mental Wellbeing Scale	QuestionnaireResponse

The subcategories which have been matched to the Observation class have been chosen based on their style of answer again linking back to the theme of if it can be physically measured or seen then it is objective data which is best recorded in the Observation class. Similarly, the Questionnaire Response subcategories are information that could not be measured in a physical way and would require the input of the patient to describe the mental aspects of their health.

The reason for separating the mappings into two classes is so that structuring of data is possible as QuestionnaireResponses could be stored in observation objects but would lose the structuring available within the class for assigning answer groups or defining the questionnaires if required in future stages of the mapping.

As an example of an Objective subcategory the daily activity class can be used, as shown in Figure 29, the data in the daily activity subcategory has been given relevant FHIR mapping in the observation class.

Daily activity	Date	String	Date	Observation.effective[effectiveDateTime]
Daily activity	Steps walked	Int	Steps	Observation.component.value[valueInteger]
Daily activity	Distance travelled (meters)	Int	Distance	Observation.component.value[valueInteger]
Daily activity	Calories burned	Int	Calories	Observation.component.value[valueInteger]
Daily activity	Floors climbed	Int	Floors	Observation.component.value[valueInteger]
Daily activity	Lightly active minutes	Int	LightlyActive	Observation.component.value[valueInteger]
Daily activity	Moderately active minutes	Int	FairlyActive	Observation.component.value[valueInteger]
Daily activity	Highly active minutes	Int	VeryActive	Observation.component.value[valueInteger]
Daily activity	Sedentary minutes	Int	Sedentary	Observation.component.value[valueInteger]
Daily activity	Total active minutes	Int	TotalActive	Observation.component.value[valueInteger]

Figure 29: Daily Activity subcategory FHIR mappings

These values are shown to be mostly general activity information such as steps walked floors climbed or general definitions of "active minutes". These values are all numerical integer values which are assigned to Observation components. These are stored alongside an observation effective mapping to show on what date these activities took place. The mappings for most of this data appears identical but will in fact be separated using identifiers within the FHIR mappings.

Similarly, as an example of the Subjective subcategories the Health Behaviour Stage of Change Questionnaire subcategory has been used to show the FHIR mappings in Figure 30.

Health Behaviour Stage of Change Questionnaire	Date time	String	DateTime	QuestionnaireResponse.authored
Health Behaviour Stage of Change Questionnaire	Smoking	Text	HBSCQ_SMOKING	QuestionnaireResponse.item.answer.value[valueString]
Health Behaviour Stage of Change Questionnaire	Nutritional Consultation	Text	HBSCQ_NUTRITION	QuestionnaireResponse.item.answer.value[valueString]
Health Behaviour Stage of Change Questionnaire	Physical Activity	Text	HBSCQ_PHYACTIVITY	QuestionnaireResponse.item.answer.value[valueString]
Health Behaviour Stage of Change Questionnaire	Alcohol Consumption	Text	HBSCQ_ALCOHOL	QuestionnaireResponse.item.answer.value[valueString]
Health Behaviour Stage of Change Questionnaire	Use of mammography and/or ultrasound	Text	HBSCQ_HEALTHUSAGE	QuestionnaireResponse.item.answer.value[valueString]

Figure 30: Health Behaviour Questionnaire FHIR Mappings

These values have been mapped to QuestionnaireResponse as they contain data which could not be physically measured or checked by anyone other than the person filling in the questionnaire. They contain text fields for giving answers around topics such as how often a person smokes or their physical activities. Similarly, to the observation class the mappings look identical but would be differentiated using IDs in the FHIR classes.

The mappings also require a data type to be associated with them which is shown in square brackets as shown in Figure 31.

QuestionnaireResponse.item.answer.value[valueString]
QuestionnaireResponse.item.answer.value[valueInteger]

Figure 31: Data types for FHIR mappings

These values generally match up with the "type" column provided in the data dictionary however it sometimes is used a little differently such as in the date mappings as they are provided as a string datatype but for proper use of this data it should be stored in a date format to make it comparable more easily. Other than this some responses are identified as being text type and have been classified into the string data type as this simply refers to a pre-set text answer which is selected by the user and is not needed to be stored in a special data type for functionality.

Other data types tend to be numerical which is simple for integer values but more complex when working with float variables as they do not have a storage type within the FHIR classes and will instead be stored as strings rather than losing the precision within the data trying to store them as integers, if necessary, this data could be parsed later for using in a functional way.

Prior to this we worked with a sample and smaller dataset for developing the necessary mapping of general data about a patient's current condition including their gender, Age Group and weight. This dataset also includes information on their nutritional intake such as vegetables, meats and dairy. Some attributes related to mental state are also included with basic agree to disagree scale answers to questions. This data has been mapped with a larger range of classes to be more specialised for the provided data and keep data about the same patient together easily.

The data has been mapped across multiple FHIR classes rather than the two used above these are: Patient, Observation, FamilyMemberHistory, Condition, NutritionIntake, QuestionnaireResponse, Activity Definition. Due to no subcategories being defined within the data and some data referring to history of the patient's family.

In this dataset the QuestionnaireResponse class is used for mappings regarding subjective data as in the previous dataset. All values are mapped to integer values due to the nature of the question and answer provided will be on a scale from 1-5 representing an agreement to disagreement based on the statement provided.

The Observation and Patient classes are used for the basic objective patient information such as their BMI, Height or Age Group. This value is stored as numerical values to represent the questions for example the Age group is represented using a value 1 to 4 which matches to 4 pre-defined age groups which the patient would be a part of.

The FamilyMemberHistory class is used for only one value regarding the History of pancreatic cancer in a relative, this is the only value in this dataset asking about information regarding a family member of the patient and seemed appropriate to include this class for properly representing that information.

The Condition class is included for information regarding the current status of known conditions diabetes and chronic pancreatitis which are stored as a Boolean value of either 1 or 0 to represent the disease being present or not within the patient.

Activity Definition is used within the dataset to represent the physical activity periods of the patient, only their moderate and vigorous physical activity are measured in a Boolean sense to determine if the patient has completed a certain time span of that category of physical activity in the past week represented as a 1 or 0.

NutritionIntake class is used for the information on types of foods eaten and the quantity consumed over a month, this again is split into classes which the patient may fall into ranging from 1 where its rarely or never eaten to 9 which is more than 6 times a day for each food group. The "NutritionIntake.consumedItem.ammount" mapping is used for this to provide a rough estimate of the number of times a piece of food has been consumed in a given period.

Figure 32 shows the different food groups and their mappings each working in the same way to give a representation of the patient's diet across a month and what food types it mostly consisted of. Which can be used to identify more health risks if their diet is particularly unbalanced.

Cereals (grains, beans, legumes)	NutritionIntake.consumedItem.ammount	1 - 9 , -9	Food consumption data based on categories valued from 1 being "never eat or less than once a month" and 9 being 6+ times a day. -9 displays missing data.
Vegetables	NutritionIntake.consumedItem.ammount	1 - 9 , -9	Food consumption data based on categories valued from 1 being "never eat or less than once a month" and 9 being 6+ times a day. -9 displays missing data
Fruits (sometimes grouped with vegetables)	NutritionIntake.consumedItem.ammount	1 - 9 , -9	Food consumption data based on categories valued from 1 being "never eat or less than once a month" and 9 being 6+ times a day. -9 displays missing data
White meat	NutritionIntake.consumedItem.ammount	1 - 9 , -9	Food consumption data based on categories valued from 1 being "never eat or less than once a month" and 9 being 6+ times a day. -9 displays missing data
Red meat	NutritionIntake.consumedItem.ammount	1 - 9 , -9	Food consumption data based on categories valued from 1 being "never eat or less than once a month" and 9 being 6+ times a day. -9 displays missing data
Processed meat	NutritionIntake.consumedItem.ammount	1 - 9 , -9	Food consumption data based on categories valued from 1 being "never eat or less than once a month" and 9 being 6+ times a day. -9 displays missing data
Dairy	NutritionIntake.consumedItem.ammount	1 - 9 , -9	Food consumption data based on categories valued from 1 being "never eat or less than once a month" and 9 being 6+ times a day. -9 displays missing data
Confectionery (aka sugary foods)	NutritionIntake.consumedItem.ammount	1 - 9 , -9	Food consumption data based on categories valued from 1 being "never eat or less than once a month" and 9 being 6+ times a day. -9 displays missing data

Figure 32: NutritionIntake FHIR Mappings

The data is provided from the University of Manchester (UNIMAN) dataset and will be provided from wearable devices which will be provided to participants in the study later in the project. This will be matched with questionnaires which are filled in from handheld devices at different intervals to keep regular updates on the patient's health. In total there are 470 different values across the two files, with the data in the larger file provided with subcategories as mentioned above. Each subcategory has its own date mapping to fit the data being provided at different dates.

7 Conclusion

This deliverable detailed secondary data collection, addressing the issues of "what" and "how" to capture. The "what" originates from the clinical needs, resulting to raw data or data from processing. The "how" is dictated by technology: attributes of the secondary data can be automatically measured, or manually reported using scheduled or ad hoc questionnaires, or widgets. The "how" is further addressed by getting the secondary data from its source, the Healthentia-powered secondary data collection system, and ingesting it into iHelp. The process is concluded by mapping all the 435 secondary data attributes across all 5 iHelp studies into the iHelp HHR.

Bibliography

[1] Jaana Sinipuro, "Statement & vision", Next Generation IoT thematic workshop, New concepts in smart health and care, available online at https://www.ngiot.eu/event/ngiot-thematic-workshop-health-and-care/

[2] Hannu Hämäläinen, Saara Malkamäki, Ilkka Räsänen, Jaana Sinipuro, Artur Olesch, "Towards trustworthy health data ecosystems", Sitra, Oct. 2020, available online at https://www.sitra.fi/en/publications/towards-trustworthy-health-data-ecosystems/

[3] US FDA, "Use of real-world evidence to support regulatory decision-making for medical devices: guidance for industry and Food and Drug Administration staff", Aug. 2017, available online at https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf

[4] OECD, "Better Life Index", available online at http://www.oecdbetterlifeindex.org

[5] Pittsburgh Sleep Quality Index, available online at http://www.opapc.com/uploads/documents/PSQI.pdf

[6] EPIC-Norfolk study, Food Frequency Questionnaire, available online at https://www.epic-norfolk.org.uk/wp-content/uploads/2020/11/CAMB-PQ-6-1205a_front.pdf and FFQ EPIC Tool for Analysis (FETA), available online at https://www.epic-norfolk.org.uk/for-researchers/feta-download/

[7] EPIC-Norfolk study, Food Diary with Rations, available online at https://epic-norfolk.org.uk/wp-content/uploads/2020/10/Food_diary_v3.pdf

[8] Cancer Care Ontario, "Edmonton Symptom Assessment System", available online at http://cancercaresoutheast.ca/sites/default/files/files/resource/edmonton_symptom_assessment_system.pdf

[9] The Warwick-Edinburgh Mental Wellbeing Scales, available online at https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/

[10] Rosenberg, M. (1965). Society and the adolescent self-image. Princeton, NJ: Princeton University Press, available online at https://fetzer.org/sites/default/files/images/stories/pdf/selfmeasures/Self_Measures_for_Self-Esteem_ROSENBERG_SELF-ESTEEM.pdf

[11] Shields, S.A., Mallory, M.E., & Simon, A. (1989). The Body Awareness Questionnaire: Reliability and validity. Journal of Personality Assessment, 53, 802-815, available online at http://hub.salford.ac.uk/psytech/body-awareness-questionnaire/

[12] S. Lyubomirsky, Subjective Happiness Scale (SHS), available online at https://ppc.sas.upenn.edu/sites/default/files/subjectivehappinessscale.pdf

[13] Locus of Control Questionnaire, available online at https://www.nwpgmd.nhs.uk/sites/default/files/locusofcontrol.pdf

[14] Gonzalez-Ramirez, L. P., De la Roca-Chiapas, J. M., Colunga-Rodriguez, C., Preciado-Serrano, M. L., Daneri-Navarro, A., Pedroza-Cabrera, F. J., & Martinez-Arriaga, R. J. (2017). Validation of Health Behavior and Stages of Change Questionnaire. Breast cancer (Dove Medical Press), 9, 199–205. https://doi.org/10.2147/BCTT.S129855, available online at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5367771/

[15] Overall health rating by participants, available online at https://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=100389

[16] Lorini C, Santomauro F, Grazzini M, et al. Health literacy in Italy: a cross-sectional study protocol to assess the health literacy level in a population-based sample, and to validate health literacy measures in the Italian language. BMJ Open 2017;7:e017812. Doi: 10.1136/bmjopen-2017-017812, available online at https://bmjopen.bmj.com/content/7/11/e017812

[17] Process of change questionnaire, available online at https://habitslab.umbc.edu/processes-of-change-questionnaire/

[18] Quality of Life (EQ-5D), available online at https://euroqol.org/eq-5d-instruments/

[19] Quality of life for patients (EORTC QLQ-C30), available online at https://qol.eortc.org/questionnaire/eortc-qlq-c30/

[20] Quality of life for pancreatic cancer patients, available online at https://qol.eortc.org/questionnaire/qlq-pan26/

[21] Mariano J, Min LC (2012) Assessment. In: Naeim A, Reuben DB, Ganz PA (eds) Management of cancer in the older patient. Elsevier, Philadelphia, pp 39–50

[22] K. Rockwood et al. A global clinical measure of fitness and frailty in elderly people. CMAJ 2005;173:489-495

List of Acronyms

API	Application Programming Interface
BLE	Bluetooth Low Energy
CSV	Comma Separated Values
DPA	Data Processing Agreement
Dx.y	Deliverable of work-package x, numbered y
eCRF	Electronic Case Report Form (a software system used to collect data in a clinical study)
ePRO	Electronic Patient-Reported Outcome (a system to collect patient reports electronically)
FDA	Food and Drug Administration
FHIR	Fast Healthcare Interoperability Resources
GDPR	General Data Protection Regulation
HCP	Healthcare Professional
HHR	Holistic Health Record (the FHIR profile designed by iHelp project)
LXS	LeanXcale
ML	Machine Learning
Mx	Month x of the iHelp project, M1 being January 2021
OECD	Organisation for Economic Co-operation and Development
PREM	Patient-Reported Experience Measure
PROM	Patient-Reported Outcome Measure
REM	Rapid Eye Movement
RWD	Real-World Data
SDK	Software Development Kit
UI	User Interface
UNIMAN	University of Manchester
UPRC	University of Piraeus Research Centre