CrowdHEALTH

Collective Wisdom Driving Public Health Policies

D6.10 – Use Cases Implementation and Experimentation v1

Project Deliverable

This project has received funding from the European Union's Horizon 2020 Programme (H2020-SC1-2016-CNECT) under Grant Agreement No. 727560

D6.10 – Use Cases Implementation and Experimentation v1

Work Package:		WP1
Due Date:		28/02/2018
Submission Date:		10/04/2018
Start Date of Project:		01/03/2017
Duration of Project:		36 Months
Partner Responsible of Deliverable:		DFKI
Version:		1.1
Status:	☒ Final ☐ Draft ☐ Ready for internal Review ☒ Task Leader Accepted ☐ WP leader accepted ☒ Project Coordinator accepted	
Author name(s):	Serge Autexier	Lucas Donat
	Anton Gradišek	Thanos Kosmidis
	Sokratis Nifakos	Christos Panagopoulos
	Jan Janssen	
Reviewer(s):	George Moldovan	Marta Patiño
Nature:	☐ R – Report ☒ D – Demonstrator	
Dissemination level:	☒ PU – Public ☐ CO – Confidential ☐ RE – Restricted	

REVISION HISTORY			
Version	Date	Author(s)	Changes made
0.1	05/02/2018	Serge Autexier, Jan Janssen (DFKI)	TOC to share with the participants and initial text about central infrastructure and UC site DFKI.
0.2	14/02/2018	Serge Autexier (DFKI)	Updated content and assignment of contribution requests to partners.
0.3	27/02/2018	Thanos Kosmidis (CRA) , Christos Panagopoulos (BIO)	Added descriptions of demonstrations from UC partners CareAcross and BioAssist
0.4	28/02/2018	Sokratis Nifakos (KI)	Added descriptions of the demonstrations at Karolinska UC partner.
0.5	01/03/2018	Anton Gradišek (JSI)	Added descriptions of demonstrations at JSI with SLOfit data.
0.6	08/03/2018	Serge Autexier (DFKI)	Added description of demonstrations at HULAFE and finalization of draft for internal reviewing.
0.7	11/03/2018	Lucas Donat (HULAFE)	Reformulations of the HULAFE use case demonstrations
0.8	23/03/2018	Marta Patiño (UPM)	Peer review
0.9	19/03/2018	George Moldovan (SIE)	Peer review
1.0	03/04/2018	Serge Autexier (DFKI)	Final version
1.1	10/04/2018	ATOS	Quality Review. Submission to EC

List of acronyms

BIO	BioAssist
BMI	Body Mass Index
CI	Continuous Integration
CRA	CareAcross
DFKI	German Research Center for Artificial Intelligence (DFKI)
FHIR	Fast Healthcare Interoperable Resources
HULAFE	Fundación para la Investigación del Hospital Universitario La Fe de la Comunidad Valenciana
JSI	Jožef Stefan Institute
KI	Karolinska Institutet
LIME	Learning, Informatics, Management and Ethics department at the Karolinska Institutet
MEB	Medical Epidemiology and Biostatistics department at the Karolinska Institutet
SLOfit	Slovenian surveillance system for physical and motor development of children and youth which was formerly known as Sports Educational Chart.
UC	Use case

Contents

1. Executive Summary

The goal of this document is to describe the first activities performed to demonstrate the implementation and experimentations with the CrowdHEALTH platform by the use case partners of the consortium. These are part of Task 6.3 and two further demonstration activities are planned to be reported in months 24 and 36.

The main demonstration activities of the first version of the CrowdHEALTH system consisted of the following aspects:

1. **Data Import from multiple sources**: At the project meeting in Rome (13-14. March 2018) a live demonstration is given of how data from use case partners is imported into the CrowdHEALTH platform; this was data provided in the specific format of the use case partners, but not data from real persons, but synthetic data or pseudonymised data, in order to avoid any privacy issues during the demonstration at the meeting. The synthetic data is provided by the use case partners BioAssist (BIO, Greece), Hospital La Fe (HULAFE) and SLOfit (JSI, Slovenia).
2. **Analytics capabilities**: At the meeting in Rome again the partners developing analytics component demonstrate the status of analytics and visualization components. These demonstrations are based on the partners own data and are specifically the analysis and predictions algorithms developed by JSI on data from the SLOfit use case.
3. **Installation and Integration with existing UC partner infrastructures and foreseen exploitation by policy makers**: All use case partners described whether they have (resp. plan to have) an on-site or assume a central installation of the CrowdHEALTH platform, how the integration with their current infrastructure is (resp. is planned) and how policy makers will be able to use the CrowdHEALTH analytics.
4. **Requirements by UC partners to foster acceptance**: In order to encourage adoption of the system use case partners foresee different additional functionalities. These consist in the SLOfit use case of providing access for teachers to collected data of individual scholars and of the usage of prediction or forecasting analytics. In the joint HULAFE-DFKI use case these include providing access for nutritionists coaching obesity patients to the tracked nutritional and activity data, and comparing it with recommended nutritional behaviour. Similarly, in the BioAssist use case these consist of allowing doctors to use CrowdHEALTH's forecasting and risk assessment tools on data of individual patients, and of comparing the results with expected values or outcomes for groups of patients.

2. Introduction

The document presents the first version of the implementation and experimentation demonstrations for the use cases. Following a brief review of the central development infrastructure for the CrowdHEALTH (see Deliverable D6.1 [1]), it details for each use case partner which installations of the CrowdHEALTH platform are used for this first demonstrator, as well as:

- if relevant, how the CrowdHEALTH platform is linked with the local infrastructures at the use cases partners institutions,
- if available, which data has been provided for testing,
- if already possible, which data has been imported to demonstrate the import into the CrowdHEALTH platform,
- and finally, how policy makers will use the planned features of the CrowdHEALTH platform.

In addition, where relevant, use case partners describe which features they envision so far would be desirable to have integrated in the CrowdHEALTH platform, such that the adoption and acceptance by local stakeholders who play a crucial role for a sustainable provision of data for the platform.

2.1. Status of the Central CrowdHEALTH System Integration and Delivery

The central integration place for the CrowdHEALTH system is the GitLab[1] provided by the *development infrastructure provider* UPRC and managed by the *deployment leader* DFKI. The components of the CrowdHEALTH system are available either as a Java library in the jar format or as a Docker image directly in the GitLab. Those components which are not provided as docker images are all integrated in a combined system Docker image, which is also always updated by the continuous integration (CI) pipeline on GitLab. The combined system Docker image has an option to change the runtime configuration, where among other settings, the endpoints for the CrowdHEALTH components can be defined. In the final stage the GitLab will have a deployment script which downloads and sets up all the Docker images needed to start a local instance of the CrowdHEALTH platform. This script is targeting Linux environments.

At the time of this deliverable, the components are under development and therefore an initial version of the CrowdHEALTH platform is available. The components that will be provided as own Docker image are Aggregator, Data Anonymization, Data Cleaner, Data Converter, Data Store and Gateway. The components that will be provided as Java libraries and hence included in the combined system docker are HHR Manager and Plug and Play sources.

[1] https://crowdhealthtasks.ds.unipi.gr/

2.2. Use Case Site BioAssist (BIO)

2.2.1. Data Import Demonstration

The BIO use case involves five conceptually distinct datasets that consist of data collected through BIO's user base throughout Greece, via the BioAssist platform. For the purposes of the deliverable, BIO used synthetic data to demonstrate data import for the following datasets:

- Bio-signals and Activity Data: This dataset includes measurements of vital signs, collected from a variety of sensors (pulse oximeter, blood pressure meter, glucometer, weighing scale). The dataset also includes data on physical activity collected from activity tracking wearables.
- BIO-PHRs: This dataset includes medical test results.
- Medication: This dataset includes prescribed medications (active ingredient and dosage.
- Allergies: This dataset includes recorded allergies of users of the BIO platform.

All of the aforementioned datasets are already structured as FHIR Observations. For the first demonstration, the first two datasets are used, as they are the most complex and useful ones. BIO's last dataset, Social Data, which includes logs of the interaction of BIO's users with the platform's social features, will be integrated in the second year of the project. The data import is showcased to the partners at the project meeting in Rome (March 12-13, 2018).

2.2.2. Installation, Integration and Exploitation by Policy Makers

For the BIO use case, an on-site installation of the CrowdHEALTH platform is planned. The platform will be deployed on public cloud infrastructure (most likely Microsoft Azure). Data collected and stored on the BIO platform will be regularly exported and pushed via a web service to the CrowdHEALTH Gateway.

There are no policy makers involved in the BIO use case at the moment. However, clinicians that use the BIO platform monitor larger groups of patients and can therefore be considered policy makers for the particular use case, in the sense that they apply clinical protocols, potentially deriving from relevant policies. These clinicians shall be given access to the CrowdHEALTH platform in order to view analytics results on their patients' data.

2.2.3. Requirements to foster acceptance

In a later version of the platform, allowing doctors to use forecasting and risk assessment tools on data of individual patients, and to compare the results with expected values or outcomes for groups of patients, shall increase the value of the platform from a clinician's perspective by providing the means to assess clinical protocols and treatment plans. In turn, this will result in a tool of greater usefulness for organisations providing healthcare or insurance services, both in the public and the private sector.

2.3. Use Case Site CareAcross (CRA)

2.3.1. Data Import demonstration

The CRA use case data consists so far of data from cancer patients. This will be extended during the runtime of the project by physical activity data. From current data input from 13 breast cancer patients has been imported into the CrowdHEALTH platform for demonstration purposes. This import has been done after following these steps:

i) Data extraction from database.
ii) Data cleaning.
iii) Data anonymization by
    a. Change of patient IDs.
    b. Removal of subset of data (to further protect anonymity and disconnect from individuals).
    c. Addition of synthetic elements (to further disconnect from individuals).

The extended data with additional physical activities will be demonstrated in the next versions of this task.

2.3.1. Installation, Integration and Exploitation by Policy Makers

For the CRA use case a central CrowdHEALTH platform hosted on different premises is planned. The integration between the CareAcross and the CrowdHEALTH infrastructures will be through file transfer. In particular, the central repository of CrowdHEALTH will receive an anonymous, de-identified, aggregated dataset through a specific file, via secure means. This will take place every six months.

The aggregation will be performed in order to further protect patients' privacy and uphold the Terms of Service and Privacy Policy of our company's services. Furthermore, the aggregation approach will aim to retain the fundamentally relevant and important elements for the purposes of CrowdHEALTH and its goals, without revealing identifying patient information or breaching our agreement with patients.

Policy makers will be able to use the CrowdHEALTH platform to identify how online patient platforms can affect patients across the following characteristics:

i) Online education and whether knowing what to expect in terms of side-effects changes the reported side-effects compared to the overall population.
ii) Online coaching and whether, when and how patients are keen on long-term engagement with it.

This will be achieved through the central policy development toolkit, from which they will be able to pick one of the above and view corresponding insights based on the data and the CrowdHEALTH analysis.

2.3.2. Requirements to foster acceptance

There must be adequate data to reach conclusions with satisfactory validity.

2.4. Use Case Site Hospital LA FE (HULAFE)

2.4.1. Data import demonstration

The Hospital La Fe has provided anonymised data extracted from their electronic health records to the CrowdHEALTH partners. Data are related to clinical use case focused on obesity and overweight patients. That data is available and has been imported into the CrowdHEALTH platform and showcased to the partners at the project meeting in Rome (March 12-13, 2018).

2.4.2. Installation, Integration and Exploitation by Policy Makers

For the HULAFE use case an installation of the CrowdHEALTH platform in a remote place outside information systems of the Hospital La Fe is planned. This is mainly due to the severe and strict regulations about which kind of software can be installed on hospitals servers and that obtaining such a permission and clearance for a software under development as CrowdHEALTH is not possible during the runtime of the project.

Policy makers will use the platform in order to have best evidence based on analytics of indicators related to obesity, to provide a personal assessment to the end user and will also be used to characterize in most efficient way beneficiaries of different possibilities of interventions through risk profile and forecast analysis

2.4.3. Requirements to foster acceptance

Generally, to give access to the clinical staff to use and apply analytics (forecasting or causality) provided by the CrowdHEALTH platform to individual patients would be desirable to foster acceptance. Also, regarding collected real time data as envisioned in the joint HULAFE and DFKI use case, to allow to analyse adherence to treatments and recommendations on an individual level would be a feature of high value to make the CrowdHEALTH system an attractive tool for doctors (see also the specific requirements in Section 2.7.2).

2.5. Use Case Site Karolinska Institutet (KI)

2.5.1. Installation, Integration and Exploitation by Policy Makers

CrowdHEALTH platform instance will be installed in one dedicated server in LIME department (Learning, Informatics, Management and Ethics). The aim of the on-premises installation is first to secure the safety process of data processing and analysis of healthcare data within Karolinska Institutet use case and second to allow wide access to state of the art

technologies, sample and data collections, CrowdHEALTH tools and scientific consultation in order to facilitate the interaction between researchers and enhance scientific productivity in CrowdHEALTH project.

CrowdHEALTH instance within KI's use case will be used supervised from the local research group and from Medical Epidemiology and Biostatistics department in KI (MEB).

The current data that will be used after the CrowdHEALTH instance installation are static; meaning that within KI's use case scenario there is no real time data. The data have been exported and pre-processed from two different quality registries in Sweden: the Swedeheart quality registry and the Stockholm CREAtinine Measurement (SCREAM) project. This has been made possible by agreements with the main laboratory service providers of Stockholm (*Unilabs, Karolinska and Aleris*), which together run the vast majority of laboratory analyses of the County

Within the data different patients' information is included such all individuals residing in Stockholm County that underwent at least one measurement of creatinine or cystatin-C assessed in either primary, in- or out-patient care during 2006-2011. Sample size surpasses 1.3 million people. Together with kidney function estimates, concurrent basic laboratory biochemistry is also available. Patient data has been, in addition, linked to National Health Care Registers including the Swedish Medication Register (Lakemedelsregister), the Inpatient Register, the Death Register, the Socioeconomic Register (Statistiska centralbyrån, SCB), and the Swedish Renal Register.

Policy makers as the end users will use the platform in order to produce the following functions:

- Risk Stratification.
- Causal and forecasting analysis.

The main aim for the policy makers is to enhance their understanding of **risk factors for healthcare implications of cardiovascular diseases and kidney dysfunction** in order to promote the implementation of preventive/management strategies in a policy making level.

Moreover, through the CrowdHEALTH platform KI aims to enrich the awareness of kidney and cardiovascular diseases where inappropriate drug prescription is likely to occur and integrated into a new policy making plan.

Finally, policy makers will be able to investigate the safety and the effectiveness of many drugs for cardiovascular and kidney patients which are currently unknown in individuals, as these individuals are usually excluded from clinical trials.

2.5.2. Requirements to foster acceptance

Within KI use case and after the first phase of CrowdHEALTH deployment on KI servers KI aims to test the applicability and the effectiveness of the platform's offerings together with public health policy bodies around cardiovascular and kidney diseases. The outcomes of this testing phase which is planned to be conducted following the focus group methodology, will generate some new insights about the used data, the data analysis and the health policies results. In parallel KI applied to collect more data from other healthcare data sources for the same patients as used in KIs current data records in order to enhance the analytics tools of CrowdHEALTH platform and generate new health policy and clinical outcomes.

2.6. Use Case Site SLOfit (JSI)

2.6.1. Data Import demonstration

The SLOfit personal data has been provided and is demonstrated to be imported into the CrowdHEALTH platform. This is demonstrated to the project partners at the project meeting in Rome March 12-13th, 2018.

In addition to the data import, the status of the following analytics components developed by JSI are demonstrated:

- First, forecasting of various parameters at the age of 18 based on the available SLOfit personal data (for example, from ages 6 to 14). In its current version the tool forecasts height, BMI, aerobic endurance. This part is based on advanced machine-learning algorithms and considers the compensation of the growth boost.
- Second, risk assessment at the age of 18 and later in adulthood. This part is based partially on the algorithms used for forecasting and partially on the literature data. These analytics are demonstrated on the original SLOfit data format because the analytics are not yet included in CrowdHEALTH.

2.6.2. Installation, Integration and Exploitation by Policy Makers

For the demonstration in Rome, the instance runs on a local server. At a later point, either there will be a local installation of CrowdHEALTH at JSI accessible from outside, or at another location agreed with the consortium.

When the central CrowdHEALTH instance will be installed at a chosen server, the SLOfit data will directly be provided to it.

For the exploitation of CrowdHEALTH by policy makes, JSI plans to develop the following aggregated overview and analytics functionalities:

i) Overview on the overall fitness assessment for children in a class/school/municipality/ region/country. The overall assessment covers cardiorespiratory fitness, muscular fitness, obesity, and potentially other categories.
ii) Performance of the above indicators over a certain period in time (likely, on the annual basis).
iii) Forecast of the assessment up to a couple of years in future.
iv) Estimated adulthood risk from obesity, low cardiorespiratory and muscular fitness in selected population compared to the average national risk (based on the estimates of these traits at age 18). This forecast is intended to give a comparable assessment of different risks. For example, policy makers could see what the risks from obesity are, cardiorespiratory and muscular fitness and how they are distributed across the state.
v) Effects of possible interventions (advanced functionality to be elaborated in following deliverables).

2.6.3. Requirements to foster acceptance

To foster uptake and acceptance JSI would like to include functionalities for teachers, parents, and physicians as well. This will not only provide benefits to a wider group of stakeholders, but will also supports further development of the SLOfit use case (collection of additional demographic and lifestyle data, linkage with medical data). Likely, this requires an additional level of access to the data where it is possible to access only personal data for a specific individual or group (e.g. parents can access only the data of their child, teachers for their class). The proposed functionalities are:

i) Performance in individual physical fitness monitoring disciplines.
ii) Forecast parameters at the age of 18, which include height, BMI, and aerobic endurance.
iii) Parents, physicians and teachers could compare the risks related to obesity, muscular fitness and cardiorespiratory fitness for an individual child. Consequently, guided exercise or lifestyle interventions could be created to minimize the most prominent risk components.

2.7. Use Case Site DFKI (DFKI)

The DFKI use case consist of two joint use cases (see deliverable [2]), one jointly with the UC partner Hospital LA FE about obesity and a second with UC partner CareAcross for cancer patients. In these use cases DFKI is collecting nutritional and activity data from patients selected by the corresponding partners. The data will be collected on a central database and use it to provide nutritional information via a web application. The collected data will be downloaded by the use case partners via a dedicated API on a regular basis and include in the HHRs of the persons at the UC partners instance of the CrowdHEALTH platform (see Sections 2.4 and 2.3).

2.7.1. Integration with UC Partners CRA and HULAFE

The database used by DFKI which provides all nutritional data and the database of collected nutritional and activity data is setup on a secured server inside the DFKI site in Bremen. Physical access to the server is restricted to the server administrators of the DFKI Bremen site and software access is secured following the state of the art secure software layers.

The UC partners can set up new patients/persons for which nutrition and/or activities should be tracked via a dedicated secured interface. Only an identifying number is used to setup new test persons and the mapping of the person identity to that number is only known to CRA and HULAFE use case partners. The tracked data collected by DFKI can be downloaded by CRA and HULAFE at any time just based on the identifying number. The tracking data is included into the health records of the respective use case partners. As the real time tracking with test persons has not yet started, the import of these extended datasets will be demonstrated in the next demonstration deliverable of this task.

To comply with ethical and legal regulations, and informed consent is provided by DFKI to inform the test persons about which data is tracked, how and where it is stored, and what will happen to the data after the end of the project. This is included into the informed consent prepared by HULAFE for their patients, which they will sign, and for CRA it is included in the Terms of Service and Privacy Policy of CareAcross.

2.7.2. Requirements to foster acceptance

In a later version of the integrated CrowdHEALTH platform the activity and nutrition tracking developed by DFKI maybe be closely integrated with the platform, in order to allow any person whose HHR is maintained in a CrowdHEALTH platform to extend it by nutritional and activity data. In that case the data will be directly pushed from the DFKI database into the HHRs maintained in the CrowdHEALTH platform using the respective gateways. The described APIs to setup a person for nutrition and activity tracking are designed such that they will support that.

3. Conclusions

The document reported the demonstration activities of the CrowdHEALTH project centred on the CrowdHEALTH platform in its current state of development. All use case partners that have data provided their data and initial tests to import it into the CrowdHEALTH platform were performed with success. All use case partners have refined there plans on how to use the CrowdHEALTH platform and this will be taken into account to direct further developments and deployment plans. Moreover, the use case partners which now are more informed about the functionalities that are or will be built into CrowdHEALTH could better assess how their envisioned policy makers will be able to use the system. Finally, requirements for functional extensions of CrowdHEALTH have been identified that would foster acceptance of CrowdHEALTH by local stakeholder to continue providing data to CrowdHEALTH and use it as a tool profitable for their own work. This will also be taken into account during the discussions directing the next development and deployment phases.

4. References

[1] CrowdHEALTH Consortium, “D6.1 Integration Plan v1,” 2018.

[2] CrowdHEALTH Consortium, “D6.7 Use Case Scenarios Definition,” 2017.