Deliverable 5.1

CDSS REQUIREMENTS

25/09/2020

Project title	Patients-centered SurvivorShIp care plan after Cancer treatments based on Big Data and Artificial Intelligence technologies
Grant Agreement number	875406
Call and topic identifier	SC1-DTH-01-2019 - Big data and Artificial Intelligence for monitoring health status and quality of life after the cancer treatment
Funding schema	RIA
Coordinator	FUNDACION CENTRO TECNOLOXICO DE TELECOMUNICACIONS DE GALICIA (GRADIANT)
Website	www.projectpersist.com
Document keywords	CDSS, inference engine, mobile app, data extraction
Document Abstract	This document specifies the conceptual and technical specifications of CDSS which include CDSS requirements, Mobile Application for New EHR and Data Extraction as main sections.

DOCUMENT	
Authors	Umut ARIOZ (EMO), Barış YILDIZ (EMO), Mert YILMAZ (EMO), Victoria M. Cal GONZALEZ (GRAD), Damien CALDY (DXC), Simon LIN (SYM), Jean-Paul Calbimonte PEREZ (HESSO)
Internal reviewers	EMO, GRAD, DXC
Work package	WP5 - Decision support system at the point of care
Task	T5.1 CDSS specification, EHR data extraction, and filtering
Nature	Report

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Dissemination Level	PU - public

VERSION	DATE	CONTRIBUTOR	DESCRIPTION
0.1	15.04.2020	Umut ARIOZ (EMO), Barış YILDIZ (EMO), Mert YILMAZ (EMO)	Inference Engine
0.1	15.04.2020	Jean-Paul Calbimonte PEREZ (HESSO)	Cohort and Trajectory Analysis
0.1	15.04.2020	Simon LIN (SYM)	EHR Data preparation and Enrichment
0.1	15.04.2020	Umut ARIOZ (EMO), Barış YILDIZ (EMO), Mert YILMAZ (EMO), Damien CALDY (DXC)	Definition of func. And non-func requirements
0.2	10.06.2020	Simon LIN (SYM)	EHR Data preparation and Enrichment
0.2	10.06.2020	Umut ARIOZ (EMO), Barış YILDIZ (EMO), Mert YILMAZ (EMO), Damien CALDY (DXC)	Definition of func. And non-func requirements
0.2	10.06.2020	Damien CALDY (DXC)	Standardization and Interoperability
0.2	10.06.2020	Damien CALDY (DXC)	CDS Hook
0.2	10.06.2020	Damien CALDY (DXC)	Overall requirements of CDS Hooks
0.3	22.07.2020	Damien CALDY (DXC)	Alert Mechanism
0.3	22.07.2020	Victoria M. Cal GONZALEZ (GRAD), Clinical Partners	Alerts: Inputs and Outputs
0.3	22.07.2020	Victoria M. Cal GONZALEZ (GRAD), Clinical Partners	Management of alerts: inputs
0.3	22.07.2020	Victoria M. Cal GONZALEZ (GRAD), Clinical Partners	Clinical information

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

0.4	10.08.2020	Umut ARIOZ (EMO), Barış YILDIZ (EMO), Mert YILMAZ (EMO)	Inference Engine
0.4	10.08.2020	Umut ARIOZ (EMO), Barış YILDIZ (EMO)	Mobile Application for New EHR - Functionalities
0.5	2.09.2020	Mert YILMAZ (EMO)	Mobile Application for New EHR - Activity Diagrams
0.5	2.09.2020	Umut ARIOZ (EMO), Barış YILDIZ (EMO), Damien CALDY (DXC)	Mobile Application for New EHR - Data Security
0.5	2.09.2020	Gaetano Manzo (HESSO)	Cohort and Trajectory Analysis
0.6	15.09.2020	Gazihan ALANKUŞ (EMO)	Mobile Application for New EHR - User Interface Designs
0.6	15.09.2020	Victoria M. Cal GONZALEZ (GRAD), GRAD	Data anonymization techniques
0.6	15.09.2020	Marta SESTELO, Lilian ADKINSON (GRAD), GRAD	Risk of reidentification techniques
0.6	15.09.2020	Victoria M. Cal GONZALEZ (GRAD), GRAD	Data models - FHIR standard for storage
0.7	20.09.2020	Umut ARIOZ (EMO)	Introduction
0.7	20.09.2020	Umut ARIOZ (EMO)	PERSIST Project
0.7	20.09.2020	Umut ARIOZ (EMO)	Scope of The Deliverable D5.1
0.7	20.09.2020	Umut ARIOZ (EMO)	Conclusion
1.0	25.09.2020	Umut ARIOZ (EMO)	Revision

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

DISCLAIMER

ACKNOWLEDGEMENT

This document is a deliverable of the PERSIST project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 875406

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

INDEX

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

- **Acronyms and abbreviations**

ACRONYM	TITLE
API	Application programming interface
BI-RADS	Breast Imaging-Reporting and Data System
BRCA	Breast Cancer Susceptibility
CDSS	Clinical Decision Support Systems
CEA	Carcinoembryonic antigen
CSV	Comma Separated Values
DRL	Drools Rule Language
ECOG	Eastern Cooperative Oncology Group
EHR	Electronic Health Record
EMR	Electronic Medical Record
EORTC	European Organisation for Research and Treatment of Cancer
ESLG	European Splenic Lymphoma Group
FHIR	Fast Healthcare Interoperability Resources
FSH	Follicle-stimulating hormone

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

GAD	Generalized anxiety disorder
GDPR	General Data Protection Regulation
GPAQ	Global physical activity questionnaire
GSES	General self-efficacy Scale
HL7	Health Level Seven
HNPCC	Hereditary No'npolipoid Colorectal Carcinoma
HRQoL	Health-related quality of life
HTTP	Hyper-Text Transfer Protocol
ICD	International Classification of Diseases
IDE	Integrated Development Environment
ISI	Insomnia Severity Index
jBPM	Java Business Process Management
JSON	JavaScript Object Notation
LOINC	Logical Observation Identifiers Names and Codes
LVEF	Left Ventricular Ejection Fraction
MSN	Multi-sensor Network

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

NCCN	National Comprehensive Cancer Network
NLM	National Library of Medicine
OHC	Open Health Connected
PHQ	Patient Health Questionnaire
PREMs	Patient reported experience
pro-BNP	proB-type Natriuretic Peptide
PROMs	Patient-reported outcome measures
QLQ-BR	Questionnaire for assessing quality of life in breast cancer patients
REST	Representational state transfer
SNOMED	Systematized Nomenclature of Medicine
SSL	Secure Sockets Layer
TLS	Transport Layer Security
UI	User Interface
WHO	World Health Organization
XML	Extensible Markup Language

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

- Executive Summary

This Project deliverable is written in the framework of WP5 – Decision support system at the point of care (Task 5.1 CDSS specification, EHR data extraction and filtering) of PERSIST project under Grant Agreement No. 875406.

The intention of this Project deliverable is to provide a report containing the specific CDSS requirements to comply with the requirements and needs defined in WP2 - PERSIST specifications and methodology (D2.5 PERSIST overall Big Data architecture). The initial version of this Project deliverable was prepared in May 2020 (M5) but it was updated throughout the duration of this task till September 2020.

The design and specification of the CDSS will be described in the following sections including:

- The definition of **CDSS requirements**, which identifies the Software Architectural Design, Definition of Inputs/Outputs, Definition of func./non-func requirements and Standardization/Interoperability issues.
- The design of the **Mobile Application for New EHR**, which specifies the Functionalities, Activity Diagrams, UI Designs and Data Security issues.
- The description of the **Data extraction**, considering the Data anonymization techniques, Risk of reidentification techniques and Data models.

Since this document contains the initial specifications of the design and CDSS architecture of the PERSIST project, it should be borne in mind that these specifications are subject to change throughout the life of the project, as the specific components become concrete during the implementation phases.

For any comments on this Handbook, please contact the WP5 Leader:

➔ **Dr. Umut ARIOZ (EMODA)**

➔ **E-mail: umut@emodayazilim.com**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

- **Introduction**

1. PERSIST Project

PERSIST aims at developing an open and interoperable ecosystem to improve the care of cancer survivors. The key results to be achieved by partners are:

- Related to **patients**: **increased self-efficacy and satisfaction** with care as well as reduced psychological stress for a better management of the consequences of the cancer treatment and the disease.
- Related to **professionals**: **increased effectiveness in cancer treatment and follow-up** by providing prediction models from Big Data that will support decision-making and contribute to optimal treatment decisions.
- Related to **healthcare providers**: improved information and evidence to advance the efficacy of management, intervention and prevention policies. The long-term result will be to **reduce the socio-economic burden** related to cancer survivors' care.

2. Scope of The Deliverable D5.1

In this document, **the conceptual and technical specifications of CDSS** which is the heart of the PERSIST project will be described. There will be 3 main sections: CDSS requirements, Mobile Application for New EHR and Data Extraction.

In the CDSS requirements section, software architectural design of the CDSS will be defined by explaining the inference engine, cohort and trajectory analysis, EHR data preparation and enrichment, and alert mechanism. At the definition of inputs and outputs part, each of the alerts will be introduced and their management procedures will be defined. Also all possible parameters of gathered clinical information will be specified at the end of this section.

In the Mobile Application for New EHR section, functionalities like the output of the CDSS, data ingestion, alerts and appointment will be detailed. Also activity diagrams of the functionalities will be shown as graphic. The developed mobile application for the New EHR section will be used mainly by clinicians. The user interfaces and data security of the app will be explained at the end of this section.

In the Data Extraction section, the data anonymization and Risk of reidentification techniques will be given. Those techniques will be applied to the dataset which will be gathered from hospitals in the project. As a last comment, the data model of those datasets will be determined in detail.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

- CDSS Requirements

1. Conceptual Design

1.1. *Software* *Architectural* *Design*

Inference Engine

An inference engine is the part of a decision support system that performs the reasoning function. It is the heart of a CDSS[1]. CDSS is a computer-based system that leverages medical knowledge and patient-specific data to respond to a request for decision support by providing recommendations with the ultimate goal of improving the quality of the service provided to the patient. Elements of a CDSS are depicted below. Inference engine uses inputs from the specific patient as well as knowledge modules and services to process machine readable rules. The results are patient-specific information and knowledge (e.g., treatment recommendation and documentation that support the recommendation) conveniently presented to users: clinicians, staff, patients or other individuals.

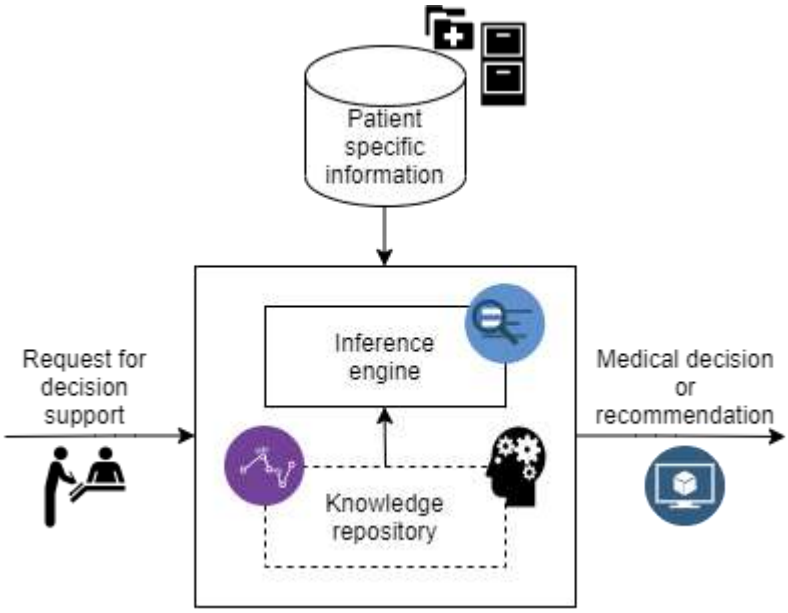


Figure 1: General structure of CDSS

The inference engine of PERSIST will combine patient-specific data input to the CDSS with the information contained in the knowledge base to generated the results expected by the CDSS: assistance with clinical decision making on the early diagnosis of

[1] **El-Gayar, O. F., Deokar, A., & Wills, M. (2008).** ***Current Issues and Future Trends of Clinical Decision Support Systems (CDSS)*** **[Chapter]. Encyclopedia of Healthcare Information Systems; IGI Global. https://doi.org/10.4018/978-1-59904-889-5.ch046**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

the recurrence of cancer, timely identification of a secondary disease (and, consequently, try to avoid a sudden death from a cardiac event), long-term and late side effects of treatment as well as treatment compliance; new evidence for cancer survivors' prognosis and treatment; and aggregate information as well as suggestions from data coming from the mobile application for patients (remote monitoring data). Primary properties of our inference engine can be listed as following:

- Drools[2] rule engine will be used for development.
- It will contain HTTP REST to communicate with other systems in PERSIST environment.
- During the decision process, it will benefit from the guidelines in the Knowledge Base part of the CDSS.
- It will get inputs from multimodal sensing network, EHR data source and CTC counting system.

Drools

Drools is a powerful hybrid reasoning system that intelligently and efficiently processes rule data. It is a part of the KIE[3] (Knowledge Is Everything) project. KIE contains the following different but related projects offering a complete portfolio of solutions for business automation and management:

- Drools is a business-rule management system with a forward-chaining and backward-chaining inference-based rules engine, allowing fast and reliable evaluation of business rules and complex event processing. A rules engine is also a fundamental building block to create an expert system which, in artificial intelligence, is a computer system that emulates the decision-making ability of a human expert. Drools core requires Java 1.5 and also provides an Eclipse-based IDE.
- jBPM[4] is a flexible Business Process Management suite allowing you to model your business goals by describing the steps that need to be executed to achieve those goals.
- OptaPlanner[5] is a constraint solver that optimizes use cases such as employee rostering, vehicle routing, task assignment and cloud optimization.
- Business Central is a full featured web application for the visual composition of custom business rules and processes.

---

2 ***Drools—Business Rules Management System (Java™, Open Source).*** **(n.d.). Retrieved July 17, 2020, from https://www.drools.org/**

3 KIE - Knowledge is Everything. (n.d.). Retrieved July 24, 2020, from https://www.kiegroup.org/

4 jBPM - Open Source Business Automation Toolkit—JBPM Business Automation Toolkit. (n.d.). Retrieved July 24, 2020, from https://www.jbpm.org/

5 Optaplanner—Constraint satisfaction solver (JavaTM, Open Source). (n.d.). OptaPlanner. Retrieved July 24, 2020, from https://www.optaplanner.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

- UberFire[6] is a web-based workbench framework inspired by Eclipse Rich Client Platform.

The Drools engine is the rules engine which uses an enhanced implementation of the Rete algorithm[7]. The Drools engine stores, processes, and evaluates data to execute the business rules or decision models that you define. The basic function of the Drools engine is to match incoming data, or facts, to the conditions of rules and determine whether and how to execute the rules[8]. The Drools engine operates using the following basic components:

- Rules: Business rules or DMN decisions that you define. All rules must contain at a minimum the conditions that trigger the rule and the actions that the rule dictates.
- Facts: Data that enters or changes in the Drools engine that the Drools engine matches to rule conditions to execute applicable rules.
- Production memory: Location where rules are stored in the Drools engine.
- Working memory: Location where facts are stored in the Drools engine.
- Agenda: Location where activated rules are registered and sorted (if applicable) in preparation for execution.

When a business user or an automated system adds or updates rule-related information in Drools, that information is inserted into the working memory of the Drools engine in the form of one or more facts. The Drools engine matches those facts to the conditions of the rules that are stored in the production memory to determine eligible rule executions. (This process of matching facts to rules is often referred to as *pattern matching*.) When rule conditions are met, the Drools engine activates and registers rules in the agenda, where the Drools engine then sorts prioritized or conflicting rules in preparation for execution.

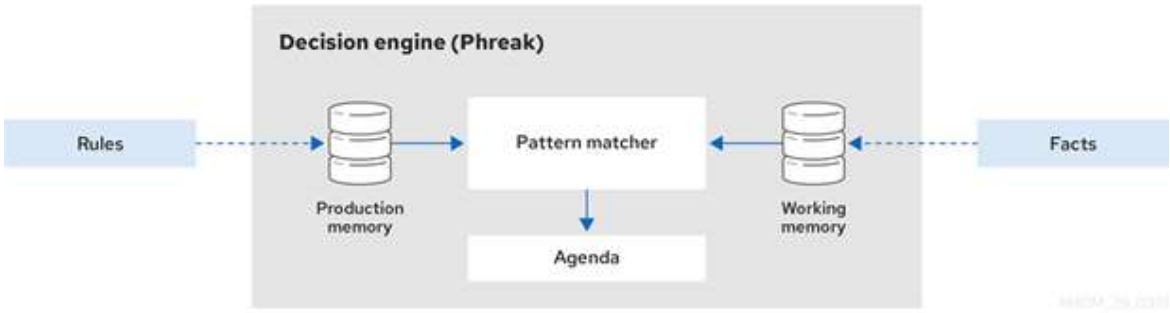


Figure 2: Basic components of the Drools engine

DRL (Drools Rule Language) rules are business rules that you define directly in .drl text files. These DRL files are the source in which all other rule assets in Business

[6] Uberfire Homepage · Uberfire Framework. (n.d.). Retrieved July 24, 2020, from https://uberfireframework.org/

[7] Forgy, C. L. (1982). Rete: A fast algorithm for the many pattern/many object pattern match problem. Artificial Intelligence, 19(1), 17–37. https://doi.org/10.1016/0004-3702(82)90020-0

[8] Drools Documentation. (n.d.). Retrieved July 25, 2020, from https://docs.jboss.org/drools/release/7.43.1.Final/drools-docs/html_single/index.html

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Central are ultimately rendered. You can create and manage DRL files within the Business Central interface, or create them externally as part of a Maven or Java project using Red Hat CodeReady Studio or another integrated development environment (IDE). A DRL file can contain one or more rules that define at a minimum the rule conditions (when) and actions (then).

Components in a DRL file

```
package

import

function  // Optional

query  // Optional

declare   // Optional

global   // Optional

rule "rule name"
    // Attributes
    when
        // Conditions
    then
        // Actions
end

rule "rule2 name"

...
```

Figure 3: DRL file components

A DRL file can contain single or multiple rules, queries, and functions, and can define resource declarations such as imports, globals, and attributes that are assigned and used by your rules and queries. The DRL package must be listed at the top of a DRL file and the rules are typically listed last. All other DRL components can follow any order.

Data flow in PERSIST ecosystem

Overview of the system is given in the figures below. Data will be obtained from data storage and applications. Then it is transferred to the big data platform which is connected to CDS and Alert systems.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Figure 4: Connectivities in PERSIST

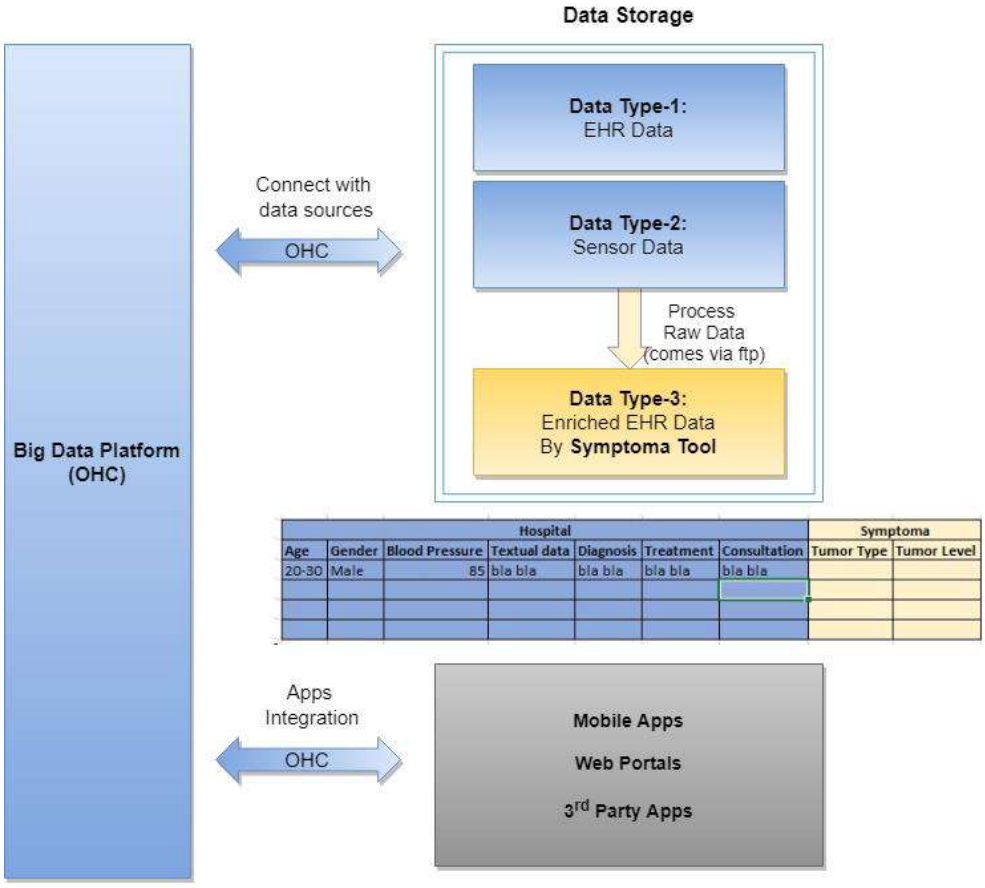


Figure 5-1: Details of Connectivities in PERSIST

Patient's data coming from hospitals is stored in Data Storage located in the OVH server. This includes EHR data and Sensor data. Following collection of data, it is transferred into the OVH server via OHC platform. Only collected EHR data is enriched at this stage by Symptoma's tool and it is converted into FHIR format by collaborative work with clinicians.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

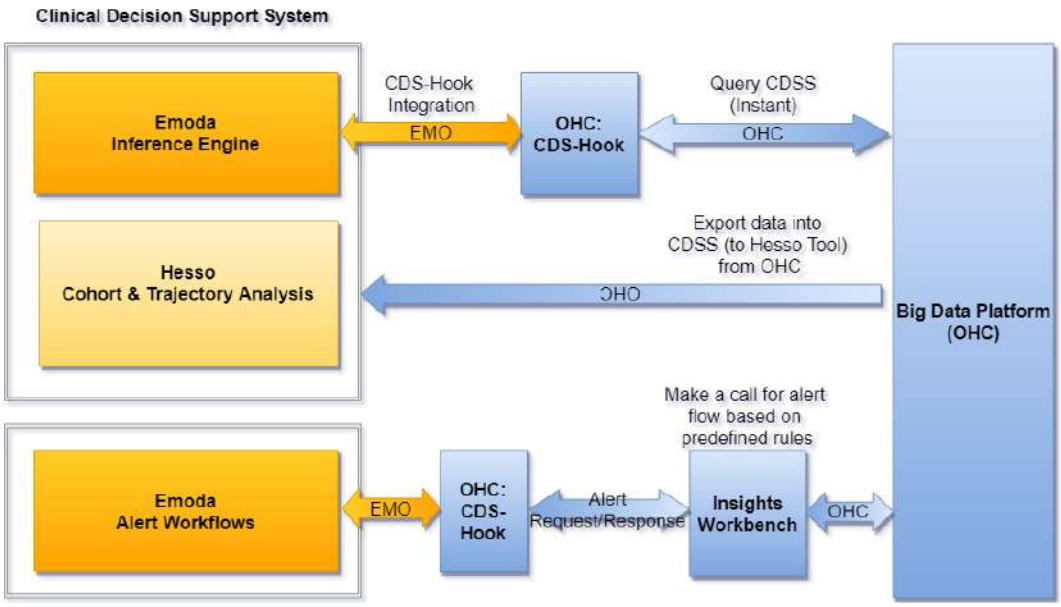


Figure 5-2: Details of Connectivities in PERSIST

All patient's data coming from hospitals is stored in Data Storage located in the OVH server. The data is transferred into the OVH server via OHC platform. The collected EHR data is enriched at this stage by Symptoma's tool and it is converted into FHIR format. Enriched and transformed data (FHIR format) is used by HESSO's tool to make Cohort and Trajectories analysis. How to fetch the FHIR data will be determined later by HESSO. The data will be exported into .csv file format and will be accessed via OHC platform. Following data wrangling and data cleansing, the data are ready for the training phase.

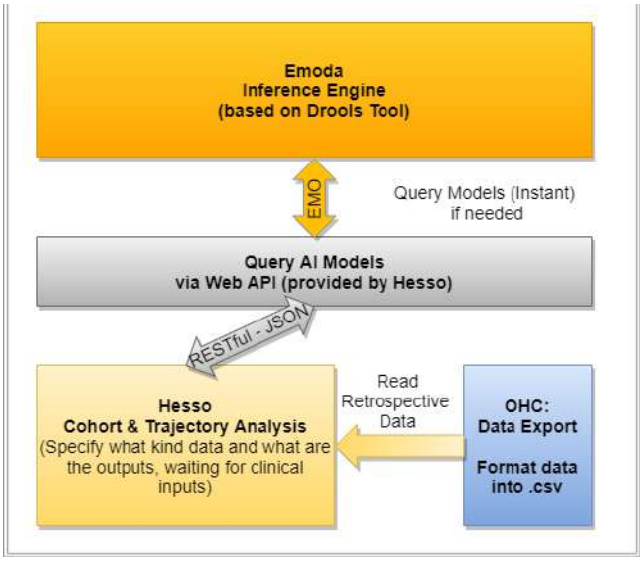


Figure 5-3: Details of Connectivities in PERSIST

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

CDSS relies on two components: a knowledge base and an Inference Engine . A knowledge base is an organized collection of facts about the system's domain. Knowledge base will be obtained from HESSO's analysis. This knowledge is then usually represented in the form of "if-then" rules (production rules): "If some condition is true, then the following inference can be made (or some action taken)." The inference engine interprets and evaluates the facts in the knowledge base in order to provide an answer. Inference Engine will be developed by EMODA using Drools rule engine. It enables the system to draw deductions from the rules in the knowledge base. For example, if the knowledge base contains the production rules "if x, then y" and "if y, then z," the inference engine is able to deduce "if x, then z." The system might then query its user, "Is x true in the situation that we are considering?" If the answer is affirmative, the system will proceed to infer z[9]. In our situation rules may include different parameters from various sources like patient's EHR data, patient's wearable data, and cohorts/trajectories.

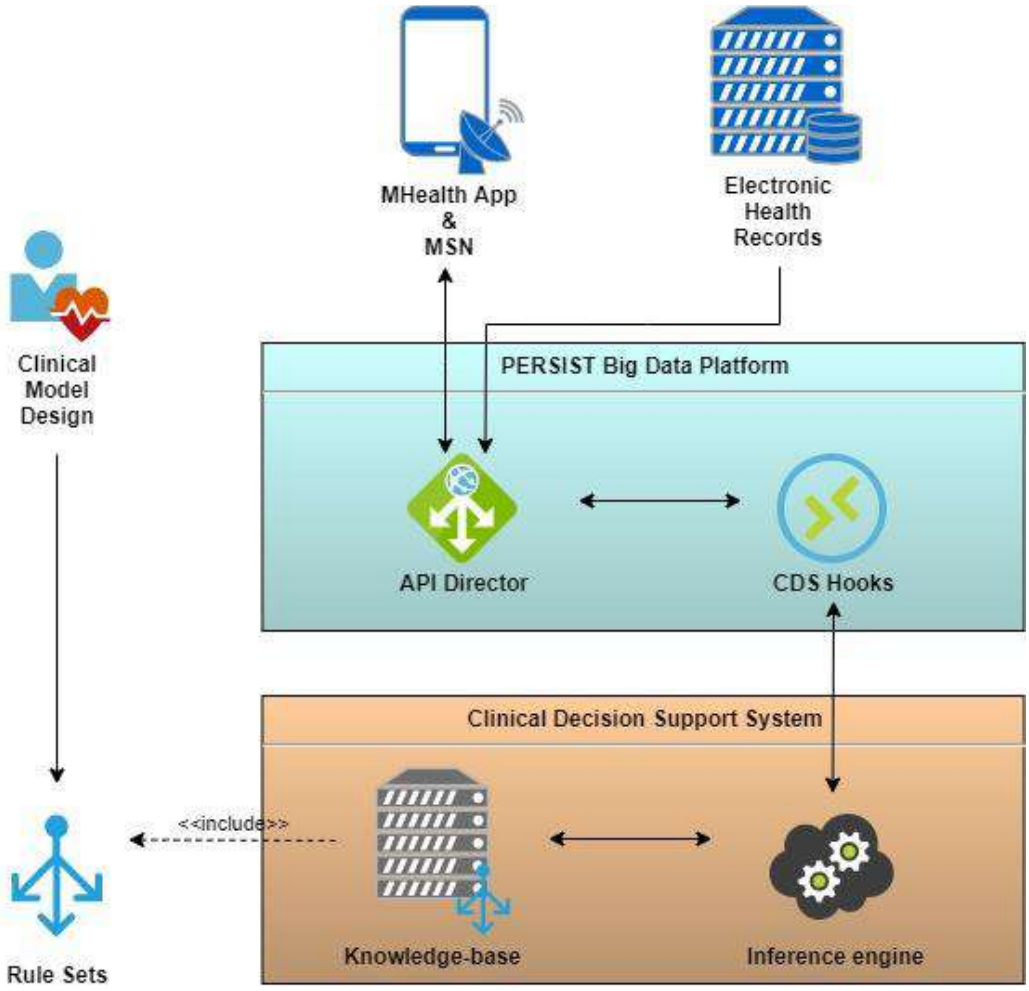


Figure 6: CDSS working schema

[9] **Artificial intelligence—Expert systems. (n.d.). Encyclopedia Britannica. Retrieved July 24, 2020, from https://www.britannica.com/technology/artificial-intelligence**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

- *The Knowledge-base* will store clinically approved rule-sets which are derived from clinical model design. These rules are utilized during the decision-making processes on the inference engine.
- *The API Director* will handle the data flow between MHealth Application / MSN and Big Data Platform. All relevant clinical data from MHealth App, MSN and EHR will be stored on the Big Data Platform.
- *The inference engin*e will not have direct access to any clinical record or personal data. Instead, they will be made available to CDSS by the Big Data Platform, using CDS Hooks for data transfer. Likewise, inference engine results will be returned to the end-user from the Big Data Platform by API Director.
- All requests are dispatched to the Inference Engine via OHC platform and the integration point is CDS-Hook[10]. This API (CDS-Hook) is probably used in clinical study with prospective data. For example, clinicians may send patient data such as lab results, physical examination etc. to the OHC platform to predict what is the best treatment for this patient when the patient comes to hospital.
- All data exchanged through the RESTful APIs will be sent and received as JSON structures, and will be transmitted over channels secured using the Hypertext Transfer Protocol (HTTP) over Transport Layer Security (TLS), also known as HTTPS and defined in RFC2818.
- Output data format will be FHIR[11]

Cohort and Trajectory Analysis

This module will use machine learning-based techniques to analyse cancer patient data extracted from the EHRs provided by the clinical partners. Cohorts are essentially clusters of patients with shared characteristics, which will be learned by analysing retrospective data. These cohorts will be employed to identify common disease and event trajectories. The disease trajectories will associate symptoms to diseases and will describe how patients progress from symptoms to disease. The event trajectories will identify associations between symptoms and events (admission, readmissions, treatments, etc.) and will be used to quantify risks. Cohorts and patients trajectories are major contributions to the CDSS as they, from data, represent the patients with similar characteristics and their progression.

The following draft diagram shows a basic description of the module. The Cohort training describes how patient information (imported from the EHR of clinical partners) is used to train different machine learning (classification) models. The same applies for the training of trajectories. While cohorts are used to classify patients in “similar” groups, trajectories allow to have a prediction model of certain patient outcomes. For example, we can create cohorts (or sub-cohorts) according to demographics (sex, age), symptoms, diagnosis, recurrence, treatments, etc. Similarly

[10] CDS Hooks. (n.d.). Retrieved July 24, 2020, from https://cds-hooks.org/

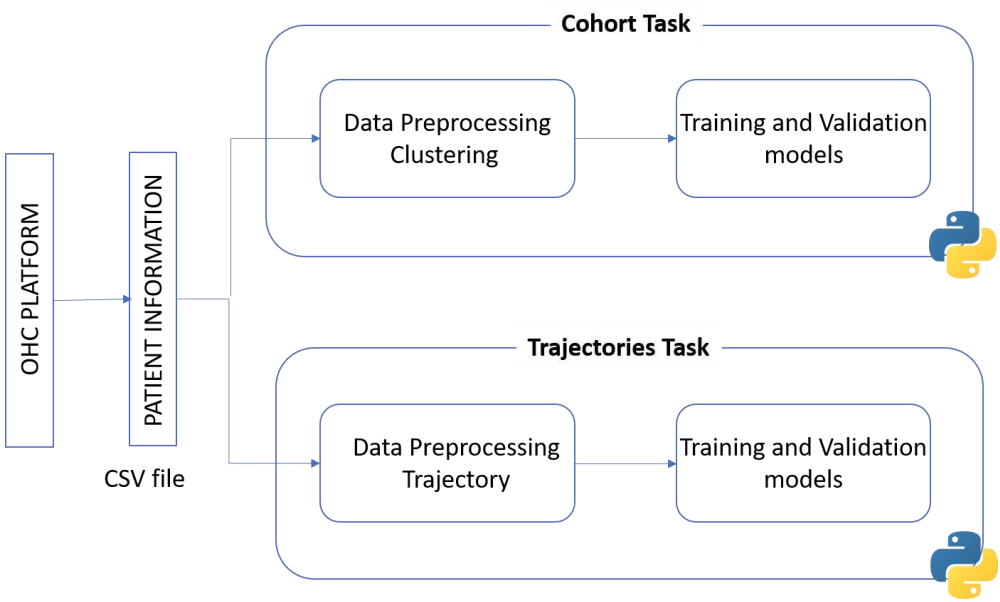
[11] Index—FHIR v4.0.1. (n.d.). Retrieved July 24, 2020, from https://www.hl7.org/fhir/

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 875406

we can have a trajectory prediction of events, and disease characteristics (e.g. probability of readmission, clinical condition). Once these models are trained they can be of use to the CDSS.

Figure 7: basic description of the Cohort and Trajectory Analysis

Via the OHC platform HESSO will fetch the data in .csv format. The retrospective raw patient data will be preprocessed by filling the missing value, cleaning outliers, and preparing for the machine learning model to be trained. For the cohort task, HESSO will employ clustering machine learning algorithms, which aggregate



patients having shared characteristics. For trajectories analysis, HESSO will employ regression machine learning algorithms, which predict how patients progress from symptoms to disease. The flow diagram is not applicable to this module, but we can sketch how it could be integrated into the CDSS. The cohorts and trajectories are one of the inputs for the CDSS. So, when a decision support request is emitted for a given patient, the CDSS can use the already trained cohort and trajectory modules, in order to: (i) classify the patient, and (ii) have a trajectory prediction. The draft diagram below illustrates this scenario. In terms of data models, HL7 FHIR will be the preferred choice for communication protocol.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

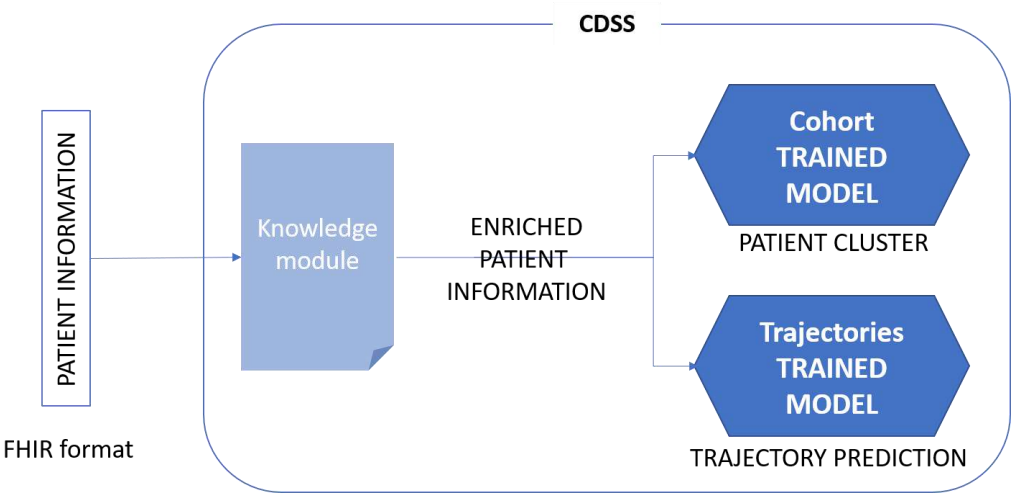


Figure 8: Scenario for Cohort and Trajectory Analysis

EHR Data preparation and Enrichment

This task will prepare data for further analysis and processing. EHR data comes from heterogeneous sources in different formats and languages. The correct diagnoses may not have been made yet, and even standardized classifications such as ICD9/10 are missing required granularity, as codes generally aggregate a host of different diagnoses. However, even with standardized EHR, which are brought into a uniform interoperable format, data is ambiguous and largely false as well as incomplete, due to the nature of medicine being a complex and chaotic science. Also subjectivity burdens all medical data with substantial bias.

In order to optimize data for the following analysis, an AI will be developed serving as “standardized second-opinion” to correct and normalize the EHRs. In this way incomplete but implicit data sets can be reconstructed through causative algorithms, and incongruous information identified. We will process scientific and clinical evidence to build this AI module. This is a breakthrough innovation enabled by building on Symptoma’s proprietary algorithms, databases and ontology.

The Data Flow Diagram shows how we get from an initially unedited EHR, an enriched version with relevant extra information. The idea is that a hospital or medical institution will send an EHR document to the OVH Server, which is XML, JSON or csv encoded, optionally it will be mapped according to the FHIR standard. On the OVH server there are instances of Symptoma AI, Gradiant, and HESSO running. These services will be responsible for the preprocessing, enrichment and further analysis of the EHR documents. After the enrichment of the EHR, detailed analysis can be performed and gained insights gathered in the knowledge base.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 875406

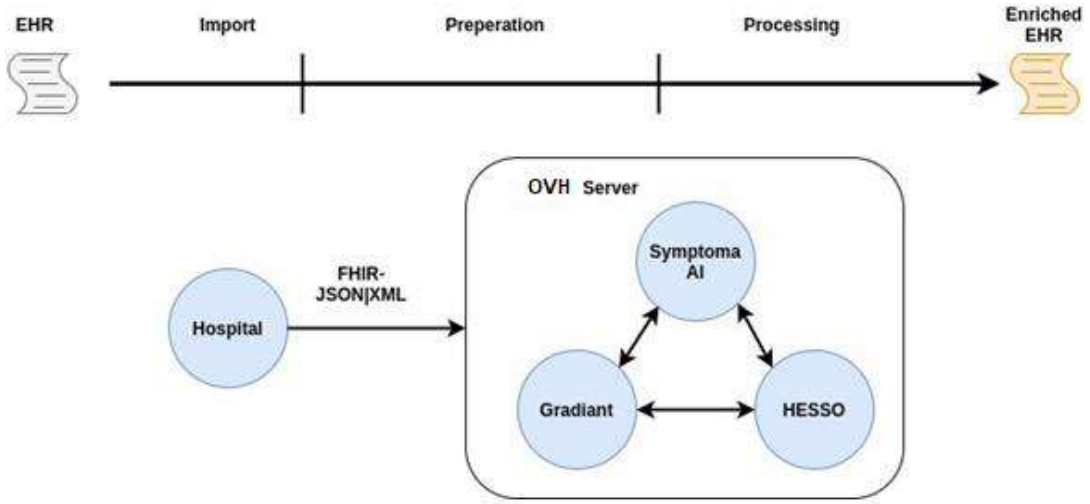


Figure 9: The Data Flow Diagram of the EHR Data preparation and Enrichment

The subsequent activity diagram shows a simple representation of the module. The Symptoma AI module receives the EHRs via an API. The Symptoma backend combines clinical and scientific evidence together with a big proprietary database to perform the enrichment process. The Symptoma backend performs regularly scheduled automated tests to ensure optimal results.

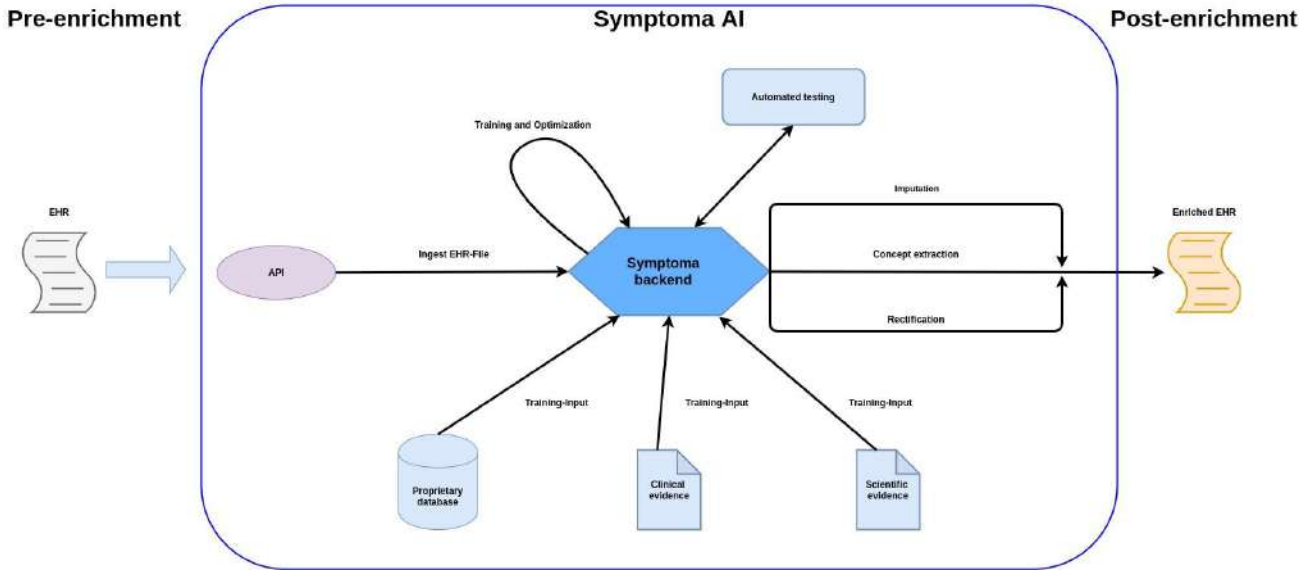


Figure 10: Simple representation of the EHR Data preparation and Enrichment

The Input data for the Symptoma AI provided via JSON-FHIR Protocol. The data is interchanged via RESTful APIs and are sent as JSON objects over Hypertext Transfer Protocol secured with Transport Layer Security (HTTPS). The output data will be JSON-FHIR.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Alert Mechanism

Alerts Management flows will also be implemented by EMODA. Probably the same flow engine (Drools) will be used for alert. All patients are monitored by the OHC platform all time. When the patient data such as blood pressure comes from wearable devices, the OHC platform first decides if an alert flow will be executed or not (e.g a "pre-alert"). If the platform decides to run alerts flow, it collects required parameters and sends a request to Alert Management Engine, which will be developed in CDSS, via CDS-Hook API with Insights Workbench component. The results are sent back to the OHC platform in FHIR format as JSON data (CDS Card).

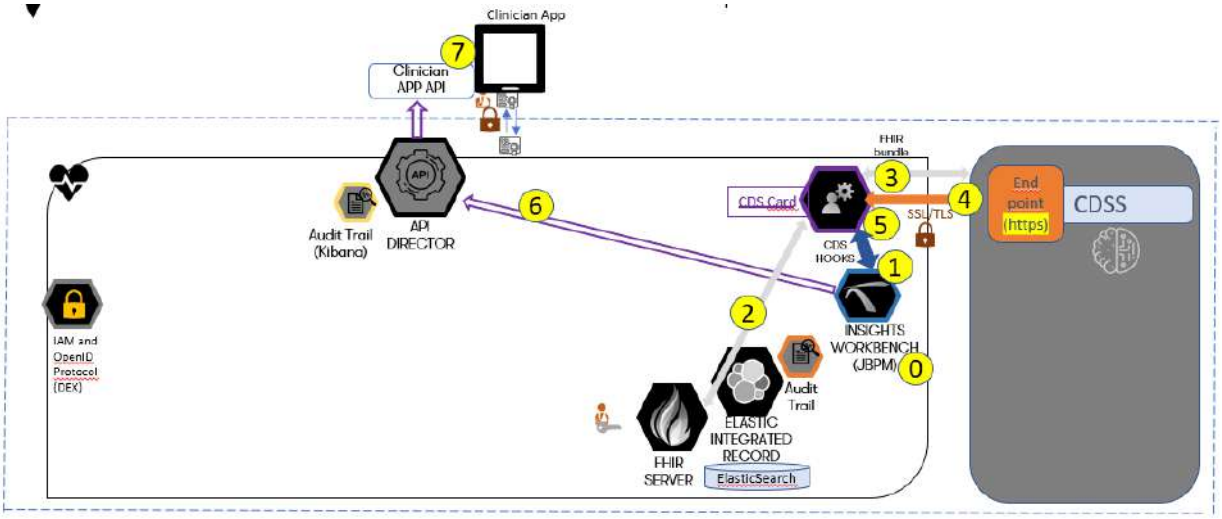


Figure 11: Alert data flow

0. Look for patients (potential) that match rules previously defined (pre-alert).
1. JBPM calls the CDS Hooks.
2. CDS Hooks get patient information (prefetch) from the FHIR server.
3. Push information to CDSS (endpoint provided by CDSS).
4. CDSS returns score for certain patients that need to trigger an alert.
5. The output of the CDSS is transformed into the right format (CDS card).
6. Sends a message to the app via a service that is exposed by the clinician app
7. The alert is displayed on the clinician app via push notification.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

1.2. Definition of Inputs and Outputs

1.2.1. Alerts: Inputs and Outputs

Because sometimes follow-up visits are not enough to timely address survivors' major issues, PERSIST will track EHR data, as well as, data generated by users on a daily basis so as to properly raise alerts. They will be mainly based on the PREMs/PROMs reported by survivors through the mHealth application or their equivalent gathered by the embodied conversational agent.

Based on the previous work performed in WP2 and improved in this task, in the following paragraphs we will describe which alerts/recommendations (outputs) may be raised by PERSIST and which is the data source of the information needed (inputs) to raise those alerts/recommendations. The questionnaires will be asked once a week. The survivor will be offered the possibility to answer the questionnaires when they consider it necessary.

Alert 1: Colorectal cancer interval

Importance Level: Red

Input 1: Cancer type

Dangerous value: colorectal

Input 2: Rectal bleeding

Question to be asked:

a. *Have you seen red blood in your feces? (YES / NO)*
   If the answer is YES:…… Question to be asked: b
   If the answer is NO: Go to the next input (Input 3)

b. *Blood in your feces, are a few drops of bright red (fresh) blood only occasionally?* (YES/NO)
   If the answer is YES: …… Question to be asked : c
   If the answer is NO:…… Question to be asked: d

c. Blood in your feces, is it mixed with it?
   If the answer is YES: Question to be asked: d
   If the answer is NO: …….. Question to be asked: e

d. Do you have recent onset rectal or perianal discomfort or pain?
   If the answer is YES: …….. Question to be asked: e
   If the answer is NO: …… ALERT

e. Has this been happening to you for more than 3 weeks?
   If the answer is YES: …… ALERT
   If the answer is NO: …….. *Recommendation*: Good rectal hygiene, sitz baths

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Go to the next input (Input 3)

Input 3: Changes in feces coloration

Question to be asked:

a. Have your feces become darker, blacker, tarry or maroon in color? (YES/NO)

If the answer is YES:……. Question to be asked: b

If the answer is NO: Go to the next input (Input 4)

b. Are you receiving iron supplements? (YES/NO)

If the answer is YES: …… Go to the next input (Input 4)

Recommendation: Don't worry, iron supplements make stools dark.

If the answer is NO:…… Question to be asked: c

c. Have you found yourself progressively more tired in the last week than in the previous weeks? (YES/NO)

If the answer is YES:…… ALERT

If the answer is NO: …….. Question to be asked: d

d. You have had changes in the intestinal rhythm or abdominal pain in the previous weeks?

If the answer is YES: …… ALERT

If the answer is NO: …… Question to be asked: e

e. Has this been happening to you for more than 2 weeks?

If the answer is YES: …… ALERT

If the answer is NO: …….. *Recommendation:* Look at the evolution of your stool coloration in the next week or the appearance of any of the symptoms we've asked you about

Go to the next input (Input 4)

Input 4: Change in bowel habits

Question to be asked:

a. Have you noticed changes in your usual bowel rhythm in recent weeks, such as diarrhea, constipation, or a combination of both? (YES/NO)

If the answer is YES:……. Question to be asked: b

If the answer is NO: *Recommendation*: A healthy diet and exercise regimen helps maintain a good bowel habit

b. Do you consider that what you have eaten in the last few weeks may have influenced the change in the intestinal rhythm? (YES/NO)

If the answer is YES: …… *Recommendation:* Look at the evolution of your bowel habit after following a healthy diet and healthy exercise guidelines (Dietary recommendations for patients with constipation and others for occasional diarrhea may be included)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

If the answer is NO:…… Question to be asked: c

c. The change in the intestinal rhythm is associated with abdominal pain or the appearance of blood in the stool?

If the answer is YES:…….. ALERT

If the answer is NO: …….. Question to be asked: d

d. Has this been happening to you for more than 3 weeks?

If the answer is YES…….. ALERT

If the answer is NO: …….. *Recommendation:* Look at the evolution of your bowel habit after following a healthy diet and healthy exercise guidelines (Dietary recommendations for patients with constipation and others for occasional diarrhea may be included)

Alert 2: Breast cancer recurrence or new primary breast cancer

Importance Level: Red

Input 1: Cancer type

Dangerous Value: Breast

Input 3: Breast lumps

Question to be asked: Have you noted a breast lump?

ALERT

Dangerous answer/value: Yes.

How many times the dangerous value should happen before raising the alert to doctor (number): For the first time, patient must be evaluated by a clinician.

RECOMMENDATION

A previous recommendation should be raisen to the patient?: Yes

Which recommendation?: If patient notes breast lump must seek medical attention

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

How many times the dangerous value should happen before raising a recommendation to the patient (number):For the first time, patient must be evaluated by a clinician.

Input 4: Skin retraction

Question to be asked:Have you noted a breast skin retraction.

ALERT

Dangerous answer/value: Yes

How many times the dangerous value should happen before raising the alert to doctor (number): For the first time, patient must be evaluated by a clinician.

RECOMMENDATION

A previous recommendation should be raisen to the patient?: Yes

Which recommendation?: If patient notes breast skin retraction must seek medical attention

How many times the dangerous value should happen before raising a recommendation to the patient (number):For the first time, patient must be evaluated by a clinician.

Input 5: Nipple discharge

Question to be asked:Have you had a nipple discharge?

ALERT

Dangerous answer/value: Yes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

How many times the dangerous value should happen before raising the alert to doctor (number): For the first time, must be evaluated by a clinician

RECOMMENDATION

A previous recommendation should be raisen to the patient?: Yes/No

Which recommendation?: if patient has had a nipple discharge must seek medical attention

How many times the dangerous value should happen before raising a recommendation to the patient (number): For the first time, patient must be evaluated by a clinician.

Alert 3: Low HRQoL

Importance Level: Yellow

Input 1: Warning values in HRQoL assessment

Data Source: PREM in mHealth app.
Selected questionnaire to be asked: preferred QLQ-30 from EORTC

ALERT

Dangerous answer/value: (will be determined after discussion with wider clinicians group)

How many times the dangerous value should happen before raising the alert to doctor? (number): (will be determined after discussion with wider clinicians group)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

RECOMMENDATION

A previous recommendation should be raisen to the patient?: Yes/No

Which recommendation?:

How many times the dangerous value should happen before raising a recommendation to the patient (number):

Alert 4: Cardiovascular risk

Importance Level: Yellow

Input 1: Warning values in Cardiovascular assessment

Data Source: PREM in mHealth app.
Selected questionnaire to be asked:
Can you maintain your normal physical activity?
1.- Yes. No further questions is needed
2.- No. Answer the next question.
Are you worsening shortness of breath with activity?
1.- YES. Patient must be evaluated by a clinician.
2.- No. Answer the next question.
Have you increased swelling of legs, feet, and ankles?
1.- YES. Patient must be evaluated by a clinician.
2.- No. If you have other symptoms, patient must be evaluated by a clinician

ALERT

Dangerous answer/value: The answers in red

How many times the dangerous value should happen before raising the alert to the doctor? (number): 1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

The alert level may be increased based on the following input variables (from 2 to 5). Whenever this information can be extracted from EHR, that would be the preferred way over asking the survivor about it.

Input 2: Treatment

Data Source: Preferably extracted from EHR. If the information is not there, it could be asked in the mApp with the following questionnaire.

1.- Did you receive chemotherapy treatment based on Antracyclines (Doxorubicin, Epirubicin), trastuzumab, pertuzumab or 5-Fluorouracil.
    1.- Yes. Risk factor for developing Heart Failure or Sudden Death. Increase alert level.
    2.- No.
2.- Did you receive left breast radiotherapy?
    1.- Yes. Risk factor for developing Heart Failure
    2.- No.

Input 3: chronic diseases

Data Source: Preferably extracted from EHR. If the information is not there, it could be asked in the mApp with the following questionnaire.

Have you diabetes or high blood pressure or hyperlipidemia?
    1.- Yes. Risk factor for developing Heart Failure. Increase alert level.
    2.- No.

Input 4: BMI

Data Source: Preferably extracted from EHR. If the information is not there, weigh and height could be asked in the mApp.

Body Mass Index > 30 .Risk factor for developing Heart Failure. Increase alert level.

Input 5: Smoking habit

Data Source: Preferably extracted from EHR. If the information is not there, it could be asked in the mApp with the following questionnaire.

Tobacco use
    1.- Yes. Risk factor for developing Heart Failure. Increase alert level.
    2.- No

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

RECOMMENDATION

A previous recommendation should be raisen to the patient?: No

Which recommendation?: --

How many times the dangerous value should happen before raising a recommendation to the patient (number): --

Alert 5: Lymphedema

Importance Level: Yellow

Input 1: Type cancer.

Dangerous value: Breast.
Don't analyse any other input if type cancer is not Breast.

Input 2: Warning values in Lymphedema assessment

Data Source: PREM in mHealth app.
Selected questions to be asked:
SIGNS of lymphedema
1-do you feel your rings have became too narrow?
2-do you feel that your sleeves tighten more on the side of the surgery?
3-do you feel your arms have different size?
4-do you think the arm near the surgery has increased its size?
SIGNS of lymphangitis:
5-do you feel readness or heat near the surgery area?
6-do you have fever?

If yes to any in (1 to 4) questions:
7- did you make an effort before your arm started to become bigger?

ALERT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Dangerous answer/value: YES to any question about lymphedema (1-4) AND YES to any question about lymphagitis (5 or 6).

How many times the dangerous value should happen before raising the alert to the doctor? (number): 1

ALERT

Dangerous answer/value: YES to any question about lymphedema (1-4)

How many times the dangerous value should happen before raising the alert to the doctor? (number): 2

RECOMMENDATION

A previous recommendation should be raisen to the patient?: Yes

Which recommendation?:

If YES to any question about lymphedema (1-4) and also YES to effort question (7), then:

Rest (decrease in production) / Raised arm (facilitation of drainage)

How many times the dangerous value should happen before raising a recommendation to the patient (number): 1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

RECOMMENDATION

A previous recommendation should be raisen to the patient?: Yes

Which recommendation?:

If Risk of lymphedema,

1-avoid to repeat the same movements.

2- avoid overweight or gaining weight.

3- avoid injections, blood tests, blood pressure measurement in the arm.

4-avoid sunburn, scuffs (gardening, pets), bites, cuts and burns

How many times the dangerous value should happen before raising a recommendation to the patient (number):

The above recommendation should be raised if axillary lymph nodes are removed during breast surgery (with sentinel node biopsy or axillary dissection) or treated with radiation therapy (info from EHR)

Alert 6: Pain

Importance Level: Yellow

Input 1: Warning values in Pain assessment

Data Source: PREM in mHealth app.
Selected questionnaire to be asked: NCCN2020 guidelines survivor pain assessment.
NCCN2020 questions for Pain assessment
1.- Are you having any pain? Yes/No
If yes, follow with question 2.
2.- How would you rate your pain on a scale of 0 (none) to 10 (extreme) over the past month? 0–10

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

ALERT

Dangerous answer/value: If YES to question 2 give a rating of > 4

How many times the dangerous value should happen before raising the alert to the doctor? (number): 1

RECOMMENDATION

A previous recommendation should be raisen to the patient?:No

Which recommendation?:

How many times the dangerous value should happen before raising a recommendation to the patient (number):

Alert 7: Hormonal disbalances

Importance Level: Yellow

Input 1: Gender

Dangerous value: female
Don't analyse any other input if gender is not female.

Input 2: Warning values in Hormonal disbalances assessment

Data Source: PREM in mHealth app.
Selected questionnaire to be asked: NCCN2020 guidelines
>Have you been bothered by hot flashes/night sweats? Yes/No
>Have you been bothered by other hormone-related symptoms (ex, vaginal dryness, incontinence)? Yes/No

ALERT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Dangerous answer/value: If YES to any question.

How many times the dangerous value should happen before raising the alert to the doctor? (number): 1

RECOMMENDATION

A previous recommendation should be raisen to the patient?: No

Which recommendation?:

How many times the dangerous value should happen before raising a recommendation to the patient (number):

Alert 8: Sexual dysfunction

Importance Level: Yellow

Input 1: Warning values in Sexual dysfunction assessment

Data Source: PREM in mHealth app.
Selected questionnaire to be asked:
1.-Are you pleased with your intimate life?
If NO, ask question 2.
2.- Is your intimacy causing you distress?

Note: for the psychologists of the CHU of Liège, it is important to know how was the sexual life of a patient before the illness. If the patient's sexuality did not go well before the disease, it will not go better afterwards and light advices through an app will have no effect because the discomfort is deeper and will worsen with the disease.

ALERT

Dangerous answer/value: If NO to question 1 or YES to question 2.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

How many times the dangerous value should happen before raising the alert to the doctor? (number): There is a risk of getting many alerts because the sexual dysfunctioning is quite spread in cancer patients for many different reasons. For that reason the alert can appear the second time it is detected. However the decision to see a psychologist/doctor must arise only from the patient.

A possibility is to show a pop-up message in the patient app to ask him/her if they want to meet a professional. For example: *"You must know that a specialist is at your disposal for talking about it."*

The patient should have the right to accept or refuse the option. In any case, the alert will be raised to the doctor explaining whether the survivor requires or not the face-to-face visit. The patient should know that even if he/she refuses to see a professional an alarm has been sent to the professionals who follow him/her. It could be explained in the information sheets during the recruitment or shown in the app when it is opened the first time. At CHU Liège, the psychologists think that the patient should be informed at the recruitment.

RECOMMENDATION

A previous recommendation should be raisen to the patient?: Yes

Which recommendation?:

The best recommendation is to talk about the problem with his/her partner or a professional (physician or psychologist) or any other person who at the eyes of the patient is emotional supportive

Nevertheless, for helping/encouraging the patient to see a professional, the app:

1- can assess the sexual life of the patient before the disease and at present: see questions QLQ-BR45 sexual activity

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

2- to help the patient to distinguish between a sexual dysfunction due to a physiological imbalance or to an emotional stress (e.g vaginal dryness, decreased libido, etc…)

How many times the dangerous value should happen before raising a recommendation to the patient (number):Since the first signs send a recommendation. Professionals should only react after a patient demands.

Alert 9: Fatigue

Importance Level: Red

Input 1: Warning values in Fatigue assessment

Data Source: PREM in mHealth app.
Selected questionnaire to be asked:NCCN2020 guidelines
1.- Do you feel persistent fatigue despite a good night's sleep? Yes/No
2.- Does fatigue interfere with your usual activities? Yes/No
3.- How would you rate your fatigue on a scale of 0 (none) to 10 (extreme) over the past week? 0–10

If < 3- mild
If 4-6- moderate
If >6 – severe – ALERT

ALERT

Dangerous answer/value: If YES to either question 1 or 2, or a rating of >6 to question 3.

How many times the dangerous value should happen before raising the alert to the doctor? (number): 1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

RECOMMENDATION

A previous recommendation should be raisen to the patient?: No

Which recommendation?:

How many times the dangerous value should happen before raising a recommendation to the patient (number):

Alert 10: Sleep disorders

Importance Level: Yellow

Input 1: Warning values in Sleep disorders assessment

Data Source: PREM in mHealth app..
Selected questionnaire to be asked (if wearable measures correlate with ISI, the questionnaire could be substituted by them): ISI Insomnia Severity Index.

NOTE: PERSIST should assess sleep disturbance before and after the disease (iatrogenic or emotional sleep imbalance)

ALERT

Dangerous answer/value: ISI Score >=15.

If measure with wearable it may having sleep problems above 3 nights/week.

How many times the dangerous value should happen before raising the alert to doctor? (number): The first time the dangerous answer/value appears.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

RECOMMENDATION

A previous recommendation should be raisen to the patient?: Yes

Which recommendation?: Maybe with ISI score between 8-14.

Recommendation:

- sleep–wake behavior hygiene (sleep and wake-up at the same time every day),
- stimulus control therapy,
- sleep restriction (avoiding naps),
- relaxation but also exercise for increasing melatonin concentration (avoiding exercise late in the evening)
- cognitive techniques,
- music therapy
- autohypnosis
- If awake for more than 15-20 min, get out of bed and return to sleep only when the desire to sleep comes back.

How many times the dangerous value should happen before raising a recommendation to the patient (number):1

Alert 11a: Psychosocial status: anxiety.

Importance Level: Yellow

Input 1: Warning values in anxiety assessment

Data Source: PREM in mHealth app.
Selected questionnaire to be asked: Generalized anxiety disorder 7-item (GAD-7)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Over the last 2 weeks, how often have you been bothered by the following problems?
1.Feeling nervous, anxious or on edge
2.Not being able to stop or control worrying
3.Worrying too much about different thing
4.Trouble relaxing
5.Being so restless that it is hard to sit still
6.Becoming easily annoyed or irritable
7.Feeling afraid as if something awful might happen

Answers values are from 0 to 3:
0 - Not at all, 1 - Several days, 2 - More than half the days, 3 - Nearly every day

The following cut-offs correlate with level of anxiety severity:
Score 0-4: Minimal Anxiety
Score 5-9: Mild Anxiety
Score 10-14: Moderate Anxiety
Score greater than 15: Severe Anxiety

ALERT

Dangerous answer/value: GAD-7 scoring > 8.

How many times the dangerous value should happen before raising the alert to doctor? (number): 1 time

NOTE: “Based on a recent meta-analysis [3], some experts have recommended considering using a cut-off of 8 in order to optimize sensitivity without compromising specificity.” [1] For this reason a strict criteria could be adopted 5 points and more being enough for alerting the doctor. If the criteria should be less strict, than 10 points and above might be sufficient.

In case of the less strict criteria - 1 time; with the more strict - 2 times

RECOMMENDATION

A previous recommendation should be raisen to the patient?: No

Which recommendation?: --

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 875406

How many times the dangerous value should happen before raising a recommendation to the patient (number): --

Alert 11b: Psychosocial status: Depression.

Importance Level: Red

Input 1: Warning values in depression assessment

Data Source: PREM in mHealth app.
Selected questionnaire to be asked: PHQ-2 (Patient Health Questionnaire). PHQ-2 consists first two questions from PHQ-9
Answers values are from 0 to 3: 0 - Not at all, 1 - Several days, 2 - More than half the days, 3 -Nearly every day [1]

Over the last 2 weeks, how often have you been bothered by the following problems?
1.Little interest or pleasure in doing things
2.Feeling down, depressed or hopeless

ALERT

Dangerous answer/value: A cut off score of 2 or higher.

How many times the dangerous value should happen before raising the alert to doctor? (number): 1.

RECOMMENDATION

A previous recommendation should be raisen to the patient?: No

Which recommendation?: --

How many times the dangerous value should happen before raising a recommendation to the patient (number): --

Alert 12: Cognitive function

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Importance Level: Yellow

Input 1: Warning values in Cognitive function assessment

Data Source: PREM in mHealth app.
Selected questionnaire to be asked: NCCN2020 questions for Cognitive Function assessment
1. Do you have difficulties with multitasking or paying attention? Yes/no
2. Do you have difficulties with remembering things? Yes/no
3. Does your thinking seem slow? Yes/no

ALERT

Dangerous answer/value: "If YES to any question"

How many times the dangerous value should happen before raising the alert to doctor? (number): 1. However there is no information on the questions and no studies which would define the values that are significant.

RECOMMENDATION

A previous recommendation should be raisen to the patient?: No

Which recommendation?: --

How many times the dangerous value should happen before raising a recommendation to the patient (number): --

Alert 13: Malnutrition.

Importance Level: Yellow

Input 1: Intake of nutrients.

Selected question to be asked:
Do you have adequate daily intake of nutrients regarding quantity and quality (vegetables and fruits half the volume on the plate, whole grains 30%, protein 20%) ?

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

ALERT

Dangerous answer/value: No.

How many times the dangerous value should happen before raising the alert to the doctor (number): 3 out of 3 subsequent weeks or 5 out of 10 non-subsequent weeks.

NOTE: The inability to do this might also be a symptom of some other illness or maybe even cancer recurrence.The patient must be evaluated by a clinician and possibly referred to a dietitian.

RECOMMENDATION

A previous recommendation should be raisen to the patient?: Yes

Which recommendation?: To increase the needed daily intake of nutrients if possible.

How many times the dangerous value should happen before raising a recommendation to the patient (number): If the daily intake is inadequate in 3 or more out of 7 days of the week.

Input 2: Intake of healthy nutrients.

Selected question to be asked:
Do you have adequate daily intake of healthy nutrients (at least five servings of fruits or vegetables daily, avoidance of processed foods, limitations of too much red meat and alcohol) ?

ALERT

Dangerous answer/value: No.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

How many times the dangerous value should happen before raising the alert to the doctor (number): 3 out of 3 subsequent weeks or 5 out of 10 non-subsequent weeks.

NOTE: The inability to do this might also be a symptom of some other illness or maybe even cancer recurrence. The patient must be evaluated by a clinician and possibly referred to a dietitian.

RECOMMENDATION

A previous recommendation should be raisen to the patient?: Yes/No

Which recommendation?: To increase the appropriate daily intake of nutrients if possible.

How many times the dangerous value should happen before raising a recommendation to the patient (number):If the daily intake is inappropriate in 3 or more out of 7 days of the week.

Input 3: Weight evolution in 3 months.

Selected question to be asked: Has your weight changed in last 3 months by more than 5%?

ALERT

Dangerous answer/value: Increase or decrease of body weight of more than 5%.

How many times the dangerous value should happen before raising the alert to the doctor (number): With no positive change in subsequent month, the patient must be evaluated by a clinician.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

RECOMMENDATION

A previous recommendation should be raisen to the patient?: Yes

Which recommendation?: With weight decrease the recommendation for intake of nutrients and healthy nutrients and with weight increase the recommendations regarding healthy lifestyle.

How many times the dangerous value should happen before raising a recommendation to the patient (number): The change of more than 5% in either direction.

Input 4: Meal frequency.

Selected question to be asked:
What is your daily meal frequency and timing of meals?

ALERT

Dangerous answer/value: Daily meal frequency less than 3 and irregular timing of meals.

"Irregular timing" is not always at the same time. For instance one day leaving out lunch and instead having a big dinner. Or having lunch one day at noon and the next in the middle of the afternoon. Having no established rhythm.

How many times the dangerous value should happen before raising the alert to the doctor (number): 3 out of 3 subsequent weeks or 5 out of 10 non-subsequent weeks.

NOTE: The patient must be evaluated by a clinician and possibly referred to a dietitian.

RECOMMENDATION

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

A previous recommendation should be raised to the patient?: Yes

Which recommendation?: To plan a regular frequency of meals.

How many times the dangerous value should happen before raising a recommendation to the patient (number): If the daily frequency and timing of meals are inappropriate in 3 or more out of 7 days of the week.

Input 5: Fluid intake frequency.

Selected question to be asked:
Do you have adequate daily intake of fluids (more than 1,5 litres)?

ALERT

Dangerous answer/value: No.

How many times the dangerous value should happen before raising the alert to doctor (number): 3 out of 3 subsequent weeks or 5 out of 10 non-subsequent weeks.

NOTE: The inability to do this might also be a symptom of some other illness or maybe even cancer recurrence. The patient must be evaluated by a clinician and possibly referred to a dietitian.

RECOMMENDATION

A previous recommendation should be raisen to the patient?: Yes

Which recommendation?: To increase the appropriate daily intake of nutrients if possible.

How many times the dangerous value should happen before raising a recommendation to the patient (number): If the ingested amount of fluids is less than 1,5 litres daily in 3 or more out of 7 days of the week.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Alert 14: Gastrointestinal conditions.

Importance Level: Red

Input 1: Swallogins issues.

Selected question to be asked: Do you experience difficulties in swallowing affecting your intake of nutrients?

ALERT

Dangerous answer/value: Yes.

How many times the dangerous value should happen before raising the alert to doctor (number): 1

RECOMMENDATION

A previous recommendation should be raisen to the patient?: No

Which recommendation?:

How many times the dangerous value should happen before raising a recommendation to the patient (number):

Alert 15: Social issues. Functional ability

Importance Level: Yellow

Input 1: Warning values in Functional ability assessment

Data Source: PREM in mHealth app.
Selected questionnaire to be asked: WHO questionnaire: 12-item Instrument Scoring Sheet .

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

ALERT

Dangerous answer/value:

For questions S1-12: if the answer is “al least severe”.

For questions H1-3 if the answer is “more than 10 days”

How many times the dangerous value should happen before raising the alert to the doctor? (number):

For S1-12: With no improvement and if this happens in 3 subsequent weeks or in 5 out 10 weeks.

For H1-3: If it happens subsequently more than twice or more than 2 months out of a 4 month period.

RECOMMENDATION

A previous recommendation should be raised to the patient?: No

Which recommendation?:

How many times the dangerous value should happen before raising a recommendation to the patient (number):

For S1-12: If it happens in 3 or more days out of 7 days of the week.

For H1-3: If it happens more than one time.

Alert 16: Social issues. Self-efficacy

Importance Level: Yellow

Input 1: Warning values in Self-efficacy assessment

Data Source: PREM in mHealth app.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 875406

Selected questionnaire to be asked: General self-efficacy Scale (GSES)

ALERT

Dangerous answer/value: --

How many times the dangerous value should happen before raising the alert to the doctor? (number): if it's possible, it might be best for the patient to raise the alert him/herself, when he/she feels in need of advice/help in this area

RECOMMENDATION

A previous recommendation should be raised to the patient?: No

Which recommendation?:

How many times the dangerous value should happen before raising a recommendation to the patient (number):

Alert 17: Healthy Lifestyle

Importance Level: Yellow

Input 1: Warning values in Healthy Lifestyle assessment

Data Source: PREM in mHealth app. Suggested questionnaires: NCCN2020 guidelines.
Questionnaire to be asked: Global physical activity questionnaire (GPAQ)

ALERT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Dangerous answer/value: Less than cumulative 150 minutes of moderate-intensity aerobic physical activity throughout the week or if bouts of activity are shorter than 10 minutes.

How many times the dangerous value should happen before raising the alert to doctor? (number): With continuing deterioration in 10 weeks. With no improvement in 15 weeks.

RECOMMENDATION

A previous recommendation should be raisen to the patient?: Yes/No

Which recommendation?:

How many times the dangerous value should happen before raising a recommendation to the patient (number): if the value happens in three subsequent weeks or 3 out 6 weeks.

1.2.2. Management of alerts: inputs

From D2.5 we know that, once an alert is raised, a professional assessment might be needed. In the following paragraphs the way in which PERSIST may support these assessments will be described, as well as, the data which may facilitate this task when it comes to professional praxis. It is worth noting that when the oncologist in charge of survivor starts to perform an assessment, this is usually aimed at finding out cancer recurrence or preventing secondary diseases.

Management: Colorectal cancer recurrence

Physicians perform an assessment of anemia:
w

- Questions aimed at finding out anemia.
- Physical exploration.
- Any multimodal sensing network data, such as heart rate from bracelets.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Based on this information, if professionals find it necessary, they could ask for a blood test to assess if a transfusion or admission is needed. Once the blood test is performed, the key information is :

- hematocrit,
- haemoglobin,
- urea,
- tumoral marker: CEA, Ca19,9 (high values are dangerous).

After a review of these results, if recurrence is suspected, an image test may be asked. Suggested image test are: TAC for abdomen, torax, pelvis. Keywords which: “dilation of the intestine”,”lesion”, “nodules”, “adenopathy”, etc.

Management: Breast cancer recurrence or new primary breast cancer

Physicians perform assessment of breast cancer:

- Physical exploration.

Based on this a mammography may be requested. Key information is BI-RADS variable which varies from 1 (no problem) to 5 (cancer detected).

If mammography is not conclusive, a blood test could be advisable. The critical information are tumoral marker: CEA, CA15,3 (high values are dangerous).

Management: Cardiovascular diseases

Physicians perform assessment for cardiovascular diseases:

- Questions aimed at finding out cardiovascular diseases.
- Physical exploration.

If any cardiovascular disease is suspected, then the survivor is given a referral to cardiology specialist.

On the contrary, If assessment is not conclusive, further tests are performed:

- Echocardiography. Key information:
  - LVEF: if LVEF< 50, then there is a risk.
  - ESLG: if ESLG has increased over time, then there is a risk. Normal values are between -19 and -21, inclusive.
- Blood test is also advisable. Key information:
  - pro-BNP. The safest value depends on the age. In average, if pro-BNP> 200, then there is a risk.

Once the additional tests are performed, the survivor is referred to cardiology specialist.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 875406

Management: Lymphedema

Physician performs an assessment of Lymphedema:

- Compares both arms.
- Check lymphadenectomy

If lymphedema is suspected, then the survivor is derived to physiotherapist.

Management: Pain

If the pain is chronic, then oncologist may contact the survivor to modify the pain treatment even before performing the visit.

If the pains is NOT chronic, an assessment for cancer (breast/colorrectal) is performed. The oncologist will adjust or recommend a treatment based on guidelines.

Management: Hormonal disbalances

When the survivor is in the a premopausic or perimenopausic phase, physician performs an assessment of hormonal disbalances, that is to say, to assess menopause.

If a hormonal study was performed, the key information is: FSH, LH, 17-beta-stradiol.

Oncologist gives referral to gynaecologist.

Management: Sexual disfunction

In case of breast cancer, oncologist gives referral to gynaecologist.
In case of colorectal, oncologist gives referral to urologist.

Management: Fatigue

Assessment for cancer (breast/colorrectal) is performed.

Management: Sleep disorders

If oncologist assesses of sleep disorders then modifies treatment
If treatment is not effective, then the survivor is referred to a specialist.
The survivor is derived to 1st psychologist. 2nd neurologist.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Management: Psychosocial status. Anxiety, depression and distress

Survivor derived to specialist

Management: Cognitive function

Physician performs an assessment of cognitive dysfunction:

- Questions aimed at finding out cognitive dysfunction.
- Specific questionnaires, such as Minimental Test.

If oncologist suspects cognitive dysfuntion, the survivor is referred to neurologist.

Management: Malnutrition.

The previously explained assessment for cancer (breast/colorectal) recurrence is performed.
If not cancer recurrence is found, then the survivor is referred to endocrinologist.

Management: Gastrointestinal conditions.

In case of motility issues, the previously explained assessment for colorectal cancer recurrence is performed.

If not cancer recurrence is found, then the survivor is referred to the digestive specialist.

Management: Social issues. Functional ability

Assessment for cancer (breast/colorrectal) is performed.
If not cancer recurrence is found, then the oncologist will evaluate other possible causes for the low functional ability, based on the general patient's information. Referral to specialist may be conducted.

Management: Social issues. Self-efficacy

Assessment for cancer (breast/colorrectal) is performed.

If not cancer recurrence is found, then the oncologist will evaluate other possible causes for the low self-efficacy, based on the general patient's information. Referral to specialist may be conducted.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Management: Healthy Lifestyle

Assessment for cancer (breast/colorrectal) is performed.

If not cancer recurrence is found, then the oncologist will evaluate other possible causes for the low adhesion to healthy habits, based on the general patient's information. The survivor may be given advice to improve her/his lifestyle.

Referral to specialists may be conducted.

1.2.3. Clinical information

In the following paragraphs we will summarize the required information for the alerts and for the clinical management of those alerts. Also, we will list other relevant information which may be initially used for cohorts and trajectories.

General Information

Gender
Age
Ethnicity
Status (Alive/Dead)
Cause of death (Cancer /Chronic/Other)
Weight
Height

Medical History

Genetic predispositions (cancer in their families)
BRCA1 positive
BRCA2 positive
HNPCC
Allergies
Previous diseases
Medications
Alcohol
Drugs
Smoking
BMI

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Diagnosis

Cancer type (C18/C19/C50)
Cancer stage (1/2/3)
Tumors information (T/N/M)
Tumor differentiation
Type of tumor
Estrogen receptor level
Progesterone receptor level
Her2 level (+/-)
Performance Status
Symptoms:

- nausea
- diarrhea
- fatigue
- skin irritation
- Swelling of all or part of the breast
- pain
- Nipple retraction (turning inward)
- The nipple or breast skin appears red, scaly, or thickened
- Nipple discharge
- constipation
- stools that appear narrower than usual
- feeling that the rectum is not completely empty after having a bowel movement
- light or very dark red blood in the stool
- bleeding from the rectum
- gas, abdominal cramps and bloating
- pain or discomfort in the rectum
- other

Laboratory:

- CA-15-3
- CA 19-9
- CEA

Treatment

Hospitalisation
Duration (days)
Type of surgery:

- mastectomy
- tumorectomy
- axillary lymph node dissection
- sentinel node biopsy
- colostomy

Breast reconstruction
Radiotherapy

- Zone
- Session modality

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Immunotherapy
Chemotherapy
- First Chemotherapy
- Second Chemotherapy

Targeted therapy
Endocrine therapy
Functional status scores (ECOG)

Relapse
- Site of relapse

Toxicity
RAS
BRAF

Suggested information for recurrence assessment

Blood test:
- hematocrit
- haemoglobin
- urea
- tumoral markers already listed in Diagnosis.

Image test reports (for colorectal)
Mammography:
- BI-RADS

Suggested information for cardiovascular assessment

Echocardiography:
- LVEF
- ESLG

Blood test:
- pro-BNP

Suggested Information for hormonal disbalances

Hormonal study:
- FSH
- LH
- 17-beta-stradiol

Questionnaires

With a pre screening at the beginning of the clinical study, PERSIST will decide which questionnaires are mandatory for certain patients.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

1.3. Definition of func. And non-func requirements

Functional Requirements of CDSS

- *Throughput Time of Restful Service (Max): 3 sec.*
  Throughput time should not be too long nor too short, so that a balance between system and network workloads can be achieved.

- *Data Size of Restful Service Payload (Max): 5 MB*
  To minimize network workloads, payload size should be as small as possible.

- *Simultaneous Request Number to be Handled (Max): 100*

  A high number of requests can decrease reliability, and can slow processing times when there are too many synchronous calls.

Non- Functional Requirements of CDSS

- All request will be handled synchronously (no need call back API in client side)
  Synchronous handling is more resource efficient than asynchronous handling, while allowing isolated error handling in the event if it occurs.

- Authentication Required (Oauth2 Standards)
- HTTPS Protocol
- TLS
  These protocols are required to ensure data security and privacy, which are essential, given the nature of data involved.

- Stable and Scalable
  Stability is required to prevent reliability decay, and scalability is required to address any reliability decay if it occurs.

- Maintainable
  Maintainability is essential to keep the system in an operational state, while allowing any modifications and updates to the software.

- Portable
  As a system operated on a cloud environment, portability has to be ensured to keep the software intact on distribution.

- Reliability (99%)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Being a critical system, reliability has to be ensured to process data as accurate as possible, even under heavy workload.

- Stateless
  A stateless system is more efficient on resource allocation, while having portability, maintainability and scalability advantages over a stateful system.

1.4. Standardization and Interoperability

CDS Hook

CDS Hooks provide services (restful API) to bring in recommendation from a CDS system (inference engine) into the patient or clinician workflow (EMR, mHealth app). CDS-hooks are built into the architecture at appropriate point to invoke to CDS services

It allows OHC to connect with clinical decision support (CDS) systems using an open industry interfacing standard. CDS-hooks allows OHC to pass data to clinical decision tools, and retrieve back recommendations for clinical actions without interrupting workflows. Supports connection of OHC to clinical decision support systems via industry standard REST APIs to enable automated decision support on data flowing through the interoperability platform.

Clinical Decision Support systems such as Deontics, allow decisions to be made and new insights discovered based on real-time data. CDS Hooks is a technology from 'Smart-On-FHIR'' that allows third-party CDS systems to register with OHC using the 'hook' pattern.

The third-party CDS system is able to provide OHC with information in the form of 'CDS cards' that may be surfaced through a system-of-engagement (e.g. Mobile app) or weaved into workflow (see the Insights Workbench). The CDS service returns 'Cards' which are typically rendered and displayed by the EHR. A typical use case is when a clinician is working with an EHR system. As they interact with the system, perhaps prescribing, the CDS system automatically returns additional information in the form of a 'card' which offers alternative suggestions. A different use-case involves the interception of data flowing across Viaduct. A call to CDS in real-time through the CDS-hooks interface can enrich the data with additional clinical context[12][13].

Overall requirements of CDS Hooks

User story

[12] https://cds-hooks.org/examples

[13] CDS-Hooks Web Site:https://cds-hooks.hl7.org/1.0/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

The clinician is using an EHR that is able to “call” a clinical decision support either automatically triggered when the clinician opens a field or manually on request when a clinician clicks on a button. This request is passed to the CDS hook passing the “service” that is requested and also any data that is needed to fulfil the request (e.g. patientID, condition,…). The CDS Hook will call the appropriate service and process the prefetch by retrieving the needed data from the FHIR server.

Functional requirements

EHR/apps need to get information from CDSS to guide and augment the decision of a clinician or patient. For instance, typically an EHR can call a CDS system to recommend a treatment based on the patient's condition or history. Similarly, a patient can be given a “risk score” based on their current medical record. The recommendation or score is worked out by the inference engine on a CDS system.

· Input-Output:

CDS Hooks is in between the system of engagement (EHR, mobile app) and the inference engine to ensure the flow of input and output is properly managed.

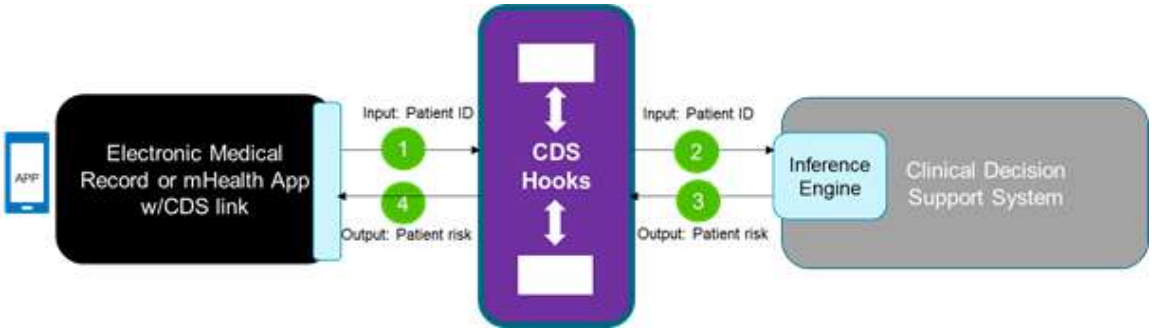


Figure 12-1: Simplified view of the process and flow of insights

CDS Hooks is managing the requests from the system of engagement (EMR, app) to the inference engine that will return an information (output) based on the input. Example above the EHR requests a diabetes risk score for a specific patient (1 and 2). The inference engine returns a score (3 and 4)

The CDS Hooks receives the input in terms of a “prefetch” with the data that the service needs and the definition of service that correspond to the inference the EMR needs.

The inference engine will produce the inference and return it (e.g. a risk score) in the shape of a CDS Card.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 875406

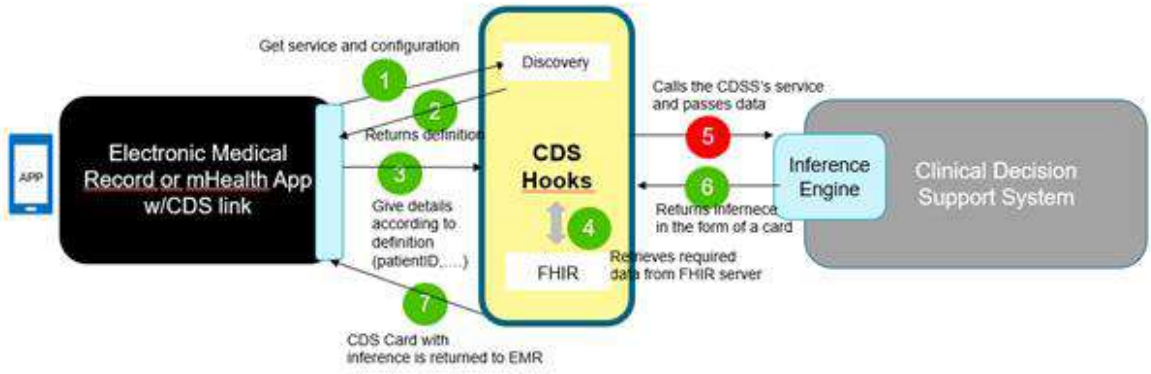


Figure 12-2: Simplified view of the process and flow of insights

Detailed requirements

The below is a more detailed picture of the steps that need to be taken with a CDS Hooks. The Word document on Basecamp details each steps[14].

Security

Every interaction between a CDS Client and a CDS Service is initiated by the CDS Client sending a service request to a CDS Service endpoint protected using the Transport Layer Security protocol. Through the TLS protocol the identity of the CDS Service is authenticated, and an encrypted transmission channel is established between the CDS Client and the CDS Service. Both the Discovery endpoint and individual CDS Service endpoints are TLS secured[15].

Formats:

All data exchanged through the RESTful APIs MUST be sent and received as JSON or XML structures, and MUST be transmitted over channels secured using the Hypertext Transfer Protocol (HTTP) over Transport Layer Security (TLS)

Input and output will use CDS Hooks standard (JSON or XML).

14 https://3.basecamp.com/3320520/buckets/14671082/uploads/2627053406

15 https://cds-hooks.hl7.org/1.0/#security-and-safety

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

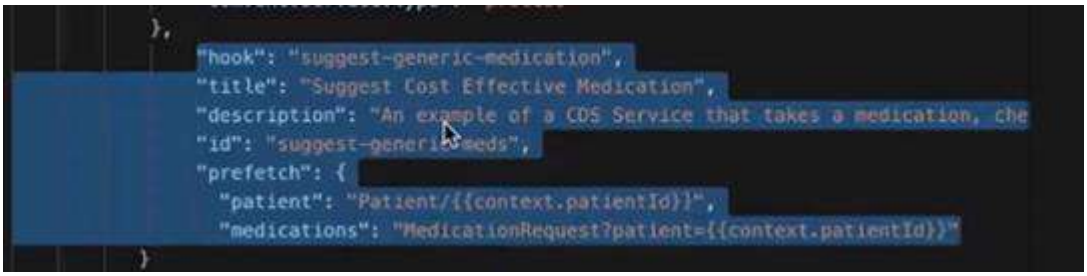


Figure 13: An example of an input (aka CDS "prefetch")

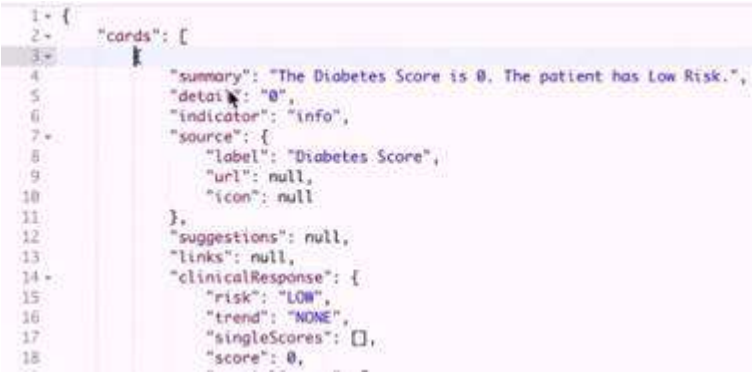


Figure 14: An example of an output- CDS Card

Request format:

as long as the CDS system/inference engine adheres to CDS hook specifications CDS Hook can make direct calls. Developers of CDS Services SHALL provide a stable endpoint for allowing CDS Clients to discover available CDS Services, including information such as a description of the CDS Service, when it should be invoked, and any data that is requested to be prefetched. A CDS Service provider SHALL expose its Discovery endpoint HTTP Request

The discovery endpoint SHALL always be available at {baseUrl}/cds-services. For example, if the baseUrl is https://example.com, the CDS Client MAY invoke: GET https://example.com/cds-services . CDS Service Response: for successful responses, CDS Services SHALL respond with a 200 HTTP response with an object containing a cards element as described below.

Terminology:

Terminology in FHIR typically uses the following but it is flexible. Depending on prefetch and CDS card content, this will need to be agreed.

SNOMED, RxNorm, ICD, LOINC

"prefetch": {

"p":{

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

```
"resourceType": "Patient",
"gender": "male",
"birthDate": "1974-12-25",
"...": "<snipped for brevity>"
},
"a1c": {
"resourceType": "Bundle",
"type": "searchset",
"entry": [{
"resource": {
"resourceType": "Observation",
"code": {
"coding": [{
"system": "http://loinc.org",
"code": "4548-4",
"display": "Hemoglobin A1c"
}]
},
"...": "<snipped for brevity>"
```


Figure 15: An example prefetch response using LOINC

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

2. Mobile Application for New EHR

2.1. Functionalities

CDSS Output

- The request from CDSS is started by a clinician for a specific patient.
- Clinicians can see related info about the patient that was collected within the project.
- Persist ID, gender, age, cancer type, stage of cancer, diagnostic date, TNM, info from wearables, radiotherapy/surgery, diagnostic performance status, tumor location
- The output of the CDSS will include results as percentage/score about groups and accuracy info.
- Previous CDSS outputs of the patient will be shown to the clinician with a timestamp.

Alerts

- Workbench will send the pre alerts in the types of data and CDSS will determine the alert with real values.
- When an alert is occured, the mobile app will give a notification through MQTT[16] which will also be used in WP4 Mhealth notifications.
- OHC platform will connect Multi modal sensing network through HTTPS and send the notification, which will be relayed to the mobile app via MQTT
- According to the importance level of the alert, notifications will be repeated in a specified time till read by the clinician.
- Reading timestamp of the alert by clinician and delivery timestamp will be stored.
- The info about the related patient will appear optionally via link for the communication.

Data Ingestion

- Mobile App will gather all data from OVH server.
- The connection between mobile app and Big data platform will be provided by API Directory.
- Data shown on mobile app will be determined by clinicians.
- Each clinician can see info about his/her patients only.
- There will be a logon page with a password for clinician entrance.

[16] **Deliverable 2.5 , Persist Project**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Appointment

- All appointment info will be taken from mHealth App.
- First appointment requests will be done by a clinician and then confirmed by a related patient or given another appointment time choices.
- After confirmation by both clinician and patient, the appointment was ended and set.
- In any case about confirmed appointment time, both clinician and patient can send a request to change.
- Clinicians can set repeated appointment times.
- All past appointments will be stored and can be seen by both clinician and patient.

2.2. Activity Diagrams

Figure 16: General Activity - 1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

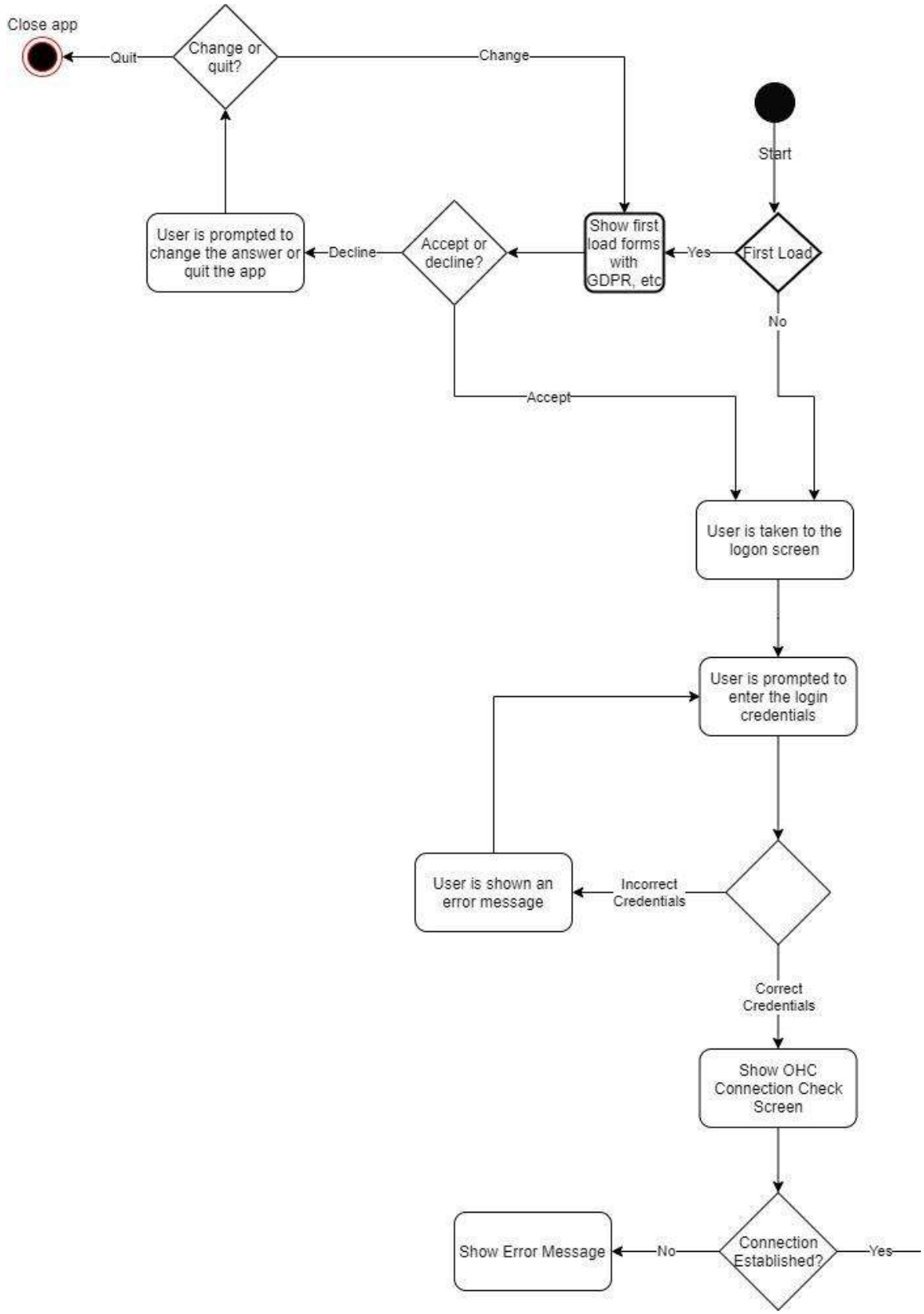


Figure 17: General Activity - 2

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

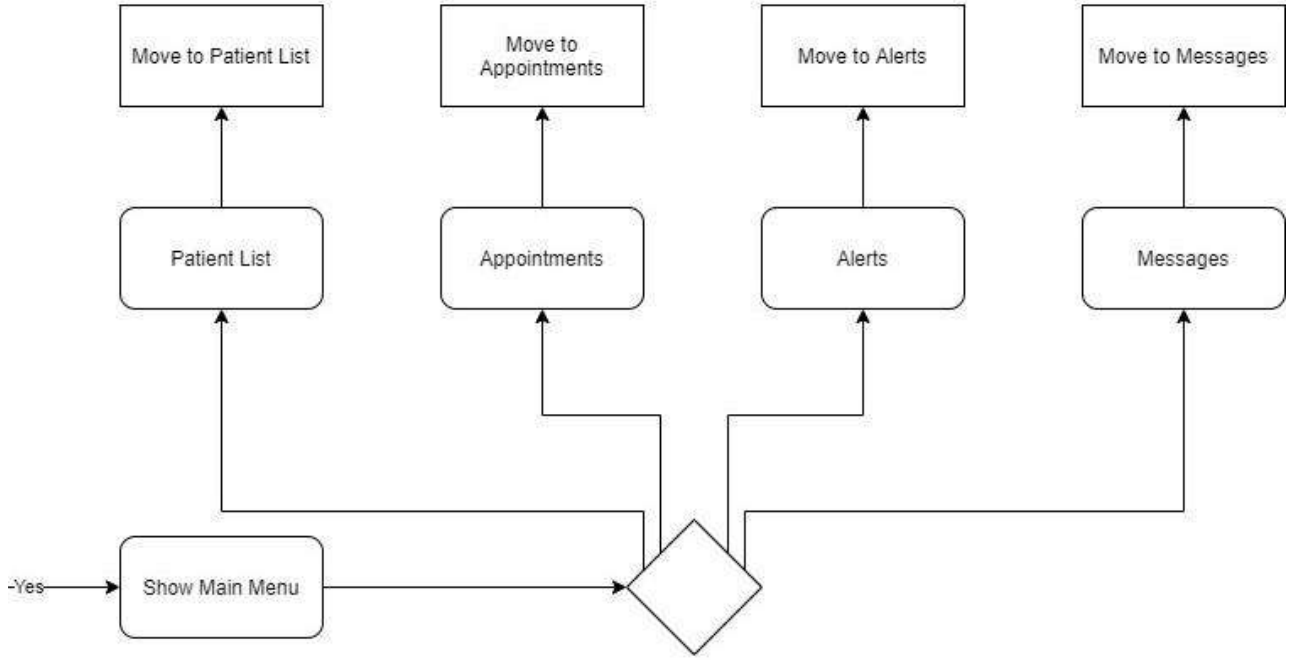


Figure 18: Patient List

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

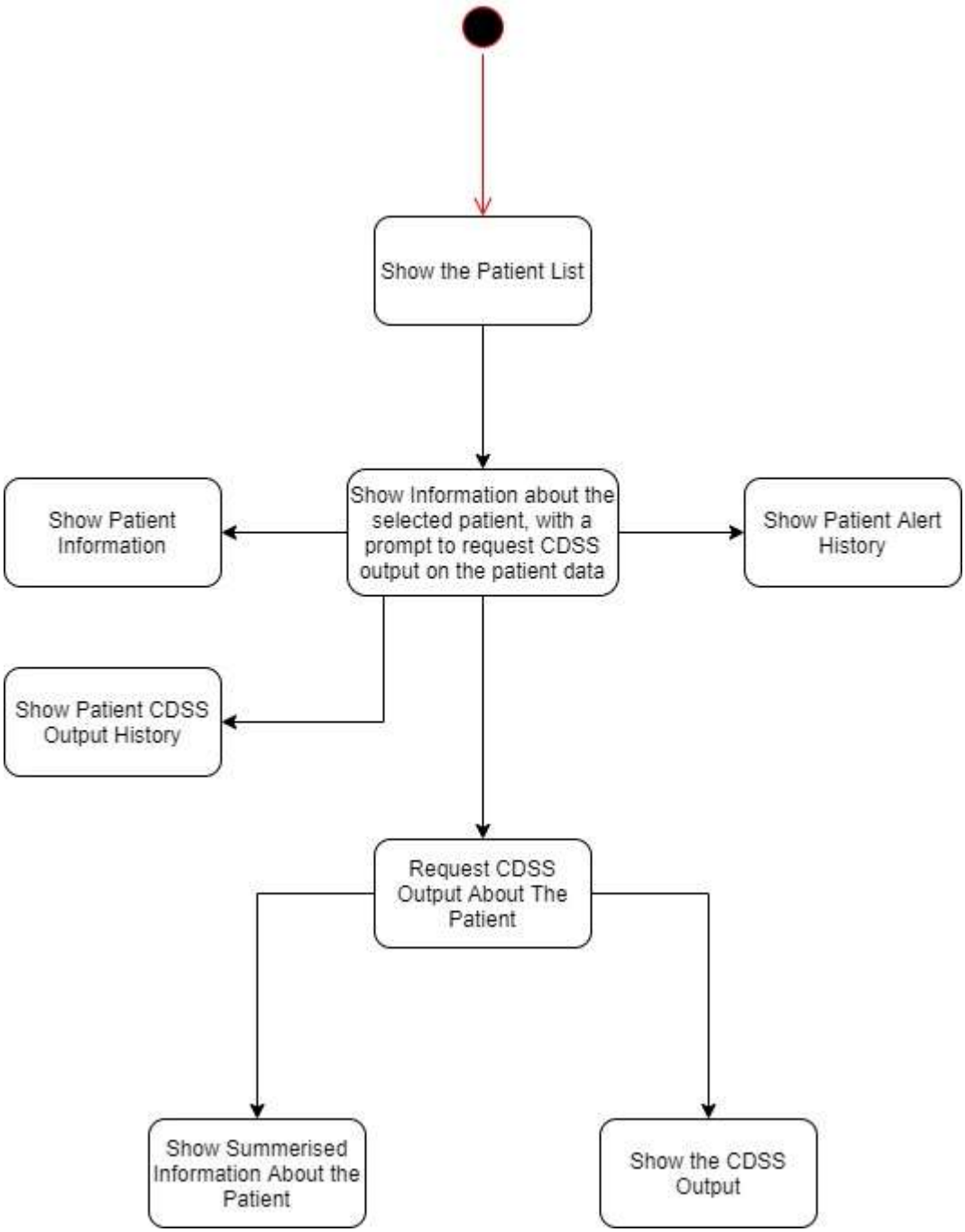
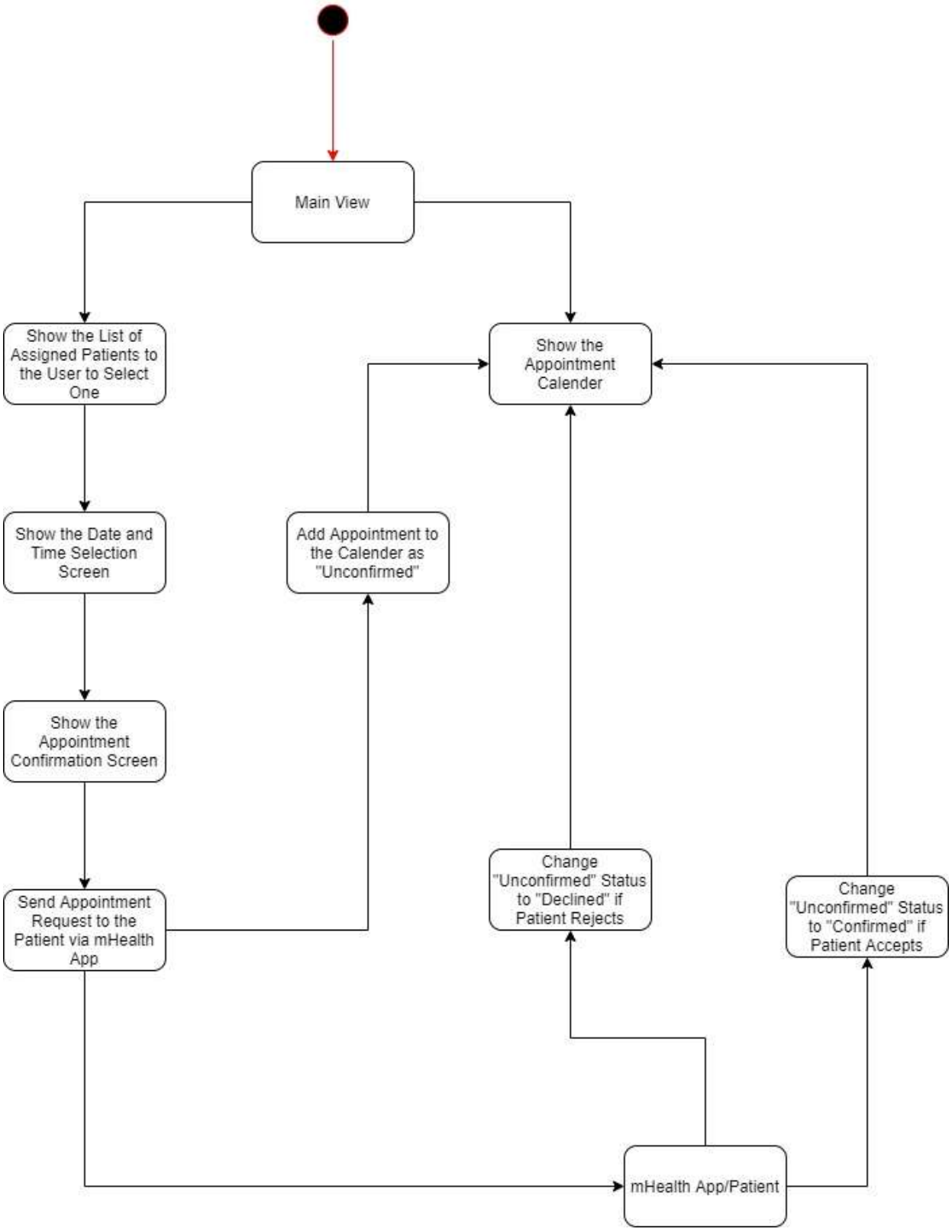


Figure 19: Appointments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Figure 20: Alerts

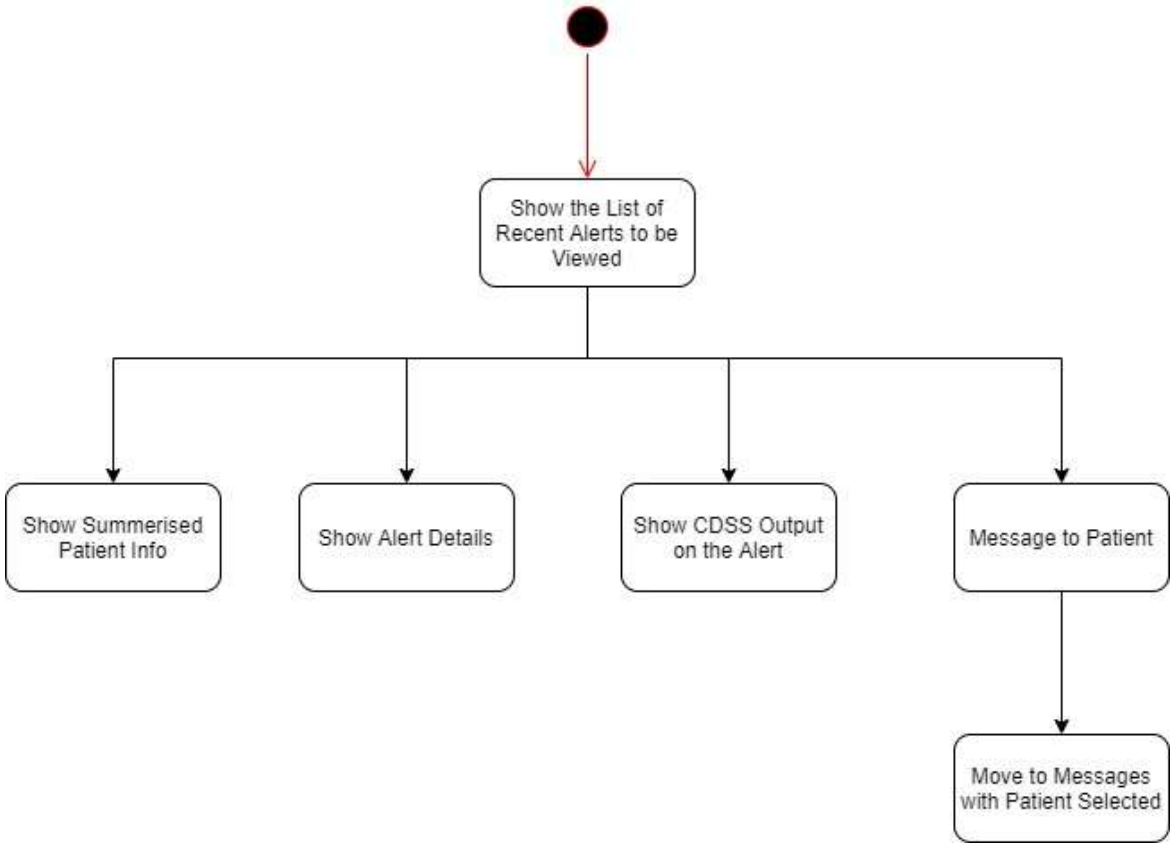
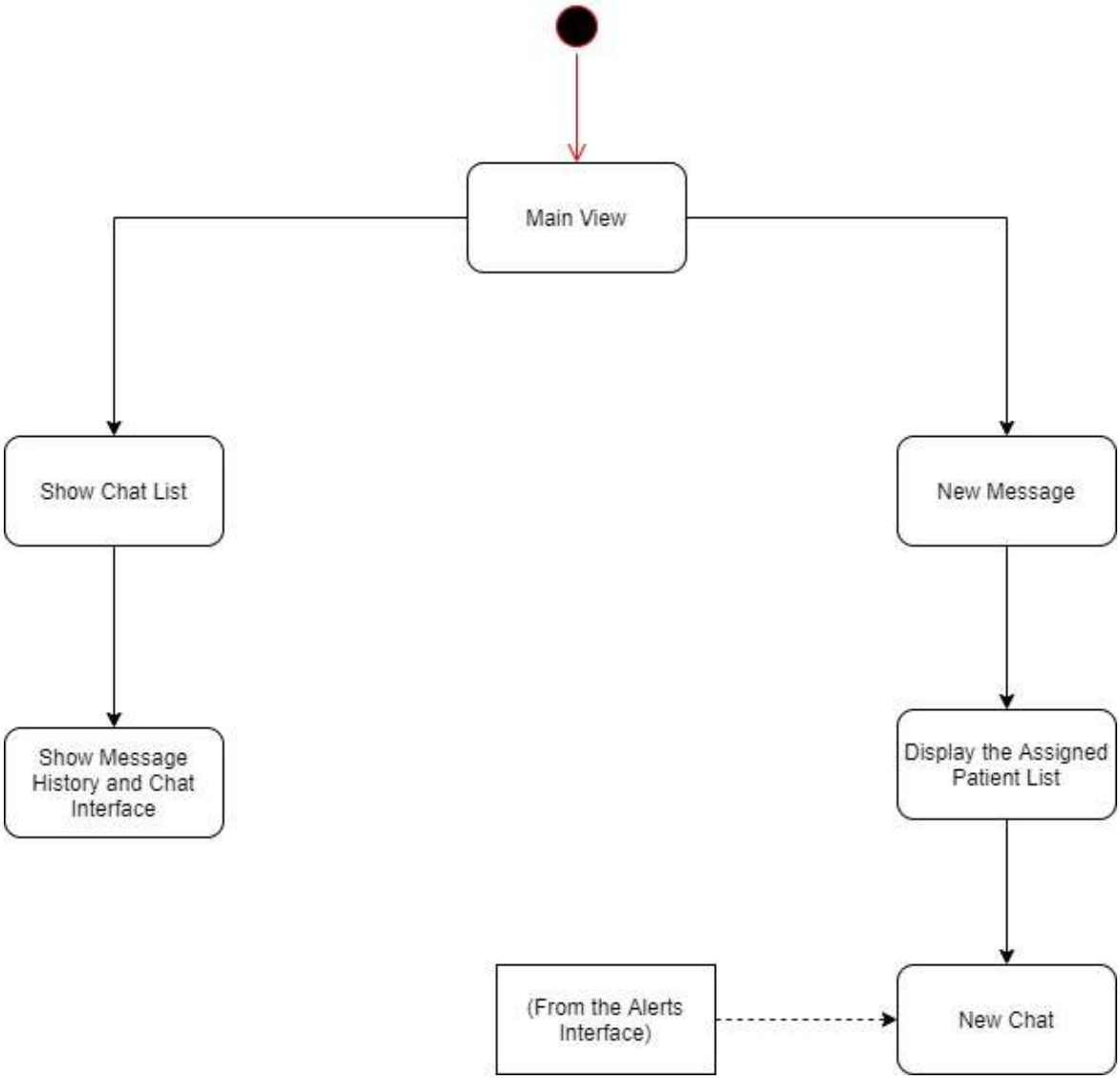


Figure 21: Messages

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406



2.3. User Interface Designs

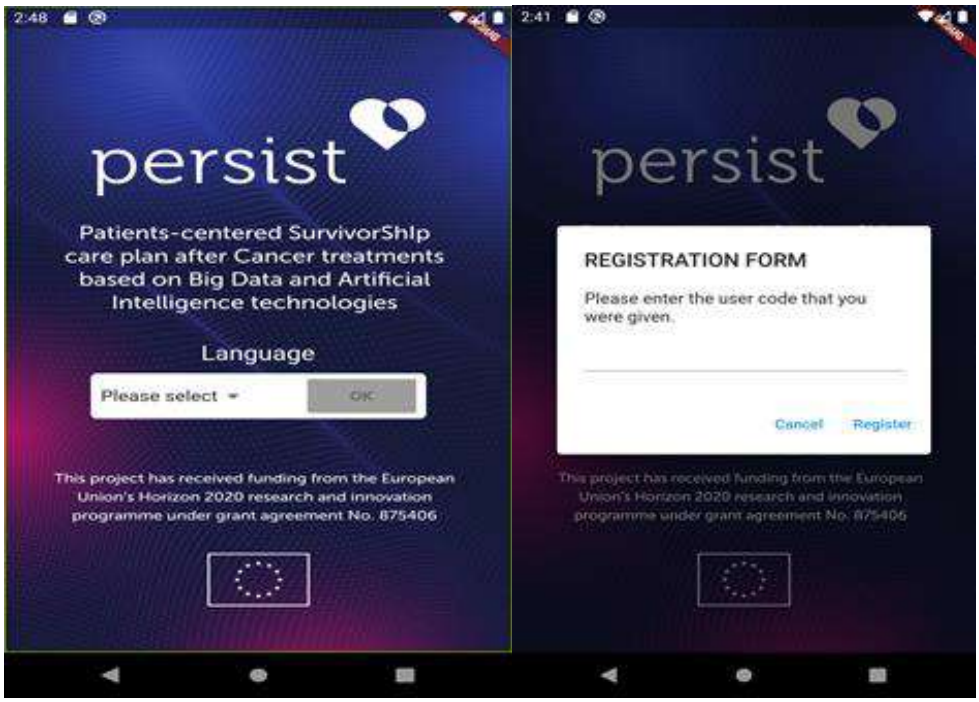
Mobile application and all user interfaces will be developed by Flutter with Dart language. Flutter is an open source project for building beautiful, natively compiled applications for mobile, web and desktop which is developed by Google.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Flutter allows developers to build cross-platform solutions and saves time. Another feature belonging to Flutter is, developers are able to use “Hot Reload”. It helps for quicker and easier experiments. Hot reload makes it possible to quickly monitor application development. The language used behind Flutter is “Dart”. Dart is a programming language also developed by Google and it is open source as well. Dart is used to write Flutter apps which are compiled to native code directly. It can be said that Flutter provides developers a strong platform especially for mobile development.

First time usage screens;

- In the first start of application, language preference will be asked.
- GDPR Form must be filled to enter user code in registration form.
- Once the user enters to given code, OHC connection will be checked, if there is no connection problem, user will be directed to main page



This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 875406

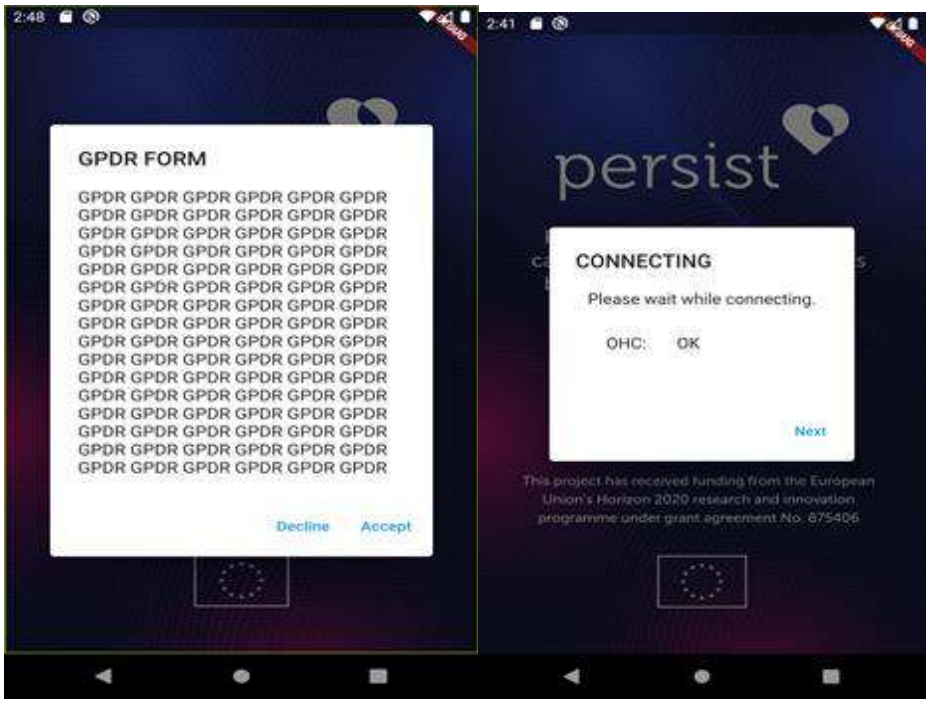


Figure 22: *First time usage screens*

Main Page and Patient List;

- In main page 4 fundamental operations will be shown.
- Those 4 operations are patient list, appointments, alerts and messages.
- Patients and their health information is available. By tapping on the related patient section, the user can reach the detailed information.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

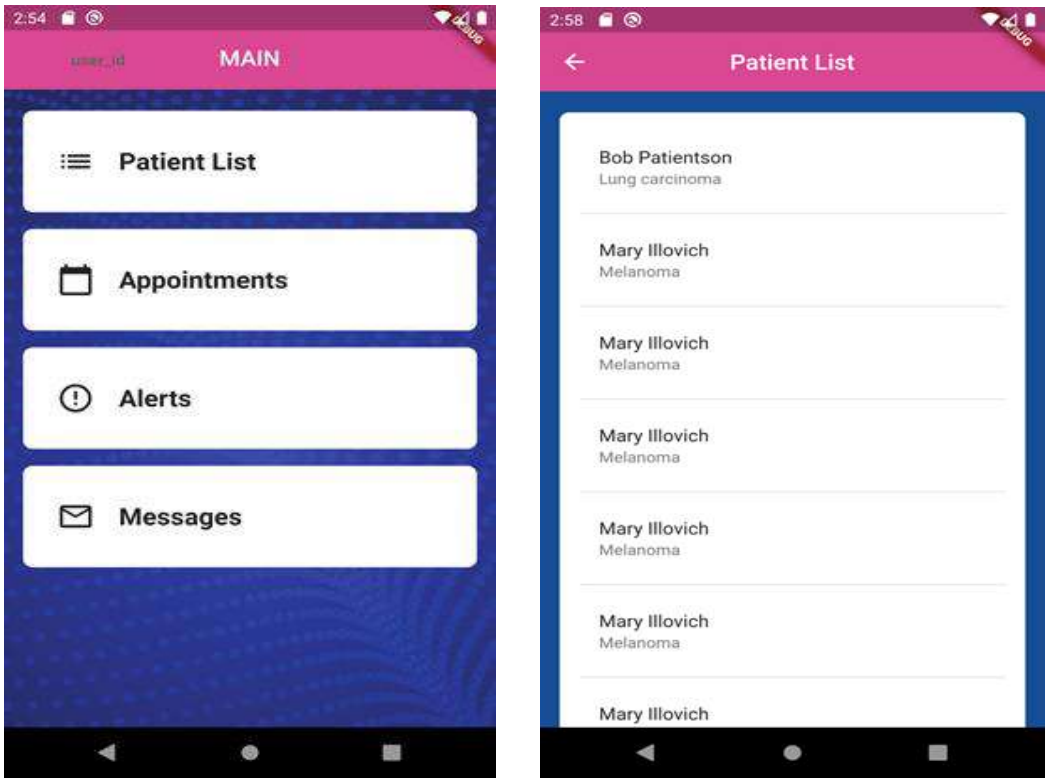


Figure 23: *Main Page and Patient List Screen*

Patient Detail;

- In patient detail, more details about the patient is available.
- Previous alerts can be seen with their timestamps.
- New alerts are marked with different colors.
- Previous CDSS outputs can be seen with their timestamps as well.
- A new CDSS request can be made by tapping on the button and the result is shown.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

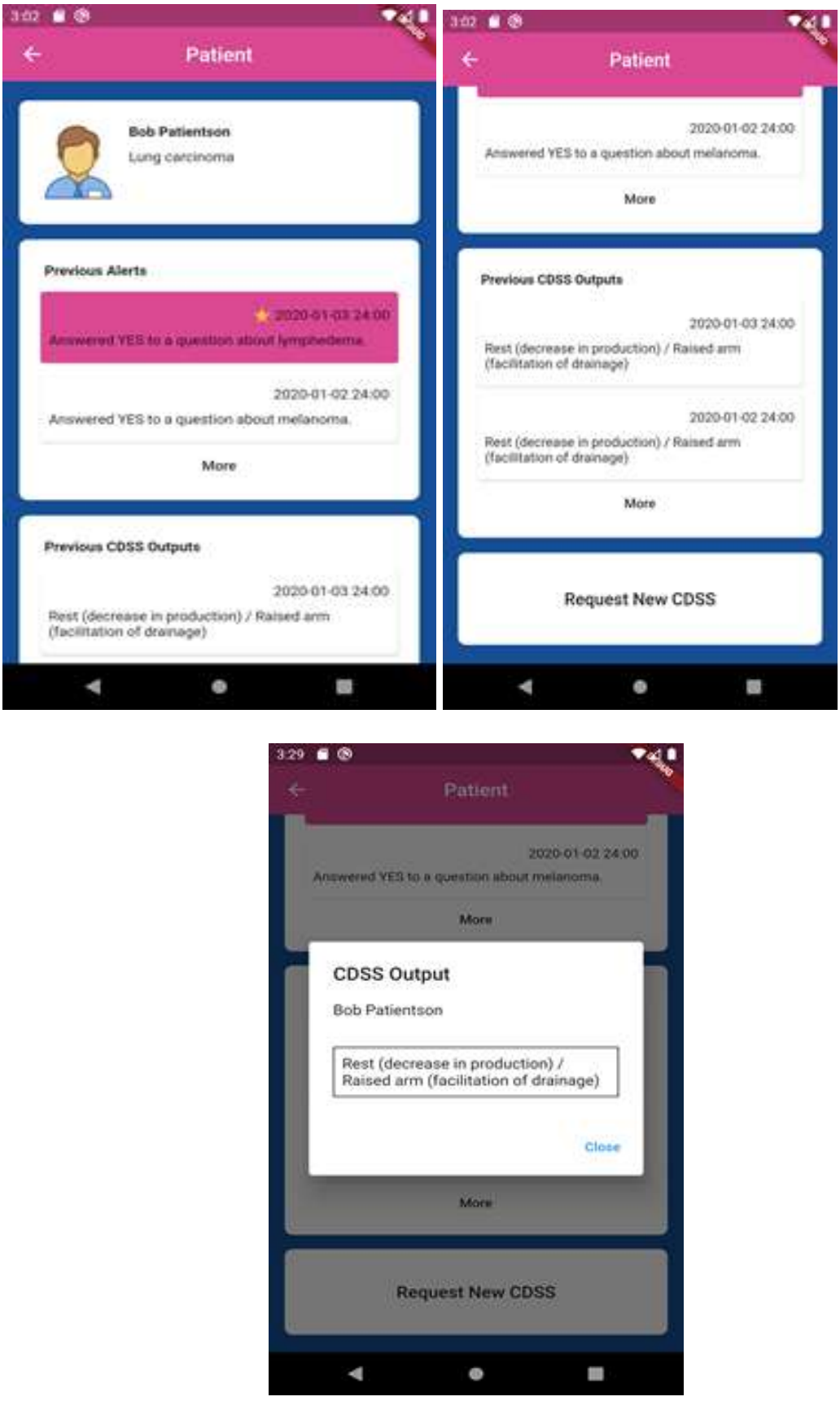


Figure 24: Patient Detail Screens

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Appointments;

- The planned appointments will be shown on the screen based on the calendar.
- By tapping on a button a new appointment can be set.

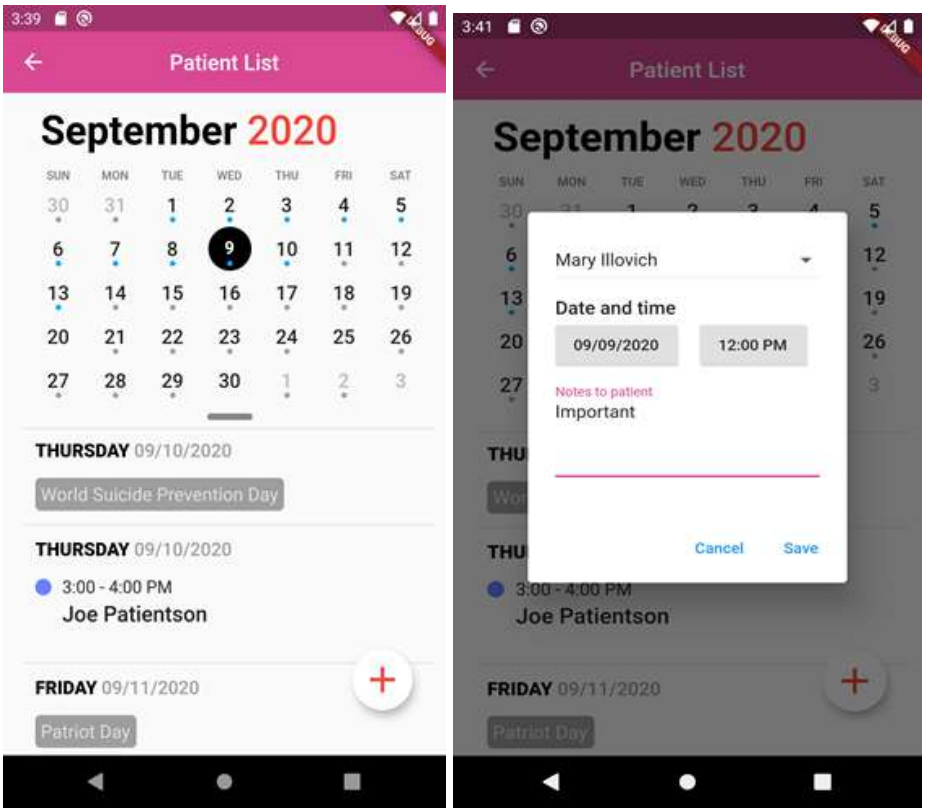


Figure 25: Appointments Screens

Alerts;

- All the alerts so far will be shown in the alerts section, unread alerts marked with different colors. For read alerts, the seen time exists.
- By tapping on a related alert, the user can view alert details.
- In alert details besides alert info, CDSS output and sending message option is available.
- After tapping send message to patient, chat page will be shown.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

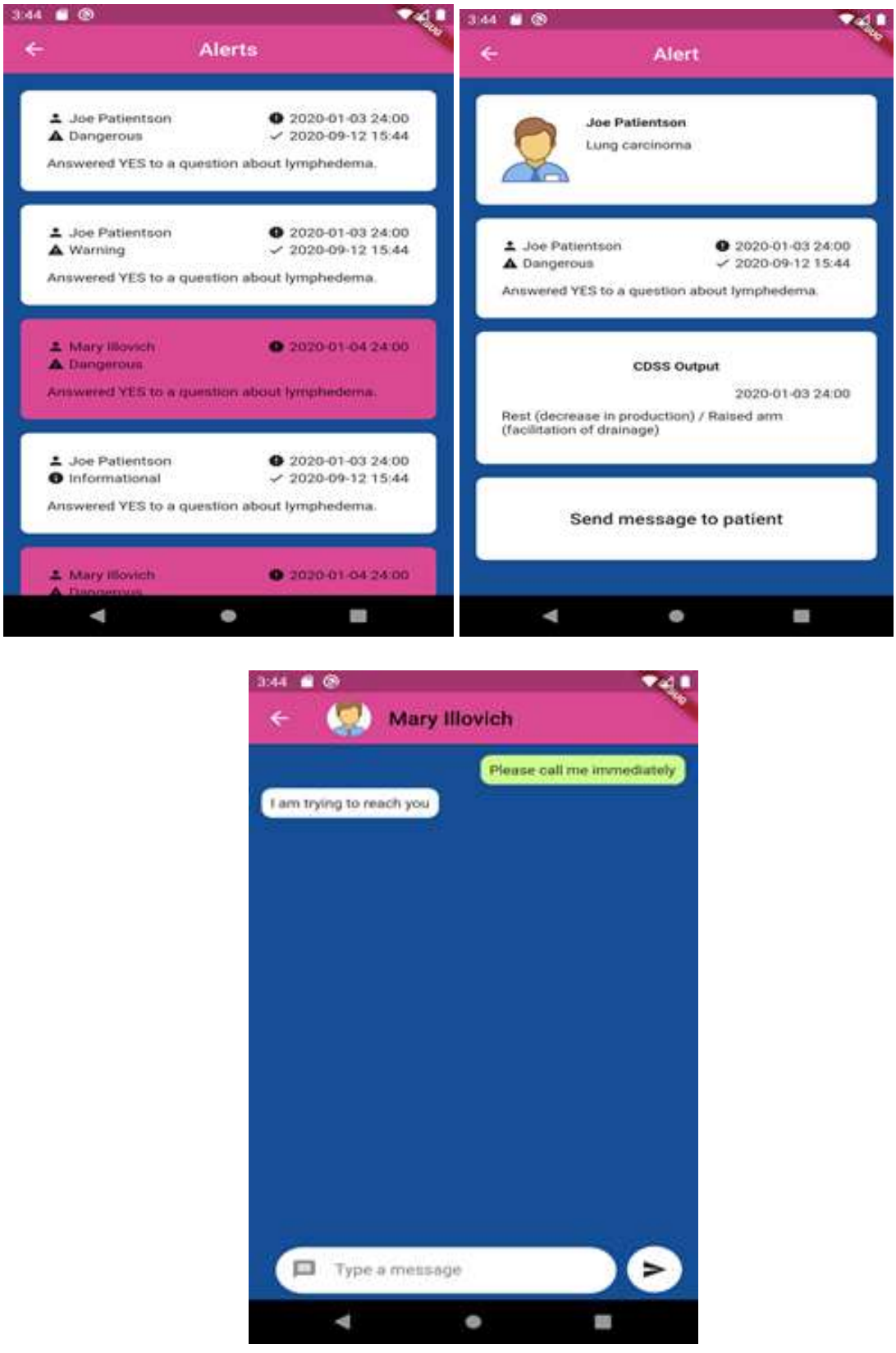


Figure 26: Alerts Screens

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Messages;

- Previous messages with patient will be shown.
- By tapping on related message, the user can reach to the conversation and send a new message as well.

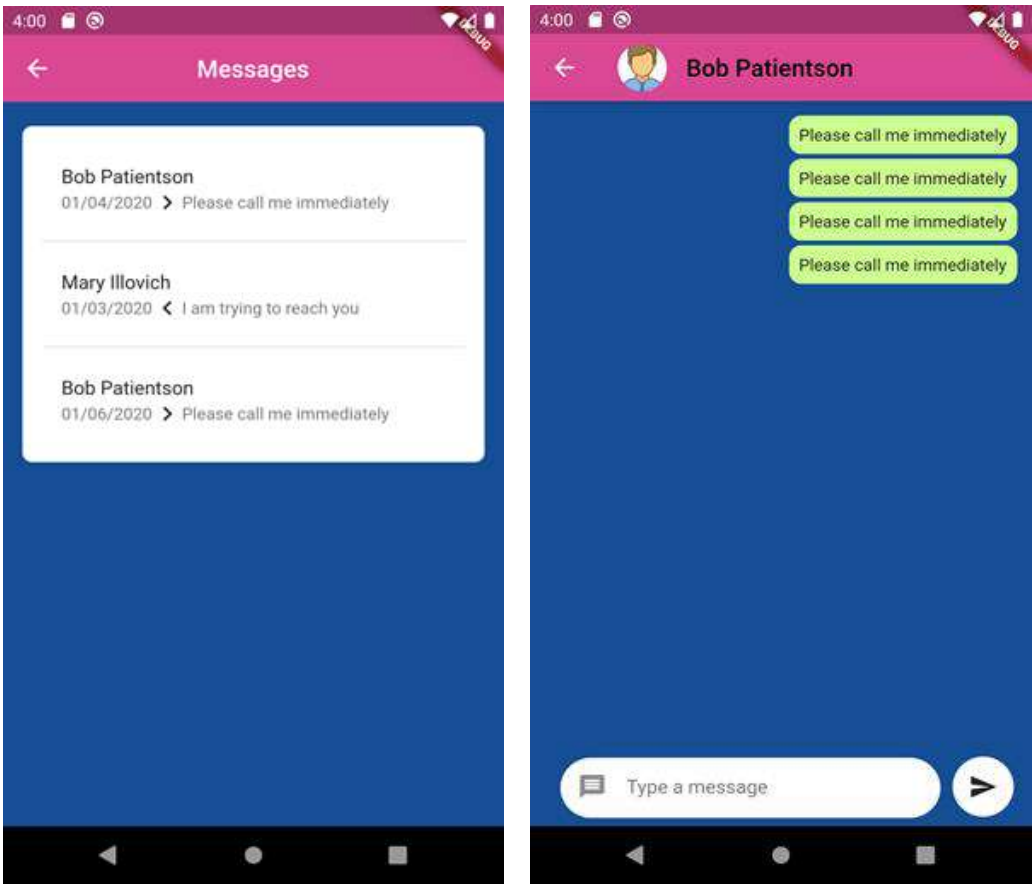


Figure 27: Messages Screens

2.4. Data Security

- Certificate (2 way SSL/TLS)- one certificate per user.
- If sensitive fields : data encryption is needed – encrypt/decrypt with keys

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Data extraction

Data anonymization techniques

In the following section we are going to analyse some anonymization tools which may be convenient for the EHR anonymization process.

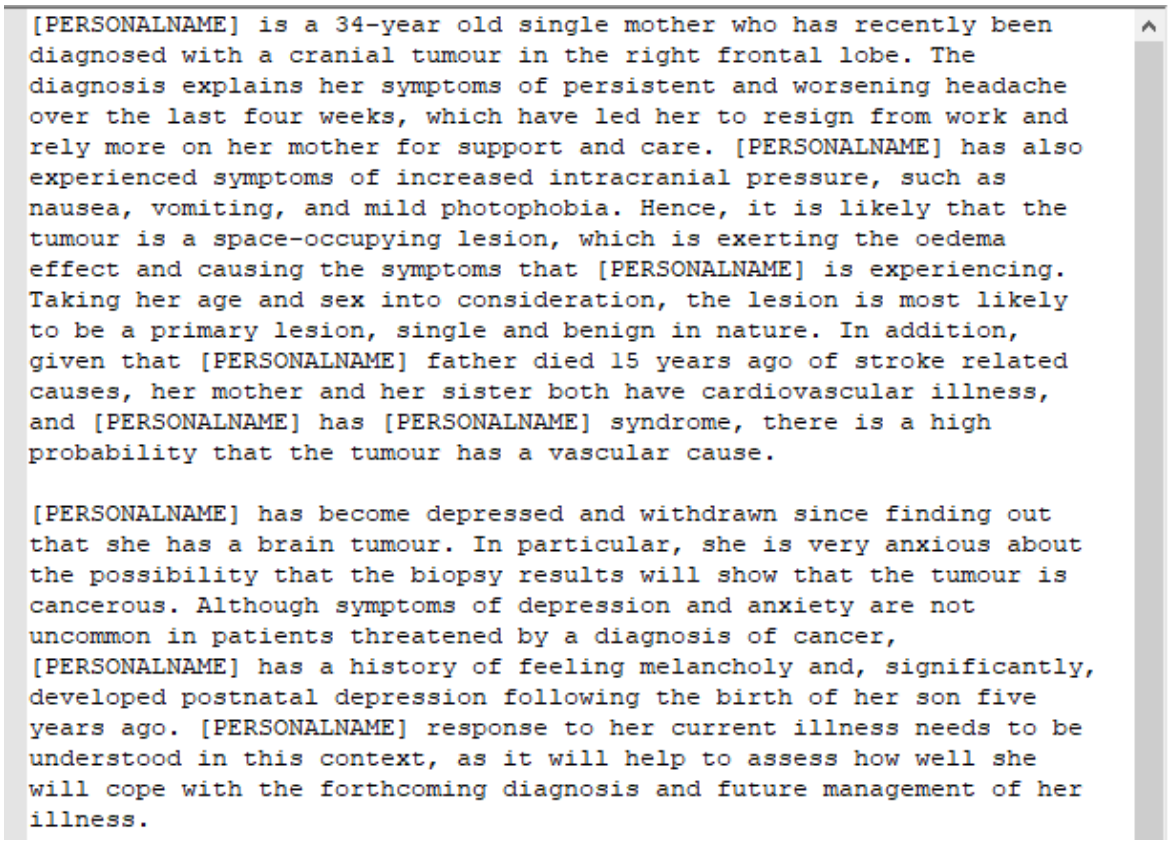
NLM-Scrubber

Description	Tool that focuses mainly in medical text. Substitute patient personal information for pseudonym, for example: Mary ----> [PERSONALNAME]
Operative System	Windows, Linux.
Owner	U.S. National Library of Medicine (NLM)
Price	Free
URL	https://scrubber.nlm.nih.gov/
Analysis	No graphic interface, only command line. The inputs and outputs of the tool are .txt files, and the path to those files should be indicated by means of a configuration file.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Figure 28: NLM Scurbber

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406



[PERSONALNAME] is a 34-year old single mother who has recently been diagnosed with a cranial tumour in the right frontal lobe. The diagnosis explains her symptoms of persistent and worsening headache over the last four weeks, which have led her to resign from work and rely more on her mother for support and care. [PERSONALNAME] has also experienced symptoms of increased intracranial pressure, such as nausea, vomiting, and mild photophobia. Hence, it is likely that the tumour is a space-occupying lesion, which is exerting the oedema effect and causing the symptoms that [PERSONALNAME] is experiencing. Taking her age and sex into consideration, the lesion is most likely to be a primary lesion, single and benign in nature. In addition, given that [PERSONALNAME] father died 15 years ago of stroke related causes, her mother and her sister both have cardiovascular illness, and [PERSONALNAME] has [PERSONALNAME] syndrome, there is a high probability that the tumour has a vascular cause.

[PERSONALNAME] has become depressed and withdrawn since finding out that she has a brain tumour. In particular, she is very anxious about the possibility that the biopsy results will show that the tumour is cancerous. Although symptoms of depression and anxiety are not uncommon in patients threatened by a diagnosis of cancer, [PERSONALNAME] has a history of feeling melancholy and, significantly, developed postnatal depression following the birth of her son five years ago. [PERSONALNAME] response to her current illness needs to be understood in this context, as it will help to assess how well she will cope with the forthcoming diagnosis and future management of her illness.

Figure 29: NLM Scurbber Example

Deid

Description	Tool developed on Perl which allows delete personal medical information. To perform this, it uses several dictionaries (medical terms, city names, hospitals, etc.)
Operative System	Multiplatform
Owner	PhysioNetWorks
Price	Free
URL	https://www.physionet.org/physiotools/deid/#notes-on-the-files-in-the-package
Analysis	No graphic interface, only command line. The tool takes quite a long time to process each file.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

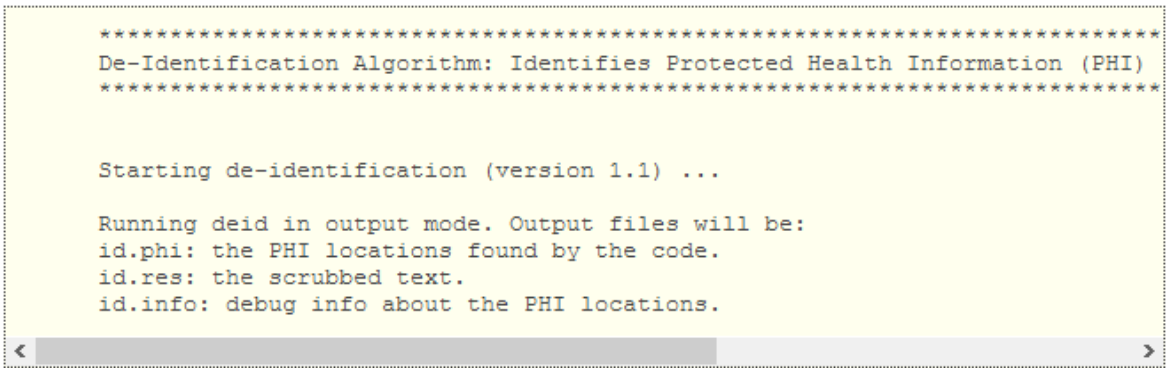


Figure 30: Deid Screen

==========

DISCHARGE SUMMARY

Name: [**Known patient lastname**], [**Known patient firstname**]

[**Unit Number 626**]

Admission Date: [**2016-11-07**]

Discharge Date: [**2016-11-22**]

Date of Birth: [**1972-09-20**]

Sex: F

HISTORY OF PRESENT ILLNESS: Patient is a 44-year-old lady status post living related kidney transplant on [**2016-10-19**], who presented at [**Hospital 36**] for end-stage renal disease secondary to type 1 diabetes mellitus.

She presented to [**Hospital1 **] on [**2016-11-07**] with increased drainage from her surgical wound and JP, increased abdominal pain, and anuria x4 days. The patient reported constipation for a week. She denies flatus. She was complaining of nausea and vomiting. Her abdominal pain had become progressively worse left lower quadrant most notable. There is no radiation to the back or elsewhere. She denied any fevers, chills. She noted decreased p.o. intake recently. Her drainage from her wound incision and JP was notable for yellowish clear urine smelling fluid.

==========

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Figure 31: Deid Example

MITRE Identification Scrubber Toolkit (MIST)

Description	Tool to substitute sensitive personal data in medical texts. MIST performes a annotation task to identity the personal information and, afterwards, a substitution task. The substitutions are based on the use of pseudonym o the change of random values with a similar meaning.
Operative System	Multiplatform
Owner	MITRE Corporation
Price	Free
URL	http://mist-deid.sourceforge.net/
Analysis	Graphic interface. BSD licence.

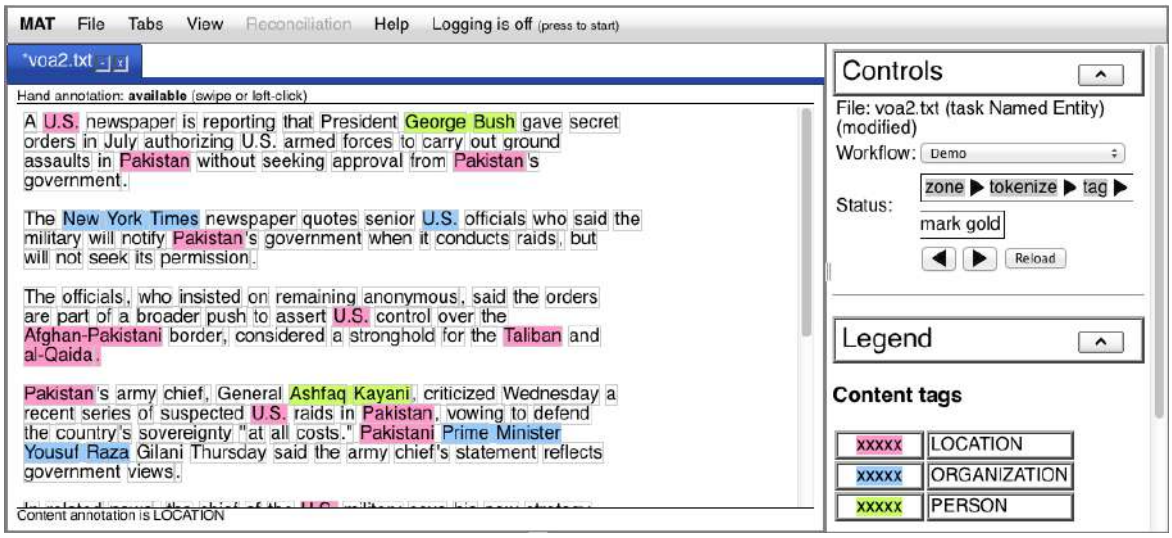


Figure 32: MITRE Identification Scrubber Toolkit Screen

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Risk of Reidentification Techniques

Data anonymization (or de-identification) is a common way to protect patient privacy when disclosing clinical data for secondary purposes, such as research. The technique puts the focus in processing personal data in order to irreversibly prevent identification[17]. This means that the information that can be used to link the data back to an individual is removed or transformed in such a way that the remaining data cannot be used to breach users' privacy. Citing the General Data Protection Regulation (GDPR)[18], "the principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person", so once the data is truly anonymized those principles will no longer apply. The problem is that applying data anonymization correctly is a challenging task and it is necessary to assess its success, typically by measuring an individual's risk of re-identification[19],[20].

Several available metrics applicable for privacy problems can be defined. In the following paragraphs their definitions and properties are exposed. Considering a general overview, the most resaltable properties that privacy metrics must fulfill are the following:

- *Understandable*: users who are not familiar with privacy-related concepts must easily understand the proposed metrics. They have to be able to define their privacy preferences in terms of information disclosure and allowable leakage.

- *Adversary effectiveness*: metrics must indicate how effective an attacker is when estimating the protected sensitive values, in terms of accuracy, correctness and uncertainty, taking into account the difficulty and the needed resources (for computationally bounded adversaries) to achieve the estimation.

- *Achieved privacy and unprotected data*: an adequate privacy metric should indicate not only the achieved privacy level, but also clearly state which portion of data is unprotected (not hidden), if any.

---

17 Article 29 Working Party. Opinion 05/2014 on Anonymisation Techniques (2014). http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216 en.pdf.

18 General Data Protection Regulation (GDPR) (2014). https://eur-lex.europa.eu/eli/reg/2016/679/oj.

19 Narayanan, A., Shmatikov, V.: De-anonymizing social networks. In: 30th IEEE Symposium on Security and Privacy, USA, pp. 173–187. IEEE Computer Society (2009)

20 Wondracek, G., Holz, T., Kirda, E., Kruegel, C.: A practical attack to de-anonymize social network users. In: 31st IEEE Symposium on Security and Privacy, USA, pp. 223–238. IEEE Computer Society (2010)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Information-theoretic metrics: Entropy

One of most well-known metrics applied to the privacy framework is entropy. This metric is typically used to measure the uncertainty of the probability distribution of a random variable and it can be applicable to several fields. When it comes to privacy preserving technologies, the entropy can also offer a measure of how much information a particular adversary gains after performing an attack over a system. Based on this, the entropy can be seen as a measure of the uncertainty that the adversary has regarding a random event *X*, where $p_i = P(X = i)$ represents the probability of the variable *X* of taking the value *i* across the set of possible outcomes *k*. Following the Shannon[21] expression it is defined as:

$$H(X) = -\sum_{i=1}^{k} p_i \log_2(p_i).$$

This metric is typically used to measure privacy on systems which are dealing with long sequences of data.

The role of the entropy as a privacy metric has been deeply analysed in the literature and different theoretical thresholds have been proposed, defining the minimum and maximum level of privacy assumable for each case. For example, if an adversary is able to deduce a confidential value after performing an attack, the sensitive information is exposed and consequently the entropy value will be zero (if $p_i = 1, H(X) = 0$). However, in some cases, it will be necessary to fix another threshold $H_{min} > 0$under which the level of privacy obtained is considered unacceptable. Oppositely, the maximum value of entropy can be achieved when an adversary has no prior knowledge and does not obtain any information additional after performing the attack. In this case, the entropy will reach the maximum value H_{max} along with the adversary's uncertainty.

There exist other situations where handling a normalized value of the entropy can be of interest. For those scenarios, the degree of anonymity can be defined as follows:

$$d = 1 - \frac{H_{max} - H(X)}{H_{max}} = \frac{H(X)}{H_{max}}.$$

The normalized entropy value ranges between 0 and 1, being the former the value corresponding to the minimum entropy (when no privacy is guaranteed) and the latter the maximum entropy (when the adversary learns no information after the attack). Note that the main advantage of this expression regarding the previous one is that it lets us compare the anonymity provided by different systems with completely different behaviours.

Other information-theoretic metrics related to absolute entropy are Hartley entropy (REF), an optimistic metric based on the entropy of the less likely outcome (max-entropy), and min-entropy, based on the entropy of the most likely outcome.

Finally, relative and conditional forms of entropy measures can also be found in the literature. The conditional entropy of *P* conditioned to *Q*, $H(P|Q)$, measures the amount of

21 Shannon, C.E., Weaver, W. (1949) The Mathematical Theory of Communication, Univ of Illinois Press. ISBN 0-252-72548-4

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

uncertainty remaining in *P* when *Q* is known; the mutual information between two variables *P* and *Q* measures the amount of information that one gives about the other $I(P,Q) = H(P) - H(P|Q)$; while the Kullback-Leibler divergence or relative entropy between two random variables *P* and *Q* measures the amount of information that is lost when approximating *P* (true sensitive value) by *Q* (adversarial estimate).

Statistical distortion

The previous metrics are based on information theory and try to measure the uncertainty that an adversary has after an attack is carried out. However, another line of research is related to the idea of using statistical methods to measure the adversary's error after performing an estimation on which can be the value of the targeted data. One of the most well known metric within this type is the Mean Square Error (MSE), an estimator that measures the average level of privacy P_{avg} that remains on a system based on how far the estimation of the adversary over the data $\hat{X}$ is from their actual value X

$$P_{avg} = E(||X - \hat{X}||^2).$$

When an attack occurs and the adversary is able to estimate the actual value of the data $(\hat{X} = X)$, the sensitive information is revealed and no privacy guarantees can be ensure $(P_{avg} = 0)$. Based on this, the higher the difference between the estimation of the adversary and the real data is, the higher the level of privacy will be. Summarizing, the value of the mean square error can be considered proportional to the level of privacy that will remain on the system after an adversary's attack.

Population uniqueness

Another option used to estimate the re-identification risk is based on the concept of uniqueness in the sample and/or in the population. The focus is on individual units that possess rare combinations of selected key variables. It is assumed that units having rare combinations of key variables can be more easily identified and thus have a higher risk of re-identification. Population uniques are records which are unique within the sample (sample uniques) and are also unique within the underlying population from which the data has been sampled. Note that not all sample uniques are also population uniques.

As the procedure requires to know data about the population, when they are not available, this number can be estimated with statistical models. Super-population models estimate the characteristics of the overall population with probability distributions using the sample to estimate the parameters needed for the distribution. Examples of these methods

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

are Hoshino (Pitman)[22,23], Zayatz[24] or the implemented on the μ-Argus software[25]. The problem of these procedures for estimating uniqueness is that they assume that the dataset is a uniform sample of the real population and if this assumption is not satify, results can could be inaccurate.

There exist some other methods that only take available frequencies in the sample into account, such as *k*-anonimity[26,27], *l*-diversity[28] or SUDA2[29]. This specific set of privacy metrics has been developed to measure the achievable anonymity for a particular dataset. The more representative methods of this family is *k*-anonymity, a metric which represents that there are at least *k* - 1 tuples on a dataset that are indistinguishable from a particular record by the adversary's perspective. To achieve this, *k*-anonymity ensures that each quasi-identifier of the dataset is generalized into a group that includes at least *k* occurrences of that value. In general terms, if the dataset meets the requirement that all the quasi-identifiers appear at least *k* times in each group, anonymity can be preserved. If the quasi-identifiers appear less than *k* times each, the privacy properties will be degraded as the number of appearances is also reduced. The worst-case scenario will occur when *k* = 1; in this case, the sensitive values can be easily linkable and, consequently, there are no anonymity guarantees for that particular data subject.

However, independently on the value of *k*, datasets that are protected with k-anonymity can be vulnerable to inference attacks. To tackle this, *l*-diversity was proposed as an extension of *k*-anonymity. This method ensures that for each group of generalized quasi-identifiers (*q**) there should be at least *l* well-represented values of each sensitive attribute (*s*). As for the previous case, anonymity can be preserved if there are at least *l* representative values of each sensitive attribute in each group *q**, it will be reduced if this value is less than *l* and it can be considered as negligible *l* = 1.

Based on the concept of entropy, it is possible to use another metric for *l*-diversity. It is used to guarantee that at least each *q** block of quasi-identifiers will contain at least *l* distinct values of each sensitive attribute. This is knowns as entropy *l*-diversity and it can be defined as:

$$H(q*) = -\sum_{s \in S} p_{(q*,s)} \log_2(p_{(q*,s)})$$

where

[22] Pitman J: Random discrete distribution invariant under size based permutation. Adv Appl Probability 1996, 28:525–539.

[23] Hoshino N: Applying Pitman's sampling formula to microdata disclosure risk assessment. J Official Stat 2001, 17(4):499–520.

[24] Zayatz L: Estimation of the percent of unique population elements on a microdata file using the sample. Washington: US Bureau of the Census; 1991.

[25] Hundepool A, de Wetering AV, Ramaswamy R, Franconi L, Polettini S, Capobianchi A, de Wolf PP, Domingo J, Torra V, Brand R, Giessing S (2008). µ-Argus. User Manual. Version 4.2.

[26] Samarati P, Sweeney L (1998). "Protecting Privacy When Disclosing Information: k- Anonymity and Its Enforcement Through Generalization and Suppression." Technical Re- port SRI-CSL-98-04, SRI International.

[27] Sweeney L (2002). "k-Anonymity: A Model for Protecting Privacy." International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5), 557–570. doi:10.1142/ s0218488502001648.

[28] Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M (2007). "l-Diversity: Privacy Beyond k-Anonymity." ACM Transactions on Knowledge Discovery from Data, 1(1). doi: 10.1145/1217299.1217302.

[29] Manning A, Haglin D, Keane J (2008). "A Recursive Search Algorithm for Statistical Disclosure Assessment." Data Mining and Knowledge Discovery, 16(2), 165–196. doi: 10.1007/s10618-007-0078-6.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

$$p_{(q*,s)} = \frac{n_{q*,s}}{\sum_{s' \in S} \quad n_{(q*,s)}}$$

represents the fraction of tuples from the generalized q^* group that contains the sensitive value s. Regarding this metric, diversity is ensured when $H(q*) \geq log_2(l)$; if this condition is not satisfied, then diversity can not be guaranteed for a particular dataset.

Last example on this section is SUDA (i.e., special uniqueness detection algorithm) which estimates disclosure risks for each observation in the dataset. SUDA2 is a recursive algorithm to find minimal sample uniques. It generates all possible variable subsets of selected categorical key variables and scans them for unique patterns in the subsets of variables. The lower the amount of variables needed to receive uniqueness, the higher the risk of the corresponding observation.

Data models - FHIR standard for storage

In PERSIST there will be two data extractions: the first one for retrospective data; the second one for prospective data which will be stored in the new PERSIST's EHR . As explained in D3.1, an ElasticSearch repository will be used for prospective data storage. Particularly, data out from the OHC Viaduct server will be sent via the FHIR server and persisted into the ElasticSearch repository, in such a way that each record in ElasticSearch is stored as a FHIR resource. For that reason, it seems convenient to **extract** prospective **data in FHIR format** and, therefore, also the retrospective data, if we intend to use **the same extraction procedure**.

In the deliverable D2.5 has been explained the data sources, flows and the data models used to exchange information. For this last purpose, FHIR standard was selected, however this standard is not intended to define a persistence layer, but a data exchange layer, that is, to define data structures used to pass information from one system to another. Even though FHIR standard has not been designed for persistence layer it has been widely used for that purpose and, therefore, there have been many discussions in this regard[30] [31].

One of the main disadvantages of using FHIR standard for storage is the **database normalisation**. Although there is some information highly normalised in FHIR, most of its resources are highly denormalized, so that granular exchanges are fairly stand alone. This is a practical consequence of FHIR being designed for exchange between systems, rather than as a database storage format. As HL7 organization has already explained[32], if an application's information has been well-defined (i.e. EHRs), it is easy to design a data storage schema which completely fits for that purpose with high efficiency compared to FHIR resources storage. On the other hand, if the application's information is not fully specified (i.e. Clinical Data Repositories), storing FHIR resources natively can help implementers to deal with whatever data comes in on an ongoing basis.

30 https://hl7.org/fhir/2018May/storage.html

31 FHIR Connectathon 20. Session: Storage and Analytics. https://wiki.hl7.org/index.php?title=201901_FHIR_Storage_and_Analytics

32 http://hl7.org/fhir/storage.html

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

Given that the data format from hospitals may differ, a FHIR storage seems adequate for retrospective data, since these data will mirror a clinical repository intended to facilate data querying and analyses for reporting and research. What is more, using one-single format will mitigate the translations needed to consume clinical data from different hospitals during the training phase (i.e. from HL7v2, HL7v3, etc.)

Another issue that may discourage from using FHIR for data storage is the management of the **referential integrity.** Many of the defined elements in a resource are **references** to other resources. Resources contain two types of references:

- **Internal** "contained" references - references to other resources packaged inside the source resource
- **External** references - references to resources found elsewhere. For example a Procedure can point to a certain patient, performer, encounter and diagnostic report.

```
{
  "resourceType": "Procedure",
  "id": "f004",
  "status": "completed",
  "code": {
    "coding": [
      {
        "system": "http://snomed.info/sct",
        "code": "48387007",
        "display": "Tracheotomy"
      }
    ]
  },
  "subject": {
    "reference": "Patient/f001",
    "display": "P. van de Heuvel"
  },
  "encounter": {
    "reference": "Encounter/f003"
  },
  "performedPeriod": {
    "start": "2013-03-22T09:30:10+01:00",
    "end": "2013-03-22T10:30:10+01:00"
  },
  "performer": [
    {
      "actor": {
        "reference": "Practitioner/f005",
        "display": "A. Langeveld"
      }   }
  ],
  "reasonCode": [
    {
      "text": "ensure breathing during surgery"
    }
  ]
  "bodySite": [
    {
      "coding": [
        {
          "system": "http://snomed.info/sct",
          "code": "83030008",
```


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

```
          "display": "Retropharyngeal area"
        }
      ]
    }
  ],
  "outcome": {
    "text": "removal of the retropharyngeal abscess"
  },
  "report": [
    {
      "reference": "DiagnosticReport/f001",
    }
  ]
}
```

Figure 33: An example for FHIR standard for storage

References are always defined and represented in one direction - from one resource (source) to another (target). The corresponding reverse relationship from the target to the source exists in a logical sense, but is not explicit. Therefore, for external references, navigating these reverse relationships requires some external infrastructure to track the relationship between resources. Moreover References are either provided as a literal URL, which may either be absolute or relative, or as a logical identifier; they may or might not follow FHIR's well defined RESTful interface pattern; and they may or might not be resolved in the local system. Given these considerations which may be required in various exchange scenarios, most applications that store resources natively find worth discussing the way references work, since, for example, it's not always possible to enforce referential integrity when storing resources directly.

Fortunately this drawback is expected to be of little consequence in the retrospective data use case. Navigate reverse relationships is very unlikely in a training scenario where most of the data will be translated into vectors while navigating the data from the root resource of a Patient. On the other hand a more strict schema will lead to a closed system approach, whereas an **open system** potentially allows to build flexible extensions of the model on which they are based, that is to say, it allows the addition of new attributes and constraints to already existing entities is an easier process. Given that we cannot ensure the data stored by the clinical partners will be homogenous among them, flexibility is an advantageous feature.

Given the above pros and cons for the adoption of FHIR for retrospective data, we ponder the main drawbacks of the database normalization and the integrity references will not pose many inconvenients for the retrospective data training phase. By contrast, the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

adoption of FHIR storage will allow to homogenize data from diverse source as well as to reuse the data extraction process in retrospective and prospective phases.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

- **Conclusion**

The conceptual and technical specifications of CDSS of the PERSIST project is intended to be a comprehensive and living document which outlines the software architectural design of the CDSS, alert mechanism, functionalities of Mobile Application for New EHR section and techniques for Data Extraction to be used for the development of the CDSS. The plan will be updated as the project develops momentum, and as further insights are acquired into the target users and developers.

Technical details given in this deliverable will determine the baseline for development of CDSS, mobile application for clinicians and data extraction approaches from the hospitals.

This deliverable was constructed with the common effort of all related partners of WP5.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

- **Appendix 1: WP5 partners list**

Participant organization name	Short name	Country	Contact person: email
FUNDACIÓN CENTRO TECNOLÓXICO DE TELECOMUNICACIÓNS DE GALICIA	GRAD	Spain	Paula Tosar: ptosar@gradiant.org
SERVIZO GALEGO DE SAUDE	SERGAS	Spain	Tania Vázquez: comunicacion@iisgaliciasur.es
EMODA YAZILIM DANISMANLIK SANAYI VE TICARET LIMITED SIRKETI	EMO	Turkey	Umut Arıöz : umut@emodayazilim.com
UNIVERZITETNI KLINICNI CENTER MARIBOR	UKCM	Slovenia	Matej Horvat: Matej.HORVAT@ukc-mb.si
HAUTE ECOLE SPECIALISEE DE SUISSE OCCIDENTALE	HES-SO	Switzerland	Jean Paul Calbimonte: jean-paul.calbimonte@hevs.ch
LATVIJAS UNIVERSITATE	UL	Latvia	Ilona Aleksandravica: ilona.aleksandravica@lu.lv
CENTRE HOSPITALIER UNIVERSITAIRE DE LIEGE	CHU	Belgium	Marcela Chavez: vchavez@chuliege.be
SYMPTOMA GMBH	SYMP	Austria	Simon Lin: lin@symptoma.com

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

IT CORPORATE SOLUTIONS SPAIN SL	DXC	Spain	Damien Caldy cdamien@dxc.com

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406

- **Appendix 2: Roles and responsibilities of partners at WP5**

Role of each partner at WP5:

EMO will lead WP5 and will be responsible for software requirements specification, software design, the implementation, integration and evaluation of CDSS.
SYM will contribute with its expertise in EHR normalization and processing to define and develop solutions to convert aggregated data records into a conclusive format.
HESSO will lead the patients' cohort and trajectory analysis that will be implemented by HESSO, GRAD, UM and SYM.
All the involved partners supported by the clinical partners will contribute to the development of CDSS and the inference engine as one of the main results of the project.

T5.1 CDSS specification, EHR data extraction and filtering [Leader: EMO; Participants: SER, UKCM, UL, CHU, HESSO, UM, GRAD] [M4-M9]

This task defines the requirements of the CDSS and provides anonymized and standardized data to the following tasks to be developed under this WP5:

- Conceptual design of CDSS. Definition of inputs and outputs. Definition of functional and non-functional requirements. Standardization and interoperability issues.
- Definition of the data sets and of which parameters will be included in the data sets.
- Definition of the anonymization methods to be employed and the methods to be used to estimate the risk of re-identification.
- Definition of the data models to be used.
- Data extraction from hospitals involved in the project and anonymization including data quality assessment.

Outputs: D5.1 Report containing the specific CDSS requirements to comply with the requirements and needs defined in WP2.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875406