Project Title	AI-based long-term health risk evaluation for driving behaviour change strategies in children and youth
Project Acronym	SmartCHANGE
Grant Agreement No.	101080965
Project Start Date:	1 May 2023
Project Duration:	48 months
Project Website:	https://www.smart-change.eu/

D6.1 – Initial technical specification

Work Package	**WP6**
Lead Partner	Engineering Ingegneria Informatica
Contributing Author(s)(Partner)	Valentina Di Giacomo, Elena Mancuso, Federica Saccà (ENG) Mitja Lustrek, Marko Jordan (JSI), Dario Fenoglio, Marc Langheinrich (USI), Marina Loulaki, Eleftheria Kouremenou (UPRC), Sharadhi Suryanarayana, Leo Christino (TUE), Lotte van der Jagt, Pieter Wolfert, Niels Cornelissen (CCARE)
Due Date	2024.10.31
Date	2024.10.29
Version	1.0

Dissemination Level

X	PU – Public, fully open
	SEN – Sensitive, limited under the conditions of the Grant Agreement
	Classified R-UE/EU-R – EU RESTRICTED under the Commission Decision No2015/444

Funded by the European Union

	Classified C-UE/EU-C – EU CONFIDENTIAL under the Commission Decision No2015/444
	Classified S-UE/EU-S – EU SECRET under the Commission Decision No2015/444

Abstract:	Initial specification of the design of the system architecture integrating all the components defined in the project to be used in the pilot study
Keyword List:	Artificial Intelligence & Decision support, chronic diseases, risk assessment, children, youth, long-term risk prediction, chronic NCD, behaviour change, artificial intelligence, federated learning, explainable AI, robustness, bias, participatory design, proof-of-concept study
Licensing information:	
Disclaimer:	This project (GA No. 101080965) has received funding from the Horizon Europe R&I programme. The information provided in this document reflects solely the author's views. The European Community, Agency, and Commission are not liable for any use that may be made of the information contained herein. The content is provided without any guarantee or warranty of fitness for a particular purpose. Users utilise the information at their own risk and liability. In the case of proprietary information of the SmartCHANGE Consortium, it shall not be used, duplicated, or communicated to third parties without prior consent.

Versioning history

Version	*Date*	*Author(s)*	*Notes &/or Reason*
0.1	08/01/2024	Valentina Di Giacomo (ENG)	TOC and V0.1
0.2	19/07/2024	Valentina Di Giacomo (ENG)	Improved TOC, expected contributions
0.3	14/08/2024	Federica Saccà (ENG)	Added contents in 'Executive Summary' and 'Data Model' sections.
0.4	30/08/2024	Federica Saccà (ENG)	Added contents in 'Risk Assessment' section.
0.5	25/09/2024	Valentina Di Giacomo, Federica Saccà (ENG)	Integration of contributions by partners, and general architecture
0.6	15/10/2024	Mitja Lustrek, Marko Jordan (JSI), Dario Fenoglio, Marc Langheinrich (USI), Marina Loulaki, Eleftheria Kouremenou (UPRC), Sharadhi Suryanarayana, Leo Christino (TUE), Lotte van der Jagt, Pieter Wolfert, Niels Cornelissen (CCARE)	FHIR Data model (UPRC, ENG) Security (USI, ENG) Requirements and deployment (CCARE) Components descriptions and interfaces (ENG, CCARE, USI, JSI, ENG)
0.7	18/10/2024	Valentina Di Giacomo, Elena Mancuso, Federica Saccà (ENG)	Detailed architecture, diagrams, consolidation of contributions, conclusions and stylistic fixes
1.0	29/10/2024	Valentina Di Giacomo, Elena Mancuso, Federica Saccà (ENG)	Final version addressing internal review comments

Quality Control (peer & quality reviewing)

Version	*Date*	*Name (Organisation)*	*Role & Scope*
0.8	21/10/2024	Tim d'Hondt, Mykola Pechenizkiy (TUe) Gabriele Dominici, Martin Gjoreski (USI)	Internal reviewers

Funded by the European Union

Table of contents

Funded by the European Union

Funded by the European Union

List of figures

List of tables

Funded by the European Union

Executive summary

The overall goal of the SmartCHANGE project is to develop trustworthy AI-based tools to support health professionals (HCPs) and citizens to improve citizens' health. Following a user-centred approach and by developing i) a mobile app designed for families (and their children) and adolescents, and ii) a web application for HCPs, it aims to minimize the long-term risk of NCDs, mainly cardiovascular and metabolic diseases, and promote the delivery of optimised risk lowering strategies.

Drawing upon these foundational concepts, this document outlines the technical specifications for the entire SmartCHANGE solution, designing the overall system architecture based on the microservices paradigm. The architecture incorporates both functional and non-functional features, including privacy and security measures, and ensuring alignment with the user requirements as reported in D3.6[1]. This also includes defining the specific services that the solution offers, establishing the protocols that governs how different components of the system communicate with each other, and mapping out the pathways through which data travel within the system. In detail:

- **Defining the specific services**: involves identifying and describing the individual functions and features that the solution provides. As example, these services could range from user authentication and data storage to personalized health recommendations and progress tracking.
- **Establishing the communication protocols**: setting the rules and standards that determine how various parts of the system exchange information. These protocols ensure that different software components can interact smoothly and effectively, facilitating seamless integration and operation.
- **Mapping out the data flows**: this involves determining how data is collected, processed, stored, and retrieved, ensuring that information flows efficiently and securely between different modules and services.

In essence, this process ensures that the SmartCHANGE solution operates as a cohesive, integrated system, with clearly defined functionalities, smooth interactions among its components, and efficient data management, according to the needs of federated learning

[1] SmartCHANGE Consortium. Deliverable D3.6 User Requirements for the mobile and web application

(WP4). Ultimately, a privacy-by-design approach has been adopted to minimize GDPR-related risks (as detailed in D2.3[2] and D2.4[3]).

[2] SmartCHANGE Consortium. Deliverable D2.3 SELP Compliance Framework

[3] SmartCHANGE Consortium. Deliverable D2.4 SELP Impact Assessment

Funded by the European Union

List of abbreviations

Abbreviation	*Definition*
AI	Artificial Intelligence
CDA	Clinical Document Architecture
CPU	Central Processing Unit
DB	Database
DBMS	Database Management System
EMR	Electronic Medical Record
ENISA	European Union Agency for Cybersecurity
FHIR	Fast Healthcare Interoperability Resources
FL	Federated Learning
GDPR	General Data Protection Regulation
GP	General Practitioner
HCP	Health Care Professional
HTTP	HyperText Transfer Protocol
HTTPS	HyperText Transfer Protocol Secure
IP	Internet Protocol
IPv	Internet Protocol version
IT	Information Technology
JDBC	Java Database Connectivity
JWT	JSON Web Token
MDCG	Medical Device Coordination Group
MIT	Massachusetts Institute of Technology

Funded by the European Union

ML	Machine Learning
NCD	Non-communicable diseases
SDK	Software Development Kit
SSD	Solid State Drive
SSH	Secure SHell
SSL	Secure Sockets Layer
SSO	Single Sign-On
TLS	Transport Layer Security
UI	User Interface
VM	Virtual Machine
VPN	Virtual Private Network
XCDS	eXplainable Decision Support

Funded by
the European Union

1 Introduction

The SmartCHANGE project is a pioneering initiative that focuses on harnessing the power of artificial intelligence to empower both healthcare professionals (HCPs) and citizens to proactively manage the health of the latter, and reduce their long-term risk of non-communicable diseases (NCDs).

Central to SmartCHANGE is a user-centric approach, reflected in the development of two AI-based tools designed to deliver optimized risk-lowering strategies, promoting healthier lifestyles and preventing the onset of NCDs, primarily cardiovascular and metabolic diseases. The outputs of the project will be delivered via the following bi-directional tools:

- **A Primary Tool for Health Professionals**: This tool is delivered as a web application (HCP Web App), providing healthcare professionals with insights and decision support. It integrates with the Trustworthy AI Framework (local), consisting of components like the Explainer, Risk Predictor, and Federated Learning (FL) Client, to offer tailored risk assessments and visual explanations of risk predictions.
- **A Citizen-Focused Tool**: SmartCHANGE also delivers its services directly to children, youth, and their families through a mobile application (Youth/Family Mobile App). This app interfaces with wearable devices to collect relevant health data and sends this information to the Execution Engine for processing.

This document serves as a blueprint for the entire SmartCHANGE solution, outlining the system architecture based on the microservices paradigm. It incorporates functional and non-functional requirements, ensuring alignment with the user needs as outlined in D3.6[4]. The document delves into the specific services offered by the solution, establishes the communication protocols governing interactions between components, and maps out the data flow within the system.

By defining the specific services, the document provides a clear understanding of the individual functions and features that the SmartCHANGE solution offers. These services may encompass tasks such as user authentication, data storage, personalized health recommendations, and progress tracking.

[4] SmartCHANGE Consortium. Deliverable D3.6 User Requirements for the mobile and web application

Funded by the European Union

Establishing communication protocols ensures that different parts of the system can interact seamlessly, facilitating efficient information exchange and integration. These protocols define the rules and standards that govern how various components communicate with each other.

Mapping out the data flows is crucial for ensuring that data is collected, processed, stored, and retrieved efficiently and securely. This process outlines the pathways through which data travels within the system, ensuring that information flows smoothly between different modules and services.

Ultimately, the document ensures that the SmartCHANGE solution operates as a cohesive, integrated system, with clearly defined functionalities, smooth interactions among its components, and efficient data management. By adopting a privacy-by-design approach[5], the project minimizes GDPR-related risks, ensuring a secure and trustworthy healthcare solution.

The document is structured as follows:

Chapter 2 (System Overview): Provides a high-level overview of the system, including its context, key components, and their roles within the SmartCHANGE ecosystem.

Chapter 3 (Requirements): Delves into the detailed requirements for the SmartCHANGE solution, encompassing user needs, hardware, software, and infrastructure specifications.

Chapter 4 (Security and Privacy): Addresses the critical aspects of security and privacy within the SmartCHANGE solution, outlining risk assessment and mitigation strategies in a federated learning architecture.

Chapter 5 (Architecture Design): Presents the architectural blueprint of the system, including high-level and detailed design, component selection, data model, and information flow.

Chapter 6 (Data Model and Flows): Delves into the different types of data exchanged among services in SmartCHANGE. It also describes the role of each service in ensuring secure, private, and efficient data exchange.

[5] SmartCHANGE Consortium. Deliverable D2.4 SELP impact assessment

Chapter 7 (Deployment and Testing): Specifies the deployment and testing procedures for the SmartCHANGE solution, ensuring its effective implementation and validation.

Chapter 8 (Conclusions and Next Steps): Summarizes the key findings and achievements to date, outlining future directions and potential areas for further development.

Funded by
the European Union

2 System Overview

2.1 System Context

The SmartCHANGE system is designed as a trustworthy AI-based decision-support solution that, as described in Figure 1, operates in a complex environment involving healthcare providers, youths as well as children and their families. It involves the use of different data sources, facilitating bidirectional communication between health professionals and citizens. The core internal components include the eXplainable Clinical Decision Support (XCDS) Engine, a Clinical Repository based on FHIR, a Trustworthy AI Frameworks that includes a Risk Prediction service, an Explainer service and services supporting Federated Learning (FL Client, and FL Server). The Security Framework ensures data integrity, privacy, and compliance with regulations. This includes user authentication via the Identity Provider, monitoring of system activities, and comprehensive logging and auditing to safeguard data transactions. All these elements work together to collect, process, analyse, and secure data.

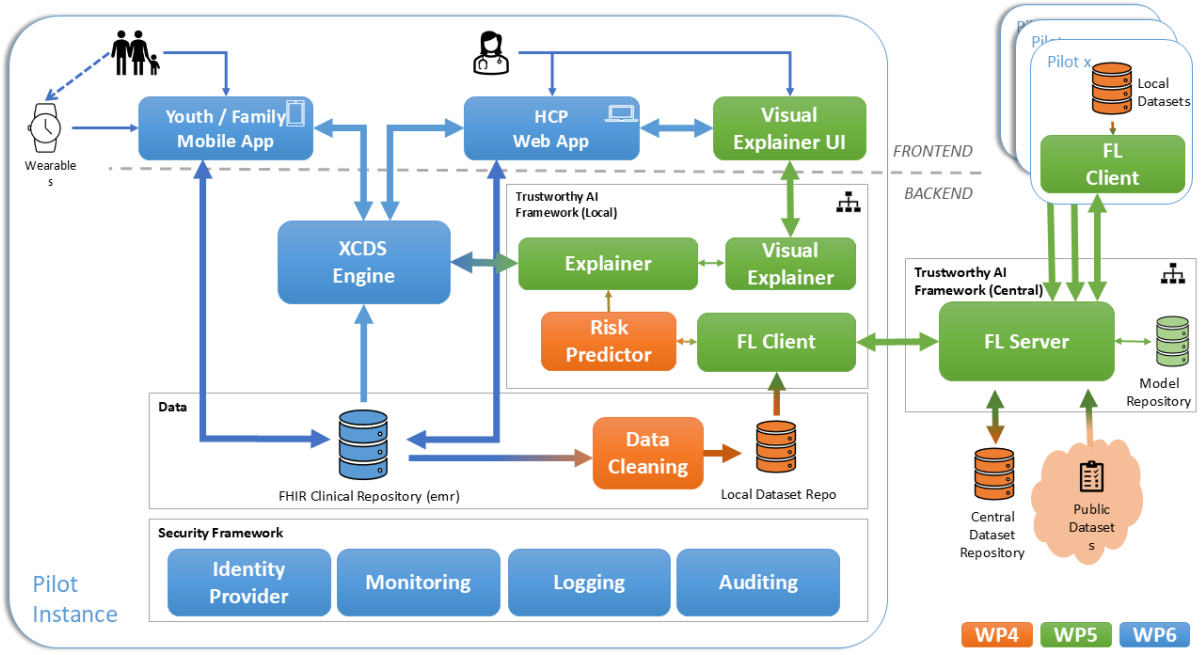

Figure 1 – Logical Architecture Diagram of the SmartCHANGE solution

Externally, the solution interacts with human actors and receives inputs from wearable devices and public health datasets. The respective roles of each of these entities within the system are as follows:

- **Health Professionals (HCP):** Use the HCP Web App to access risk assessments, receive recommendations, and input user data. They interact with the system to monitor subjects, understand AI-generated insights, and make informed decisions based on personalized recommendations.
- **Children, Youth, and Families**: Engage with the system via the Youth/Family Mobile App. This tool allows users to provide data (e.g., health metrics from wearables), receive feedback on risk levels, and obtain personalized health advice.
- **Wearable Devices**: Serve as external data sources, capturing health-related information such as activity levels, heart rate, and other vital metrics. Data from these devices are processed by the system and contribute to compute individual risk profiles.
- **Public Datasets**: Used by the central Trustworthy AI Framework for training and enhancing the AI models. These datasets provide additional context and benchmarks for improving predictive accuracy.

Data flows in the system according to the main scenarios are listed below:

1. **Data Collection**: Data is collected from the Youth/Family Mobile App and wearable devices. This data includes personal health information, which is sent to the Clinical Repository, to make it available to the authorized services in the platform.
2. **Risk Assessment**: The HCP requests a risk prediction through the Webapp. Processed data flows to the Local Trustworthy AI Framework, using the XCDS Engine as a broker. The Explainer service orchestrates the collection of the data from the Risk Predictor and provides explanations to support the prediction.
3. **Goal Setting**: The HCP consults the dashboard of the child or adolescents to assess their status. Using this information, they set goals to improve behaviour and potentially reduce the risk of developing diseases in the future.
4. **Federated Learning**: Local data contributes to federated learning processes, allowing the FL Client to train models without sharing raw data externally. The FL Server aggregates these updates, enhancing the system's predictive capabilities.

Each scenario will be described in more detail in section 6.3 . The solution is based on the microservice architectural paradigm. In this context, a microservice is a distinct, independently deployable unit of software that focuses on performing a single, specific business function, typically communicating with other microservices through lightweight protocols like HTTP or messaging queues [62]. The system is designed to operate in a cloud-based and distributed environment, allowing scalability, flexibility, and the ability to integrate with multiple pilots and datasets. It supports real-time data processing and secure model training across distributed locations.

Funded by the European Union

2.2 Key Components and Their Roles

This section gives a high-level description of the components of the SmartCHANGE solution and their role, whereas a more in-depth description will be provided in section 5.2

- **XCDS (eXplainable Clinical Decision Support) Engine**: Central to the system, the Execution Engine processes data from the Trustworthy AI Framework, and makes them available to the HCP Web App and the Youth/Family Mobile App. It has the responsibility to retrieve the information needed to the AI services from the FHIR Repository, where health data is securely stored and managed.
- **A FHIR-based Clinical Repository**: Central to the system, the Execution Engine processes data from both the HCP Web App and the Youth/Family Mobile App. It communicates with the FHIR Repository, where health data is securely stored and managed.
- **Trustworthy AI Framework**: The framework includes all the services that enable the training of the AI models, following a Federated Learning architecture, and that provide the AI predictions. The central framework coordinates federated learning across multiple pilots, aggregating knowledge from various FL Clients. It communicates with a centralized FL Server that hosts a model repository and integrates public datasets for model improvement. The local framework includes several sub-components:
    - **Explainer**: Provides interpretable insights into AI-generated risk predictions.
    - **Risk Predictor**: Utilizes health data to predict individual risk levels.
    - **FL Client**: A federated learning component that enables collaborative model training using local data without sharing raw data.
- **Security Framework**: To ensure the system's trustworthiness, security measures are embedded, including identity management, monitoring, logging, and auditing.
- **Data Cleaning**: Incoming data is cleaned and pre-processed before being used by the Risk Predictor, ensuring high-quality input for reliable risk assessment.

This microservice-based architecture facilitates secure, bi-directional communication among the components, ensuring that both HCPs and citizens receive accurate, personalized, and actionable risk assessments. The overall system supports a scalable, federated learning approach, enabling the integration of diverse data sources while maintaining data privacy.

Funded by the European Union

3 Requirements

For the components of the SmartCHANGE system to be built and connected correctly, several sets of system requirements need to be considered, as drafted in earlier deliverables. A summary of user requirements, hardware, software as well as infrastructure requirements is provided in this chapter.

3.1 User Requirements

Deliverable D3.6[6] outlines the user requirements for both the web application and mobile application, being designed and developed for HCP and families and adolescents respectively. For the complete list of these requirements refer to D3.6.

The SmartCHANGE system's usage scenario, as described in D3.6, is based on the inputs gathered from the co-design process. The usage scenario includes two main usage paths, one initiated via school communities and the other via HCPs' recommendation (i.e. paediatricians). Within the school-initiated scenario, families and adolescents can still decide (while being guided with this) whether to involve an HCP in the process. Both tools, mobile and web app, have been designed in order to support both scenario paths. Details on how the user requirements have been translated in a consistent usage experience have been described in the prototypes and then validated by the co-creation process. Detailed description of the process along with the prototypes have been reported in the D3.7[7] Deliverable document.

For the mobile application key requirements are:

- **Risk Communication:** The app should not directly communicate risk predictions to the user.
- **Behaviour Change:** The app should facilitate and focus on behaviour change, while being experienced as fun and using a positive tone of voice.
- **Multiple Health Areas:** Users should be able to work on various health areas.

[6] SmartCHANGE Consortium. Deliverable D3.6 User Requirements for the mobile and web application

[7] SmartCHANGE Consortium. Deliverable D3.7 UX design and prototypes of the mobile and web application

Funded by the European Union

- **Use small and achievable goals[8]:** Provide the user with multiple small and achievable goals to increase the feeling of capability and motivate towards healthy behaviour change.
- **HCP Guidance:** The system or HCP should provide guidance on which health area is most important for the user.
- For adolescents give the user a **feeling of autonomy:** Users should feel in control on their behaviour change process, for example by being able to decide which goals to focus on.
- For families, **focus on the entire family:** The mobile application for the family should address all family members within the household, rather than focussing on one child.

For the web application key requirements are:

- **Data Visualization:** Subject data (on activity and physiological parameters), risk predictions, and counterfactuals should be visualized in a clear and accessible manner for HCPs.
- **Data Entry:** The app design should allow HCPs to easily input structured clinical data into the subject's record.
- **Risk Explanation:** The HCP should be provided with an "explanation" of the AI model or of the individual prediction.
- **Straightforward and relevant information:** Only relevant information and results should be presented within the web-app in a concise and straightforward manner.

D3.6[9] provides a list of requirements for the entire system to ensure a secure application, following current security and privacy guidelines. Such requirements have been analysed in better details and are presented in chapter 4, where the resulting security measures are listed.

[8] while D3.6 was written, the co-design process was unfinished. Therefore, this requirement was still described as an idea direction by that time. Later co-design sessions showed it should be listed as a requirement.

[9] SmartCHANGE Consortium. Deliverable D3.6 User Requirements for the mobile and web application

3.2 Hardware, Software and Infrastructure Requirements

In deliverable D6.6[10], hardware, software and infrastructure requirements are described for the deployment of the SmartCHANGE containerized microservices at each pilot site (for the complete list of requirements, please consult D6.6 document). To support the 2-phase development cycle, the deployment will also undergo a 2 phases process. Hence, D6.6 is an initial plan for deployment, integration and testing, and later a final plan will be presented in D6.7[11].

Briefly, D6.6 outlines a blueprint for a virtual machine (VM) configuration to ensure consistent deployment across all pilot sites. This standardization facilitates easier management and troubleshooting throughout the project.

Regarding the VM specifications, the most important requirements are:

- Operating System: Ubuntu 24.04 LTS
- Minimum of 2 vCPUs and 16GB RAM
- Storage: 100 GB SSD for system disk and 500 GB SSD for data storage
- Static IPv4 address, reachable via a (sub)domain

For security and connectivity, the key requirements are:

- Open ports 80, 443 (HTTPS only), and 2022 (SSH)
- Key-based authentication for SSH and disabled root login over SSH
- Encryption of data storage partition by default
- Usage of SSL certificates for micro service communication

Regarding Docker and container configuration, the essential requirements are:

- Installation of the latest stable version of Docker Engine, configured to start on boot
- Pull necessary Docker images from the central container registry
- Configure Docker Compose for multi-container applications

[10] SmartCHANGE Consortium. Deliverable D6.6 Initial Integration and Test Plan

[11] SmartCHANGE Consortium. Deliverable D6.7 Final Integration and Test Plan

- Ensure container configurations adhere to project standards

The blueprint emphasizes the importance of communicating the public IP address and relevant access details securely within the project network to facilitate smooth deployment and operation of the SmartCHANGE system.

Funded by
the European Union

4 Security and Privacy Considerations

This section describes how the SmartCHANGE solution will protect its data, users and resources, following the constraints posed by the GDPR regulations and a privacy by design approach. A specific focus will be placed on the Security and Privacy implications of adopting a Federated Learning approach. This chapter then reports the results of the risk assessment performed according to the ENISA guidelines, and the corresponding security measures that will be adopted in designing, developing and maintaining the SmartCHANGE solution, applying the recommendation provided in D2.4[12], section 2.4.1.

4.1 Security in a Federated Learning Architecture

In the context of Federated Learning (FL), **security** refers to the protection of the FL process from malicious attacks, erroneous inputs, and other threats that could compromise the integrity, and availability of the data, models, and communication channels involved. Conversely, **privacy** is concerned with protecting the confidentiality of the participants' data throughout the FL process, ensuring that sensitive information remains secure and is not exposed to other participants or external entities.

4.1.1 Security Risks and Countermeasures in SmartCHANGE

The SmartCHANGE project operates within a cross-silo FL framework, where trusted institutions collaborate to train a machine learning model without directly sharing their sensitive data. Given that all participating institutions are validated, bounded by legal contracts (e.g., Data Transfer Agreements and the Consortium Contract) and motivated to contribute effectively to the training process, and no external entities are permitted to participate in the process (i.e., the FL framework is accessible only to the consortium members), the inherent security risks that could degrade the final model's performance are significantly reduced (Figure 2). Moreover, robust server-side aggregation methods can be implemented to detect and exclude suspicious updates from unverified sources, further enhancing security [20][21][22][23]. These countermeasures are designed to operate entirely

[12] SmartCHANGE Consortium. Deliverable D2.4 SELP Impact Assessment

on the server side, without necessitating modifications to the underlying FL system architecture.

4.1.2 Privacy Risks and Countermeasures in SmartCHANGE

While FL inherently enhances privacy by not directly sharing sensitive data across institutions, recent studies have revealed that trained machine learning models can inadvertently leak sensitive information about the data on which they were trained, particularly under conditions such as model overfitting, reduced batch sizes, and other specific scenarios [24][25][26][27]. The potential attack vectors in a FL system include the communication channels, client entities (i.e., institutions), and the server.

To mitigate these risks, the SmartCHANGE project implements several key countermeasures. Firstly, gRPC, a widely used communication framework in distributed systems, is employed [28]. This framework uses SSL/TLS protocols to secure data transmission between clients and the server, ensuring that all communication is encrypted and authenticated, thus protecting against eavesdropping and unauthorized access.

Secondly, the SmartCHANGE FL system requires rigorous institution or user verification for participation in the training phase or when submitting queries during the inference phase in real-world deployment (Figure 2). This verification process distinguishes our system from public AI-based services, allowing only verified SmartCHANGE entities — such as consortium members, medical professionals, or citizens — to interact with the model. Entities unable to provide SmartCHANGE verification are denied access to both contributing data to the training process and querying the system for predictions.

Furthermore, the project considers the integration of advanced privacy-preserving techniques, including Differential Privacy [29][30][31][32], Secure Multiparty Computation [33][34], or Homomorphic Encryption [35][36][37]. While these techniques may involve trade-offs, such as reduced model accuracy or increased computational demands, they offer additional privacy protections without requiring fundamental changes to the FL framework and can be implemented with minimal adjustments on the client or server side. In particular, Differential Privacy can provide formal privacy guarantees by ensuring that the inclusion or exclusion of any single data point has a limited impact on the overall model, thereby protecting individual data from inference attacks.

Funded by
the European Union

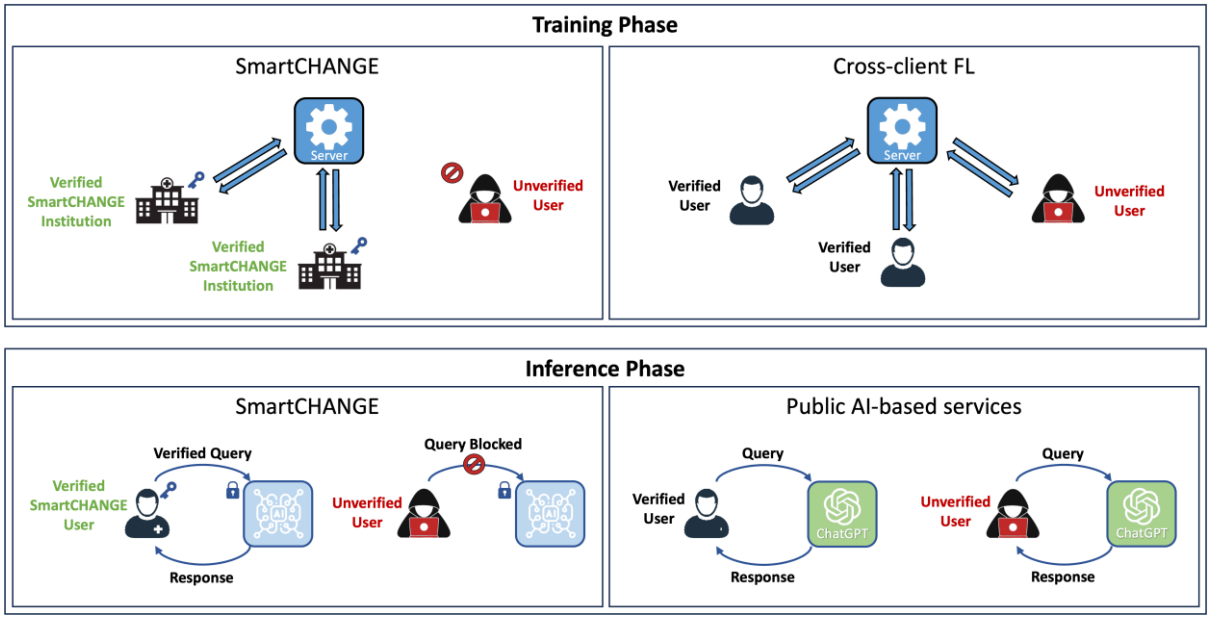

Figure 2 - Overview of the SmartCHANGE system during the training and inference phases. In the training phase, only verified SmartCHANGE institutions participate in the model training, while in the inference phase, the final model only responds to queries from verified clients. This approach enhances security compared to cross-client federated learning and public AI-based services

4.2 Risk Assessment

Risk assessment is a systematic process of identifying, evaluating, and prioritizing potential risks that could adversely affect an organization, project, or system, considering the specific operational context. Within the cybersecurity context, the process generally includes the following steps [38][39]:

1. **Identification of Risks**: Recognizing and listing all possible threats and vulnerabilities associated with data processing activities, including unauthorized access, data breaches, loss, or misuse of personal information.

2. **Evaluation of Risks**: Analyzing the likelihood and potential impact of each identified risk considering factors such as the nature of the data, processing activities, and existing security measures.

3. **Prioritization of Risks**: Ranking the risks based on their assessed impact and probability to determine which ones require immediate attention or mitigation.

Funded by the European Union

4 

Mitigation Strategies: Developing strategies and actions to reduce or eliminate the identified risks, thereby minimizing their potential impact, such as enhancing security controls, establishing data protection policies, and ensuring compliance with legal and regulatory requirements.

5 

Monitoring and Review: Continuously monitoring the effectiveness of risk mitigation measures and regularly reviewing and updating the risk assessment to address new threats and changes in the operational environment.

In the context of the **SmartCHANGE solution**, the process outlined above has been developed through three key phases:

- **Data Analysis**: Determining the level of sensitivity of the clinical and personal data being processed (details in sub-section 4.2.1).
- **Impact Analysis**: Assessing the potential harm to individuals and the organization if risks materialize (see details in sub-section 4.2.2).
- **Security measures**: Establishing measures such as encryption, access controls, and regular security audits to protect data integrity and confidentiality (details in section 4.3).

4.2.1 Data Analysis

The data integrated into the **SmartCHANGE solution** have been identified through the platform's use requirements and co-creation sessions, as detailed in deliverables D3.6[13] and D4.1[14] and reported in chapter 3 and section 6.1 of this document. Due to the nature of this data, which also includes sensitive personal information from adolescents and children, the related analysis indicated a high level of risk. This activity ensured that the appropriate security measures were implemented throughout the process to protect the data, as will be further discussed in the Impact Analysis section (4.2.2).

[13] SmartCHANGE Consortium. Deliverable D3.6 User requirements for the mobile and web application

[14] SmartCHANGE Consortium. Deliverable D4.1 Datasets and framework of health determinants

In addition, following the **Privacy by Design** principles outlined in D2.3[15] and D2.4[16], SmartCHANGE aims to minimize risks of not being compliant to GDPR regulations. Clinical sensitive data, when personal (i.e. not anonymous) clearly fall under the scope of the GDPR. SmartCHANGE, therefore, adheres to core principles such as lawfulness, data minimization, and confidentiality to safeguard data subjects' rights.

4.2.2 Impact Analysis

The potential impacts on subjects, whose data would be processed in the **SmartCHANGE solution,** were assessed in D2.4, while the related outcomes are reported below in relation to the CIA (Confidentiality, Integrity, Availability) concepts (also known as CIA triad), as illustrated in Figure 3 [45][46].

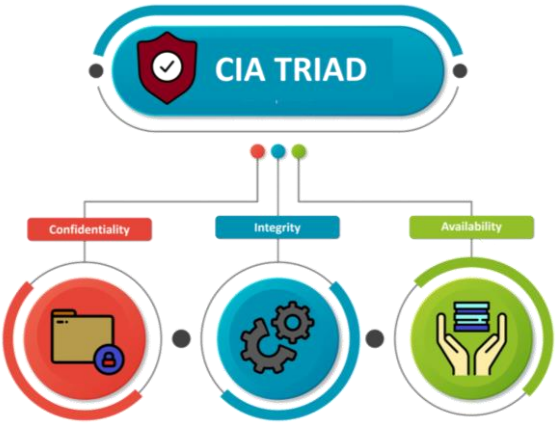

Figure 3 - CIA triad

The assessment results are categorized into three possible outcomes - low, medium, and high impact levels - based on the severity of potential consequences. A *low impact* indicates minimal data disruption with limited effects, a *medium impact* reflects noticeable but manageable consequences on data, and a *high impact* suggests severe, long-term damage to data confidentiality, integrity, or availability.[59]

[15] SmartCHANGE Consortium. Deliverable D2.3 SELP Compliance Framework

[16] SmartCHANGE Consortium. Deliverable D2.4 SELP Impact Assessment

These impact levels provide a framework for understanding the potential risks associated with data processing, particularly concerning the CIA concepts, which are critical for evaluating the security and privacy implications of the SmartCHANGE solution, as detailed below.

- **Confidentiality** means protecting information from unauthorized access.
  The **loss of confidentiality of personal and sensitive data** could have a significant impact on the subjects' rights and freedoms. In particular, the following factors were considered:
  - the **main impacts on those concerned if the risk materialized**: loss of dignity, identity theft, loss of control over their data, loss of trust, psychological damage, loss of reputation.
  - the **main threats that could materialize the risk**: illicit dissemination of personal data, illegal access to computer systems, interception of data on the web portal.
  - the **sources of risk**: internal human sources, external human sources, natural sources.
  - the **measures identified that help to mitigate the risk**: encryption, traceability, logical access control, vulnerability control, physical access control, security of computer channels, personnel management, hardware security, sign a contract with the data controller, fight against malware.

  Based on the confidentiality factors analysed, the **level of impact on loss of confidentiality was estimated as high**.
- **Integrity** means ensuring data are trustworthy, complete, and unaltered.
  The **loss of integrity** could compromise the accuracy, consistency and reliability of the data processed within the SmartCHANGE solution. The following factors were considered to define the level of impact:
  - the **main impacts on those concerned if the risk materialized**: if data is altered, either intentionally or accidentally, it could lead to incorrect assessments, decisions, or actions based on false information, loss of control over the data, psychological damage.
  - the **main threats that could have materialized the risk**: illegal access to computer systems, interception of data on the web portal and app, unlawful dissemination of personal data
  - the **sources of risk**: internal human sources, external human sources, natural sources.
  - the **measures identified that help to mitigate the risk**: backup, data minimization, fight against malware, hardware security, control of physical access.

  Based on the integrity factors analysed, the **level of impact on loss of integrity was estimated as high**.
- **Availability** means making data accessible when needed.
  The **loss of availability** refers to the inability to use the data when needed. This risk is closely related to the **loss of integrity** and for this reason, the same considerations were applied in estimating the level of impact of loss of availability risk. Thus, the **level of impact on loss of availability was estimated as high.**

The results of this initial assessment of the **risk impact level** were used as input for the subsequent analysis phase, focusing on the **probability of threat occurrences** in the **SmartCHANGE solution**. In particular, this analysis included the **identification of main threats**, **risk sources**, and the corresponding **mitigation measures**. Even at this stage, the classification was made based on the CIA triad.

- **Confidentiality**
  The **probability** for a breach of data confidentiality is estimated as **low** in view of the planned technical and organisational measures.
- **Integrity**
  The **probability** for data integrity to be compromised is estimated as **low** in view of the planned technical and organisational measures.
- **Availability**
  The **probability of the risk of data loss** was estimated to be **low** in view of planned security measures and in particular backup procedures.

FINAL CONSIDERATIONS

Based on the outcomes of the previous impact analysis, it was possible to produce the **Impact Analysis Matrix** (Table 1) that reports the values of (i) risk impact levels and (ii) likelihood of threat occurrences.

Table 1 - Impact Analysis Matrix

		Risk Impact Level		
		LOW	MEDIUM	HIGH
Probability Of Threat Occurrences	LOW			X
	MEDIUM			
	HIGH			

In particular, the *Risk Impact Level* was determined as HIGH due to significant potential impacts on confidentiality, integrity, and availability, considering the *consequences of risk materialization*, the *relevant threats* and *risk sources*. Conversely, the *Probability of Threat Occurrences* was assessed as LOW, based on factors such as *relevant threats*, *risk sources*, and the *effectiveness of mitigation security measures*.

Funded by the European Union

Therefore, while the **Impact Analysis Matrix** indicates a high-risk impact level, the probability of threat occurrences is expected to remain low, provided that the appropriate security measures detailed in Section 4.3 are implemented to mitigate/prevent the risks identified in the previous analysis.

4.3 Security Measures

This section outlines the security measures for the protection of personal data, based on the previous analysis. These measures are organized according to the two macro-categories defined by ENISA, organizational and technical, which are further divided into specific subcategories i) Access control and authentication; ii) Logging and monitoring; iii) Server/Database security; iv) Workstation security; v) Network/Communication security; vi) Backups; vii) Mobile/Portable devices; viii) Application lifecycle security; ix) Data deletion/disposal; x) Physical security, following the specifications described in ISO/IEC 27001 Annex A and ISO/IEC 27002. In this specific context, *the focus is on the technical ones.*

Under each subcategory, measures are assigned:

1. a **unique identifier** in the format: *<SECTION_LETTER>.<NUMBER>*, where
    - the SECTION_LETTER indicates the category to which the measure belongs
    - the NUMBER represents the specific position of the measure in that category

    Both tags (i.e., SECTION_LETTER and NUMBER) follow an alphabetical and numerical order.
2. a **risk assessment level** - low (L), medium (M), high (H) - based on its potential impact according to the CIA triad, as described in the previous section.

ACCESS CONTROL AND AUTHENTICATION

The security measures for access control and authentication will be implemented through mechanisms that will restrict access to personal data, ensuring that only authorized users will be able to access the system. The system will identify users with administrative roles who will be responsible for creating and deleting accounts, while the approval/review process will take place externally to the medical records system. Generic accounts will not be supported by the application, so each user will have a personal account. Authentication will require the use of personal usernames and passwords, with password security levels being configurable, though multi-factor authentication will be preferred.

Risk Level	Measure identifier	Measure description
L	**K.1**	An access control system applicable to all users accessing the IT system should be implemented. The system should allow creating, approving, reviewing and deleting user accounts.
L	**K.2**	The use of common user accounts should be avoided. In cases where this is necessary, it should be ensured that all users of the common account have the same roles and responsibilities.
L	**K.3**	An authentication mechanism should be in place, allowing access to the IT system (based on the access control policy and system). As a minimum a username/password combination should be used. Passwords should respect a certain (configurable) level of complexity.
L	**K.4**	The access control system should have the ability to detect and not allow the usage of passwords that don't respect a certain (configurable) level of complexity.
M	**K.5**	A specific password policy should be defined and documented. The policy should include at least password length, complexity, validity period, as well as number of acceptable unsuccessful login attempts.
M	**K.6**	User passwords must be stored in a "hashed" form.
H	**K.7**	Two-factor authentication should preferably be used for accessing systems that process personal data. The authentication factors could be passwords, security tokens, USB sticks with a secret token, biometrics etc.
H	**K.8**	Device authentication should be used to guarantee that the processing of personal data is performed only through specific resources in the network.

LOGGING AND MONITORING

The use of log files is an essential security measure that will be implemented by the SmartCHANGE solution to enable tracking and monitoring, particularly the access to data and hosting systems. Every data access will be monitored by creating specific application logs that include a timestamp, user information, and descriptors of the access (e.g., type of operation, type of data). These application logs will be stored in a dedicated database, whose security will be ensured by using technology that guarantees their integrity.

Funded by the European Union

Risk Level	Measure identifier	Measure description
L	L.1	Log files should be activated for each system/application used for the processing of personal data. They should include all types of access to data (view, modification, deletion).
L	L.2	Log files should be timestamped and adequately protected against tampering and unauthorized access. Clocks should be synchronised to a single reference time source
M	L.3	Actions of the system administrators and system operators, including addition/deletion/change of user rights should be logged.
M	L.4	There should be no possibility of deletion or modification of log files content. Access to the log files should also be logged in addition to monitoring for detecting unusual activity.
M	L.5	A monitoring system should process the log files and produce reports on the status of the system and notify for potential alerts.

SERVER/DATABASE SECURITY

Security measures related to servers and applications will be integrated into the system design using a pseudonymization and data encryption.

Risk Level	Measure identifier	Measure description
L	M.1	Database and applications servers should be configured to run using a separate account, with minimum OS privileges to function correctly.
L	M.2	Database and applications servers should only process the personal data that are actually needed to process in order to achieve its processing purposes.
M	M.3	Encryption solutions should be considered on specific files or records through software or hardware implementation.
M	M.4	Encrypting storage drives should be considered.
M	M.5	Pseudonymization techniques should be applied through separation of data from direct identifiers to avoid linking to data subject without additional information

Funded by the European Union

M	**M.6**	Techniques supporting privacy at the database level, such as authorized queries, privacy preserving data base querying, searchable encryption, etc., should be considered.

WORKSTATION SECURITY

Security for devices will be governed by the internal procedures of the organizations involved in data processing. Additional rules and policies for managing workstations used for data processing will be also subject to the internal procedures of each organization.

Risk Level	Measure identifier	Measure description
L	**N.1**	Users should not be able to deactivate or bypass security settings.
L	**N.2**	Anti-virus applications and detection signatures should be configured on a weekly basis
L	**N.3**	Users should not have privileges to install or deactivate unauthorized software applications.
L	**N.4**	The system should have session timeouts when the user has not been active for a certain time period.
L	**N.5**	Critical security updates released by the operating system developer should be installed regularly.
M	**N.6**	Anti-virus applications and detection signatures should be configured on a daily basis.
H	**N.7**	It should not be allowed to transfer personal data from workstations to external storage devices (e.g. USB, DVD, external hard drives).
H	**N.8**	Workstations used for the processing of personal data should preferably not be connected to the Internet unless security measures are in place to prevent unauthorised processing, copying and transfer of personal data on store.
H	**N.9**	Full disk encryption should be enabled on the workstation operating system drives

NETWORK/COMMUNICATION SECURITY

Network security measures will be implemented during the deployment phase of the solution, ensuring that all incoming connections will use secure protocols and valid TLS certificates. Additionally, it has been planned to deploy firewalls in the environments hosting

the solution and, optionally, intrusion detection solutions in environments exposed via public networks.

Risk Level	Measure identifier	Measure description
L	**O.1**	Whenever access is performed through the Internet, communication should be encrypted through cryptographic protocols (TLS/SSL).
M	**O.2**	Wireless access to the IT system should be allowed only for specific users and processes. It should be protected by encryption mechanisms.
M	**O.4**	Traffic to and from the IT system should be monitored and controlled through Firewalls and Intrusion Detection Systems.
H	**O.6**	The network of the information system should be segregated from the other networks of the data controller.
H	**O.7**	Access to the IT system should be performed only by pre-authorized devices and terminal using techniques such as MAC filtering or Network Access Control (NAC)

BACK-UPS

Backup procedures will be aimed at ensuring business continuity in case of data loss or damage to the software systems that store them. The availability of up-to-date backups will allow for the restoration of such data.

Risk Level	Measure identifier	Measure description
L	**P.1**	Backup and data restore procedures should be defined, documented and clearly linked to roles and responsibilities.
L	**P.2**	Backups should be given an appropriate level of physical and environmental protection consistent with the standards applied on the originating data
L	**P.3**	Execution of backups should be monitored to ensure completeness.
L	**P.4**	Full backups should be carried out regularly.
H	**P.9**	Copies of backups should be encrypted and securely stored offline as well.

Funded by the European Union

MOBILE/PORTABLE DEVICES

Security for mobile/portable devices and wearables will be governed by the internal procedures of the organizations involved in data processing.

Risk Level	Measure identifier	Measure description
L	**Q.1**	Mobile and portable device management procedures should be defined and documented establishing clear rules for their proper use.
L	**Q.2**	Mobile devices that are allowed to access the information system should be pre-registered and preauthorized.

APPLICATION LIFECYCLE SECURITY

During the design of the solution, well-known protocols and technological frameworks (e.g., OIDC protocols and Spring Security) will be employed to ensure a secure development environment. Practices such as secure coding, code review, testing, and access control will be considered during the code development phases. The databases will be configured to ensure that personal data will always remain encrypted.

Risk Level	Measure identifier	Measure description
L	**R.1**	During the development lifecycle best practices, state of the art and well acknowledged secure development practices, frameworks or standards should be followed.
L	**R.2**	Specific security requirements should be defined during the early stages of the development lifecycle.
L	**R.3**	Specific technologies and techniques designed for supporting privacy and data protection (also referred to as Privacy Enhancing Technologies (PETs)) should be adopted in analogy to the security requirements.
L	**R.4**	Secure coding standards and practices should be followed.
L	**R.5**	During the development, testing and validation against the implementation of the initial security requirements should be performed.
M	**R.6**	Vulnerability assessment, application and infrastructure penetration testing should be performed by a trusted third party prior to the operational adoption. The application shall not be adopted unless the required level of security is achieved.

Funded by the European Union

M	**R.8**	Information about technical vulnerabilities of information systems being used should be obtained.
M	**R.9**	Software patches should be tested and evaluated before they are installed in an operational environment.

DATA DELETION/DISPOSAL

Measures in this category will pertain to the irreversible deletion and/or destruction of personal data so that it cannot be recovered. During data processing, the hosting system provided by the health partner will not use any local disk copies, nor any removable or portable devices to store the personal data.

Risk Level	Measure identifier	Measure description
L	**S.1**	Software-based overwriting should be performed on all media prior to their disposal. In cases where this is not possible (CD's, DVD's, etc.) physical destruction should be performed.

PHYSICAL SECURITY

Physical security of the systems will refer to measures designed to protect the systems hosting the SmartCHANGE solution and the managed data (e.g., server machines). This level of protection will reduce the risk of direct—malicious and/or unauthorized—access by unauthorized personnel to these machines. The following measures will need to be managed by the health partners.

Risk Level	Measure identifier	Measure description
L	**T.1**	The physical perimeter of the IT system infrastructure should not be accessible by non-authorized personnel.
M	**T.2**	Clear identification, through appropriate means e.g. ID Badges, for all personnel and visitors accessing the premises of the organization should be established, as appropriate.
M	**T.3**	Secure zones should be defined and be protected by appropriate entry controls. A physical logbook or electronic audit trail of all access should be securely maintained and monitored.

Funded by the European Union

M	**T.4**	Intruder detection systems should be installed in all security zones.
M	**T.5**	Physical barriers should, where applicable, be built to prevent unauthorized physical access.
M	**T.6**	Vacant secure areas should be physically locked and periodically reviewed.
M	**T.7**	An automatic fire suppression system, closed control dedicated air conditioning system and uninterruptible power supply (UPS) should be implemented at the server room.
M	**T.8**	External party support service personnel should be granted restricted access to secure areas.

4.3.1 Mapping ENISA Requirements to MCDG Cybersecurity Guidance

Based on the recommendations provided in section 2.4.1 of D2.4[17], this section compares the ENISA requirements described in section 4.3 of this document with the Medical Device Coordination Group's guidance (MCDG 2019-16 [40]).

While ENISA uses *macro-categories* and *subcategories*, MDCG guidance adopts *domains* and *security categories* to group the measures: i) Governance and Ecosystem; ii) Protection; iii) Defence; iv) Resilience. Therefore, to facilitate comparison, a customised table (see Table 2) that includes the MDCG 2019-16 domains and how these were implemented through the ENISA requirements was produced.

The 'D/N' column of the table is written as a **unique identifier** in the format: *<DOMAIN_NUMBER>.<NUMBER>*, where:

a. the DOMAIN_NUMBER indicates the domain to which the measure belongs
b. the NUMBER represents the specific position of the measure in that domain

Both tags (i.e., DOMAIN_NUMBER and NUMBER) follow a numerical order.

[17] SmartCHANGE Consortium Deliverable D2.4 SELP Impact Assessment

Table 2 - ENISA requirements vs MCDG 2019-16: a cross-check

D/N	MDCG Domain Name	ENISA Measures
1.1	Information System Security Governance & Risk Management	Covered by the impact analysis in section 4.2.2
1.2	Ecosystem Management	Covered by the security measures in section 4.3
2.1	IT Security Architecture	Covered by the security measures in subsections **NETWORK/COMMUNICATION SECURITY**, Workstation security, Mobile/Portable devices
2.2	IT Security Administration	Covered by the security measures in subsections Server/Database security, Data deletion/disposal
2.3	Identity and access management	Covered by the security measures in subsections Access control and authentication, Mobile/Portable devices
2.4	IT security maintenance	Covered by the security measures in subsection Application lifecycle security
2.5	Physical and environmental security	Covered by the security measures in subsection **PHYSICAL SECURITY**
3.1	Detection	Covered by the security measures in subsections Logging and monitoring, Mobile/Portable devices
3.2	Computer security incident management	Covered by the security measures in subsections Back-ups, Logging and monitoring
4.1	Continuity of Operations	Covered by the security measures in subsection Back-ups
4.2	Crisis Management	Covered by the security measures in section 4.3

5 Architecture Design

This section provides an overview of the system architecture, including a detailed description of the main components and their dependencies. The main interactions among them will be analysed in section 6.3 .

The SmartCHANGE architecture, depicted in Figure 4, consists of multiple microservices, each exposing a well-defined set of APIs. The core system runs in a local platform, that will be deployed separately in each pilot site (see section 7 for details on deployment), while external clients such as the HCP Webapp or the mobile app services can interact through an HTTPS REST API managed by an API gateway. This serves as the entry point for clients to interact with the backend services and guarantees. The security of the external connections will be guaranteed by the usage of HTTPS and the OIDC based identity provider, which will provide single sign-on for all the clients and a role-based access to the services.

In relation to the core system, this is composed of several microservices, both SmartCHANGE specific and support services (messaging, identity management, metrics collection, and auditing). All microservices interact through REST APIs and rely on various repositories for storing data like identities, audits, and FHIR-compliant electronic medical records (EMR).

The AI framework is split into local and central subsystems. The local framework has components for data cleaning, visual explanation, federated learning, and risk prediction. These rely on repositories for model storage and local datasets. The federated learning setup includes both local (FLClient) and central (FLServer) components, allowing decentralized training on private data while leveraging public datasets for collaborative model improvement.

The principles that have driven the architecture design are modularity, with clear separation of concerns between microservices, scalability via the federated learning model, and privacy-preserving machine learning techniques. The use of REST APIs ensures interoperability between components.

In the next sections, each service will be described in more detail, and its interfaces and dependencies will be defined.

Funded by
the European Union

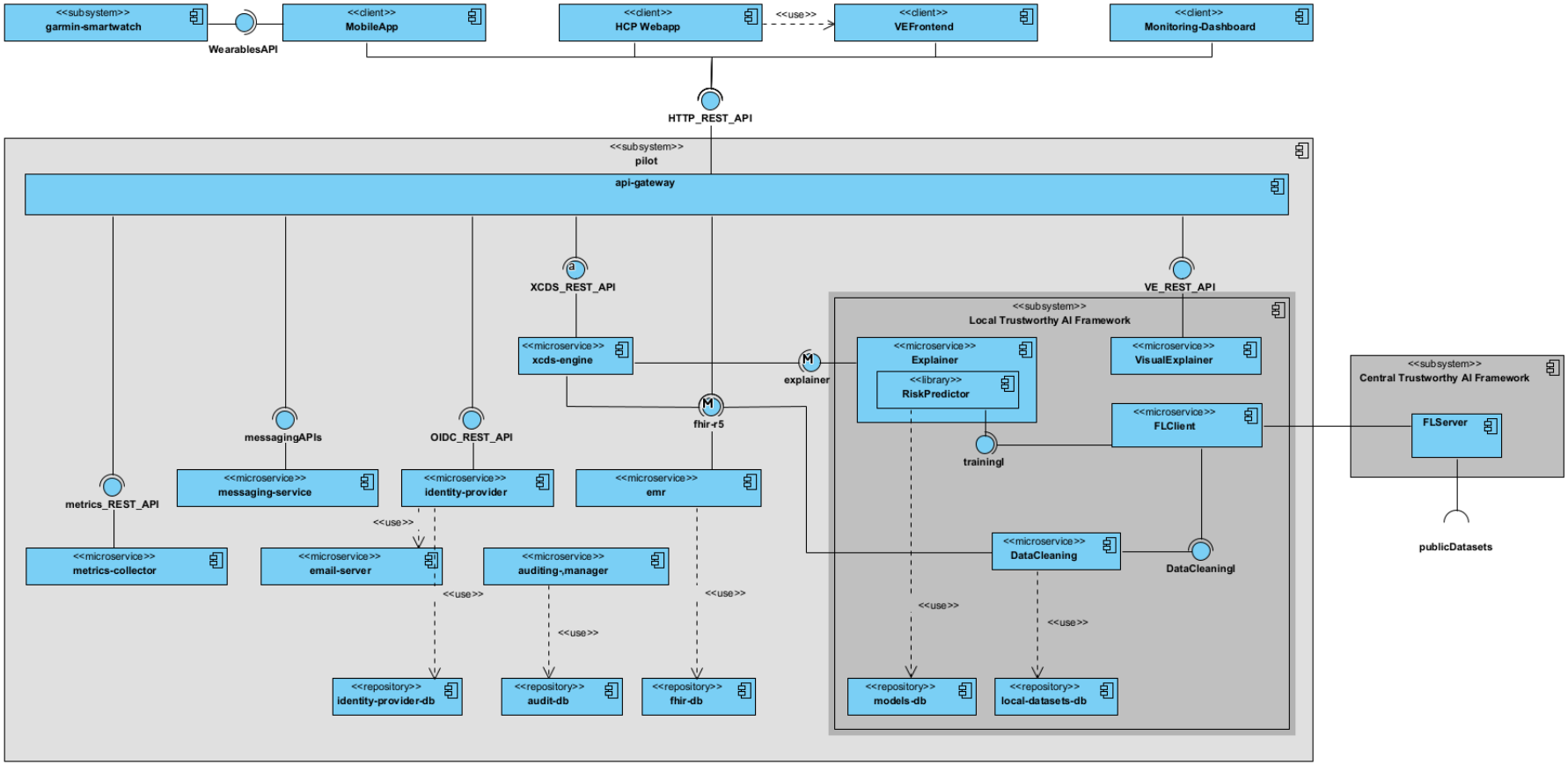

Figure 4 - SmartCHANGE solution component diagram

Funded by
the European Union

5.1 Clients

SmartCHANGE solution includes a set of client applications that interact with the backend microservices. These clients serve different platforms and user roles, providing interfaces for accessing the system's functionality securely and efficiently. Each client is tailored to the needs of its target audience, through a comprehensive co-design process carried out in WP3.

The following sections provide an overview of the clients from a technical point of view, their intended use and their connection with backend services.

5.1.1 HCP Web Application

The Web Application will be a browser accessible application developed in Angular used by Healthcare Practitioners for monitoring the health status of children and adolescents and access the risk predictions on the probability to develop certain diseases in the future. The Webapp will be deployed through a microservice that contains the Angular based application. Its container will be used to serve static contents (e.g., HTML, JavaScript, images) to the client browser that will host the code and run it.

DEPENDENCIES

- **Clinical Repository**, to read and write clinical data from the children and adolescents
- The **XCDS Engine**, to retrieve the risk predictions from the AI services
- The **messaging** service, to allow exchanging chat messages with users of the mobile app
- **Keycloak** service, to provide authentication and authorization based on OIDC.

5.1.2 Mobile app for adolescents and families

The mobile app will be a cross-platform (Android and iOS) application used by families and adolescents for tracking and improving their lifestyle. It makes use of data from the user' wearable and communicates this data with the respective healthcare practitioners.

DEPENDENCIES

- **Clinical Repository**, to read and write clinical data from the children and adolescents
- The **messaging** service, to allow exchanging chat messages with the HCPs using the webapp
- **Keycloak** service, to provide authentication and authorization based on OIDC

5.2 SmartCHANGE Services

This section outlines the microservices that will compose the SmartCHANGE solution. Each microservice is responsible for a specific capability, communicating with others through well-defined APIs. This architecture facilitates easier maintenance, parallel development, and scaling of individual components based on demand.

The following sections detail the various microservices that comprise the solution, their responsibilities, communication patterns, and dependencies. Each microservice is designed to operate autonomously, offering well-defined APIs to interact with other components. Together, they work to deliver the core functionality of the system, ensuring high availability, reliability, and performance across the distributed architecture.

A dedicated section (5.3 will cover the backend services that are based on existing open-source solutions, which have been configured and tailored for use in the project. These open-source services can provide robust, scalable solutions while minimizing development effort, not being core to the project innovation goals, accelerating project timelines, and leveraging community-supported software.

5.2.1 Clinical Repository

The Clinical Repository service (based on FHIR repository) is the backbone of the solution as it is the tool enabling the integration of the elements of the system. In particular, its main capabilities consist of managing the health and sensitive data of the subject. The service regulates, read and write access to clinical data managed by the platform. It implements the HL7 FHIR standard.[1] This service is accessible by the other components of the system, which must be authorized in advance to perform operations on the data. The solution is designed to be modular, platform independent and compliant with the GDPR by applying the privacy-by-design approach. Given those guidelines the solution can be deployed to any cloud commercial solution as long as a specific Data Processing Agreement is in place. The FHIR data server delegates control over the user's authorization to the Security manager module, which checks the user requesting read or write access to the database and allows (or prevents) their access.

Finally, the FHIR data server uses the services provided by the auditing service (see section 5.3.5) to track all data reading and writing operations and then store them into an immutable auditing database as to be further compliant with GDPR requirements.

INTERFACES

The service exposes an interface that can be accessed by clients through the HTTPS protocol. The main services that are supported can be summarized as:

- Read
- Vread
- Create
- Delete
- Update
- Search
- Batch / Transaction

These operations comply with the HL7 / FHIR standard, described on the standard's website [43].

DEPENDENCIES

- **Keycloak** service, to provide authorization based on JWT tokens.
- **Postgres DB.** The Clinical Repository will use a relational database to store data. It will leverage on a JDBC access protocol and therefore supports any DBMS that can be connected via that protocol. In particular, the prototype created for the purposes of the project will use a Postgres server, which provides adequate guarantees on the management of security in compliance with privacy requirements. In fact, Postgres allows the tracing of the operations carried out directly through the DBMS.

5.2.2 XCDS Engine

Within the SmartCHANGE solution, the role of the XCDS Engine will be to act as a broker between the HCP Web App (the Client) and the services providing risk prediction and explanation (ML services). The broker is based on the CDSHooks 2.0 specification [41], and extends some of the standard data structures to enrich the information that can be exchanged, including a richer description of the models and support the exchange of explanations to the outputs of the AI models. The data structures used and the extension to the standard will be described in the interface section.

The XCDS Engine will be responsible for retrieving all the necessary clinical data and forwarding it to the AI services. In particular, in the current architecture the Explainer service will act as the service that aggregates the information on both risk predictions and explanations. Details on the service are provided in section 5.2.4.

The service is designed to be flexible, so that new AI services can be added or modified without changing the client implementation. To do so, the XCDS Engine maintains a list of all available services providing decision support, their endpoints, supported hooks, and required prefetch templates.

Being responsible to prefetch the necessary clinical data ensures that the AI services have all the information needed to generate meaningful guidance or recommendations. Once the engine has pre-fetched the required data, it forwards the entire request, including both the original trigger information and the pre-fetched data, to the Explainer. The Explainer in turn processes the request and returns a response, as described in the corresponding section.

After receiving the response, the engine aggregates and formats the information as per the CDS Hooks standard before sending it back to the webapp. The webapp can then present this information to the healthcare provider.

The usage as a broker and the compliance to a standard such as CDS Hooks provides a clean separation between the webapp and potentially multiple decision support services while also optimizing the flow of clinical data to provide timely and relevant decision support.

INTERFACES

Service Registry:

- **DESCRIPTION:** Get a description of all CDS services offered by this CDS Provider
- **PATH:** GET /cds-services
- **INPUT**: none
- **OUTPUT**: List of *XCDSService*

Get Prediction

- **DESCRIPTION:** Send a request for a prediction
- **PATH:** POST /cds-services
- **INPUT**: *XCDSServiceRequest*
- **OUTPUT**: List of *XCDSCard*

The specification of the data structures referred in this list are available in Appendix A2.I.

DEPENDENCIES

- The XCDS Engine depends on the **Explainer** service to provide risk predictions and explanations, conforming to the description provided in a XCDSService instance.
- The XCDS also depends on a FHIR R5 compliant **Clinical Repository** to retrieve health data for the AI services.

5.2.3 Risk Predictor

The Risk Predictor is a tool to predict the health risk of a person. It operates by analysing the variables describing the person's current health. The inputs to the Risk Predictor are some of the health variables that the machine learning and public risk models need. They are expected at ages between 6 and 14 (one age only). In case some of the inputs are not provided by the user, for example because the general practitioner does not have the necessary equipment at that moment or because the children cannot have their blood test at the moment of the visit, they are instead imputed from the remaining inputs - from those that the user did provide. The Risk Predictor then infers the values of additional variables required for the prediction: these are the variables that are useful for the analysis of risk, but are never provided by the users (it is not practical to require that, for example, 30 different inputs/variables be measured for a child), such as the amount of triglycerides in blood and the hours of sleep per night. The Risk Predictor then forecasts the values of all variables at age 55, when non-communicable diseases become common.

It inputs the variables at the age indicated into one or more existing risk-prediction models. Specifically, it uses the SCORE2 and Health Heart Score models for cardiovascular disease (the first relying more on biological variables, the second on lifestyle), and the Test2Prevent model for diabetes. In addition to the risk and health variables (risk factors) at age 55, the Risk Predictor can also provide the same at lower ages, which can be used for plotting the evolution of risk (factors) over time. The detailed methodology of the Risk Predictor is described in D4.2[18].

[18] SmartCHANGE Consortium. Deliverable D4.2 Baseline prediction pipeline

Funded by the European Union

While not yet implemented in the initial prototype, the Risk Predictor has the potential to incorporate federated learning for model training.The specific models involved may evolve as the risk prediction methodology is refined.

INTERFACES

The Risk Predictor is contained in the Explainer, since the Explainer already heavily depends on the Risk Predictor and constructing additional, separate module does not offer any advantages.

Predict risk

- **DESCRIPTION:** Invoke the service to get a risk prediction
- **INPUT**:
  - a list of key / value pairs representing the values of all relevant input data for the risk prediction. The actual list of keys is dependent on the type of prediction. Some values can be null, in this case the model will infer those values.
  - the disease it is predicting among a set (e.g. “DIABETES”, “CVD”)
- **OUTPUT**: a list of risk values, each composed of a pair `<age, risk>`.

Train

- **DESCRIPTION:** Additionally train trainable models inside the risk predictor
- **INPUT**: The model to be trained, training data in a format compatible with the models
- **OUTPUT**: Updated model

DEPENDENCIES

The Risk Predictor does not call any further component.

5.2.4 Explainer

The Explainer module is designed to provide insights into the Risk Predictor’s outputs by generating counterfactuals. These counterfactuals highlight the necessary modifications to a subject’s profile that would alter their risk of developing a specific disease. In addition to enhancing interpretability, these counterfactuals can serve as actionable recommendations for doctors to propose to the participants. When implemented in real-life scenarios, these recommendations encourage healthier behaviours, ultimately reducing the risk of disease development.

Our Explainer module takes as input the individual's current condition, predicts the risk using the Risk Predictor, and utilizes the processed information, along with the target risk factor, to generate the counterfactuals. Moreover, it offers temporal flexibility: counterfactuals can be generated not only for the present moment but also for a future time frame (e.g., one year from now) to support long-term recommendations and goals for the individual.

The Explainer requires training, which can be carried out in a federated learning setting. There are two training approaches:

1. **Joint Training**: The Explainer and Risk Predictor are trained simultaneously. This approach allows the Explainer to influence the training process of the Risk Predictor, leading to potentially better integrated models.
2. **Post-Hoc Training**: The Risk Predictor is trained independently, and once its parameters are fixed (frozen), the Explainer is trained on top of it.

Both training approaches are compatible with federated learning scenarios. However, we are still evaluating which strategy will be most effective for our use case.

INTERFACES

When these functions are called, the service will query the Explainer which will interact directly with the Risk Predictor.

Predict and explain:

- **DESCRIPTION:** Invoke the service to get a risk prediction and the corresponding explanation
- **PATH:** POST /execute
- **INPUT**: ServiceInput: Includes the user data, how many counterfactuals to generate, the desired risk of the counterfactuals and the year of the subject for which we want to generate counterfactuals (see appendix A2.I)
- **OUTPUT**: ServiceOutput: The predicted risk and the generated counterfactuals and their risk (see appendix A2.I)

Explain:

- **DESCRIPTION:** Invoke the service to get the corresponding explanation of an already generated risk prediction
- **PATH:** POST /explain
- **INPUT**: *ExplanationRequest*

- **OUTPUT**: list of *XCDSExplanation*: The generated counterfactuals

The specification of the data structures referred in this list are available in Appendix A2.I.

DEPENDENCIES

- Risk Predictor

5.2.5 FL Services

The FL Services component is responsible for enabling secure and efficient distributed learning across multiple machines or institutions. Its primary purpose is to facilitate the concurrent training (or fine-tuning) of a predictor and explainer model without transferring the clients' raw data to external servers, thereby ensuring data privacy and compliance with regulations such as the GDPR. The component establishes a client instance on each machine that owns the data, while the server coordinates the aggregation of locally computed updates. Secure protocols are employed to enhance robustness against adversarial contributions and minimize the risk of data leakage.

The key functionalities of this component include:

- **Model Initialization**: The server initializes the FL process by distributing the initial parameters of the pre-established models (i.e., predictor and explainer) to each client.
- **Client-Server Communication**: The component coordinates communication between the clients and the central server, with clients transmitting local updates and metrics to the server, and the server sharing global model parameters and aggregated results back to the clients, ensuring communication integrity and security.
- **Aggregation**: The server aggregates the local updates from all clients to generate a global model, which is then shared back with the clients for the next iteration of training.
- **Performance Monitoring and Adaptation**: The component continuously monitors the performance of the global model and adapts the learning process to improve accuracy, reduce overfitting or underfitting, and maintain model robustness.
- **Secure Aggregation**: The component might include additional privacy-enhancing technologies to further ensure data confidentiality and integrity.

INTERFACES

Invoke:

- **DESCRIPTION:** This interface is used to initiate the FL process, supporting both the training of new models and the fine-tuning of pre-trained predictor and explainer models.
- **PATH:** POST /execute
- **INPUT**: The current model parameters for the predictor and explainer (if available, otherwise initialized with random values), as well as the locations of client datasets
- **OUTPUT**: The trained model parameters for both predictor and explainer

DEPENDENCIES

Dependencies from Other Components in SmartCHANGE

- **Client Dataset Locations**: Information on where the client datasets are stored, along with any necessary preprocessing steps.
- **Model Architectures**: Specifications of the predictor and explainer model architectures to be used in the training process.

Libraries

- **Preprocessing Libraries**: Required libraries for data preprocessing, if applicable.
- **Model Training Libraries**: Libraries used to create and train the predictor and explainer models.
- **Flower Library**: A federated learning library used to manage the FL process, coordinate client-server interactions, and perform model aggregation.

5.2.6 Visual Explainer

The counterfactual visual analytics tool, usually named Visual Explainer, is a tool to allow health practitioners to follow-up on explorative or investigative analysis beyond what they can do with the HCP Web Application. Assuming a doctor uses the HCP Web Application and have follow up questions, they click a link within the HCP Web Application forwarding them along with the user's information to the Visual Explainer. There, the doctor can ask "what if" questions like "what if this patient had less HDL, would the diabetes risk be different?", explore different forecasting possibilities like "how will my patient look like in 38 years?", and investigate the subject's counterfactual cohort to better understand their situation and the various potential directions they could take. Though flexible, it is important to note that the Visual Explainer is not a tool which suggests plans of action for the subject or the doctor, but rather with the extra information seen though the visualizations, the doctor can make a more informed decision. The Visual Explainer is also aimed at providing the machine learning experts a glimpse of how their models are interacting with each other and to audit if the

results are as expected, or if further development is needed for a given model. Additional information can be found in the D5.2[19] document.

INTERFACES

URL Connection: The Visual Explainer will be available through a http(s) URL, which can be accessed via a URI to be defined (e.g. https://future_link/). The HCP Web Application can forward a health practitioner to the Visual Explainer by using the link with parameters. At the moment of writing, the supported parameters are the following:

- Age: current age of subject (number)
- Gender: subject gender (0 for male and 1 for female)
- BMI: subject's bmi (number)
- Systolic_blood_pressure: subject's systolic blood pressure (number)
- Total_cholesterol: subject's cholesterol (number)
- HDL: subject's hdl (number)
- LDL: subject's ldl (number)
- Smoking: smoking pattern of the subject (any value lower than 4: smoking, 4 or higher: not smoking.). It is important that this variable is still under discussion due to variation on how it is stored and used by the models.

A sample request can be the following:

```
https://future_link/patient?Age=18&Gender=1&BMI=23&Systolic_blood_pressure=128&Total_cholesterol=6&HDL=1.28&LDL=4.4&Smoking=1
```

SECURITY CONSIDERATIONS:

No personal identifiable information is passed to the tool, but the tool can potentially support an oauth2 authentication protocol for the transmission of the user's data.

[19] SmartCHANGE Consortium. Deliverable D5.2 Initial trustworthy AI suite

Funded by the European Union

DEPENDENCIES

Currently all interface with existing models and data is done within, not requiring any other API interfaces to the rest of the SmartCHANGE systems, but in a future release this may be revisited so that WP5 and WP6 can consolidate where the models are located and queried from.

Although the Visual Explainer tool can live independently, it depends on consuming the forecast model and counterfactual model made by JSI and USI. A connection to the HCP Web Application as an external link is also valuable.

5.2.7 Data Cleaning Service

The Data Cleaning Service is designed to handle the preprocessing, cleaning, and harmonization of diverse datasets. The primary purpose of this component is to ensure that all datasets, regardless of their origin, are standardized and cleaned before being merged into a unified dataset. The component addresses various data quality issues, including missing values, outliers, irrelevant data, and duplicates. It offers flexible cleaning methods adaptive to the specific needs of the datasets and a harmonization process that aligns the data structure and format across all datasets to facilitate integration.

INTERFACES

Retrieve all databases:

- **DESCRIPTION:** Retrieves a list of all available databases for the user to select
- **PATH:** GET /database
- **INPUT**: -
- **OUTPUT**: List of *Database* objects

Selection of databases:

- **DESCRIPTION:** Users can select the datasets that want to continue with the process of cleaning through this interface.
- **PATH:** POST /selection
- **INPUT**: list of string. (Selected databases by the user)
- **OUTPUT**: *SelectionReport*: a report on any inconsistencies found in the dataset and then will be prompted to select one or more cleaning methods from the available options to begin the cleansing process

Data Cleaning

- **DESCRIPTION:** This operation applies selected cleaning methods to the data and prepares them for download
- **PATH:** POST /cleaning
- **INPUT**: list of strings (cleaning methods to apply)
- **OUTPUT:** *CleaningReport*: Returns the link where the user can download the database

Cleaned Dataset & Log Download:

- **DESCRIPTION:** Once the cleaning process is complete, this interface allows users to download the cleaned dataset. Additionally, a detailed log file documenting all changes made during the cleaning process is available.
- **PATH:** POST /download
- **INPUT**: -
- **OUTPUT**: *DownloadResponse*: Returns the link where the user can download the database and the logs

The data structures referred in this list are available in Appendix A3.III, together with the OpenAPI specification Swagger [48].

DEPENDENCIES

The Data Cleaning Service has no external dependencies.

5.3 Services based on existing solutions

5.3.1 Wearables

The Garmin VivoSmart 5 fitness tracker was selected as the wearable of choice for this project because of its minimal design, low price, and the possibility to use the wearable without subscribing to any proprietary cloud service or third-party application.

Garmin offers a fully featured, out-of-the-box solution in their Garmin Connect API and cloud service. However, this project's target audience is too young to make use of the Garmin Connect cloud for storage and processing of data, it would not be compliant with Garmin's terms of service. Instead, the project will use the official Garmin Software Development Kit (SDK). The SDK will be integrated with the Android and IOS application and will take care of the Bluetooth pairing and data synchronization with the wearable.

The wearable has local storage, where it can store up to 14 days' worth of tracked data. Upon synchronization, which should happen preferably each day, our mobile application will receive the data from the wearable in Garmin's API format, at which point the wearable forgets the data. The mobile application will translate this to be FHIR compliant before sending it to our API for processing and storage. At no point is any of the data that is collected by the wearable sent to Garmin.

5.3.2 API Gateway

In a microservices architecture, the client apps usually need to consume functionality from more than one microservice. If that consumption is performed directly, the client needs to handle multiple calls to microservice endpoints. Therefore, having an intermediate level or tier of indirection (Gateway) can be convenient for microservice-based applications. The API-gateway is a server that is the single-entry point into the system. It is similar to the Façade pattern from object-oriented design. It encapsulates the internal system architecture and provides an API that is tailored to each client. It might have other responsibilities such as authentication, monitoring, load balancing, caching, request shaping and management, and static response handling. Moreover, it is responsible for request routing, composition, and protocol translation. All requests from clients first go through the api-gateway. It then routes requests to the appropriate microservice. The api-gateway will often handle a request by invoking multiple microservices and aggregating the results. It can translate between web protocols such as HTTP and WebSocket and web-unfriendly protocols that are used internally. Moreover, the api-gateway is also responsible of managing cross-cutting concerns such the management of SSL. The api-gateway is based on the open-source product **Traefik** - released with MIT license.[42]

5.3.3 Identity provider

The SmartCHANGE platform provides sign up and login functionality through **Keycloak**, an open-source Identity and Access Management solution that simplifies the process of securing applications and services. It supports standard protocols like OpenID Connect (OIDC), OAuth 2.0, and SAML 2.0, making it versatile for various authentication needs [19].

Keycloak provides features such as Single Sign-On (SSO), user federation, and fine-grained authorization, which streamline user management and enhance security. Its web-based admin console allows for easy configuration and management, reducing the complexity for developers. The centralization of authentication and authorization processes in Keycloak ensure a consistent and secure user experience across multiple applications. This makes

Keycloak an ideal choice for organizations looking to improve security, reduce administrative overhead, and provide a seamless user experience.

In Figure 5, a schema illustrating how Keycloak works is reported. It acts as a central authentication server, managing user identities and access permissions. Applications and services (CLIENT side) delegate authentication to Keycloak, which verifies user credentials and issues tokens for access (SERVER side). This centralized approach not only simplifies the integration of security features but also ensures that all applications adhere to the same security standards, making Keycloak a robust and scalable solution for modern identity and access management.

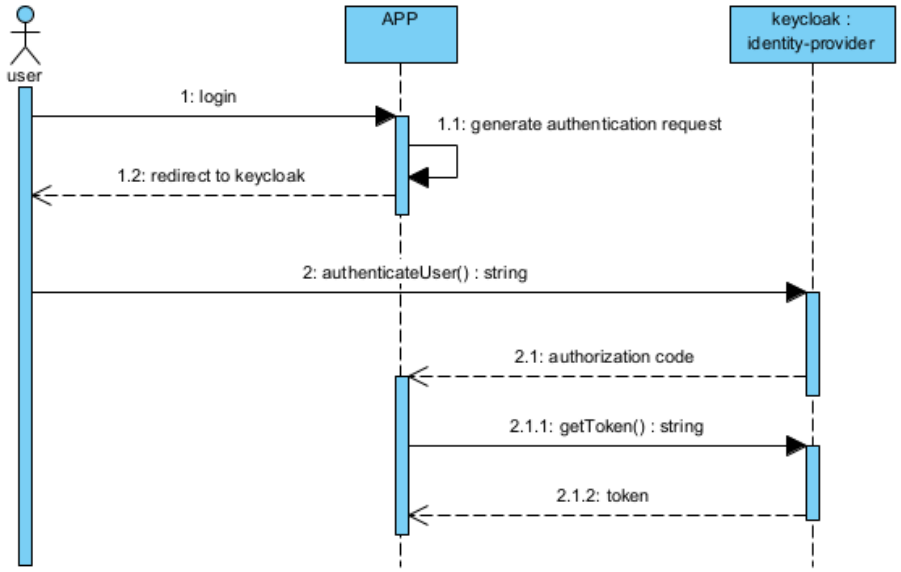

Figure 5 - Keycloak: OIDC Authorization flow

5.3.4 Monitoring

Monitoring is the practice of tracking and analysing the operations and performance of individual microservices within a larger application architecture (as the SmartCHANGE platform). This involves gathering, reporting, and storing data (monitoring the API calls) to ensure each microservice is functioning correctly and efficiently and is able to interact with each other to deliver the expected results. In this way, monitoring provides a comprehensive view of the application's key areas and performance during development, enabling targeted decisions in later phases. The monitoring component will be made of the following modules:

- A *call-tracing* module, based on the product **Zipkin** [43], which is released with an Apache License 2.0. In particular, the module enables the distributed tracing of the transaction by

helping to gather timing data needed to troubleshoot latency problems in service architectures. Features include both the collection and lookup of this data. To be able to monitor all transactions within the solution, services of the whole platform will have to be instrumented to notify it whenever a transaction is conducted.

- A *metrics-collector* module, acting as a probe for the collection and structuring of infrastructure monitoring data. This module will scrape metrics from instrumented applications, either directly or via an intermediary push gateway. This module will be based on the open-source product **Prometheus**, released with *Apache License 2.0*.[57]
- A *monitoring dashboard*, offering the capability of defining custom charts and alerts by leveraging the information collected by the *metrics-collector* module. The dashboard will be based on the open-source product **Grafana**, released with *Apache License 2.0*.[56]

5.3.5 Logging & Auditing

Log components are meant to centralize the container logs as to support log analyses. A log collector component, based on the lightweight **Grafana Loki** stack, will provide support for logs collection and visualization. The service will parse the containers logs as to extract structured data collector like timestamp, log level and message. Those data will then be pushed to a *log aggregator.* The *log-aggregator* component will be based on the open-source product **Loki** - released with *Apache License 2.0*. The logging components will be customized with logs retention policies, according to the project needs.[55]

To prevent tampering, and enabling auditing, logs will be stored in an **immudb** based repository. *immudb* is an open-source database with built-in cryptographic proof and verification. It can track changes in sensitive data and the integrity of the history will be protected by the clients, without the need to trust the server repository.

5.3.6 Messaging service

RabbitMQ was selected as the message broker, aimed at enabling the web app and the mobile to exchange chat messages. RabbitMQ, known for its ability to handle asynchronous communication using a publisher-subscriber model, is well-suited for decoupling services, which is essential in a microservice architecture.[61] This enables independent scaling, deployment, and updates of each component.

To facilitate communication, the clients will use STOMP (Simple Text Oriented Messaging Protocol) over WebSockets, allowing the web and mobile apps to interact with RabbitMQ in a lightweight, text-based protocol. STOMP simplifies the integration between services,

making it easier to implement while maintaining a reliable, asynchronous communication pattern.

RabbitMQ also supports OpenID Connect (OIDC) to ensure that only authorized services and users can publish or subscribe to queues. The availability of a RabbitMQ as a Docker image will also ease the deployment.

6 Data Model and Flows

This section defines the types of data exchanged among services within the SmartCHANGE solution. It outlines the data categories, their sources, and the flow of information between system components. Additionally, the section discusses the role of each service in the data exchange process, ensuring that the interactions are aligned with security, privacy, and performance requirements.

6.1 Health data managed the SmartCHANGE Solution

6.1.1 Data Exchange with the Mobile Application

The mobile application primarily shares activity data measured by the wearable and qualitative data gathered via short questions presented to the user. This data allows to calculate the user's future health risks and allows HCP to monitor their young participants.

WEARABLE DATA SHARED VIA THE MOBILE APP:

- (Resting) heart rate
- Physical exercise data
  - Type
  - Duration
  - Frequency
- Hours of sleep/night

Additional characteristics that are measured by the wearable and might be used by the SmartCHANGE system are:

- Oxygen saturation
- VO2max

INDICATION OF QUESTIONNAIRE DATA SHARED VIA THE MOBILE APP:

- Socio-demographic information:
  - Age
  - Sex
- Physical and vital measurements:
  - Height
  - Weight

Funded by the European Union

    - Waist circumference
- Basic dietary information
    - Servings of fruit/day
    - Servings of vegetables/day
    - Servings of sugary drinks/day
    - Servings of nuts/day
    - Servings of red and processed meats/day
    - Grams of cereal fibres/day
- Average Physical activity (hours/week)

In addition, the mobile application *receives* some data from other system components:

- Initial and updated risk prediction, influencing the behaviour of the app.
- Goal adjustments by HCP.

Lastly there will be a chat exchange with HCP and the user of the mobile app.

6.1.2 Data exchanged by the Webapp for HCPs

The webapp for HCPs will mainly consume data stored in the Clinical Repository, to populate the health dashboards of the individuals in care with the HCPs.

The webapp will produce health data, that can be entered by HCPs during the encounters with the persons they have in care, through dedicated forms. In particular, the webapp will produce or update data regarding:

- Socio-demographic information:
    - Age
    - Sex
- Physical and vital measurements:
    - Height
    - Weight
    - Waist circumference
    - BMI (automatically computed from height and weight)
    - Blood pressure (systolic and diastolic)
- Blood test results
    - Glucose (fasting glucose)
    - Cholesterol (LDL, HDL, Total)
- Basic dietary information
    - Pieces of fruit/day
    - Pieces of vegetables/day

    - Number of sugary drinks/day
    - Servings of nuts/day
    - Servings of red and processed meats/day
- Socio-economic status
    - Parents' education level
- Fitness level
    - Cardiorespiratory fitness

The webapp will also produce *goal assignments* and *risk predictions*.

The webapp will also provide a functionality to register the families and adolescents that will use the SmartCHANGE solution, thus it will produce registration information.

6.2 FHIR: the standard for clinical data

Fast Healthcare Interoperability Resources (FHIR) [1] is a REST-ful based approach developed by HL7 for modelling data structures as Healthcare Resources and services as REST-APIs to provide a solution for **health interoperability**. This ensures that healthcare information can be exchanged and integrated seamlessly, securely, and efficiently across different systems and organizations. Some key aspects are discussed here below.

INTEROPERABILITY:

- **Simplified Implementation:** FHIR aims to simplify the implementation process without sacrificing information integrity. It leverages existing logical and theoretical models to provide a consistent and easy-to-implement mechanism for exchanging data between healthcare applications.
- **Modular Components:** FHIR uses modular components called "*resources*" to represent different types of healthcare data, such as patients, observations, medications, and procedures. These resources can be combined in various ways to meet different healthcare needs, enhancing flexibility and scalability.
- **Semantic Interoperability:** FHIR addresses semantic interoperability by standardizing the terminology used within healthcare systems. This is achieved through FHIR Profile Resources, which define the specific health terminologies (e.g., SNOMED-CT [2], ICD [4], LOINC [3]) used by different adopters. By mapping these terminologies, different systems can understand and use the data consistently.

DATA EXCHANGE:

- **Standardized Formats:** FHIR supports multiple data formats, including JSON, XML, and RDF, making it versatile for different systems and applications. This standardization ensures that data can be easily exchanged and understood across different platforms
- **RESTful API:** FHIR uses RESTful APIs, which are widely adopted in web services, to facilitate efficient and secure data exchange. This approach allows for seamless communication between systems, enabling real-time data sharing and integration.

SECURITY AND PROVENANCE:

- **Robust Security Protocols:** While FHIR itself is not a security protocol, it defines exchange protocols and content models that need to be used with various security protocols. This includes the use of TLS for secure communications, OAuth for authentication, and a security label infrastructure for access control.
- **Audit and Provenance:** FHIR includes resources for tracking the origins, authorship, history, status, and access of data, ensuring transparency and accountability.

6.2.1 Usage of FHIR standard in SmartCHANGE

Building upon the above discussion, here below some details on the FHIR R5 resources that were used to model the data gathered by the mobile app, the wearables and the web app in the SmartCHANGE project.

6.2.1.1 FHIR resources used in the project

1. **Patient** is a foundational element that represents key demographic and administrative information about an individual receiving care. This includes data such as name, age, gender, contact information, and other personal identifiers. The Patient resource **is central to almost all healthcare interactions in FHIR**, serving as a link between the individual and the various health records, observations, assessments, and treatments that pertain to them.[7]
   In the SmartCHANGE project this resource has been used to map key demographic information, such as the individual's sex and age, which are critical for context when interpreting other health data. These demographic factors are often essential for tailoring health assessments, risk predictions, and treatment plans to the individual's specific needs.
2. **Related Person** represents Information about a person that is involved in an individual's health or the care for a subject, but who is not the target of healthcare. In SmartCHANGE, the family members will be modelled with both the Patient resource, as they may be the subject of some activities in the behaviour change strategies, and through the use of the RelatedPerson resource, to represent their relationship with the mail subject of care. The strategy used is the one described in [59].

3. **Practitioner** represents an individual who is involved in the delivery of healthcare services. This can include various healthcare professionals such as **doctors**, **nurses**, pharmacists, therapists, and other clinical or **non-clinical personnel** who contribute to subject care. The resource captures essential information about the practitioner, including their qualifications, certifications, specialties, and contact details. It serves as a critical component for documenting who is responsible for delivering specific aspects of care and linking that information to individual health records.[11] The specific roles or functions (e.g., general practitioner, specialist, IT specialist) a Practitioner may hold within an organization or healthcare setting is defined by the **PractitionerRole**. This resource links a Practitioner to his/her responsibilities, location, services provided, and the healthcare organization he/she is associated with. PractitionerRole is particularly useful for managing complex healthcare settings where professionals may fulfil multiple roles across different contexts or organizations.[12] In the **context of the SmartCHANGE project**, these resources were used to map both clinical and non-clinical experts: the school nurse, the general practitioner (GP), and the AI expert. In particular, the Practitioner resource captured essential details about their qualifications, certifications, and contact information ensuring that each professional's identity and expertise were documented, regardless of their specific role or specialization. By utilizing the PractitionerRole resource, the system effectively highlighted the distinct responsibilities and areas of expertise that each professional brought to their respective fields:
    a. the school nurse's role was defined within the educational setting, focusing on adolescent health, first aid, and health screenings, and linking him/her to the school where he/she worked.
    b. the general practitioner's role was tailored to providing primary care for children aged 6-10, with a focus on paediatric care, routine check-ups, immunizations, and growth monitoring, usually in association with a clinic or hospital.
    c. the AI expert's role, while non-clinical, was pivotal in healthcare innovation, specifically in developing and applying ML and AI algorithms for risk prediction of cardiovascular and metabolic diseases. This role emphasized his/her contribution to data analysis, model development, and collaboration with clinical professionals, defining his/her unique position within the healthcare framework.
4. **RiskAssessment** is used to **evaluate and document the likelihood of a particular outcome or condition occurring, based on specific factors or risk indicators**. This resource typically includes the risk prediction, the factors contributing to this assessment, and any recommendations or interventions that might mitigate the risk. RiskAssessment is crucial for preventive care, as it helps in identifying individuals who may benefit from early interventions or monitoring to prevent the development of serious conditions.[6]
   In the SmartCHANGE project this resource has been used to assess and document the likelihood that a person will develop cardiovascular or metabolic diseases in adulthood. This risk is calculated based on data collected through both web and mobile applications,

encompassing lifestyle factors, medical history, physical measurements, and other relevant health indicators. The RiskAssessment provides a predictive model that can inform both participants and healthcare providers about potential future health risks, allowing for early intervention or lifestyle modifications.

5. **Goal** captures specific health objectives or targets set for the subject, often based on clinical assessments or risk predictions. It reflected measurable challenges, such as managing disease risk or improving health outcomes, and is used to guide treatment plans and individual care strategies over time[53]. A **CarePlan** describes the intention of how a practitioner intend to deliver care for a particular individual for a period of time. They may describe education campaigns [58].
6. **Observation** is designed to capture and represent measurements, assertions, and assessments regarding an individual's health status. This resource is highly versatile and can represent a wide variety of clinical and non-clinical observations, including **vital signs**, laboratory results, **physical measurements**, and even **lifestyle indicators** such as physical activity or dietary habits. The Observation resource is essential for documenting the subject's condition over time, enabling clinicians to track changes and trends in health-related data.[5] **In the SmartCHANGE project this resource has been used to map various types of health-related data**, including:
   - *Physical and Vital Measurements*, capturing metrics like heart rate, blood pressure, height, weight, etc.
   - *Blood Biomarkers*, recording laboratory results such as glucose levels, cholesterol, etc
   - *Lifestyle Behavior*, monitoring lifestyle choices like smoking, and exercise habits.
   - *Physical Fitness*, tracking fitness levels through measurements such as BMI, body fat percentage, or performance on fitness tests.
7. **Questionnaire** is designed to define structured sets of questions intended for data collection. These questions can range from simple yes/no formats to more complex input types, such as multiple-choice or free-text answers.[8] **Questionnaire Response** captures the actual responses provided by participants or healthcare providers. Together, **these resources facilitate the collection of structured information, which can then be analysed, stored, or converted into other relevant FHIR resources**.[9]
   In the SmartCHANGE project these resources have been used to capture initial individual information through structured surveys. These questionnaires might gather data on medical history, lifestyle habits, symptoms, or general health concerns. The responses provided were then stored as QuestionnaireResponse resources and were used directly for analysis or converted into other FHIR resources, such as Observations, when the information was linked to measurable health data.
8. **Condition** includes clinical details about an individual's health condition including chronic/rare diseases, acute/ongoing diseases or other medical conditions. This resource documents key aspects such as the clinical status, verification status, onset date, severity, and outcomes. By recording these details, the Condition resource provides a comprehensive overview of the

person's medical history and current health status, ensuring accurate tracking and management of their health conditions.[50]

In the SmartCHANGE solution, this resource has been used to map the medical history of the participant, including past medical conditions or surgical procedures. This ensured a comprehensive record of the subject's health background, aiding in informed clinical decision-making.

9. **Organization** represents entities involved in the provision of healthcare services, such as hospitals, clinics, and other healthcare facilities, including details about the organization's name, type, contact information, address, and the services provided.[51]

   In the SmartCHANGE solution this resource has been used to map the healthcare structures where participant had encounters or was under care, contributing to the coordination of care across different healthcare entities.

10. **Encounter** captures interactions between a subject and healthcare provider for the purpose of providing healthcare services. This resource includes information about the type and reason for the appointment, the date and time of the encounter, and the involved participants. By documenting the specifics of each healthcare interaction, the Encounter resource facilitates better coordination of care and a clearer understanding of the individual's healthcare journey.[54]

    In the SmartCHANGE project this resource has been used to map details of healthcare interactions, such as the type and reason for the appointment, or instances when the healthcare provider or school nurse interacted with the subject. This setup facilitated the systematic gathering and organization of user data throughout the care process.

11. **NutritionIntake** details the dietary intake of a participant, including types and amounts of food and drink consumed. This resource also captures information on the individual's nutritional habits and any related observations, supporting nutritional assessments and interventions.[52]

    In the SmartCHANGE project this resource has been used to map the subject's dietary habits, including their consumption of food and beverages, as well as alcohol consumption. By documenting these details, the NutritionIntake resource provides valuable insights into the young participant's nutritional status and potential health risks.

12. **Device** represents any physical item that is used in the provision of healthcare, such as medical devices, **wearable technology**, implants, or even software that is used for diagnosis or monitoring. This resource includes details about the device's identity, type, and operational status, as well as information about its manufacturer, model, and usage. In contexts where devices are used to monitor person health, the **Device resource is critical for linking the data generated by these tools to the participant's health records**.[10]

    In the SmartCHANGE project these resources have been used to map wearable devices that monitor subject parameters like physical activity, sleep patterns, or other health metrics. These devices play a crucial role in continuous health monitoring, providing real-time data that can be integrated with other FHIR resources like Observation to give a comprehensive

view of the young individual's health status. This integration allows for a more detailed and accurate assessment of the subject's health, which can inform personalized care strategies.

Appendix 1 reports the details of the FHIR Data model and how the health data produced in the SmartCHANGE solution (and described in the previous sections) will be persisted. The same data model is used to send the relevant clinical data to the AI services, and to obtain a risk assessment value, thus it is crucial in guiding the development phase.

6.3 Data Information Flow

This subsection includes diagrams that depict the interactions among system components, users, and external systems. The flows are described using UML Sequence Diagrams.

6.3.1 Data Collection

Data is collected from the Youth/Family Mobile App and wearable devices. This data includes personal health information, which is sent to the Clinical Repository, to make it available to the authorized services in the platform. Figure 6 depicts the three data collection flows within SmartCHANGE solution.

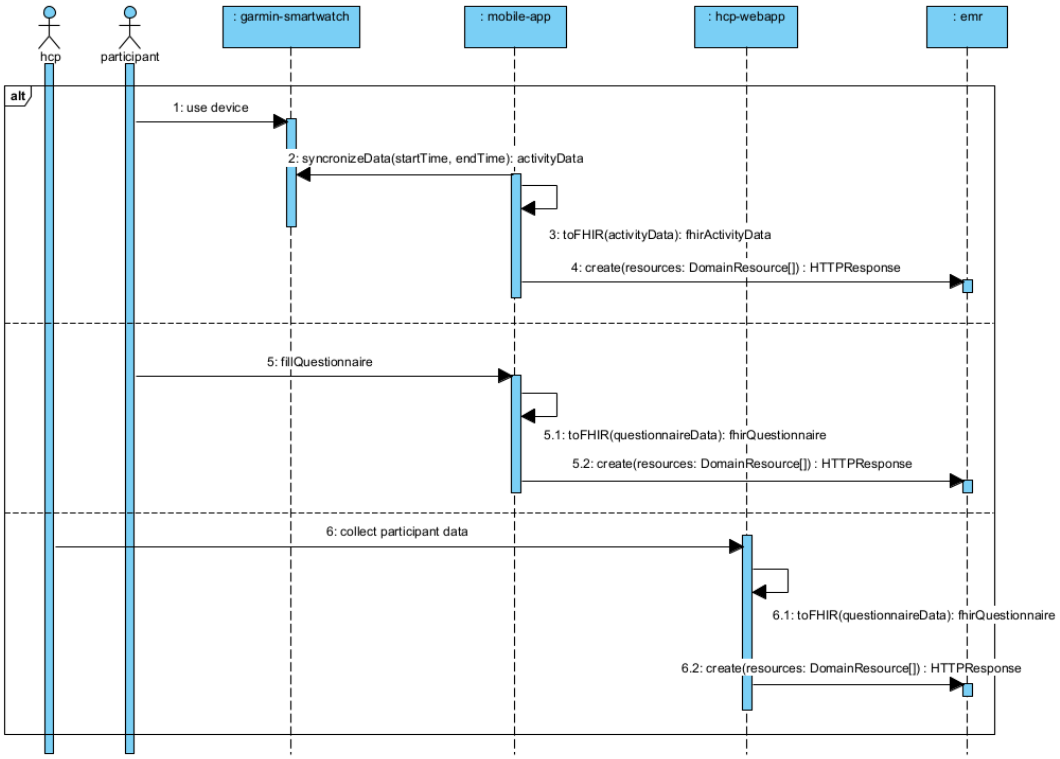

Figure 6 - Data Collection Sequence Diagram

Data Collection Flow 1:

1. The Participant uses the device (i.e., 'garmin-smartwatch') to gather his / her data through
2. The Mobile App periodically starts / ends a process to synchronize with the data collected by the device
3. The Mobile App converts this data to the corresponding FHIR resources and sends it to the Clinical Repository for storage, making them available to the authorised services within the SmartCHANGE platform.

Data Collection Flow 2:

1. The Participant is prompted through the Mobile App to fill a questionnaire regarding his / her health status, habits and other personal information
2. The Mobile App converts this data to the corresponding FHIR resources and sends it to the Clinical Repository for storage, making them available to the authorised services within the SmartCHANGE platform.

Data Collection Flow 3:

1. The HCP collects participant's data through the Web App, including questionnaire
2. The Web App converts this data to the corresponding FHIR resources and sends it to the Clinical Repository for storage, making them available to the authorised services within the SmartCHANGE platform.

6.3.2 Risk Assessment

The HCP requests a risk prediction through the Webapp. The XCDS Engine acts as a broker to collect the clinical data needed for the prediction and requests the risk to the Local Trustworthy AI Framework services. The Explainer service orchestrates the collection of the data from the Risk Predictor and provides explanations to support the prediction. The following sequence diagram (Figure 7) depicts the interactions needed between the platform components to present the risk prediction to the HCP.

Data flows as follows:

1. Each time the HCP opens the health dashboard of a person, or completes an encounter, the webapp triggers the corresponding hook on the XCDSEngine.
2. The Engine retrieves the list of the services that are available using the list of XCDSService entries that are in the internal memory
3. For each service, using the information contained in the XCDSService object:
    a. Retrieves the clinical data needed for the prediction
    b. Sends a request to get the prediction to the RiskPredictor service.
    c. Receives the prediction and explanation

4. The engine combines the results of all services and sends them back to the web app.
5. The web app uses the information provided to render the webpage on risk and presents the information to the HCP.

Funded by
the European Union

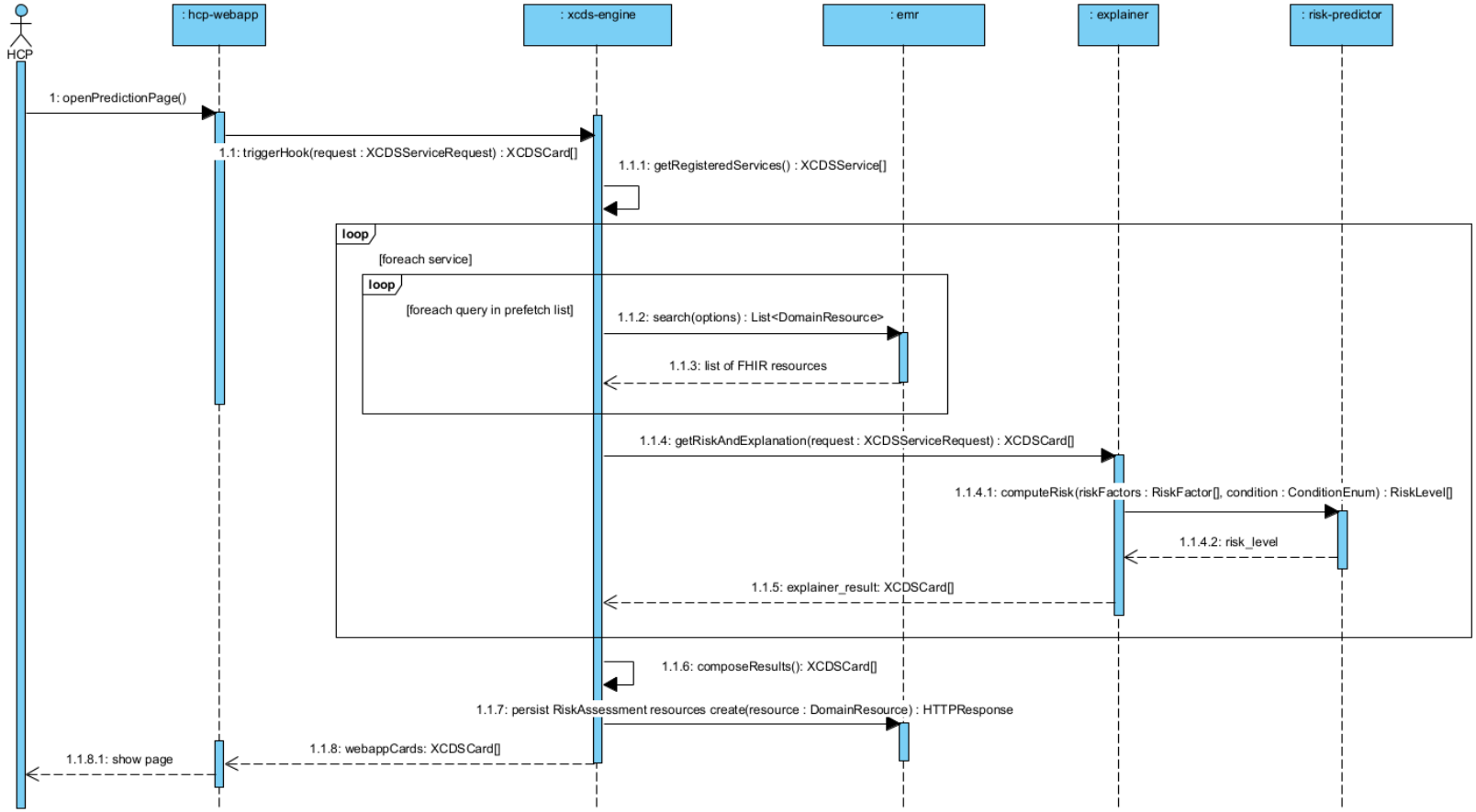

Figure 7 – Risk Prediction sequence diagram

Funded by
the European Union

6.3.3 Data Visualization and Goal Setting

The flow in Figure 8 illustrates the process of setting participant goals through the HCP's web application. The HCP consults the dashboard of the child or adolescents to assess their status. Using this information, they set goals to improve the person behaviour. The flow also demonstrates how the mobile app can access the goals defined by the HCPs to create personalized health challenges.

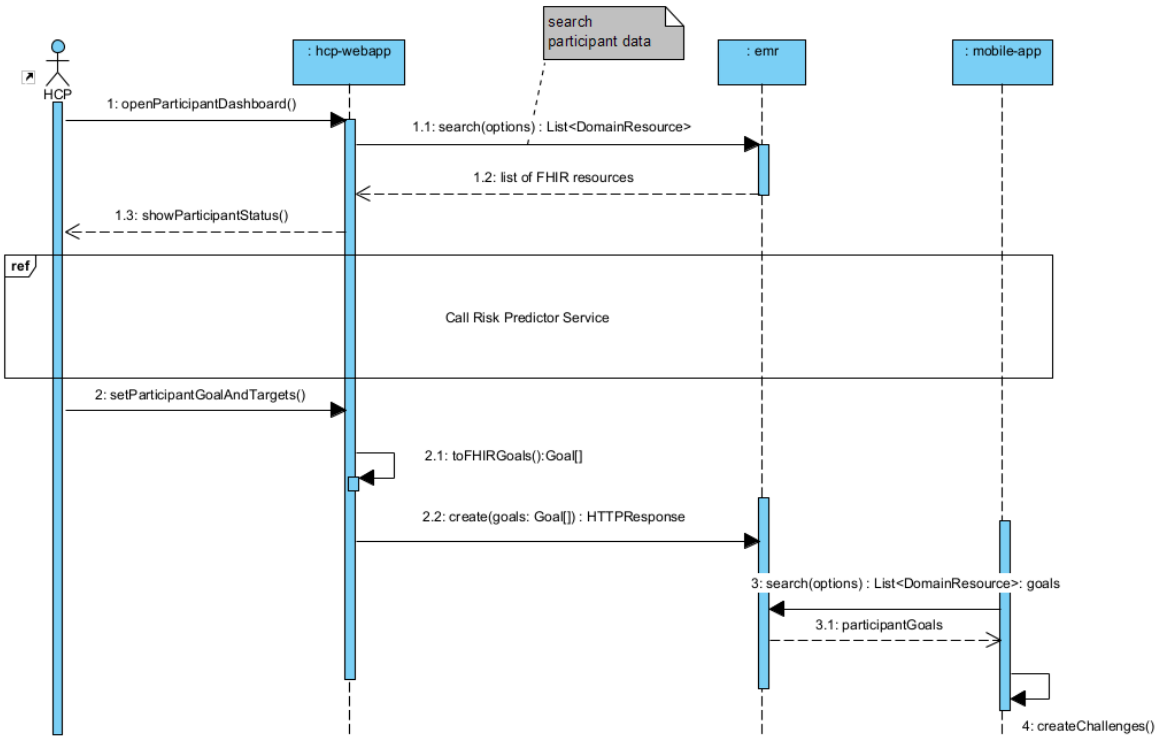

Figure 8 – Data Visualization and Goal Setting Sequence diagram

1. Each time the HCP selects an individual, the Web application searches for the health data for that person and retrieves a list of FHIR resources.
2. The webapp organizes the health data and presents them to the HCP.
3. The HCP visualizes the Risk information of the individual (for details refer to the flow in section 6.3.2)
4. The HCP sets goals and targets using the web app
5. The web app converts these goals to FHIR-compliant goals and sends them to the clinical repository for storage
6. The mobile app can retrieve the updated goals accessing the Clinical Repository
7. The mobile app uses the retrieved goals to create challenges for the individual.

6.3.4 Federated Learning

The diagram in Figure 9 describes how the central FL Server coordinates the FL Clients running in each pilot site, allowing the FL Clients to train models without sharing raw data externally. The FL Server aggregates these updates, enhancing the system's predictive capabilities.

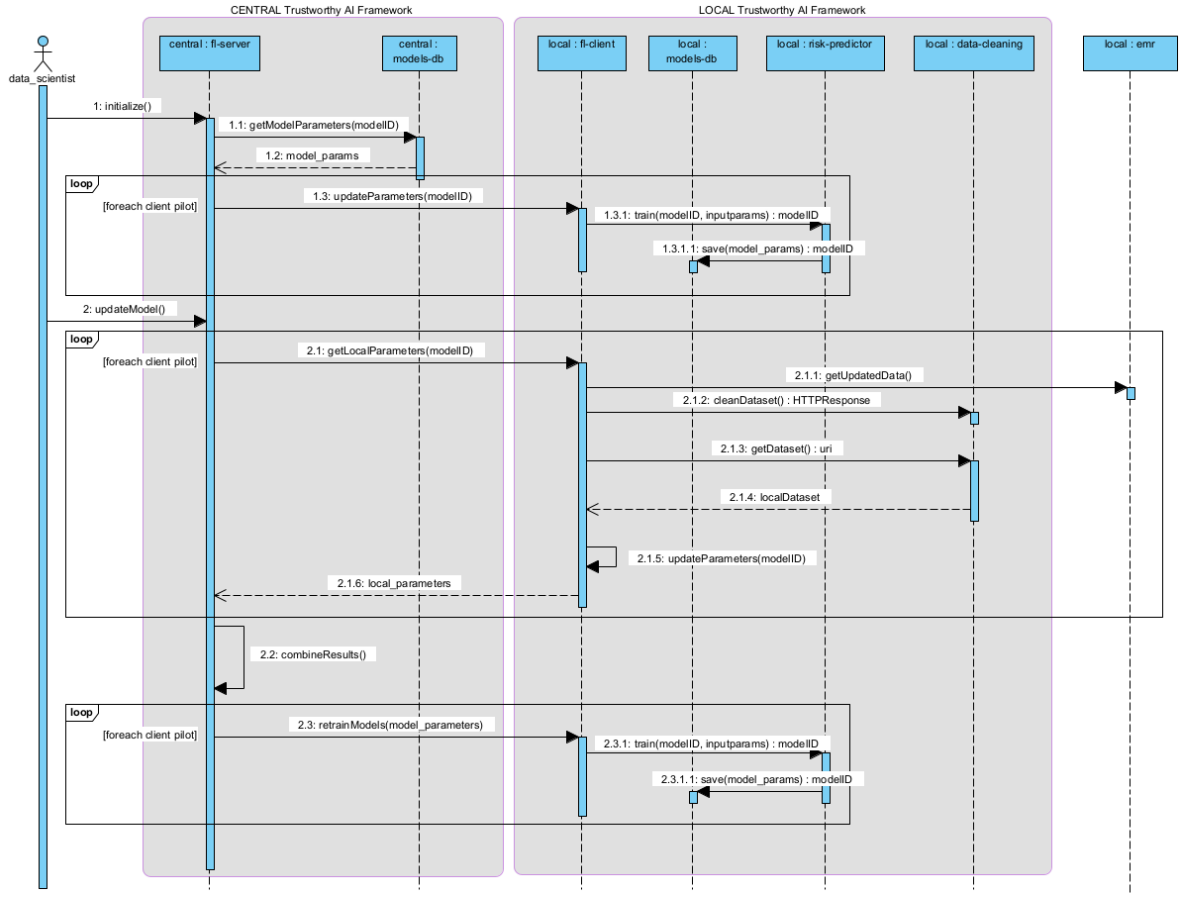

Figure 9 - Federated Learning Sequence Diagram

The diagram also highlights the interaction with the clinical repository (EMR in the diagram) and the Data Cleaning service.

For details on the mechanisms of federated learning refer to deliverable D5.2[20].

[20] SmartCHANGE Consortium. Deliverable D5.2 Initial trustworthy AI suite

The top part of this sequence diagram represents the first training phase (initialization) of the models across clients:

1. The Federated Learning (FL) Server sends the model parameters to the FL clients.
2. Each client passes the global parameters to its Prediction Services, which fine tune their models using local data and save the results.

The FL Server also manages the collection of updated parameters from each client:

1. The FL Server requests updated parameters from the clients.
2. Each FL Client retrieves new clinical data from the local repository and uses the Data Cleaning Service to clean the data, producing a usable dataset.
3. The client then updates its local model parameters and sends them to the FL Server.
4. Once the server receives all local updates, it computes the globally updated model parameters and initiates local model updates.

7 Deployment and testing

The Initial Integration and Test Plan (D6.6[21]) establishes a comprehensive framework for software development, integration, and testing within the SmartCHANGE project. This plan introduces a three-tiered environment strategy, encompassing Development, Test, and Production stages, to ensure stability and integrity throughout the development lifecycle. More in details:

- **Development Environment -** This is where developers create and test code. It's a flexible environment for experimentation and debugging.
- **Test Environment -** This environment replicates the production environment as closely as possible, allowing for rigorous testing of the application under realistic conditions.
- **Production Environment -** This is the live environment where the application is deployed for end-users. It should be highly stable and secure.

Central to this approach is the mandated use of Docker [47] as a popular platform for containerization, which involves packaging applications and their dependencies into a single unit (a container) that can be run consistently across different environments. This has several advantages:

- **Portability:** Containers can be easily moved between different machines and environments.
- **Consistency:** The same container will behave the same way regardless of the underlying infrastructure.
- **Efficiency:** Containers are lightweight and start up quickly.

In the SmartCHANGE project, using Docker for containerizing microservices facilitates consistent deployment across various environments and pilot sites, reducing the risk of compatibility issues. A schematic representation of the approach is schematically presented in Figure 10.

[21] SmartCHANGE Consortium. Deliverable D6.6 Initial Integration and Test Plan

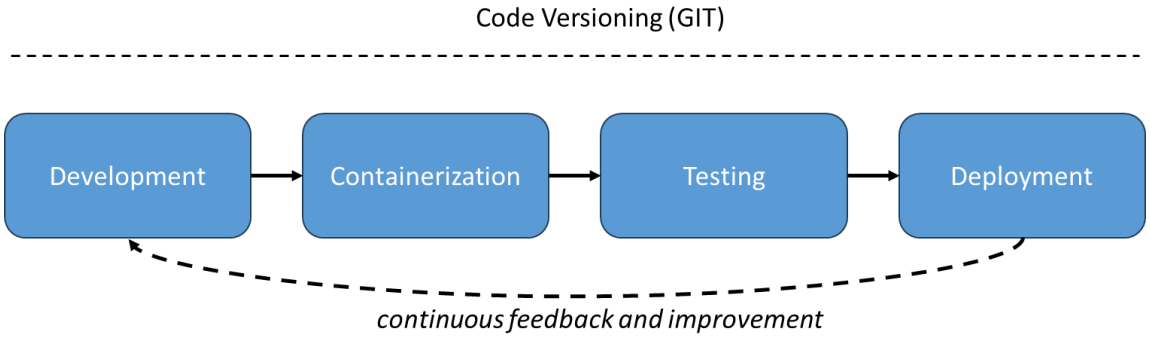

Figure 10 - Development to deployment cycle, based on continuous feedback and improvement.

For effective code management, the plan adopts GitHub [49] as the central code repository and container registry, implementing a specific branching strategy aligned with the project's environmental structure. By using GitHub as the central code repository, the followings will be ensured:

- **Track changes -** Every change to the code is recorded, making it easy to revert to previous versions if necessary.
- **Collaborate -** Multiple developers can work on the same project simultaneously.
- **Manage issues -** Bugs and feature requests can be tracked and assigned to specific developers.

Moreover, to support uniform deployment across pilot sites, Deliverable 6.6[22] provides a detailed blueprint for virtual machine configuration capable of running the containerized microservices. Additionally, API documentation is streamlined through the implementation of Swagger [48], generating interactive and up-to-date documentation that enhances communication and accessibility for all stakeholders. The plan also outlines a comprehensive testing strategy spanning different environments, coupled with a structured bug reporting system utilizing GitHub's issue tracking features.

In Figure 11, it becomes visible that for each user in a different pilot site, their data is saved and processed on a server owned by the respective pilot site.

[22] SmartCHANGE Consortium. Deliverable D6.6 Initial Integration and Test Plan

Moreover, in D6.6[23] an integration roadmap that outlines a phased approach for component integration, progressing from stub implementation to final system delivery is also presented. This ensures a smooth transition from development to production stages. Collectively, these elements form a robust framework for managing the technical aspects of the SmartCHANGE project, promoting efficient development, seamless integration, and reliable deployment of the project's software components.

Ultimately, in order to support the 2-phase pilot plan, the deployment strategy - here broadly reported - will also undergo 2 cycles. Hence, D6.6 will have a final version that is planned to be released at month 24.

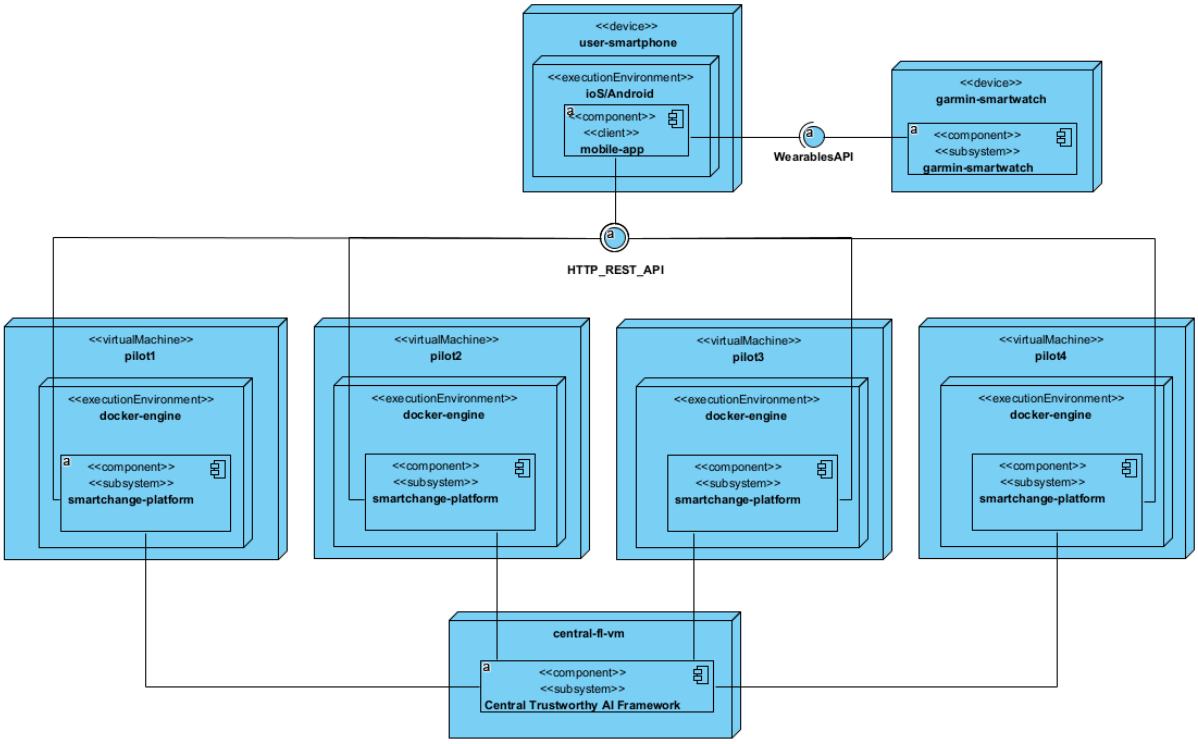

Figure 11 – Deployment diagram showing the interaction of one mobile application with the pilot locations, where data is saved.

[23] SmartCHANGE Consortium. Deliverable D6.6 Initial Integration and Test Plan

8 Conclusions and next steps

This document reports on the technical specifications for the SmartCHANGE solution, aiming to be a platform designed to reduce the long-term risks of NCDs through the innovative application of artificial intelligence. The solution's architecture is built upon a microservices paradigm and incorporates both functional and non-functional requirements, ensuring that the system not only meets its intended objectives but also addresses critical aspects such as privacy, security, and user experience.

As detailed in previously submitted deliverables, the project has prioritized a user-centric approach that resulted into the design of two innovative tools: the HCP Web App and a Mobile App for families/children and youth. These tools are tailored to meet the specific needs of HCP and individuals, respectively, providing them with valuable insights and personalized recommendations.

Moreover, the solution is being built on a robust Trustworthy AI Framework, ensuring that the AI models are transparent, explainable, and reliable. This framework aims to incorporate federated learning techniques to enhance model performance as well as protecting data privacy.

Importantly, the document provides insights on how the SmartCHANGE solution intends to protect its data, users and resources, following the constraints posed by the GDPR regulations and a privacy by design approach. Security and privacy are crucial measures to be considered throughout the development phase. Indeed, the findings of a comprehensive system/solution safety risk assessment conducted in accordance with the ENISA guidelines are presented. These will inform the specific security measures that will be implemented throughout the design, development, and maintenance phases of the project.

In addition to this, a detailed description of the main components of the system architecture and their dependencies is provided in this document. In particular, the adoption of a microservices-based architecture whose design is driven by the principle of scalability, flexibility, and efficient integration of the various components within the SmartCHANGE solution is described in detail.

In conclusion, by focusing on user-centric design, robust system architecture, and stringent privacy measures this deliverable will guide the development of the initial version of the SmartCHANGE solution, which will be tested in the first pilot trials.

Funded by
the European Union

Moving forward, feedback from these trials, along with progress in the technical work packages, will inform a second iteration of the design and development. During this phase, user requirements will be refined, and the prototypes for the HCP web app and mobile app will be further advanced. Any change resulting from updated requirements will be incorporated into a second version of the technical specifications, which will include revisions to the components' descriptions and data model, towards the second version of the SmartCHANGE solution.

Funded by
the European Union

9 References

[1] FHIR R5: https://www.hl7.org/fhir/r5/
[2] SNOMED-CT: SNOMED CT - Home (ihtsdotools.org)
[3] LOINC: https://loinc.org/
[4] ICD: International Classification of Diseases (ICD) (who.int)
[5] Observation FHIR R5: Observation - FHIR v5.0.0 (hl7.org)
[6] RiskAssessment FHIR R5: RiskAssessment - FHIR v5.0.0 (hl7.org)
[7] Patient FHIR R5: Patient - FHIR v5.0.0 (hl7.org)
[8] Questionnaire FHIR R5: Questionnaire - FHIR v5.0.0 (hl7.org)
[9] QuestionnaireResponse FHIR R5: QuestionnaireResponse - FHIR v5.0.0 (hl7.org)
[10] Device FHIR R5: Device - FHIR v5.0.0 (hl7.org)
[11] Practitioner FHIR R5: Practitioner - FHIR v5.0.0 (hl7.org)
[12] PractitionerRole FHIR R5: PractitionerRole - FHIR v5.0.0 (hl7.org)
[13] GDPR: https://eur-lex.europa.eu/eli/reg/2016/679/oj
[14] Personal Data: https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en
[15] ENISA official site: ENISA (europa.eu)
[16] ENISA data pseudonymisation: Recommendations on shaping technology according to GDPR provisions - An overview on data pseudonymisation — ENISA (europa.eu)
[17] ENISA data protection: Recommendations on shaping technology according to GDPR provisions - Exploring the notion of data protection by default — ENISA (europa.eu)
[18] Risk management: https://www.enisa.europa.eu/publications/risk-management-for-data-protection
[19] Keycloak: https://www.keycloak.org/
[20] Fenoglio, D., Dominici, G., Barbiero, P., Tonda, A., Gjoreski, M., & Langheinrich, M. (2024). Federated Behavioural Planes: Explaining the Evolution of Client Behaviour in Federated Learning. arXiv. https://arxiv.org/abs/2405.15632
[21] Pillutla, K., Kakade, S. M., & Harchaoui, Z. (2022). Robust Aggregation for Federated Learning. IEEE Transactions on Signal Processing, 70, 1142-1154. https://doi.org/10.1109/TSP.2022.3153135
[22] Yin, D., Chen, Y., Kannan, R., & Bartlett, P. (2018). Byzantine-robust distributed learning: Towards optimal statistical rates. In International Conference on Machine Learning (pp. 5650-5659). PMLR.
[23] Blanchard, P., El Mhamdi, E. M., Guerraoui, R., & Stainer, J. (2017). Machine Learning with Adversaries: Byzantine Tolerant Gradient Descent. In Advances in Neural Information Processing Systems (Vol. 30). Curran Associates, Inc.
[24] Hu, H., Zhang, X., Salcic, Z., Sun, L., Choo, K.-K. R., & Dobbie, G. (2023). Source Inference Attacks: Beyond Membership Inference Attacks in Federated Learning. IEEE Transactions on Dependable and Secure Computing, 1-18. https://doi.org/10.1109/TDSC.2023.3321565

Funded by the European Union

[25] Yin, H., Zhang, L., Xu, Y., Fan, L., & Ren, K. (2021). See Through Gradients: Image Batch Recovery via Gradinversion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[26] Geiping, J., Bauermeister, H., Dröge, H., & Moeller, M. (2020). Inverting Gradients - How Easy Is It to Break Privacy in Federated Learning? In Proceedings of the 34th International Conference on Neural Information Processing Systems (Article No. 1421, pp. 1-11). Curran Associates Inc.

[27] Jeon, J., Kim, H., Shin, J., & Moon, S. (2021). Gradient Inversion with Generative Image Prior. Advances in Neural Information Processing Systems, 34, 29898-29908.

[28] Beutel, D. J., Topal, T., Mathur, A., Qiu, X., Fernandez-Marques, J., Gao, Y., Sani, L., Kwing, H. L., Parcollet, T., Gusmão, P. P. B. de, & Lane, N. D. (2020). Flower: A Friendly Federated Learning Research Framework. arXiv preprint arXiv:2007.14390.

[29] Dwork, C. (2006). Differential Privacy. In International Colloquium on Automata, Languages, and Programming (pp. 1-12). Springer.

[30] Sun, L., Qian, J., & Chen, X. (2021). LDP-FL: Practical Private Aggregation in Federated Learning with Local Differential Privacy. In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization.

[31] Wang, Z., Yang, Z., Azimi, I., & Rahmani, A. M. (2024). Differential Private Federated Transfer Learning for Mental Health Monitoring in Everyday Settings: A Case Study on Stress Detection. arXiv preprint arXiv:2402.10862.

[32] Miao, Y., Xie, R., Li, X., Liu, X., Ma, Z., & Deng, R. H. (2022). Compressed Federated Learning Based on Adaptive Local Differential Privacy. In Proceedings of the 38th Annual Computer Security Applications Conference (pp. 159-170).

[33] Xia, Y., Hofmeister, C., Egger, M., & Bitar, R. (2024). Byzantine-Resilient Secure Aggregation for Federated Learning Without Privacy Compromises. arXiv preprint arXiv:2405.08698.

[34] Mohassel, P., & Zhang, Y. (2017). SecureML: A System for Scalable Privacy-Preserving Machine Learning. In 2017 IEEE Symposium on Security and Privacy (SP) (pp. 19-38). https://doi.org/10.1109/SP.2017.12

[35] Ma, Z., Ma, J., Miao, Y., Li, Y., & Deng, R. H. (2022). ShieldFL: Mitigating Model Poisoning Attacks in Privacy-Preserving Federated Learning. IEEE Transactions on Information Forensics and Security, 17, 1639-1654. https://doi.org/10.1109/TIFS.2022.3169918

[36] Fang, H., & Qian, Q. (2021). Privacy Preserving Machine Learning with Homomorphic Encryption and Federated Learning. Future Internet, 13(4), 94. MDPI.

[37] Ma, J., Naas, S.-A., Sigg, S., & Lyu, X. (2022). Privacy-Preserving Federated Learning Based on Multi-Key Homomorphic Encryption. International Journal of Intelligent Systems, 37(9), 5880-5901. Wiley Online Library.

[38] Korff, Douwe, GDPR Requirements on Data Protection Impact Assessments & Methodologies for DPIAs (July 20, 2020). Available at SSRN: https://ssrn.com/abstract=3656234 or http://dx.doi.org/10.2139/ssrn.3656234.

[39] Aven, T., Andersen, H.B., Cox, T., Droguett, E.L., Greenberg, M., Guikema, S., Kroger, W., McComsa, K., Renn, O., Thompson, K.M., Zio, E., (2018). Risk Analysis: Fundamental Principles. (link: Risk Analysis Fundamental Principles from the Society for Risk Analysis - Society for Risk Analysis (sra.org))
[40] MDCG 2019-16 Rev.1 Guidance on Cybersecurity for medical devices: b23b362f-8a56-434c-922a-5b3ca4d0a7a1_en
[41] CDS Hooks Specification: Current (draft) - CDS Hooks (cds-hooks.org)
[42] Træfik: Traefik Proxy Documentation - Traefik
[43] FHIR RESTful API: Http - FHIR v5.0.0 (hl7.org)
[44] Zipkin: OpenZipkin · A distributed tracing system
[45] CIA triad – 1: The CIA Triad: Key to Modern Cybersecurity Strategies (veeam.com)
[46] CIA triad – 2: What Is the CIA Triad? (f5.com)
[47] Docker: Docker: Accelerated Container Application Development
[48] Swagger: API Documentation & Design Tools for Teams | Swagger
[49] Github: GitHub: Let's build from here · GitHub
[50] Condition FHIR R5: Condition - FHIR v5.0.0 (hl7.org)
[51] Organization FHIR R5: Organization - FHIR v5.0.0 (hl7.org)
[52] NutritionIntake FHIR R5: NutritionIntake - FHIR v5.0.0 (hl7.org)
[53] Goal FHIR R5: Goal - FHIR v5.0.0 (hl7.org)
[54] Encounter FHIR R5: Encounter - FHIR v5.0.0 (hl7.org)
[55] Grafana Loki: Grafana Loki | Grafana Loki documentation
[56] Grafana: Grafana OSS and Enterprise | Grafana documentation
[57] Prometheus: Overview | Prometheus
[58] CarePlan FHIR R5: CarePlan - FHIR v5.0.0
[59] Family Relationship in FHIR: https://www.hl7.org/fhir/patient.html#maternity
[60] NIST impact level: "Guide for Mapping Types of Information and Information Systems to Security Categories," NIST Special Publication 800-60, vol. 1, 2008. [Online]. Available: https://csrc.nist.gov/publications/detail/sp/800-60/rev-1/final
[61] RabbitMQ: RabbitMQ: One broker to queue them all | RabbitMQ
[62] Newman, S. (2015). Building Microservices: Designing Fine-Grained Systems. O'Reilly Media.

Appendix 1. Data model to HL7-FHIR R5 mapping rules and terminologies

The appendix document lists three types of tables, each addressing a different aspect of the profile:

- The **FHIR Resources** section details how general concepts managed by the SmartCHANGE platform are mapped in FHIR resources, detailing which of the optional fields will be used and any constraints specific to the project.
- The **FHIR / Health Data Mapping** section maps the health concepts used in this project (clinic variables collected by the application and used in the risk predictions) to FHIR, adding details on the codes used for each resource type and the admissible values.
- The **FHIR-SC-VALUESETS** tables list all the codes that are used in the FHIR / HEALTH DATA MAPPING tables, either from standard vocabularies or custom to the SmartCHANGE project.

A1.I. FHIR Resources

The tables in this section are structured as follows:

- the *Attribute* column corresponds to the *Items* column of the FHIR resource.
- the *Type* column indicates the FHIR data type that corresponds to the data type of the original variable.
- the *Description* column provide a textual description.
- The *Constraint* column details any conditions or constraints that need to be met.
- the *FHIR Attribute Mapping* column specifies the item of the resource that corresponds to the original variable.
- the *FHIR Note* column provides additional information.

Funded by the European Union

RISK PREDICTION (FHIR::RISKASSESSMENT)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
identifier		an identifier assigned to the specific assessment	Required (1..1) **use**: is a code that specifies the purpose of the assessment. Allowed values are: usual, official, temp, secondary, old (If known)	RiskAssessment.identifier	we can assume to use the '**use: official'**
status	Code	The status of the result value	Required (1..1) status = "final"	RiskAssessment.status	we can assume that the status of the risk assessment is complete.
subject	Reference	represent who and/or what the observation is about	Required (1..1) reference to Patient resource, i.e. **Patient/id**	RiskAssessment.subject	we can assume that the subject of the risk assessment is the specific participant
encounter	Reference	where the assessment was performed	reference to Encounter resource, i.e. **Encounter/id**	RiskAssessment.encounter	
occurence[x]	DateTIme	Date consists of Date and/or time.		RiskAssessment.occurence.occurenceDateTime	we can assume that the risk assessment occurs in a specific date time
prediction	BackboneElement	it is a composite element. For more details see the FHIR specification.		RiskAssessment.prediction.outcome RiskAssessment.prediction.probability.probabilityDecimal RiskAssessment.prediction.when.whenRange	we can assume to use the (i) outcome also to include a code for identifying the potential predicted condition, (ii) probabilityDecimal and (iii) whenRange

Funded by the European Union

BEHAVIOUR CHANGE PLAN (FHIR::CAREPLAN)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
Identifier	Identifier	Business identifier for the plan	Required (1..1)	CarePlan.identifier	Represents an external id for the plan
status	code	The status of the plan throughout its lifecycle	Required (1..1)	CarePlan.status	Based on FHIR careplan request status code -> draft | active | on-hold | revoked | completed | entered-in-error | unknown
intent	Code	Indicates the level of authority/intentionality associated with the care plan and where the care plan fits into the workflow chain	Required (1…1)	CarePlan.intent	One of proposal | plan | order | option | directive
related risk predictions	Reference	The reference to the RiskAssessment resources that were consulted	Required (1…*)	CarePlan.supportingInfo	Profiling from the general FHIR resource, binds to Reference to RiskAssessment resources
encounter	Reference	Reference to the encounter when this plan was defined	Optional (0…1)	CarePlan.encounter	
subject	Reference	The person this plan refers to	Required (1…1)	CarePlan.subject	
contributor	Reference	Who provided the content to this plan. Usually, the HCP	Optional (0…1)	CarePlan.contributor	

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
custodian	Reference	Who is responsible to supervise the achievement. It can the individual itself (Patient resource) or the parent (RelatedPerson resource)	Optional (0…1)	CarePlan.custodian	
goals	Reference	List of goals that are part of this plan	Required (1…*)	CarePlan.goal	For the details of goals
activities	Planned Activities	The list of activities included in the plan	Optional (1…*)	CarePlan.activity	They may refer to the challenges from the mobile app

PLANNED ACTIVITIES (FHIR::CAREPLAN.ACTIVITY)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
code	Code	The concept describing the activity	Optional (0…1)	CarePlan.activity.performedActivity.code	
activity	Reference	The reference to the FHIR resource describing the activity	Required (1…1)	CarePlan.activity.performedActivity.reference	
progress	Annotation	Annotation containing info on the progress	Optional (0…*)	CarePlan.activity.progress	Each can contain the author, the time and a text describing the progress

Funded by the European Union

GOAL (FHIR::GOAL)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
identifier	Identifier	Business identifier for the goal	Required (1..1)	Goal.identifier	Represents an external id for the goal
goal status	code	The status of the goal throughout its lifecycle	Required (1..1)	Goal.lifecycleStatus	Based on FHIR goal status code -> proposed, planned, accepted, active, on-hold, completed, cancelled, entered-in-error, rejected
achievementStatus	CodeableConcept	Describes the progression, or lack thereof, towards the goal against the challenge	Required (1..1)	Goal.achievementStatus	Based on FHIR goal valueset -> in-progress, improving, worsening, no-change, achieved, sustaining, not-achieved, no-progress, not-attainable
description	CodeableConcept	Coded description of the desired objective	Required (1..1)	Goal.description	
subject	Reference	The person this goal refers to	Required (1…1)	Goal.subject	The goal might refer to the Mother / father or Family Group
addresses	Reference	Reference to the RiskAssessment	Optional (0…*)	Goal.addresses	
start	Date	Start date for this goal	Optional (0…*)	Goal.start	
target	Goal Target	List of targets included in the goal	Optional (0…*)	Goal.target	

Funded by the European Union

GOAL TARGET (FHIR::GOAL::TARGET)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
measure	CodeableConcept		Required (1…1)	Goal.target.measure	
detail	Boolean | quantity | number | Range | string | code	The target value to be achieved	Optional (0…1)	Goal.target.detail	Depending on the measure, use the appropriate detail type
due	Date | duration	Due date or period of time before achieving it	Optional (0…1)	Goal.target.due	

CHILD OR ADOLESCENT (FHIR::PATIENT)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
identifier	Identifier	Business identifier for the participant	Required (1..*)	Patient.identifier	Represents the participant identifier Multiple identifiers can be used to integrate external systems (e.g., slofit)
gender	CodeableConcept	The gender of the participant	Required (1..1)	Patient.gender	Based on FHIR Administrative Gender. -> male, female, unknown, other
birthdate	Date	The participant's date of birth	Required (1..1)	Patient.birthDate	Year and month of birth could be used for privacy
communication	CodeableConcept	Preferred language of the participant	Optional (0..*)	Patient.communication.language	Can include multiple languages

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
email	String	The main email address of the individual.	Required (1..1) telecom.system = 'email' telecom.value = email address	Patient.telecom	
general practitioner	Reference	Reference to the participant's general practitioner	Optional (0..*)	Patient.generalPractitioner	Can include multiple practitioners
deceased	Boolean	Whether the participant is deceased	Optional (0..1)	Patient.deceasedBoolean	True if the participant has passed away
multiple birth	Boolean/Integer	Indicator if participant was part of a multiple birth	Optional (0..1)	Patient.multipleBirthBoolean	Can also include the birth order if applicable
managing organization	Reference	Organization managing the participant's health information	Optional (0..1)	Patient.managingOrganization	Reference to the organization in charge of the participant (e.g., school or healthcare institution)

FAMILY MEMBER (FHIR::PATIENT)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
identifier	Identifier	Business identifier for the family member	Required (1..*)	Patient.identifier	Represents the family member 's unique ID
gender	CodeableConcept	The gender of the family member	Required (1..1)	Patient.gender	Based on FHIR Administrative Gender. -> male, female, unknown, other
birthdate	Date	The family member's date of birth	Required (1..1)	Patient.birthDate	Year and month of birth could be used for privacy

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
communication	CodeableConcept	Preferred language of the family member	Optional (0..*)	Patient.communication.language	Can include multiple languages
email	String	The main email address of the individual.	Required (1..1) telecom.system = 'email' telecom.value = email address	Patient.telecom	
general practitioner	Reference	Reference to the family member's general practitioner	Optional (0..*)	Patient.generalPractitioner	Can include multiple practitioners
marital status	CodeableConcept	The marital status of the family member	Optional (0..1)	Patient.maritalStatus	Used in case of family members
deceased	Boolean	Whether the family member is deceased	Optional (0..1)	Patient.deceasedBoolean	True if the family member has passed away
link	Reference/RelatedPerson	Link to the child or adolescent	Required (1..1) Patient.link.type='see-also' Patient.link.other=Reference/RelatedPerson	Patient.link.other	Reference to RelatedPerson linking to the child

FAMILYMEMBER RELATIONSHIP (FHIR::RELATEDPERSON)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
patient	Reference	The participant the family member is related to	Required (1..1) reference to Patient resource, i.e. **Patient/id**	RelatedPerson.patient	Reference to Patient resource linking to the family member

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
relationship	CodeableConcept	The relationship of the related person to the participant	Required (1..1) FAMMEMB (see https://terminology.hl7.org/5.1.0/CodeSystem-v3-RoleCode.html)	RelatedPerson.relationship	If the family member is a parent, sibling or other relative

MEDICAL HISTORY (FHIR::CONDITION)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
identifier	Identifier	Business identifier for the condition	Optional (0..*)	Condition.identifier	A unique identifier that may be assigned to the condition.
clinicalStatus	CodeableConcept		Required (1..1)	Condition.clinicalStatus	
verificationStatus	CodeableConcept	The verification status (e.g., confirmed, unconfirmed)	Required (1..1)	Condition.verificationStatus	Describes if the condition is confirmed, differential, or refuted.
subject (patient)	Reference (Patient)	The participant associated with the condition	Required (1..1) reference to Patient resource, i.e. **Patient/id**	Condition.subject	Refers to the participant involved in the condition.
encounter	Reference	Encounter during which the condition was asserted	Optional (0..1)	Condition.encounter	Refers to the encounter during which the condition was asserted.
practitioner	Reference (Practitioner)	The practitioner associated with the condition		Condition.participant.actor	The practitioner who is asserting the condition.
code	CodeableConcept	The code representing the condition (disease/diagnosis)		Condition.code	SNOMED or ICD codes for the condition (e.g., diabetes).
severity	CodeableConcept	The severity of the condition		Condition.severity	Indicates the severity level of the condition.

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
active	Boolean	Whether the condition is active		Condition.clinicalStatus	Whether the condition is currently active.
period	Period	The time period the condition was present		Condition.onset[x], Condition.abatement[x]	Defines the period when the condition occurred or was resolved.
meta	Meta	Metadata about the resource		Condition.meta	Metadata related to the resource, including version ID and last updated timestamp.

ORGANIZATION (FHIR::ORGANIZATION)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
identifier	Identifier	Business identifier for the organization		Organization.identifier	Identifies the organization uniquely.
name	String	Full name of the organization		Organization.name	Stores the full name of the organization.
type	CodeableConcept	The type of organization	Required (1..1)	Organization.type	
contact	ExtendedContactDetail	Contact information for the organization		Organization.telecom	Contact details, such as phone or email.

ENCOUNTER (FHIR::ENCOUNTER)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
status	Code	Describes the current status of the encounter	Required (1..1) status = 'completed'	Encounter.status	The status of the encounter is set to completed.

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
subject	Reference	Reference to participant involved	Required (1..1) it carries the reference to Patient resource, i.e. **Patient/id**	Encounter.subject	Refers to the participant involved in the encounter.
period	Period	Time period of the encounter		Encounter.actualPeriod.start, Encounter.actualPeriod.period.end	Represents the time period of the encounter.
type	CodeableConcept	Type of encounter		Encounter.type	The type of the encounter, such as consultation, check-up.
reason	BackboneElement	Reason for the encounter		Encounter.reason.value	The reason for the encounter.
practitioner	Reference	The practitioner who participated in the encounter	Required (1..1) Reference to the HCP reording the encounter	Encounter.participant.actor	

FAMILY HISTORY (FHIR::FAMILYMEMBERHISTORY)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
identifier	Identifier	Business identifier for the family history	Optional (0..*)	FamilyMemberHistory.identifier	A unique identifier assigned to the family history record.
status	CodeableConcept	The clinical status of the family history record	Required (1..1)	FamilyMemberHistory.status	The status of the family history record (e.g., active, completed).
subject	Reference	Reference to the participant whose family history is described	Required (1..1)	FamilyMemberHistory.patient	The participant whose family history is being described.

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
relationship	CodeableConcept	Relationship to the participant (e.g., parent, sibling)	Required (1..1)	FamilyMemberHistory.relationship	The relationship of the family member to the participant.
born	Period/Date/String	The date of birth or estimated age of the family member	Optional (0..1)	FamilyMemberHistory.born[x]	The birth date or age of the family member.
condition	BackboneElement	A list of conditions that the family member has	Optional (0..*)	FamilyMemberHistory.condition	Conditions that the family member had, affecting the family history of the participant.
deceased	Boolean/CodeableConcept/Date	Whether the family member is deceased and cause	Optional (0..1)	FamilyMemberHistory.deceased[x]	Specifies if and how the family member died.
reason	CodeableReference	Reason for capturing family history	Optional (0..*)	FamilyMemberHistory.reason	Reasons for recording the family history (e.g., genetic risk).
note	Annotation	General notes about the family history	Optional (0..*)	FamilyMemberHistory.note	Any additional comments about the family history.
meta	Meta	Metadata about the resource	Optional	FamilyMemberHistory.meta	Metadata such as version ID, last updated timestamp, etc.

PRACTITIONER (FHIR::PRACTITIONER)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
identifier	Identifier	Business identifiers for the practitioner	Optional (0..*)	Practitioner.identifier	A unique identifier assigned to the practitioner.
name	HumanName	The name(s) associated with the practitioner	Optional (0..*)	Practitioner.name	Human names associated with the practitioner.

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
address	Address	Address(es) associated with the practitioner	Optional (0..*)	Practitioner.address	The addresses associated with the practitioner.
qualification	CodeableConcept	The practitioner qualification	Optional (0..1)	Practitioner.qualification.code	Code representing the qualification of the practitioner.
communication	CodeableConcept	Communication preferences	Optional (0..*)	Practitioner.communication	Languages or communication preferences for the practitioner.
email	String	Email address to contact the practitioner	Required (1..1) Practitioner.telecom.system='email'	Practitioner.telecom.value	Valid email address

PRACTITIONER ROLE (FHIR::PRACTITIONERROLE)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
identifier	Identifier	Business identifiers for the practitioner's role	Optional (0..*)	PractitionerRole.identifier	A unique identifier assigned to the role of the practitioner.
period	Period	The time period this role is active	Optional (0..1)	PractitionerRole.period	Time period during which the role is active.
practitioner	Reference	Reference to the practitioner	Required (1..1)	PractitionerRole.practitioner	Reference to the practitioner who holds this role.
organization	Reference	The organization where the practitioner provides services	Optional (0..1)	PractitionerRole.organization	The organization the practitioner is associated with in this role.
code	CodeableConcept	The practitioner role code (e.g., surgeon, physician)	Optional (0..*)	PractitionerRole.code	Code representing the role of the practitioner.

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
specialty	CodeableConcept	The area of specialization of the practitioner	Optional (0..*)	PractitionerRole.specialty	The area of specialty in which the practitioner operates (e.g., Cardiology).
location	Reference	The locations where the practitioner provides services	Optional (0..*)	PractitionerRole.location	The locations associated with the practitioner role.
healthcareService	Reference	The healthcare services provided under this role	Optional (0..*)	PractitionerRole.healthcareService	The healthcare services offered by the practitioner in this role.
contact	ExtendedContactDetail	Contact details associated with the practitioner role	Optional (0..*)	PractitionerRole.contact	Contact points for the practitioner role (e.g., phone, email).

MEDICATION (FHIR::MEDICATION)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
identifier	Identifier	Business identifiers for the medication	Optional (0..*)	Medication.identifier	A unique identifier assigned to the medication.
status	Code	The status of the medication (active, inactive)	Required (1..1)	Medication.status	Indicates whether the medication is active, inactive, or entered in error.
manufacturer	Reference	Manufacturer of the medication	Optional (0..1)	Medication.marketingAuthorizationHolder	Organization responsible for manufacturing the medication.
form	CodeableConcept	Form of the medication (e.g., tablet, liquid)	Optional (0..1)	Medication.doseForm	The form in which the medication is administered (e.g., capsule, powder).

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
ingredient	BackboneElement	Active ingredients of the medication	Optional (0..*)	Medication.ingredient.item Medication.ingredient.isActive	Information about the ingredients of the medication and if it is active in medication
amount	Ratio	Quantity of ingredient in the medication	Optional (0..1)	Medication.ingredient.strength	Strength of the ingredient in the medication.
batch	BackboneElement[0..1]	Information about a batch of the medication	Optional (0..1)	Medication.batch.lotNumber Medication.batch.expirationDate	Batch information for the medication (e.g., lot number, expiration date).
code	CodeableConcept	The medication code (e.g., RxNorm code)	Required (1..1)	Medication.code	The code representing the medication (e.g., RxNorm or ATC codes).
meta	Meta	Metadata about the resource	Optional	Medication.meta	Metadata such as version ID, last updated timestamp, etc.

MEDICATION STATEMENT (FHIR::MEDICATIONSTATEMENT)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
identifier	Identifier[0..*]	Unique identifiers for the medication statement	Optional (0..*)	MedicationStatement.identifier	A unique identifier for this medication statement.
status	Code	The status of the medication usage (active, completed, etc.)	Required (1..1)	MedicationStatement.status	Indicates whether the statement is active, completed, entered in error, etc.
category	CodeableConcept[0..1]	The category of the medication statement	Optional (0..1)	MedicationStatement.category	Category of medication usage (e.g., inpatient, outpatient, community).

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
subject	Reference (Patient)	The patient to whom the medication is related	Required (1..1)	MedicationStatement.subject	Reference to the subject who is taking the medication.
medication[x]	CodeableConcept/Reference	The medication being administered (could be a code or a reference to another resource)	Required (1..1)	MedicationStatement.medication[x]	The medication that is being taken or was taken by the subject.
effective[x]	DateTime/Period	The time or period during which the medication was taken	Optional (0..1)	MedicationStatement.effective[x]	Can be a single date/time or a period to specify when the medication was taken.
dateAsserted	DateTime	Date when the medication usage was asserted	Optional (0..1)	MedicationStatement.dateAsserted	Date on which the statement was asserted.
informationSource	Reference (Practitioner/Patient/RelatedPerson)	Source of information about the medication usage	Optional (0..1)	MedicationStatement.informationSource	Person who provided the information about the medication usage.
reasonCode	CodeableConcept[0..*]	Reason for taking the medication	Optional (0..*)	MedicationStatement.reasonCode	The reason for the medication (e.g., diabetes, hypertension).
reasonReference	Reference[0..*]	Condition or observation that justifies the medication	Optional (0..*)	MedicationStatement.reasonReference	A reference to the reason for the medication usage (e.g., a Condition resource).
dosage	BackboneElement[0..*]	Details of the medication dosage	Optional (0..*)	MedicationStatement.dosage	Details on how the medication was taken, including dosage, rate, and timing.
dosage.text	String	Text representation of the dosage	Optional (0..1)	MedicationStatement.dosage.text	Free text description of the dosage (e.g., “Take 1 tablet daily”).

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
dosage.timing	Timing	Timing information about when the medication is/was taken	Optional (0..1)	MedicationStatement.dosage .timing	How frequently the medication was taken.

QUESTIONNAIRE (FHIR::QUESTIONNAIRE)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
identifier	Identifier[0..*]	Unique identifier	Optional (0..*)	Questionnaire.identifier	A unique identifier assigned to the questionnaire.
status	Code	Publication status	Required (1…1) Status='active'	Questionnaire.status	
name	String	Name of the questionnaire	Requider(1…1)	Questionnaire.name	
code	CodeableConcept	Concept describing the questionnaire	Optional(0…1)	Questionnaire.code	The codes used will be in the FHIR/ health data mapping tables
questions	QuestionItem	See the QuestionItem table	Required [1…*]	Questionnaire.item	

QUESTION (FHIR::QUESTIONNAIRE::ITEM)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
id	String	Unique identifier of the question	Required	Questionnaire.item.linkId	Unique for item in the questionnaire
type	Code	Type of response expected	the type of response	Questionnaire.item.type	
text	String	The question	There must be at least one question per questionnaire	Questionnaire.item.text	
required	Boolean	Whether the question is required		Questionnaire.item.required	

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
repeats	Boolean	The question may repeat	True	Questionnaire.item.repeats	Whether the item may repeat
code	CodeableConcept	Concept describing the item		Questionnaire.item.code	
answerOptions	List of values	List of permitted answers	Required if type='coding' and answerValueSet is undefined	Questionnaire.item.answerOption[]	Used to constrain the possible answers
answerValueSet	URI	To use if a valueset of all permitted answers exist	Required if type='coding' and answerOptions is undefined	Questionnaire.item.answerValueset	
items	QuestionItem	Question group	Required if type = 'group'	Questionnaire.item.item	Used if this resource represents a group of questions (e.g. the section of a questionnaire, a group of related answers)

QUESTIONNAIRE RESPONSE (FHIR::QUESTIONNAIRERESPONSE)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
identifier	Identifier	Unique identifier for the response	Optional (0..*)	QuestionnaireResponse.identifier	A unique identifier assigned to the response
questionnaire	Reference	Reference to the questionnaire this response derives from	Required (1…1) it carries the reference to Questionnaire resource, i.e. **Questionnaire/id**	QuestionnaireResponse.questionnaire	
status	CodeableConcept	The status of the questionnaire response	Required (1..1)	QuestionnaireResponse.status	Active, completed, amended, etc.

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
subject (patient)	Reference	Reference to the subject of the questionnaire (i.e., participant)	Required (1…1) it carries the reference to Patient resource, i.e. **Patient/id**	QuestionnaireResponse.subject	The participant who is filling out the questionnaire
encounter	Reference	Encounter during which the questionnaire was completed	Optional (0..1)	QuestionnaireResponse.encounter	Links the questionnaire response to a particular encounter
authored	DateTime	Date when the questionnaire response was completed	Required (1..1)	QuestionnaireResponse.authored	The date and time of the response
author	Reference	The individual who completed the questionnaire	Optional (0..1)	QuestionnaireResponse.author	Could be the participant or a healthcare provider
source	Reference	The origin of the information	Optional (0..1)	QuestionnaireResponse.source	The source providing the information in the response
item	QuestionnaireAnswer	Grouped collection of questions and answers	Required (1..*)	QuestionnaireResponse.item	See the QuestionnaireAnswer structure for details

QUESTIONNAIRE ANSWER (FHIR::QUESTIONNAIRERESPONSE::ITEM)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
linkId	String	Required (1…) Refers to the item.linkid in the referenced Questionnaire resource		QuestionnaireResponse.item.linkId	
value	Boolean | number | date | string | code	Answer to the question in its format	Required (1..1)	QuestionnaireResponse.item.answer.value	Based on the questionnaire.item.type, use the appropriate value field

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
items	QuestionnaireResponse.item	Used to group answers when questionnaire had a question group		QuestionnaireResponse.item.item	

OBSERVATION (FHIR::OBSERVATION)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
identifier	Identifier	Unique identifiers for this observation	Optional (0..*)	Observation.identifier	A unique identifier assigned to the observation.
status	Code	The status of the observation	Required (1..1)	Observation.status	Status is set to “final” indicating the observation is complete.
subject	Reference	The subject the observation is about	Required (1..1)	Observation.subject	Refers to the Patient resource associated with this observation.
effectiveDateTime	DateTime	The date and time when the observation was made	Optional (1..1)	Observation.effectiveDateTime	Date and time when the observation was taken.
code	CodeableConcept	The code that represents the type of observation	Required (1..1)	Observation.code	The code that represents the observation type. SNOMED/LOINC (e.g., weight: LOINC/29463-7)
category	CodeableConcept	A code that classifies the general category of the observation	Optional (0..1)	Observation.category	A classification of the observation (e.g., vital signs).
value	*	The value of the observation, depending on the type of	Optional (1..1)	Observation.value	The actual measurement value (e.g., weight, height).

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
		value different sub fields can be used			
method	CodeableConcept	The method used to perform the observation	Optional (0..1)	Observation.method	The method or technique used for the observation.
interpretation	CodeableConcept	A clinical interpretation of the observation's value	Optional (0..1)	Observation.interpretation	The clinical interpretation of the observation value (e.g., normal, abnormal).
performer	Reference	Who performed the observation	Optional (0..*)	Observation.performer	Refers to the practitioner or organization who performed the observation.
device	Reference	The device used to generate the observation	Optional (0..1)	Observation.device	Refers to the device used to capture the observation (e.g., garmin fitness tracker).
referenceRange	BackboneElement	Provides the normal range for the observation's value	Optional (0..*)	Observation.referenceRange	The normal or expected range of values for the observation.
note	Annotation	Additional notes or comments about the observation	Optional (0..*)	Observation.note	Any additional information about the observation.

Funded by the European Union

NUTRITION INTAKE (FHIR::NUTRITIONINTAKE)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
status	code	The status of the intake	Required (1..1)	NutritionIntake.status	Based on EventStatus valueset -> *preparation, in-progress, not-done, on-hold, stopped, completed, entered-in-error, unknown*
code	CodeableConcept	Represents an overall type of nutrition intake		NutritionIntake.code	
subject	Reference	The individual the intake is related to	Required (1..1) Reference to Patient resource, e.g. **Patient/id**	NutritionIntake.subject	Identifies the participant the dietary information is related to
encounter	Reference	Encounter associated with NutritionIntake	Required (1..1) Reference to Encounter resource, e.g. Encounter/id	NutritionIntake.encounter	
recorded	dateTime	The date when the Nutrition Intake was asserted by the participant	Required (1..1)	NutritionIntake.recorded	

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
reported	Reference	Who provided the information about the consumption of this food or fluid	Required (1..1) reported[x] = reported.reportedReference Reference to Patient or RelatedPerson resource, e.g. **Patient/id or RelatedPerson/id**	NutritionIntake.reported.reportedReference	Identifies the person who provides the dietary information in the app
consumedItem	BackboneElement	What food or fluid product or item was consumed	Required (1..1) NutritionIntake.consumedItem.type NutritionIntake.consumedItem.nutritionProduct.concept NutritionIntake.consumedItem.amount	NutritionIntake.consumedItem	Identifies the type, quantity o number of servings and / or beverages consumed by the participant

WEARABLE (FHIR::DEVICE)

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
identifier	Identifier	Unique identifiers for this device	Optional (0..*)	Observation.identifier	A unique identifier assigned to the device.
deviceName	String	The name or model of the device	Optional (0..1)	Device.deviceName	The device name assigned to the specific participant, e.g. deviceName = ‘garmin001’
manufacturer	String	Manufacturer of the wearable device	Optional (1..1) manufacturer = ‘Garmin’	Device.manufacturer	

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Attribute Mapping	Note
name	BackboneElement	The name or names of the device as known to the manufacturer and/or individual	Optional (0..1) name.value = 'Garmin VivoSmart 5' name.type = 'registered-name'	Device.name	The official name provided by the manufacturer
owner	Reference	Organization responsible for device		Device.owner	
contact	ContactPoint	Details for human/organization for support	Optional (0..1) contac.system = 'email' contact.value = email address	Device.contact	Who contact for support

Funded by the European Union

A1.II. FHIR / Health Data Mapping

The section lists all the concepts used in the project, grouped by type, and how those concepts will be mapped as FHIR Resources. The final list of health data to be modelled, gathered from the information detailed in section 6.1 , is the following:

- **Socio-demographic information:** Age, Sex
- **Physical and vital measurements:** Height, Weight, Waist circumference, BMI, Blood pressure (systolic and diastolic),
- **Blood test results:** Glucose (fasting glucose), Cholesterol (LDL, HDL, Total)
- **Physical exercise data:** Type, Duration, Frequency
- **Fitness level & Lifestyle:** Average Physical activity (hours/week), Hours of sleep/night, (Resting) heart rate, Cardiorespiratory fitness, Oxygen saturation, VO2max
- **Basic dietary information:** Servings of fruit/day, Servings of vegetables/day, Servings of sugary drinks/day, Servings of nuts/day, Servings of red and processed meats/day, Grams of cereal fibres/day
- **Socio-economic status:** Parents' education level

The mapping tables are structured using the following convention:

- The **columns in blue** represent information from the original variables list. Specifically:
  - the title of the subsection, such as "*Socio-demographic information*," corresponds to the *Categories* column of the original variables. The text in parentheses indicates the FHIR resource to which the variable should be mapped.
  - the *Attribute* column corresponds to the *Items* column of the original variables.
  - the *Type*, *Description* and *Constraint* columns refer respectively to the data type, a textual description and any specific constraint related to the original variables.
- The **columns in red** represent the information related to the corresponding FHIR resource:
  - the *FHIR Mapping* column specifies the item of the resource that corresponds to the variable.

- the *FHIR Assumption* column outlines the codes that need to be used to identify the specific variable. Codes are expressed as constants, whose value can be found in the table FHIR-SC-VALUESETS.
  For example, for the code of the weight *Observation.code = SC-VS.OBSERVATION-CODE.WEIGHT where SC-VS.OBSERVATION-CODE.WEIGHT = LOINC/29463-7*
- The *FHIR Note* column provides additional information

SOCIO-DEMOGRAPHIC INFORMATION (FHIR::PATIENT|OBSERVATION)

Attribute	Type	Description	Constraint	FHIR Mapping	FHIR Assumption	FHIR Note
Gender	CodeableConcept	The gender of the individual	FHIR Patient-Administrative gender value set	Patient.gender	AdministrativeGender -> male, female, unknown, other	
Age	Date	The individual's date of birth	dd/mm/yyyy	Patient.birthDate		Year and month of birth could be used for privacy
Ethnicity	CodeableConcept	The ethnicity of the individual	Value must be one of LOINC answers in 46463-6 (Race)	Observation.value	Observation.status ='final' Observation.subject=Reference (participant_id) Observation.category='social-history' Observation.code=LOINC/46463-6 Observation.value is code	Ethnicity represented as an Observation resource

Funded by the European Union

RECORDS OF PHYSICAL ACTIVITY (FHIR::OBSERVATION)[24]

Attribute	Type	Description	Constraint	FHIR Mapping	FHIR Assumption	FHIR Note
Daily Physical Activity from wearable	Codeable Concept	Physiscal acrivity (minutes per day)	Observation.code must belong to SC-VS. SC-PA-ACTIVITY valueset Device should refer to source device	Observation.value.value Quantity	Observation.category= VITAL_SIGN Observation.value is Quantity Observation.valueQuantity.unit = UCUM. MIN_PER_DAY	Codes represent moderate / vigorous / total PA in minutes per day
Fitness measures	CodeableConcept	Muscular fitness score	Observation.code must belong to SC-PHYSICAL-ASSESSMENT-VS	Observation.value.value Quantity	Observation.category= 'activity' Observation.value is Quantity	Represented as an Observation resource

[24] Based on Welcome to the Physical Activity IG - Physical Activity Implementation Guide v1.0.1

Funded by the European Union

PHYSICAL AND VITAL MEASUREMENTS (FHIR::OBSERVATION)

Attribute	Type	Description	Constraint	FHIR Mapping	FHIR Assumption	FHIR Note
weight	float	Weight in kilograms. Unit of measure: kg		Observation.value	Observation.category= 'vital-signs' Observation.value is Quantity Observation.valueQuantity.unit = UCUM.KG Observation.code = SC-MEASUREMENTS-VS.WEIGHT	Weight represented as an Observation resource
height	float	Height in centimeters. Unit of measure: cm		Observation.value	Observation.category= 'vital-signs' Observation.value is Quantity Observation.valueQuantity.unit = UCUM.CM Observation.code = SC-MEASUREMENTS-VS.HEIGHT	Height represented as an Observation resource
waist circumference	float	Waist circumference in centimeters. Unit of measure: cm		Observation.value	Observation.category= 'exam' Observation.value is Quantity Observation.valueQuantity.unit = UCUM.CM Observation.code= SC-MEASUREMENTS-VS.WAIST_CIRCUMFERENCE	WaistCircumference represented as an Observation resource
bmi	int	Body mass index. Unit of measure: %		Observation.value	Observation.category= 'vital-sign' Observation.value is Quantity Observation.valueQuantity.unit=UCUM.PERCENT Observation.code= SC-MEASUREMENTS-VS.BMI	Bmi represented as an Observation resource

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Mapping	FHIR Assumption	FHIR Note
systolic blood pressure	int	Systolic blood pressure in mmHg. Unit of measure: mmHg		Observation.component[0].value	Observation.category= 'vital-signs' Observation.value is Quantity Observation.valueQuantity.unit=UCUM.MILLIMETER_OF_MERCURY Observation.code= SC-MEASUREMENTS-VS.SYSTOLIC_BLOOD_PRESSURE	For these two measurements is created a unique Observation resource
diastolic blood pressure	int	Diastolic blood pressure in mmHg. Unit of measure: mmHg		Observation.component[1].value	Observation.category= 'vital-signs' Observation.value is Quantity Observation.valueQuantity.unit=UCUM.MILLIMETER_OF_MERCURY Observation.code= SC-MEASUREMENTS-VS.DIASTOLIC_BLOOD_PRESSURE	
oxygen saturation	Int	oxygen saturation in blood in %. mmHg. Unit of measure: %		Observation.value	Observation.category= 'vital-signs' Observation.value is Quantity Observation.valueQuantity.unit=UCUM.PERCENT Observation.code= SC-MEASUREMENTS-VS. OXYGEN_SATURATION	Oxygen saturation represented as an Observation resource

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Mapping	FHIR Assumption	FHIR Note
fasting glucose	int/float	Fasting glucose level in mmol/L or mg/dL. Unit of measure: mmol/L or mg/dL		Observation.value	Observation.category= 'laboratory' Observation.valueQuantity.unit=UCUM. MILLIGRAM_PER_DECILITER/MILLIMOLE_PER_LITER Observation.code=SC-MEASUREMENTS-VS.FASTING_GLUCOSE	
Total Cholesterol	int/float	Total cholesterol level in mg/dL or mmol/L. Unit of measure: mg/dL or mmol/L		Observation.value	Observation.category= 'laboratory' Observation.valueQuantity.unit=UCUM. MILLIGRAM_PER_DECILITER/MILLIMOLE_PER_LITER Observation.code=SC-MEASUREMENTS-VS.TOTAL_CHOLESTEROL	
HDL Cholesterol	int/float	HDL cholesterol level in mg/dL or mmol/L. Unit of measure: mg/dL or mmol/L		Observation.value	Observation.category= 'laboratory' Observation.valueQuantity.unit=UCUM. MILLIGRAM_PER_DECILITER /MILLIMOLE_PER_LITER Observation.code=SC-MEASUREMENTS-VS.HDL_CHOLESTEROL	

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Mapping	FHIR Assumption	FHIR Note
LDL Cholesterol	int/float	LDL cholesterol level in mg/dL or mmol/L. Unit of measure: mg/dL or mmol/L		Observation.value	Observation.category= 'laboratory' Observation.valueQuantity.unit=UCUM. MILLIGRAM_PER_DECILITER /MILLIMOLE_PER_LITER Observation.code=SC-MEASUREMENTS-VS.LDL_CHOLESTEROL	

LIFESTYLE AND FITNESS (FHIR::OBSERVATION)

Attribute	Type	Description	Constraint	FHIR Mapping	FHIR Assumption	FHIR Note
Hours of sleep/night	Int/float	Hours of sleep every night in hours. Unit of measure: h/day		Observation.value	Observation.category='activity' Observation.value is Quantity Observation.valueQuantity.unit=UCUM.HOUR_PER_DAY Observation.code= SC-MEASUREMENTS-VS.SLEEP_DURATION	Sleep duration represented as an Observation resource
(Resting) heart rate	Int	Resting heart rate in bpm. Unit of measure: bpm		Observation.value	Observation.category= 'exam' Observation.value is Quantity Observation.valueQuantity.unit=UCUM.BPM Observation.code= SC-MEASUREMENTS-VS. RESTING_HEART_RATE	Oxygen saturation represented as an Observation resource

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Mapping	FHIR Assumption	FHIR Note
Average Physical activity (hours/week)	Int/float	Average Physical activity per week. Unit of measure: h/wk		Observation.value	Observation.category='activity' Observation.value is Quantity Observation.valueQuantity.unit=UCUM.HOUR_PER_WEEK Observation.code= SC-MEASUREMENTS-VS.	Physical activity represented as an Observation resource
VO2max	Int/float	Maximum oxygen uptake. Unit of measure: respiratory volume rate per body mass		Observation.value	Observation.category= 'activity' Observation.value is Quantity Observation.valueQuantity.unit=UCUM.MILLILITER_PER_KG_PER_MIN Observation.code=SC-MEASUREMENTS-VS.VO2_MAX	Vo2max represented as an Observation resource

SOCIO ECONOMIC STATUS QUESTIONNAIRE (FHIR::QUESTIONNAIRE)

Attribute	Type	Description	Constraint	FHIR Mapping	FHIR Assumption	FHIR Note
Status	Code	Publication status	'active'	Questionnaire.status		
name	String	Name of the questionnaire	'EDUCATION'	Questionnaire.name		
code	CodeableConcept	Concept describing the questionnaire	QUESTIONNAIRECODESVALUESET. SOCIO_ECONOMIC_STATUS_QUESTIONNAIRE	Questionnaire.code		

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Mapping	FHIR Assumption	FHIR Note
Question1	QuestionItem	Mother Highest Educational Level		Questionnaire.Item [0]	Item.type = 'code' Item.linkIId='SE_1' Item.text='education of mother' Item.code= SC-QUESTIONNAIRE-VS. MOTHER_EDUCATION Item.answerValueset = SC-EDUCATION-VS Item.required=false	
Question2	QuestionItem	Father Highest Educational Level		Questionnaire.Item [1]	Item.type = 'code' Item.linkIId='SE_2' Item.text='education of father Item.code= SC-QUESTIONNAIRE-VS. FATHER_EDUCATION Item.answerValueset = SC-EDUCATION-VS Item.required=false	

Funded by the European Union

NUTRITION INFORMATION (FHIR::NUTRITIONINTAKE)

Attribute	Type	Description	Constraint	FHIR Mapping	FHIR Assumption	FHIR Note
Pieces of fruit/day	Int/float	The number of fruits consumed per day		NutritionIntake.consumedItem	NutritionIntake.consumedItem.type=SC-NUTRITIONINTAKE-VS.FOOD NutritionIntake.consumedItem.nutritionProduct.concept= SC-NUTRITIONINTAKE-VS.FRUIT	
Pieces of vegetables/day	Int/float	The number of vegetable portions consumed per day		NutritionIntake.consumedItem	NutritionIntake.consumedItem.type=SC-NUTRITIONINTAKE-VS.FOOD NutritionIntake.consumedItem.nutritionProduct.concept= SC-NUTRITIONINTAKE-VS.VEGETABLES	

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Mapping	FHIR Assumption	FHIR Note
Number of sugary drinks/day	Int/float	The total number of beverages with added sugars consumed per day.		NutritionIntake.consumedItem	NutritionIntake.consumedItem.type=SC-NUTRITIONINTAKE-VS.DRINK NutritionIntake.consumedItem.nutritionProduct.concept= SC-NUTRITIONINTAKE-VS.SUGARY_DRINK	
Servings of nuts/day	Int/float	The number of portions of nuts consumed per day		NutritionIntake.consumedItem	NutritionIntake.consumedItem.type=SC-NUTRITIONINTAKE-VS.FOOD NutritionIntake.consumedItem.nutritionProduct.concept= SC-NUTRITIONINTAKE-VS.NUT	
Servings of red and processed meats/day	Int/float	The number of portions of red and processed meat consumed per day		NutritionIntake.consumedItem	NutritionIntake.consumedItem.type=SC-NUTRITIONINTAKE-VS.FOOD NutritionIntake.consumedItem.nutritionProduct.concept= SC-NUTRITIONINTAKE-VS.MEAT	

Funded by the European Union

Attribute	Type	Description	Constraint	FHIR Mapping	FHIR Assumption	FHIR Note
Grams of cereal fibres/day	Int/float	The grams of cereal fibres consumed per day		NutritionIntake.consumedItem	NutritionIntake.consumedItem.type=SC-NUTRITIONINTAKE-VS.FOOD NutritionIntake.consumedItem.nutritionProduct.concept= SC-NUTRITIONINTAKE-VS.CEREAL	

Funded by the European Union

A1.III. FHIR Valuesets for SmartCHANGE

In the following section are reported the codes of used the project's data model from standard clinical vocabularies - i.e. LOINC (http://loinc.org/) and SNOMED CT (http://snomed.info/sct) - select international edition), and the units of measure from the international unit of measure - i.e. UCUM (http://unitsofmeasure.org).

- All the column NAME is a unique coded name in the context of the valueset
- the System, Code and Display columns correspond to the coding from the one of standard vocabularies

SC-MEASUREMENTS-VS

Name	System	Code	Display
ALCOHOL_USE_DRINKS_PER_DAY	http://loinc.org/	74013-4	Alcoholic drinks per day
SLEEP_DURATION	http://loinc.org/	93832-4	Sleep duration
ETHNICITY	http://loinc.org/	46463-6	Race
BMI	http://loinc.org/	39156-5	Body mass index (BMI)
EDUCATION	http://loinc.org/	72226-2	Educational achievement
WEIGHT	http://loinc.org/	29463-7	Body weight
HEIGHT	http://loinc.org/	8302-2	Body height
WAIST_CIRCUMFERENCE	http://snomed.info/sct	276361009	Waist circumference (observable entity)
SYSTOLIC_BLOOD_PRESSURE	http://loinc.org/	8480-6	Systolic blood pressure
DIASTOLIC_BLOOD_PRESSURE	http://loinc.org/	8462-4	Diastolic blood pressure
FASTING_GLUCOSE	http://loinc.org/	76629-5	Fasting glucose [Moles/volume] in Blood
OXYGEN_SATURATION	http://loinc.org/	59408-5	Oxygen saturation in Arterial blood by Pulse oximetry
TOTAL_CHOLESTEROL	http://loinc.org/	9830-1	Cholesterol [Mass/volume] in Blood
HDL_CHOLESTEROL	http://loinc.org/	2085-9	High-density lipoprotein (HDL) cholesterol
LDL_CHOLESTEROL	http://loinc.org/	2089-1	Low-density lipoprotein (LDL) cholesterol
RESTING_HEART_RATE	http://snomed.info/sct	444981005	Resting heart rate (observable entity)
PHYSICAL_ACTIVITY	http://snomed.info/sct	68130003	Physical activity (observable entity)

VO2_MAX	http://snomed.info/sct	251898000	Maximum oxygen uptake (observable entity)

SC-FAMILY-ANM

Name	System	Code	Display
CARDIOVASCULAR_DISEASES	http://loinc.org/	88021-1	History of cardiovascular disease
KIDNEY_DISEASES	http://loinc.org/	86906-5	History of kidney disease
LIVER_DISEASES	http://loinc.org/	LA25810-5	Liver disease diagnosis
DEAD_FROM_HEART_ATTACK_ AGE_OF_55	http://snomed.info/sct	419099009	Family history of sudden death due to myocardial infarction
OVERWEIGHT_FATHER	http://snomed.info/sct	414916001	Family history of overweight
DIABETES_IMMEDIATE_FAMILY	http://snomed.info/sct	44054006	Family history of diabetes mellitus
HYPERTENSION_BEFORE_AGE_OF_55	http://snomed.info/sct	59621000	Family history of hypertension before 55
HIGH_CHOLESTEROL_IMMEDIATE_FAMILY	http://snomed.info/sct	13644009	Family history of hypercholesterolemia
STROKE_BEFORE_AGE_OF_55	http://snomed.info/sct	230690007	Family history of stroke before the age of 55
HEART_ATTACK_BEFORE_ AGE_OF_55	http://snomed.info/sct	22298006	Family history of myocardial infarction
CORONARY_HEART_DISEASE	http://snomed.info/sct	53741008	Coronary heart disease

SC-PA-ACTIVITY

Name	System	Code	Display
TOTAL_PHYSICAL_ACTIVITY	http://loinc.org/	77594-0	IPAQ total score
MODERATE_PA	http://loinc.org/	77592-4	Moderate physical activity [IPAQ]
VIGOROUS_PA	http://loinc.org/	77593-2	Vigorous physical activity [IPAQ]

SC-PHYSICAL-ASSESSMENT-VS

Name	System	Code	Display
MUSCULAR_FITNESS_SDM	http://smartchange.eu/	mf-sdm	Muscular Fitness SDM
MUSCULAR_FITNESS_VZG	http://smartchange.eu/	mf-vzg	Muscular Fitness VZG
N_OF_STEPS_IN_24H_MEAN_MEASURED	http://loinc.org/	41951-5	Number of steps in 24 hour mean Measured
MUSCULAR_FITNESS_PRE	http://smartchange.eu/	mf-pre	Muscular Fitness PRE

Funded by the European Union

DOES_SIT_UP	http://snomed.info/sct	282921005	Does sit up (finding)
JUMPING	http://snomed.info/sct	9251003	Jumping
DISTANCE_PER_TIME	http://snomed.info/sct	870597004	Distance per time (property) (qualifier value)
EXERCISE_DURATION_EXERCISE_FREQUENCY	http://loinc.org/	74009-2	Exercise duration/Exercise frequency
DURATION_OF_VIGOROUS_ACTIVITY	http://loinc.org/	101690-6	Duration of vigorous activity
HEART_RATE_10_MINUTES_MEAN	http://loinc.org/	66440-9	Heart rate 10 minutes mean [palpation]
NUMBER_OF_PUSH-UPS_COMPLETED	http://loinc.org/	66212-2	Number of push-ups completed
NUMBER_OF_MODIFIED_PULL-UPS_COMPLETED	http://loinc.org/	66223-9	Modified pull-ups [#] [PhenX]
LENGTH_OF_WALK_TEST_COURSE	http://loinc.org/	66254-4	Length of walk test course
PARTICIPATION_IN_SPORTS	http://loinc.org/	89266-0	Participation in sports

SC-PSYCOLOGICALQUESTIONNAIRE-VS

Name	System	Code	Display
EASILY_STARTLED	http://loinc.org	100242-7	Easily startled
TRIED_TO_NOT_THINK_OF_ADVERSE_EVENT	http://loinc.org	100243-5	Tried to not think of adverse event
AWARE_OF_FEELINGS_ABOUT_ADVERSE_EVENT	http://loinc.org	100244-3	Aware of feelings about adverse event, but did not address them
FEELINGS_ABOUT_ADVERSE_EVENT_WERE_NUMB	http://loinc.org	100245-0	Feelings about adverse event were numb
ACTED_OR_FELT_THE_SAME_AS_DURING_ADVERSE_EVENT	http://loinc.org	100246-8	Acted or felt the same as during adverse event
TROUBLE_FALLING_ASLEEP	http://loinc.org	100247-6	Trouble falling asleep
WAVES_OF_STRONG_FEELINGS_ABOUT_ADVERSE_EVENT	http://loinc.org	100248-4	Waves of strong feelings about adverse event
TRIED_TO_REMOVE_ADVERSE_EVENT_FROM_MEMORY	http://loinc.org	100249-2	Tried to remove adverse event from memory
TROUBLE_CONCENTRATING	http://loinc.org	100250-0	Trouble concentrating
REMINDERS_OF_ADVERSE_EVENT_CAUSED_PHYSICAL_REACTIONS	http://loinc.org	100251-8	Reminders of adverse event caused physical reactions
HAD_DREAMS_ABOUT_ADVERSE_EVENT	http://loinc.org	100252-6	Had dreams about adverse event
FELT_WATCHFUL_AND_ON-GUARD	http://loinc.org	100253-4	Felt watchful and on-guard
TRIED_NOT_TO_TALK_ABOUT_ADVERSE_EVENT	http://loinc.org	100254-2	Tried not to talk about adverse event
TOTAL_SCORE_IMPACT_OF_EVENT_SCALE	http://loinc.org	100255-9	Total score [Impact of Event Scale-Revised]
FALLEN_IN_LAST_6_MONTHS	http://loinc.org	100256-7	Fallen in last 6 months
FEEL_UNSTEADY_WHEN_STANDING_OR_WALKING	http://loinc.org	100257-5	Feel unsteady when standing or walking
HISTORY_OF_FALL-RELATED_INJURY	http://loinc.org	100258-3	History of fall-related injury
AWARE_OF_TACTILE_SENSATIONS	http://loinc.org	100259-1	Aware of tactile sensations

Funded by the European Union

EASILY_DISTRACTED	http://loinc.org	100263-3	Easily distracted
AWARE_OF_DIETARY_INTAKE_IMPACT_ON_SELF	http://loinc.org	100265-8	Aware of dietary intake impact on self
DO_THINGS_WITHOUT_PAYING_ATTENTION	http://loinc.org	100269-0	Do things without paying attention
RECOVER_QUICKLY_FROM_DISTRESSING_THOUGHTS_IMAGES	http://loinc.org	100270-8	Recover quickly from distressing thoughts and images
AWARE_OF_VISUAL_STIMULI	http://loinc.org	100272-4	Aware of visual stimuli
AWARE_OF_EMOTIONS_EFFECT_ON_THOUGHTS_AND_BEHAVIOR	http://loinc.org	100273-2	Aware of emotions' effect on thoughts and behavior
SUBSCORE_SDQ_EMOTIONAL_PROBLEMS	http://loinc.org	100242-7	Emotional problems subscore [SDQ]
SUBSCORE_SDQ_CONDUCT_PROBLEMS	http://loinc.org	100243-5	Conduct problems subscore [SDQ]
SUBSCORE_SDQ_HYPERACTIVITY_ATTENTION_DEFICIT	http://loinc.org	100244-3	Hyperactivity/Attention Deficit subscore [SDQ]
SUBSCORE_SDQ_PEER_PROBLEMS	http://loinc.org	100245-0	Peer problems subscore [SDQ]
SUBSCORE_SDQ_PROSOCIAL_BEHAVIOR	http://loinc.org	100246-8	Prosocial behavior subscore [SDQ]
TOTAL_SDQ_SCORE	http://smartchange.eu/	total-sdq	Total SDQ score (calculated)
SUBSCORE_DASS21_DEPRESSION	http://loinc.org	100248-4	Depression subscale [DASS21]
SUBSCORE_DASS21_ANXIETY	http://loinc.org	100249-2	Anxiety subscale [DASS21]
SUBSCORE_DASS21_STRESS	http://loinc.org	100250-0	Stress subscale [DASS21]

SC-QUESTIONNAIRE-VS

Name	System	Code	Display
SE_STATUS_QUESTIONNAIRE	http://smartchange.eu/	SOCIO	Socio Economic Status Questionnaire
SE_EDUCATION_LEVEL	http://loinc.org/	LL5338-0	Highest Education Level
MOTHER_EDUCATION	http://loinc.org/	57712-2	Highest level of education Mother
FATHER_EDUCATION	http://loinc.org/	87300-0	Highest level of education Father
HEAD_OF_HOUSEHOLD_EDUCATION	http://loinc.org/	99796-5	Highest level of education completed for head of household

SC-EDUCATION-VS

Name	System	Code	Display
NO_SCHOOLING	http://loinc.org/	LA35-1	No schooling
8_GRADE_LESS	http://loinc.org/	LA36-9	8th grade/less
9_11_GRADES	http://loinc.org/	LA12456-2	9-11 grades
HIGH_SCHOOL	http://loinc.org/	LA12457-0	High school

Funded by the European Union

TECHNICAL_OR_TRADE_SCHOOL	http://loinc.org/	LA39-3	Technical or trade school
SOME_COLLEGE	http://loinc.org/	LA40-1	Some college
ASSOCIATE_DEGREE	http://loinc.org/	LA12459-6	Associate degree (e.g., AA, AS)

SC-NUTRITIONINTAKE-VS

Name	System	Code	Display
FOOD	http://snomed.info/sct	255620007	Food (substance)
DRINK	http://snomed.info/sct	226465004	Drinks (substance)
FRUIT	http://snomed.info/sct	72511004	Fruit (substance)
VEGETABLES	http://snomed.info/sct	22836000	Vegetable (substance)
SUGARY_DRINK	http://snomed.info/sct	818989004	Sugar sweetened beverage (substance)
NUT	http://snomed.info/sct	13577000	Nut (substance)
RED_MEAT	http://snomed.info/sct	226915003	Red meat (substance)
PROCESSED_MEAT	http://snomed.info/sct	227031001	Processed meat (substance)
MEAT	http://snomed.info/sct	28647000	Meat (substance)
ALCOHOLIC_BEVERAGE	http://snomed.info/sct	53527002	Alcoholic beverage (substance)
FIBER	http://snomed.info/sct	37202001	Plant fiber (substance)
CEREAL	http://snomed.info/sct	23182003	Cereal grain (substance)

UCUM

Name	System	Unit	code
KG	http://unitsofmeasure.org	kilogram	kg
MILLIMOLE_PER_MOLE	http://unitsofmeasure.org	millimole per mole	mmol/mol
MILLIMOLE_PER_LITER	http://unitsofmeasure.org	millimole per liter	mmol/L
MILLIGRAM_PER_DECILITER	http://unitsofmeasure.org	milligram per deciliter	mg/dL
MILLIGRAM_PER_LITER	http://unitsofmeasure.org	milligram per liter	mg/L
GRAM_PER_DECILITER	http://unitsofmeasure.org	gram per deciliter	g/dL

Funded by the European Union

Name	System	Unit	code
MILLIMETER_OF_MERCURY	http://unitsofmeasure.org	millimeter of mercury	mm[Hg]
METER	http://unitsofmeasure.org	meter	m
CENTIMETER	http://unitsofmeasure.org	centimeter	cm
PERCENT	http://unitsofmeasure.org	Percent	%
HOUR_PER_WEEK	http://unitsofmeasure.org	hour/week	h/wk
HOUR_PER_DAY	http://unitsofmeasure.org	hour/day	h/d
MILLILITER_PER_KG_PER_MIN	http://unitsofmeasure.org	milliliter per kilogram and minute	mL/kg/min
NUMBER_PER_DAY	http://unitsofmeasure.org	/day	/d
GRAM	http://unitsofmeasure.org	gram	G
MIN_PER_DAY	http://unitsofmeasure.org	min/d	Minutes per day

Funded by
the European Union

Appendix 2. Inputs and outputs of SmartCHANGE service interfaces

A2.I. XCDSEngine and Explainer inputs and outputs

XCDSSERVICE

Field	Cardinality	Type	Description
hook	1...1	*string*	The field describes the relevant moment when the service should be called. For risk prediction services, the hook could be whenever a user views a patient's information (patient-view), or when the user ended a new encounter (encounter-end)
title	0...1	*string*	The human-friendly name of this service.
description	1...1	*string*	The description of this service.
id	1...1	*string*	The {id} portion of the URL to this service which is available at {baseUrl}/cds-services/{id}
prefetch	0...1	*Prefetch*	An object containing key/value pairs of FHIR queries that this service is requesting to perform and provide on each service call. The key is a string that describes the type of data being requested and the value is a string representing the FHIR query. See https://cds-hooks.hl7.org/2.0/#prefetch
usage Requirements	0...1	*string*	Human-friendly description of any preconditions for the use of this Service.
extension	1...*	*XCDSModelDescription*	Extension to the CDSHooks specification for the SmartCHANGE project

XCDSMODELDESCRIPTION

Field	Cardinality	Type	Description

Funded by the European Union

`name`	0...1	*string*	The human readable name of the model, also used as a local id to the service (in particular, it must be unique for each service, so that it is not replicable in other services; in future versions we will consider whether to distinguish the id from the name). It matches baseModelDescription.name (see below) and must be indicated in at least one of the two fields.
`algorithmType`	0...1	*string*	type of algorithm used.
`learningMethodType`	1...1	*string*	type of learning used.
`inputFeatures`	1...*	*string*	list of all features used by the model, specified by dedicated FHIR CodeSystem using the shorthand defined in the *@context* attribute.
`outputClasses`	1...	*string*	list of M classes that the model is able to output; each class must belong to one of the defined FHIR CodeSystem.
`performance`	1...1	*object*	JSON structure containing: • as **key**, one of the M+1 possible *classes* (as defined by the class attribute plus the special class "*.microAvg*"[25]) • as **value**, in turn a JSON structure, with one of the possible evaluation metrics as key and the corresponding measurement (in range [0,1] for the metrics indicated below) as value. Possible metrics are specified by dedicated FHIR CodeSystem and referenced by the *@context* attribute.

[25] Special label indicating that the metrics refer to the overall performance of the model, not related to a specific class, calculated by micro-average.

`baseModelDescription`	0...1	*object*	full description of the model, according to the MLDCAT-AP[26] specification. Fields of interest are: • **created**: date of creation (mandatory) • **creator**: model maker • **description:** overall description of the model • **hasBibliographicReference:** bibliographic reference • **hasOutputFilePrediction**: references to the file containing the trained model (mandatory) • **intendedUse**: purpose of the model • **limitations**: known limitations • **name:** intelligible model name (mandatory) • **trainedOn**: information about the dataset on which the model was trained on (mandatory) • **trainingProcess:** information about the training process • **testedOn**: information about the dataset on which it was tested • **validatedOn**: information on the dataset on which it has been validated • **version:** version of the product model (mandatory)

[26] https://semiceu.github.io/MLDCAT-AP/releases/2.0.0/#overview

XCDSServiceRequest

The input to the service, compliant to the CDSHooks standard [41]. In the SmartCHANGE project the following fields will be used:

Field	Cardinality	Type	Description
`hook`	1...1	*string*	The hook that triggered this call. See hooks in XCDSService.
`hookInstance`	1...1	*string*	A universally unique identifier (UUID) for this particular hook call
`context`	1...1	*object*	Hook-specific contextual data that the CDS service will need. For example, with the patient-view hook this will include the FHIR id of the Patient being viewed.
`prefetch`	0...1	*object*	The FHIR data that was prefetched by the XCDS Engine, as detailed in the prefetch field of the XCDSService data structure. In SmartCHANGE, this field will not be used in requests by clients, only in calls from the XCDSEngine to the Explainer

XCDSServiceResponse

The Explainer service responds to prediction requests with a list of XCDSCard objects

Field	Cardinality	Type	Description
`cards`	1...*	XCDSCard	Each service can respond with potentially multiple XCDSCards to provide a combination of information. The specification of the XCDSCard object is in the following table.

Funded by the European Union

XCDSCard

The structure of the XCDSCard is compliant to the CDSHooks 2.0 Card specification, extending it to include information on the explanation. The extended information is included in the extension field, as suggested by the specification[27].

Field	Optionality	Type	Description
`uuid`	0...1	*string*	Unique identifier of the card. It may be used for auditing and logging cards.
`summary`	1...1	*string*	One-sentence message for display to the user inside of this card.
`indicator`	1...1	*string*	Urgency/importance of what this card conveys. Allowed values, in order of increasing urgency, are info, warning, critical. The CDS Client MAY use this field to help make UI display decisions such as sort order or coloring.
`source`	1...1	*object*	**Label [1…1]:** A short, human-readable label to display for the source of the information displayed on this card **url [0…1]:** An optional absolute URL to load (via GET, in a browser context) when a user clicks on this link to learn more about the organization or data set that provided the information on this card. Note that this URL should not be used to supply a context-specific "drill-down" view of the information on this card. For that, use card.link.url instead
`suggestions`	0...*	XCDSSuggestion	In SmartCHANGE they contain the updated RiskAssessment resource and the explanation, in the extension field

[27] https://cds-hooks.hl7.org/2.0/#extensions

Funded by the European Union

Field	Optionality	Type	Description
`selectionBehavior`	0...1	*string*	Describes the intended selection behavior of the suggestions in the card. Allowed values are: • "*at-most-one*", indicating that the user may choose none or at most one of the suggestions; • "*any*", indicating that the end user may choose any number of suggestions including none of them and all of them.
`link`	0...*	*Link*	Allows a service to suggest links to an app that the user might want to run for additional information or to help guide a decision. See https://cds-hooks.hl7.org/2.0/#link

XCDSSUGGESTION

Field	Cardinality	Type	Description
`label`	1...1	*string*	Human-readable label to display for this suggestion.
`uuid`	0...1	*string*	Unique identifier, used for auditing and logging suggestions.
`isRecommended`	0...1	*boolean*	Boolean value allowing service to indicate that a specific suggestion is recommended from all the available suggestions on the card. Multiple suggestions MAY be recommended.
`actions`	0...1	*array* of XCDSAction	Array of objects, each defining a suggested action.
`extension`	1...1	XCDSExplanation	Extension to the CDSHooks Suggestion data format to hold the explanation

Funded by the European Union

XCDSACTION

Field	Cardinality	Type	Description
`type`	1...1	*string*	The type of action being performed. Allowed values are *create*, *update*, *delete*. In the context of the project, the create or update options will be used
`description`	1...1	*string*	Human-readable description of the suggested action
`resource`	0...1	*object*	A FHIR resource. For risk prediction, the suggested resource to be updated is the RiskAssessment resource. Details on the usage within the project is in the FHIR profile in Appendix 1

XCDSEXPLANATION

Field	Cardinality	Type	Description
`confidence`	1...1	*double*	probability (in the range [0,1]) of how confident the algorithm is in giving that particular suggestion, that is, in predicting the connected class.
`class`	1...1	*string*	class linked to the suggestion provided, between the possible M that the model is able to produce in output (as defined by the extension *classes* within the *XCDSService* description). The class is identified using a standard vocabulary such as loinc or snomed
`Features Impact`	1...*	*object*	*<key, value>* pairs defined as follow: • **key**: the name of one of the n most important features for the class in question (defined by the *class* attribute) • **Value:** a JSON structure that indicates the LIME value (key *weight*), the id of the resource FHIR received in input that contained the value of the feature in question (key *resourceId*), the same value assumed by the feature (key *value*, can be omitted if the value is boolean and equal to true).
`counterfactuals`	1...*	*object*	*<key, value>* pairs defined as follow: • **key**: a risk factor name, or the keyword 'risk', • **Value:** is the projected value for that risk factor or risk prediction in that counterfactual

Funded by the European Union

EXPLANATIONREQUEST

Field	Cardinality	Type	Description
input	1...1	*XCDSServiceRequest*	Includes the participant data
variants	1...1	*number*	how many counterfactuals to generate
projectedRisk	1...1	*number*	the desired risk of the counterfactuals
age	1...1	*number*	the age of the participant for which we want to generate counterfactuals

A2.II. Data Cleaning

DATABASE

Field	Cardinality	Type	Description
id	1...1	*string*	Id of the database
name	1...1	*string*	Name of the database
description	1...*	*string*	Description of the database

SUBMITREPORT

Field	Cardinality	Type	Description
report	1...1	*string*	Report on cleaning
inconsistencies	1...*	*string*	List of inconsistencies found

CLEANINGREPORT

Field	Cardinality	Type	Description
message	1...1	*string*	Output message
Download_link	1...1	*URI*	Link to download the clean dataset

Funded by the European Union

DOWNLOADRESPONSE

Field	Cardinality	Type	Description
`dataset_link`	1...1	*URI*	Link to download the clean dataset
`log_file_link`	1...1	*URI*	Link to the complete log file

Funded by the European Union

Appendix 3. OpenAPI

A3.I. XCDSEngine

```yaml
openapi: 3.0.1
info:
  title: SmartCHANGE XCDS Engine API
  version: 1.0.0
  description: API for the XCDS Engine, providing CDS services for risk
prediction and explanations, compliant with CDS Hooks 2.0.

servers:
  - url: https://api.smartchange.example.com

paths:
  /cds-services:
    get:
      summary: Get all CDS services
      description: Retrieve a list of all available CDS services.
      responses:
        '200':
          description: A list of available CDS services
          content:
            application/json:
              schema:
                type: object
                properties:
                  services:
                    type: array
                    items:
                      $ref: '#/components/schemas/XCDSService'

  /cds-services/{id}:
    post:
      summary: Get prediction from a specific CDS service
      description: Submit a request to the XCDS Engine for a risk prediction
using a specific CDS service.
      parameters:
        - in: path
          name: id
          required: true
          schema:
```

```
          type: string
        description: The ID of the CDS service
    requestBody:
      required: true
      content:
        application/json:
          schema:
            $ref: '#/components/schemas/XCDSServiceRequest'
    responses:
      '200':
        description: List of risk prediction cards
        content:
          application/json:
            schema:
              $ref: '#/components/schemas/XCDSServiceResponse'

components:
  schemas:
    XCDSService:
      type: object
      properties:
        hook:
          type: string
          description: The hook that triggers the service.
        title:
          type: string
          description: Human-readable name of the service.
        description:
          type: string
          description: Description of the service.
        id:
          type: string
          description: Unique ID of the service.
        prefetch:
          type: object
          description: Prefetch queries for required data.
        usageRequirements:
          type: string
          description: Human-friendly description of any preconditions for the
use of this Service
        extension:
          $ref: '#/components/schemas/XCDSModelDescription'

    XCDSModelDescription:
```

```
      type: object
      properties:
        name:
          type: string
          description: The human-readable name of the model, used as a unique
identifier for the service.
        algorithmType:
          type: string
          description: The type of algorithm used by the model (e.g., neural
network, decision tree, etc.).
        learningMethodType:
          type: string
          description: The type of learning employed by the model (e.g.,
supervised, unsupervised, reinforcement).
        inputFeatures:
          type: array
          description: List of all features used by the model.
          items:
            type: string
        outputClasses:
          type: array
          description: List of possible output classes that the model can
predict.
          items:
            type: string
        performance:
          type: object
          description: Performance metrics of the model for each output class.
          additionalProperties:
            type: object
            description: A JSON structure containing the performance metrics
for each output class.
            properties:
              class:
                type: string
                description: One of the M+1 possible output classes, including
"microAvg".
              metrics:
                type: object
                description: Evaluation metrics and corresponding values.
                additionalProperties:
                  type: number
                  description: Value of the metric (in range [0, 1]).
        baseModelDescription:
```

```
          type: object
          description: Detailed metadata about the model according to the
MLDCAT-AP specification.
          properties:
            created:
              type: string
              format: date-time
              description: The date and time the model was created.
            creator:
              type: string
              description: The name or organization that created the model.
            description:
              type: string
              description: Overall description of the model.
            hasBibliographicReference:
              type: string
              description: Any bibliographic reference relevant to the model.
            hasOutputFilePrediction:
              type: string
              description: Reference to the file containing the trained model.
            intendedUse:
              type: string
              description: The purpose or intended use of the model.
            limitations:
              type: string
              description: Known limitations of the model.
            name:
              type: string
              description: Intelligible name of the model.
            trainedOn:
              type: string
              description: Information about the dataset used to train the
model.
            trainingProcess:
              type: string
              description: Information about the model's training process.
            testedOn:
              type: string
              description: Dataset used to test the model.
            validatedOn:
              type: string
              description: Dataset used to validate the model.
            version:
              type: string
```

```
            description: Version of the model.


  XCDSServiceRequest:
    type: object
    properties:
      hook:
        type: string
        description: The hook that triggered the request.
      hookInstance:
        type: string
        description: Unique identifier for the hook call.
      context:
        type: object
        description: Hook-specific contextual data.
      prefetch:
        type: object
        description: Prefetched FHIR data, if available.

  XCDSCard:
    type: object
    properties:
      uuid:
        type: string
        description: Unique identifier of the card.
      summary:
        type: string
        description: One-sentence message summarizing the card.
      indicator:
        type: string
        description: Urgency level of the card (info, warning, critical).
      link:
        type: array
        items:
          type: string
        description: Urgency level of the card (info, warning, critical).
      selectionBehavior:
        type: string
        enum:
         - 'at-most-one'
         - 'any'
      source:
        type: object
        properties:
```

```
            label:
              type: string
            url:
              type: string
        suggestions:
          type: array
          items:
            $ref: '#/components/schemas/XCDSSuggestion'

    XCDSSuggestion:
      type: object
      properties:
        uuid:
          type: string
          description: Unique identifier of the suggestion
        isReccomended:
          type: boolean
          description:  allowing service to indicate that a specific
suggestion is recommended from all the available suggestions on the card
        label:
          type: string
          description: Human-readable label for the suggestion.
        actions:
          type: array
          items:
            $ref: '#/components/schemas/XCDSAction'
        extension:
          $ref: '#/components/schemas/XCDSExplanation'

    XCDSAction:
      type: object
      properties:
        type:
          type: string
          description: Type of action (create, update, delete).
        description:
          type: string
          description: Human-readable description of the action.
        resource:
          type: object
          description: FHIR resource being acted on.

    XCDSExplanation:
      type: object
```

```
      properties:
        confidence:
          type: number
          description: Probability of the suggestion being accurate.
        class:
          type: string
          description: The class of risk the explanation is related to.
        featuresImpact:
          type: object
          description: Key/value pairs of features and their impact.


    XCDSServiceResponse:
      type: object
      properties:
        cards:
          type: array
          items:
            $ref: '#/components/schemas/XCDSCard'
```

A3.II. Explainer

```
openapi: 3.0.1
info:
  title: SmartCHANGE Explainer API
  version: 1.0.0
  description: API for the Explainer service providing risk prediction and
explanations, compliant with CDS Hooks 2.0.

servers:
  - url: https://api.smartchange.example.com/explainer

paths:
  /execute:
    post:
      summary: Predict and explain risk
      description: Send a prediction request and receive both the prediction
and explanation.
      requestBody:
        required: true
        content:
```

```
          application/json:
            schema:
              $ref: '#/components/schemas/XCDSServiceRequest'
      responses:
        '200':
          description: Predicted risk and generated counterfactuals
          content:
            application/json:
              schema:
                $ref: '#/components/schemas/XCDSServiceResponse'
  /explain:
    post:
      summary: Explain an existing risk prediction
      description: Send a request to get the explanation of a previously
generated risk prediction.
      requestBody:
        required: true
        content:
          application/json:
            schema:
              $ref: '#/components/schemas/XCDSExplanationRequest'
      responses:
        '200':
          description: List of generated explanations (counterfactuals)
          content:
            application/json:
              schema:
                type: array
                items:
                  $ref: '#/components/schemas/XCDSExplanation'

components:
  schemas:
    XCDSExplanationRequest:
      type: object
      properties:
        input:
          $ref: '#/components/schemas/XCDSServiceRequest'
        variants:
          type: number
          description: Number of counterfactuals to generate.
        projectedRisk:
          type: number
          description: Desired risk value of the counterfactuals.
```

```
      age:
        type: number
        description: Age for generating counterfactuals.


  XCDSServiceRequest:
    type: object
    properties:
      hook:
        type: string
        description: The hook that triggered the request.
      hookInstance:
        type: string
        description: Unique identifier for the hook call.
      context:
        type: object
        description: Hook-specific contextual data.
      prefetch:
        type: object
        description: Prefetched FHIR data, if available.

  XCDSCard:
    type: object
    properties:
      uuid:
        type: string
        description: Unique identifier of the card.
      summary:
        type: string
        description: One-sentence message summarizing the card.
      indicator:
        type: string
        description: Urgency level of the card (info, warning, critical).
      link:
        type: array
        items:
          type: string
        description: Urgency level of the card (info, warning, critical).
      selectionBehavior:
        type: string
        enum:
          - 'at-most-one'
          - 'any'
      source:
```

```
          type: object
          properties:
            label:
              type: string
            url:
              type: string
        suggestions:
          type: array
          items:
            $ref: '#/components/schemas/XCDSSuggestion'

    XCDSSuggestion:
      type: object
      properties:
        uuid:
          type: string
          description: Unique identifier of the suggestion
        isReccomended:
          type: boolean
          description:  allowing service to indicate that a specific
suggestion is recommended from all the available suggestions on the card
        label:
          type: string
          description: Human-readable label for the suggestion.
        actions:
          type: array
          items:
            $ref: '#/components/schemas/XCDSAction'
        extension:
          $ref: '#/components/schemas/XCDSExplanation'

    XCDSAction:
      type: object
      properties:
        type:
          type: string
          description: Type of action (create, update, delete).
        description:
          type: string
          description: Human-readable description of the action.
        resource:
          type: object
          description: FHIR resource being acted on.
```

Funded by
the European Union

```
    XCDSExplanation:
      type: object
      properties:
        confidence:
          type: number
          description: Probability of the suggestion being accurate.
        class:
          type: string
          description: The class of risk the explanation is related to.
        featuresImpact:
          type: object
          description: Key/value pairs of features and their impact.
    XCDSServiceResponse:
      type: object
      properties:
        cards:
          type: array
          items:
            $ref: '#/components/schemas/XCDSCard'
```

A3.III. Data Cleaning

```
openapi: 3.0.0
info:
  version: "1.0"
  title: Data Cleaning Mechanism
  description: The API for the data cleaning-harmonization process
servers:
  - url: '{{baseUrl}}'
paths:
  /database:
    get:
      summary: Get all databases
      tags:
        - Databases
      description: Retrieves a list of all available databases for the user to
select
      operationId: getDatabases
      responses:
        '200':
          description: Successfully retrieved the list of databases
          content:
```

```
            application/json:
              schema:
                type: array
                items:
                  type: object
                  properties:
                    id:
                      type: string
                    name:
                      type: string
                    description:
                      type: string
  /selection:
    post:
      summary: Submit selected databases
      tags:
        - Selection
      description: Receives the user-selected databases and returns reports of
inconsistencies
      operationId: postSelectedDatabases
      requestBody:
        description: Selected databases by the user
        required: true
        content:
          application/json:
            schema:
              type: array
              items:
                type: string
      responses:
        '200':
          description: Reports generated successfully
          content:
            application/json:
              schema:
                type: object
                properties:
                  report:
                    type: string
                  inconsistencies:
                    type: array
                    items:
                      type: string
        '400':
```

```
          description: Bad request if the database selections are invalid
  /cleaning:
    post:
      summary: Apply cleaning methods
      tags:
        - Cleaning
      description: Applies selected cleaning methods to the data and prepares
them for download
      operationId: applyCleaningMethods
      requestBody:
        description: Cleaning methods selected by the user
        required: true
        content:
          application/json:
            schema:
              type: object
              properties:
                method_ids:
                  type: array
                  items:
                    type: string
      responses:
        '200':
          description: Data cleaned successfully, ready for download
          content:
            application/json:
              schema:
                type: object
                properties:
                  message:
                    type: string
                  download_link:
                    type: string
  /download:
    get:
      summary: Download cleaned data
      tags:
        - Download
      description: Provides links to download the cleaned datasets and log
files
      operationId: downloadCleanedData
      responses:
        '200':
          description: Files ready for download
```

```
        content:
          application/json:
            schema:
              type: object
              properties:
                dataset_link:
                  type: string
                log_file_link:
                  type: string
      '404':
        description: Files not found
```

Funded by
the European Union