ONCOSCREEN

D4.3 ONCOSCREEN INTEGRATED INTELLIGENT PLATFORM FOR CITIZENS, CLINICIANS & POLICYMAKERS (FIRST VERSION)

31.01.2024

Funded by the European Union

Submission date: M13

Due date: M13

DOCUMENT SUMMARY INFORMATION

Grant Agreement No.	101097036	**Acronym**	ONCOSCREEN
Full Title	A European "shield" against colorectal cancer based on novel, more precise and affordable risk-based screening methods and viable policy pathways		
Start Date	1 January 2023	**Duration**	48 months
Project URL	https://oncoscreen.health/		
Deliverable	D4.3: ONCOSCREEN Integrated Intelligent Platform for Citizens, Clinicians & Policy Makers (First version)		
Work Package	WP4 – ONCOSCREEN Intelligent Platform & Tools for Citizens, Clinicians & Policy Makers		
Type	DEM — Demonstrator, pilot, prototype	**Dissemination Level**	SEN - Sensitive
Lead Beneficiary	CERTH		
Contributions	EXUS, iSPRINT, KT, BEIA, TECHNION, TLBG, AINIGMA, MUI, GERCOR, VITO, CTL, CSIC		
Authors	Thanassis Mavropoulos, Athina Tsanousa (CERTH) Aristodemos Pnevmatikakis, Spyros Spanos, Chris Tasatzis, Artemis Kim, Danae Lekka (iSprint);		
Co-authors	N/A		
Reviewers	VITO, TLBG		

DISCLAIMER

COPYRIGHT MESSAGE

DOCUMENT HISTORY

Version	Date	Changes	Contributor(s)	Comments
V0.1	20.09.2020	First draft: ToC Released	CERTH	Proposed changes were discussed with the Project Coordinator
V0.2	23.11.2020	Second draft: Sections updated based on review comments	CERTH	Initial input was included by CERTH
V0.3	15.01.2021	Third draft: input from partners is incorporated	CERTH	Requests were sent to partners and received input was included.
V0.4	23.01.2021	Pre-final version: peer-review comments were addressed	CERTH	Incorporated comments from peer reviewers.
V1.0	31.01.2021	Final version: submitted to EC	EXUS	Quality Check and submission to the EU portal

PROJECT PARTNERS

Partner	Acronym
EXUS SOFTWARE MONOPROSOPI ETAIRIA PERIORISMENIS EVTHINIS	EXUS
UNIVERSITAETSMEDIZIN DER JOHANNES GUTENBERG-UNIVERSITAET MAINZ	UMC-Mainz
INSTITUTE OF COMMUNICATION AND COMPUTER SYSTEMS	ICCS
FIRALIS	Firalis
UNIVERSITATSKLINIKUM SCHLESWIG-HOLSTEIN	UKSH
UNIVERSITAET zu LUEBECK	UzL
LIETUVOS SVEIKATOS MOKSLU UNIVERSITETAS	LSMU
MEDIZINISCHE UNIVERSITAT GRAZ	MUG
INSTITUTO PORTUGUES DE ONCOLOGIA DO PORTO FRANCISCO GENTIL, EPE	IPO
INSTITUTUL ONCOLOGIC PROF. DR. ALEXANDRU TRESTIOREANU BUCURESTI	IOB
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY	TECHNION
UNIVERSIDADE DO MINHO	UMINHO
UNIVERSITEIT VAN TILBURG	TLBG
VLAAMSE INSTELLING VOOR TECHNOLOGISCH ONDERZOEK N.V	VITO
ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS	CERTH
INNOVATION SPRINT	iSPRINT
SCIENTIFIC ACADEMY FOR SERVICE TECHNOLOGY EV	SERVTECH
AINIGMA TECHNOLOGIES	AINIGMA
CATALINK LIMITED	CATALINK
KONNEKT ABLE TECHNOLOGIES LIMITED	KT
BEIA CONSULT INTERNATIONAL SRL	BEIA
UNIVERSIDAD DE LA RIOJA	URIOJA
TIME.LEX	time.lex
CARR COMMUNICATIONS LIMITED	CARR
MINISTRY OF HEALTH	MoHGR
PAGALBOS ONKOLOGINIAMS LIGONIAMS ASOCIACIJA	POLA LT
EUROPACOLON PORTUGAL- ASSOCIACAO DE LUTA CONTRA O CANCRO DO INTESTINO	ECPT
ELLINIKI ETAIREIA OGKOLOGIAS PEPTIKOU	HSGO
EUROPEAN SOCIETY OF DIGESTIVE ONCOLOGY	ESDO
FUNDATIA YOUTH CANCER EUROPE	YCE
MEDIZINISCHE UNIVERSITAT INNSBRUCK	MUI
LIETUVOS RESPUBLIKOS SVEIKATOS APSAUGOS MINISTERIJA	MoH-LT
EY ADVISORY SPA	EY
AGENCIA ESTATAL CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS	CSIC
UNIVERSITE DE FRANCHE-COMTE	UFC
ROZENBAUM KONSULTING	ROSENBAUM
GIE AXA	GIE AXA
ASSOCIATION GERCOR	GERCOR

CC ASSURED	CC ASSURED

LIST OF ABBREVIATIONS

Abbreviation	Description
CA	Consortium Agreement
CRC	ColoRectal Cancer
D	Deliverable
GA	Grant Agreement
IPR	Intellectual Property Rights
JWT	JSON Web Token
LIT	Live Integration Test
SSL	Secure Sockets Layer
TR	Technical Requirement
UR	User Requirement
WP	Work Package

Executive Summary

The first version of D4.3 includes information regarding the development of the WP4 tools ONCO-RISTE, ONCO-CAWA, ONCO-CLIDE, ONCO-EVIDA and data fusion. D4.3 describes the iterative Co-Designed System activities with the end users, the analysis and architecture, and the performance in lab tests of the corresponding tools within the first year of the project.

This document offers a comprehensive description of the project's actions on addressing end user requirements, and it explores the tools' innovation aspects along with its functionalities and designs. Specifically, the tools reported in D4.3 present a significant achievement during the first year since a large set of end user and technical requirements has been fulfilled. Over the next period, the final updates of the tools will be included in the second iteration of this deliverable (D4.4), which is due at M29. Moreover, the overall trustworthiness is examined along with the risks and the ability to mitigate them. The last part involves the next actions that will be followed for each tool.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

1.1 The need for D4.3 in the project level

This deliverable unveils a significant milestone in the project's evolution, presenting the initial version that delves into the development progress of key tools, namely ONCO-RISTE, ONCO-CAWA, ONCO-CLIDE, ONCO-EVIDA, and data fusion.

This documentation encapsulates the iterative Co-Designed System activities, involving end users in the analysis and architecture phases, and rigorously assessing the tools' requirements through two lab tests during the project's inaugural year. The project commenced with workshops facilitating the early gathering and fulfilment of the end-users' requirements in the tool's development. Furthermore, in the two Live Integration Tests (LITs) conducted for this deliverable, end-user feedback and inquiries after tool demonstrations played a pivotal role in enhancing the tool's current quality and shaping the subsequent developmental steps. However, it should be noted that the tools will not undergo clinical trial testing since samples of blood, saliva etc. that are collected are not relevant for the use cases of these tools.

Within D4.3 the project's dedication to meet end user needs and the steps towards implementation of the respective technical requirements is shown. Its subsections analyse each ONCOSCREEN module by describing the innovation aspects, functionalities, and complex design of the tools. The report delves into an examination of trustworthiness for each tool, methodically scrutinizing associated risks and proposing strategies for mitigation.

In chapter 8 the performance in lab tests is examined and the improvements from one pilot to the other is examined reporting major updates for each tool. Finally, in chapter 9 a forward-looking perspective is also presented, outlining the subsequent actions to be followed for each tool.

In essence, D4.3 is providing insights into the project's current roadmap, achievements, and future steps.

1.1 Deliverable objectives

Table 1 Description of Action: Task 4.3

ONCOSCREEN DoA requirements	Deliverable addressing the requirements	Brief description
WP4: ONCOSCREEN Intelligent Platform & Tools for Citizens,	D4.3: ONCOSCREEN Integrated Intelligent Platform for Citizens,	This deliverable reports on the co-design activities with the end-users on the development and architecture of the ONCO-RISTE, ONCO-CAWA,

Clinicians & Policy Makers	Clinicians & Policy Makers (First version)	CDSSM, ONCO-EVIDA, and data fusion tools.

1.2 Relationship with other deliverables and tasks

Table 2 Linkages between D4.3 and other ONCOSCREEN deliverables

Deliverable	Description of the deliverable	Link to D4.3
D4.1	Co-Designed System Architecture (First Version)	The collection of user and technical requirements are implemented in D4.1 which are addressed in this deliverable.
D2.1	Clinical Knowledge Database (First Version)	D2.1 provides a description of data that are needed for the various tools of this deliverable to operate.
D5.6	ONCOSCREEN Laboratory tests and evaluation	D5.6 closely relates to the current deliverable, since it contains the results and the adjustments/updates from the lab tests.

1.3 Deliverable structure

This deliverable is sectioned as follows: section 1 introduces the reader to the concept of the document, while section 2 reports on the steps taken towards an IIterative Co-Design approach with the participation of Technical Partners and End Users and the respective activities that have been adopted for each tool. Sections 3 to 7 elaborate on the architecture and provide an analysis for each tool, addressing End User and Technical Requirements, presenting the respective Innovation Aspects, the Functionalities and Design, Trustworthiness, while, lastly, reporting on Risks and Mitigation Measures. Section 8 focuses on the tools' performance in Lab Tests/Clinical Trials, while Section 9 is dedicated to the next steps for each tool. The document is concluded with a respective section, where an outlook of the work done is summarised.

2 Iterative Co-Design Activities with End Users

At the beginning of the project, workshops were held with the end-users so that their requirements could be collected and satisfied from the beginning of the development of the tool.

On the other hand, within the two LITs carried out to date for this deliverable, the feedback from the end-users as well as their questions after having shown the development, has been key to improving the current quality of the tool, and defining the next steps to follow.

Bilateral meetings with individual end-users were held in case not all feedback could be processed during the workshops and LITs.

A more analytical description in regards to the overall co-designed methodology that was followed is described in D4.1 Co-Designed System Architecture.

ONCO-RISTE

End-user requirements were collected at the beginning of the project along with the technical requirements. From that point, the development of the tool as well as the next steps have been subject to compliance with these requirements.

Communication with clinical-oriented end-users has been established to understand and improve how ONCO-RISTE would work for the LITs. Risk factor lists and classification of the users in groups have been reviewed.

Following the 2nd LIT, communication has been consistently upheld with various end-users with the objective of comprehending and enhancing the outcomes that have been showed up to now. Furthermore, rule base creation, obtained from expert insights, has been embarked upon. Consequently, this will lead to the development of two distinct versions of the tool: one that is semi-empirical and another that operates in an unsupervised way.

ONCO-CAWA

Each end user's requirements for ONCO-CAWA have been collected in a series of workshops with the clinical and the user organization partners. They have been tabulated in D4.1, while their fulfilment is discussed in section 4.1.1. It is obvious from the information there that many of the requirements have already been fulfilled to some degree. The resulting first version of ONCO-CAWA has been demonstrated in both the first and second laboratory integration tests.

Beyond the demonstration, the current version of ONCO-CAWA is now accessible via both Apple Store and Google Play. Currently a demo study is being implemented and users will be invited into it from the clinical and the user organization partners, for them to have a hands-on experience with the software and come back with suggestions.

Finally, a workshop has been carried out regarding the data collection requirements of ONCOSCREEN that ONCO-CAWA needs to cover, but these requirements are not finalized yet, neither are they implemented in the app, apart from some generic ones on capturing physical activities, sleep, heart and nutrition that need to be specialized and augmented.

ONCO-CLIDE

The collection of end-user requirements for the AI-based Clinical Decision Support System (cDSS) as part of the T4.5 initiative has been a thoughtful and measured process, focusing on aligning the tool's development with the practical needs of clinicians and medical experts in colorectal cancer (CRC) screening.

At the outset of the project, a foundational step involved the collection of end-user requirements alongside the technical specifications. This initial phase was crucial in setting a baseline understanding of what the medical professionals anticipated from the cDSS, ensuring that the system's development was in line with real-world clinical needs and expectations.

Throughout the development of ONCO-CLIDE, the team has maintained open channels of communication with various end-users. This has allowed for a steady flow of information, albeit limited in scope, regarding how the system might best serve in clinical settings. The insights gathered have been valuable in guiding the refinement of the system.

In this context, the development has proceeded with a focus on adaptability and responsiveness to potential user needs. Recognizing the variability in feedback, the approach has been to build a system that is flexible and capable of evolving as more user interactions and experiences are gathered over time.

The team has also been proactive in considering various scenarios and potential user preferences, leading to the development of different operational modes of ONCO-CLIDE. This includes versions that cater to diverse clinical environments and requirements such as roles and different versions for each clinical center, ensuring that the system remains versatile and applicable across various CRC screening contexts.

Overall, the co-design activities have been characterized by a careful balance between user-driven insights and proactive development strategies. The aim has been to create a cDSS that is not only technologically advanced but also prepared to adapt and evolve as more user feedback becomes available, ensuring long-term relevance and effectiveness in the clinical setting.

ONCO-EVIDA

All end-user requirements and suggestions were tabulated and compiled (see 5.2.1) by the ONCO-EVIDA technical partners.

VITO hosted three meetings (30/11/23, 15/12/23 and 19/01/24) with end-user partners designated to ONCO-EVIDA to clarify some of the URs, affirm or revise priorities, and leave room for additional wishes and feedback.

All end-users also received time to provide feedback during the Q&A sessions of the two LIT meetings. Moreover, a first version of the ONCO-EVIDA dashboard is available at http://91.184.203.22:88/.The end-users have the possibility to register and experiment with the initial functionalities that the dashboard provides. In addition, a set of datasets has been also included to demonstrate how the interaction with the graphs can be facilitated.

DATA FUSION

Data fusion has not received any specific end-user requirements yet. Currently data fusion is a research oriented task that aims to discover underlying risk factors from open source data. Any possible interesting results will be discussed with and evaluated from the end users. As the project progresses, the data fusion may apply advanced ensemble techniques to combine outputs of different tools and the respective user requirements may be adapted accordingly.

Regarding data lake that will be developed during the data fusion task, the development of which has just started, thus it has not received yet user requirements.

3 ONCO-RISTE Tool analysis and architecture

ONCO-RISTE is a semi-empirical stratification engine that aims to identify dependencies and reveal correlations among heterogeneous features such as demographic, behavioural, clinical and other variables, enabling the effective classification of individuals, both citizens and patients, into their respective risk-level.

The tool employs fuzzy and machine learning techniques to handle uncertainty, impreciseness, and vagueness in data interpretation. This approach allows for the finite interpretation of empirical rules from clinical experts in a probabilistic manner. This was the main reason on why fuzzy-based techniques were chosen.

ONCO-RISTE dynamically generates a 5 risk-level classification considering all factors, stressors, diagnostic results, and expert rules.

Addressing End-user and Technical Requirements

3.1.1 Addressing End User Requirements

Successful software development hinges on a thorough understanding of the end user's needs, preferences, and challenges. Through the end-user requirement process we identify, analyse, and incorporate users' requirements into the design and development phases. By prioritizing the end-user's perspective ONCO-RISTE tool ensures a seamless and user-centric experience.

Table 3 ONCO-RISTE end-user requirements

ID	Description	Priority	Fulfilment
RISTE_UR1	The tool should be able to classify the risk in a 5-level stratification.	High	Partially fulfilled. At the current state of the approach, 3-level stratification is completed due to the amount of available data. A 5-level is planned for the 2nd iteration.
RISTE_UR2	As user, I expect the tool to give me the time distribution of the risk based on my risk level (i.e., short vs long-term).	Medium	Planed for the 2nd Iteration.

RISTE_UR3	As an HCP user/patient, I expect to know the importance of the factors in the given risk level.	High	Expected to be in the next version of the approach.
RISTE_UR4	Tool should dynamically change the output if the patient changes the factors given.	High	Already satisfied. When the input of the user changes in ONCO-CAWA, a new calculation is made, and the risk level changes accordingly if needed.
RISTE_UR5	Tool must show the rules to the medical partners and give them the option to change them.	High	Expected to be in the next version of the approach.
RISTE_UR6	Tool must communicate the high-level information of risk levels with demographic data.	High	Planed for the 2nd Iteration. Connection with Data Fusion is prerequisite.

A total of six user requirements were collected, with the one related with the dynamic change of the output having been already satisfied at this state of the project. Moreover, ONCO-RISTE in the first year of the project managed to fulfil in total 2 of the requested requirements (**RISTE_UR01** and **RISTE_UR04**). More detailed information related with user requirements can be found in Deliverable D4.1.

3.1.2 Addressing Technical Requirements

Table 4 ONCO-RISTE technical requirements

ID	Description	Fulfilment
RISTE_TR1	ONCO-RISTE will be integrated with ONCO-CAWA	Fulfilled via Kafka SSL in LIT 2. ONCO-RISTE gathers patient's answers and then based on them calculates the risk score which then sends back to ONCO-CAWA.
RISTE_TR2	ONCO-RISTE will receive the semi-empirical rules from Retrospective CRC screening Data Collection & Analysis	Planed for the 2nd Iteration

RISTE_TR3	ONCO-RISTE will be integrated with data fusion service	Fulfilled via Kafka SSL in LIT 2. Risk calculation score is sent to Data Fusion.
RISTE_TR4	ONCO-RISTE will implement a classification process based on ML algorithms	Fulfilled. A fuzzy version of a clustering method is used as current approach.
RISTE_TR5	ONCO-RISTE output will be a risk-based score for each individual	Fulfilled. Json output with risk-based score for each patient
RISTE_TR6	ONCO-RISTE will be integrated with all the previous tools via the virtual data lake and not directly	Planed for the 2nd Iteration
RISTE_TR7	An aggregator service will be implemented in order to collect all the data and transform them to an ONCO-RISTE input	Fulfilled. A python back-end service was implemented that acts as a middleware that stands in the middle of the ml module and other tools
RISTE_TR8	The ONCO-RISTE output will be saved to a dedicated database	Planed for the 2nd Iteration
RISTE_TR9	Authentication and password management will be implemented	Planed for the 2nd Iteration
RISTE_TR10	ONCO-RISTE will integrate with ONCO-CLIDE	Fulfilled via Kafka SSL in LIT 2. Risk calculation score is sent to ONCO-CLIDE and is depicted in the UI
RISTE_TR11	All components will communicate via a secure network	Planed for the 2nd Iteration
RISTE_TR12	All components will communicate with secure communication protocols	Fulfilled. Use of Kafka SSL when for exchanging messages which is considered a secure communication protocol and secures data transmission over a computer network

A total of twelve technical requirements were collected and linked with the user requirements. Six of them have been classified as high priority, and 7 of them are already fulfilled (**RISTE_TR1,**

RISTE_TR3, RISTE_TR4, RISTE_TR5, RISTE_TR7 , RISTE_TR10, RISTE_TR12). More information about the technical requirements, their link with user requirements, priority, and other comments can be found in-depth in Deliverable D4.1.

ONCO-RISTE Analysis

Software tool analysis refers to the process of evaluating and assessing ONCO-RISTE capabilities and features. This analysis is crucial for making informed decisions about which technologies to use in software development, testing or any other aspect of the software development life cycle.

Key aspects of ONCO-RISTE include:

Functionality

- Identifying dependencies and correlations among heterogeneous features.
- Effective classification of individuals into risk levels.
- Utilizes fuzzy and machine learning techniques to handle uncertainty and impreciseness.
- Allows for finite interpretation of empirical rules in a probabilistic manner.
- Dynamically generates a 5 risk-level classification considering various factors.

Ease of Use

- ONCO-RISTE tool coding language is Python, which is widely known and used in the data science community.
- Employs Django rest framework for the ONCO-RISTE backend, providing a robust and scalable architecture.

Performance

- Dynamic risk-level classification suggests adaptability to various scenarios.
- The combination of fuzzy logic and machine learning techniques may enhance the accuracy of the final response.

Reliability and Stability

- Fuzzy logic and machine learning techniques contribute to robust risk-level calculations.

Overall, ONCO-RISTE analysis helped to make decisions about the selection and adoption of tools in the software development process, aiming to enhance productivity, efficiency, and the overall success of the task.

ONCO-RISTE Innovation Aspects

ONCO-RISTE provides a 3 risk-level classification (and soon will offer a 5-level classification) using heterogeneous information from different sources such as the mobile app (ONCO-CAWA), the Data Fusion tool, and other tools of the ONCOSCREEN ecosystem.

In terms of innovative features, the tool extends beyond providing a single-value classification. It assigns a probabilistic score to each category, thereby offering clinicians a more comprehensive understanding of the individual's condition. For the first time, the planned stratification will incorporate environmental and socio-economic data based on the proposed methodology suggested from the end user experts in D2.1.

Moreover, this classification is not solely dependent on data derived from other tools. It also incorporates the expert knowledge base of the clinicians. This knowledge base can be dynamically adjusted to disregard certain factors in the final computation or to assign more (or less) weight to specific factors. These adjustments do not necessitate subsequent model training, distinguishing this approach from traditional Machine Learning algorithms.

ONCO-RISTE Functionalities and Design

A high level view of the system architecture is shown in Figure 1. ONCORISTE consists of two components 1) the backend module and 2) the fuzzy component.

ONCO-RISTE backend component

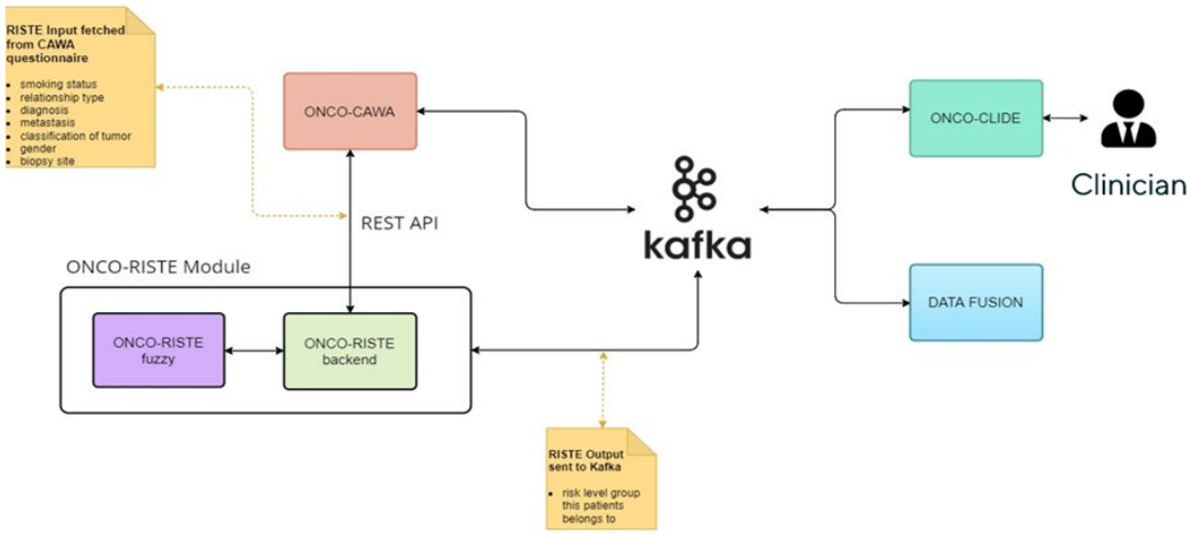


Figure 1 ONCOSCREEN System architecture.

The backend, implemented using the Django Rest Framework, serves as the middleware of ONCO-RISTE, orchestrating functions in data management, processing, and communication. The backend component gathers info from other tools, sourcing information from REST APIs provided by collaborative partners. It processes these incoming data, ensuring their integrity and compatibility for subsequent analysis by the fuzzy component. Functioning as the communication hub, the backend manages seamless interactions between the ONCO-RISTE tools and the fuzzy component. It facilitates a flow of information, ensuring that data from the tools are efficiently transmitted to the fuzzy algorithms for risk-level calculation. Upon receiving the risk calculation from the fuzzy component, the backend undertakes the task of transforming

the results into the agreed JSON format. Subsequently, it utilizes Kafka, a distributed event streaming platform, to transmit the processed results back to other ONCOSCREEN tools. This seamless integration through Kafka ensures efficient and real-time communication between ONCO-RISTE and its partner systems. Emphasizing security, the Django Rest Framework integrated into the backend ensures data protection during transmission and storage. Compliance with industry standards and best practices is prioritized to safeguard sensitive healthcare information. Furthermore, designed with scalability in mind, the backend ensures that the system remains responsive and efficient as the demand for ONCO-RISTE grows.

ONCO-RISTE fuzzy component

The fuzzy component of the tool is developed using Python. For the unsupervised version of the tool, Fuzzy C-means algorithm has been applied. Data is procured from the ONCORISTE backend module and undergoes a format verification prior to classification. Upon completion of the calculation the final outcome, which could be either a score-only output or a probabilistic output, is relayed back to the backend in JSON format. The backend module then assumes responsibility for facilitating communication with the remaining tools.

3.1.3 Functionalities

3.1.3.1 Receive patients' questionnaire answers from ONCO-CAWA

ONCO-RISTE is integrated with the ONCO-CAWA mobile application to consider user input in the risk calculation algorithms. This integration facilitates the extraction of valuable questionnaire answers from ONCO-CAWA, empowering ONCO-RISTE's fuzzy algorithm with pertinent input data for risk-level assessments.

Authorization Process

ONCO-RISTE initiates the integration by authorizing itself through a dedicated API, acquiring a JWT (JSON Web Token) for secure communication. The use of JWT tokens ensures secure authorization, safeguarding the integrity of data exchanged between ONCO-RISTE and ONCO-CAWA.

Token Utilization

Leveraging the obtained JWT token, ONCO-RISTE makes a subsequent request to the ONCO-CAWA API.

Questionnaire Data Retrieval

ONCO-RISTE gathers relevant questionnaire answers from ONCO-CAWA, extracting crucial information pertaining to the user's health status.

Fuzzy Algorithm Input

These gathered questionnaire answers serve as vital input for ONCO-RISTE's fuzzy algorithm, enhancing the precision and contextuality of the risk-level assessments.

3.1.3.2 Risk-level calculation

In the present methodology, the classification of individuals or patients into groups based on risk levels is accomplished through the use of a fuzzy clustering algorithm. This algorithm considers all pertinent information related to behavioural and clinical aspects gathered from the questionnaires. The output can be either a score-only output, which provides a singular risk level, or a probabilistic output, which offers the probability of the individual belonging to each of the groups.

Communication of risk-based score to other tools using Kafka-SSL

ONCO-RISTE extends its integration capabilities by securely transmitting its output to other tools, like ONCO-CLIDE or DATA-FUSION through Kafka SSL. This integration facilitates real-time collaboration by enabling the immediate transmission of the risk-level calculation and it establishes a reliable and encrypted communication channel. Kafka SSL ensures that all data transmitted are encrypted, safeguarding the confidentiality and integrity of sensitive health-related information. Furthermore, it ensures end-to-end encryption, significantly enhancing data security during transmission and minimizing the risk of unauthorized access, adhering to industry-standard security practices. Design

In Figure 1, along with the high level system architecture, an illustration of the architecture of the ONCO-RISTE tool is also presented. This visual representation provides a comprehensive overview of the system's structure, highlighting key components and their interactions. Referencing this figure will enhance understanding and facilitate discussions regarding the tool's design and functionality.

The specifics of ONCO-RISTE's design are intentionally omitted, acknowledging the intrinsic character of ONCO-RISTE as a backend tool. For a more detailed exposition of its design elements, please direct attention to Chapters 8.1 (ONCO-RISTE in the 1st LIT) and 8.6 (ONCO-RISTE in the 2ND LIT). These sections contain additional materials and print screens that comprehensively illustrate the tool's functionality.

3.1.4 ONCO-RISTE Data Flows

Please refer to deliverable D4.1 for a more detailed description of ONCO-RISTE data flow and how the tool stands to the whole ONCOSCREEN architecture.

In adherence to the 'Security by Design' philosophy, Kafka implementation and deployment completed by EXUS incorporates robust security measures, prominently featuring the integration of SSL (Secure Sockets Layer) for enhanced data protection and privacy. SSL serves as a cryptographic protocol, ensuring a secure communication channel between our Kafka brokers and clients.

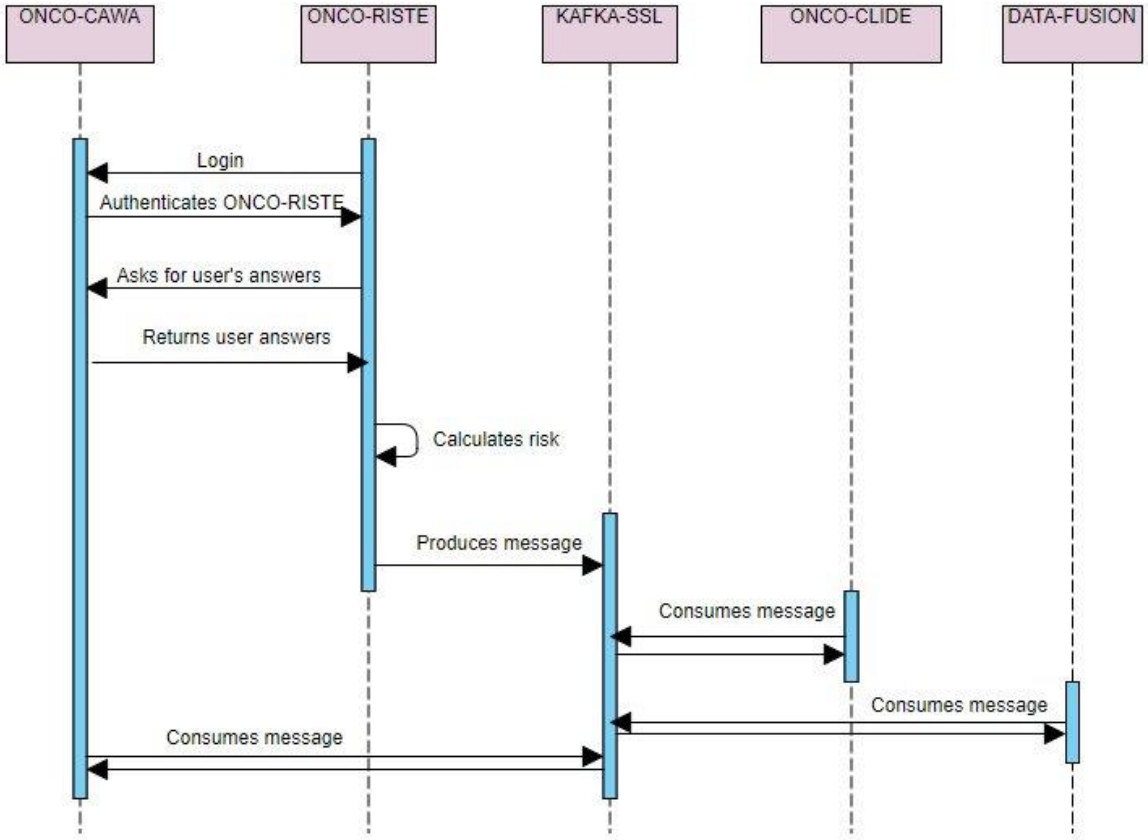


Figure 2 The ONCO-RISTE data flow.

By implementing Kafka SSL, we fortify our data pipelines against unauthorized access and eavesdropping, guaranteeing the confidentiality and integrity of the information exchanged within the Kafka ecosystem. This strategic inclusion aligns with our proactive approach to embed security measures at the core of our architecture, reinforcing our commitment to safeguarding sensitive data throughout the entire data streaming process.

In chapter 3.3 of deliverable D4.1 you can find details about Kafka SSL implementation, certificate management, and encryption protocols employed. This 'Security by Design' approach ensures that the data infrastructure not only meets industry standards but also establishes a resilient foundation for the secure and reliable exchange of information within ONCOSCREEN project.

ONCO-RISTE Trustworthiness

Dynamic Parametrisation of Fuzzy C-Means Algorithm

In the current approach a fuzzy clustering algorithm called Fuzzy C-Means is used. This algorithm is a fuzzy version of the C-Means clustering method where an instance, instead of belonging only to one cluster, could belong to one or more clusters with a membership degree. The higher the membership degree, the more common the instance is with other instances inside the same cluster. Advantages such as the flexibility among the results, having a membership degree to

each center and not a crisp value, and the ambiguous nature of the data, Fuzzy C-means is chosen as clustering mehod, surpassing other more known clustering methods such as K-means.

In the context of Fuzzy C-means, trustworthiness refers to the reliability of the clustering results, and it is determined by the optimal values of algorithm parameters. The most important parameter of the algorithm is the number of possible clusters. Although in the final approach, the number of clusters or groups to be used will be five, corresponding each one-to-one risk level, in the approach showed in the integration tests, in order not to set the number to a fixed value from the beginning, the Silhouette score is calculated. This value gives how good (or bad) an object has been classified, and can take values between -1 and 1, 1 being the best value possible.

Risks and Mitigation Measures

During all periods of the project, it is important to show that you have been aware of potential risks and that you have implemented mitigation measures to address them.

In this section, please align the risks with the ONCOSCREEN Risk Registry.

3.1.5 Risk 1 – Lack of initial data

Table 5 ONCO-RISTE Risk 1.

Risk Title	Lack of initial data	**Status**	Closed
Risk Category	Technical	**Related Task**	4.3
Risk Description			
The risk pertains to the potential inability to develop a reliable tool due to insufficient initial data. This could lead to inaccurate outputs, hindering the tool's effectiveness and usability.			
Responsible Partner	EXUS	**Period Identified**	Q2|2023
Probability		**Impact**	
Medium		Medium	
Mitigation Plan (Indicative)			
To address the initial data scarcity, we aim to identify open data registries associated with CRC. This will facilitate the creation of a preliminary tool, providing a foundation for further development.			
Comments			

Open data registry used: https://portal.gdc.cancer.gov/ . If further delays are observed the creation of synthetic data in collaboration with end user is planned.

4 ONCO-CAWA Tool analysis and architecture

ONCO-CAWA is the subjects' companion app, to be used in ONCOSCREEN studies for prospective data collection as well as analytics, feedback, and advice provision.

Addressing End-user and Technical Requirements

The end-user and technical requirements are detailed in D4.1. In the following sections we discuss their fulfilment.

4.1.1 Addressing End User Requirements

The fulfilment of the end-user requirements is tabulated in Table 5.

Table 5: Fulfilment of the end-user requirements.

ID	Description	Fulfilment
CAWA_UR1	As an ONCOSCREEN app user, I expect to have access to the devices necessary for the needed measurements.	Support for Garmin and Fitbit activity trackers, as well as any connected to Apple Health Kit has been implemented. Same for scales and other devices from iHealth.
CAWA_UR2	As an ONCOSCREEN app user, I expect the app translated to my native language.	The application supports translations to any language. Already translations for Bulgarian, Chinese Traditional, Dutch, English, Estonian, French, Greek, Italian and Spanish are implemented.
CAWA_UR3	As an ONCOSCREEN app user, I expect to be able to consent to the use of the app.	A mechanism of consents is already implemented that can be extended with as many different items as needed.
CAWA_UR4	As an ONCOSCREEN app user, I expect to consent to the processing of my data by ONCOSCREEN partners.	A mechanism of consents is already implemented that can be extended with as many different items as needed.
CAWA_UR5	As an ONCOSCREEN app user, I expect to read a privacy policy,	A mechanism of consents is already implemented that can be extended

	explaining to me why my data is being collected, who it is going to be shared with, how it is protected, what are my rights in this regard, before I give my consent.	with as many different items as needed.
CAWA_UR6	As an ONCOSCREEN app user, I need to be certain that my data is securely protected from unauthorized access, loss, destruction, modification.	The app is a Class I medical device, certified under the Medical Device Directive.
CAWA_UR7	As an ONCOSCREEN app user, I expect to read EULA/terms of use.	A configurable mechanism of consents has been implemented. It can be extended with as many different items as needed.
CAWA_UR8	As an ONCOSCREEN app user, I expect personalized recommendations to be delivered to me.	Mechanisms have been implemented for notifications (short messages) and dialogue (long interactive messages) delivery.
CAWA_UR9	As an ONCOSCREEN app user, I expect it available both for Android and iOS.	CAWA can be downloaded from both the Apple Store and Google Play.
CAWA_UR10	As an ONCOSCREEN app user, I expect it to have an intuitive UI.	CAWA is being implemented on top of the Healthentia mobile app that has been used by more than 1,500 users.
CAWA_UR11	As an ONCOSCREEN app user, I expect its UI to adapt to the expectations of my age group / personality.	No such mechanism is currently in place. Planned for the 2nd iteration.
CAWA_UR12	As an ONCOSCREEN app user, I expect it to offer gamification features.	Currently users receive encouraging messages when they reach goals or answer questionnaires. This is to be expanded into a proper gamification mechanism.

CAWA_UR13	As an ONCOSCREEN app user, I expect it to show data from wearable activity trackers, either devices (e.g., Fitbit, Garmin, Polar) or services (e.g., Apple Health Kit, Health Connect).	Activity tracking is supported from Fitbit, Garmin and mobile phone's sensors. Also, any source integrated with Apple Health Kit is supported. Extensions are planned for more sources.
CAWA_UR14	As an ONCOSCREEN app user, I expect it to be using questionnaires for self-reporting PROM/PREM.	A questionnaire delivery and answer mechanism is already implemented. Questionnaires can be entered and updated using the Healthentia portal web app.
CAWA_UR15	As an ONCOSCREEN app user, I expect it to visualise data from diagnostic devices.	A screening tool visualization widget is being designed.
CAWA_UR16	As an ONCOSCREEN app user, I expect it to show CRC prevalence analysis from ONCO-EVIDA.	EVIDA analysis results will be shown in the screening tool visualization widget.
CAWA_UR17	As an ONCOSCREEN app user, I expect the app to present to me my risk results from ONCO-RISTE.	RISTE analysis results will be shown in the screening tool visualization widget.

4.1.2 Addressing Technical Requirements

The fulfilment of the technical requirements is tabulated in Table 6.

Table 6: Fulfilment of the end user requirements.

ID	Description	Fulfilment
CAWA_TR1	CAWA will be written in React Native, to support both iOS and Android devices from a single code base.	CAWA is written in React Native and already downloadable for Android and iOS from the relevant app stores.
CAWA_TR2	The app will use minimum text. All messages will be translated to the necessary languages.	CAWA is implemented following the guidelines for mobile applications supporting translations. It is already offered in Bulgarian, Chinese Traditional, Dutch, English, Estonian,

		French, Greek, Italian and Spanish. Any other language can be supported.
CAWA_TR3	CAWA will be integrated with devices for automatic data entry via the manufacturers' SDKs.	The integration of CAWA with the Apple Health Kit, iHealth and Fora via their respective SDK integrations.
CAWA_TR4	CAWA will be integrated with devices for automatic data entry via the manufacturers' APIs.	The communication between CAWA and Garmin & Fitbit via their respective APIs has been implemented.
CAWA_TR5	CAWA will be accepting manual input for devices that do not offer any radio communications.	Widgets have been implemented in CAWA to accept manual input for weight, blood pressure, blood glucose and SPO2 measurements.
CAWA_TR6	CAWA will be offered with an intuitive UI comprising single-purpose widgets.	The CAWA home page is already implemented as a collection of widgets.
CAWA_TR7	Both the main screen and the widgets of CAWA will have minimum text, based mainly on icons.	The CAWA homepage is implemented featuring icon-heavy widgets, and the widgets themselves have graphs and data entry controls.
CAWA_TR8	Application branding based on ONCOSCREEN.	A mechanism for branding studies is being finalized. This will allow CAWA to sport the banner, logo and color theme of the ONCOSCREEN project
CAWA_TR9	CAWA UI will adapt to age group expectations.	Currently not supported. Planned for the 2nd iteration.
CAWA_TR10	Gamification features (goals & rewards) will be promoting CAWA usage.	Currently a notification-based mechanism has been implemented. This is to be expanded into a proper gamification mechanism.
CAWA_TR11	CAWA will be contacting the users with the recommendations	Both notification and dialogue delivery mechanisms have been

	generated by CLIDE. These should be easy to understand, trying different approaches for the same end-goal, attempting to match users' personality.	implemented. They can be used for delivering information derived by other ONCOSCREEN components. The content of the information is the responsibility of those components.
CAWA_TR12	CAWA will be providing optimum timing of messages to users, based on their app usage history.	CAWA asks users to accept having their use of it tracked. Only in that case can models of interaction been built and are consequently used to optimize interaction time.
CAWA_TR13	CAWA will be able to forward questionnaires to users, preferably validated ones.	The mechanism to present rich-text questionnaires to the users, with the questions being answered using a multitude of widgets has been implemented. The questionnaires used are the responsibility of the study design side, not CAWA.
CAWA_TR14	CAWA will be offering proper visualizations for the information collected from the screening tests.	A new widget will be designed and implemented, offering the information from the screening tests.
CAWA_TR15	CAWA will be offering proper visualizations for the information collected from activity tracking.	The physical activity and sleep widgets are implemented, offering visualizations of the data from activity tracking sources.
CAWA_TR16	CAWA should be able to get and offer users the information generated by RISTE in a comprehensive way.	The same widget for the screening tests will accommodate the risk assessment.
CAWA_TR17	CAWA should be offering the users the terms of use, EULA, privacy policy & consents, recording their acceptance.	A configurable mechanism for consents has been implemented. It will be tailored to the specific needs of ONCOSCREEN.
CAWA_TR18	CAWA should follow medical device standards to ensure data security.	CAWA is built upon the Healthentia mobile app. Healthentia is a CE-marked medical device class I,

		certified under MDD. This certification covers all data security concerns.
CAWA_TR19	CAWA is offering the collected information to the data fusion via the Healthentia API.	An API exposed by CAWA is implemented, offering all collected information to an authorized 3rd party software via its endpoints.

ONCO-CAWA Analysis

To be a usable subject companion app that is capable of collecting information and displaying advice, ONCO-CAWA has not been designed from scratch. Instead, it is based on Innovation Sprint's Healthentia. Healthentia is a certified software device, partially fulfilling both the data collection and advice delivery needs of ONCOSCREEN.

Thus, the implementation followed the process outlined below:

- Matching of the technical requirements of ONCO-CAWA to the functionalities offered by Healthentia and performing a gap analysis.
- Modifying Healthentia to fully support functionality that is now partially supported.
- Extending Healthentia to provide new functionality.
- Repeating the previous two steps while involving ONCOSCREEN end users to maximize ONCO-CAWA acceptance probability.

Having followed this process thus far has led to a version of ONCO-CAWA presented via its functionalities in section 4.4. While still not at a level to be used in the ONCOSCREEN pilots, it is a mature app soon to be distributed to the partners for usage and feedback.

ONCO-CAWA Innovation Aspects

ONCO-CAWA is a subject companion app, addressing the dual need of collecting information from the subject and informing them. Although there is quite a bit of software engineering involved in its implementation, the real innovation lies in two aspects. Firstly, the decisions reached by ONCOSCREEN on what data is needed by the different systems and is thus collected by ONCO-CAWA. Secondly, on the advice it can deliver to its users via other ONCOSCREEN components.

ONCO-CAWA Functionalities and Design

The modules, implementation, and functionalities of ONCO-CAWA have been outlined in D4.1. In this section we will show how the functionalities have been addressed in the current version of ONCO-CAWA.

4.1.3 Functionalities

The different ONCO-CAWA functionalities are demonstrated by including actual screenshots from the working version of the application.

4.1.3.1 Registration & login

There are two ways to register to Healthentia, and hence ONCO-CAWA. The simplest way is to have a study investigator already pre-enrolled a subject, whereupon the subject proceeds logging in using the credentials they are provided. The process is shown in Figure 3. The subject taps on "Login" and find themselves in the credentials page.

Alternatively, the subject uses a registration code. Either the subject is already given a code from the investigators inviting them to use ONCO-CAWA, or the subject requests one. The process is shown in Figure 4. The subject taps on the "Request Code" link and lands on the request page with a form to ask for one, making sure to indicate ONCOSCREEN. If they have a code or receive one following the request process, they enter it and tap "Continue." Then, they land on the account creation page. Following account creation, the email verification process kicks in, as shown in Figure 5. The subject lands on the email verification page, whereupon they need to check their email client (and the spam folder) and click on the "Confirm email" link, to see in their browser the successful verification page. They can then proceed with the login.

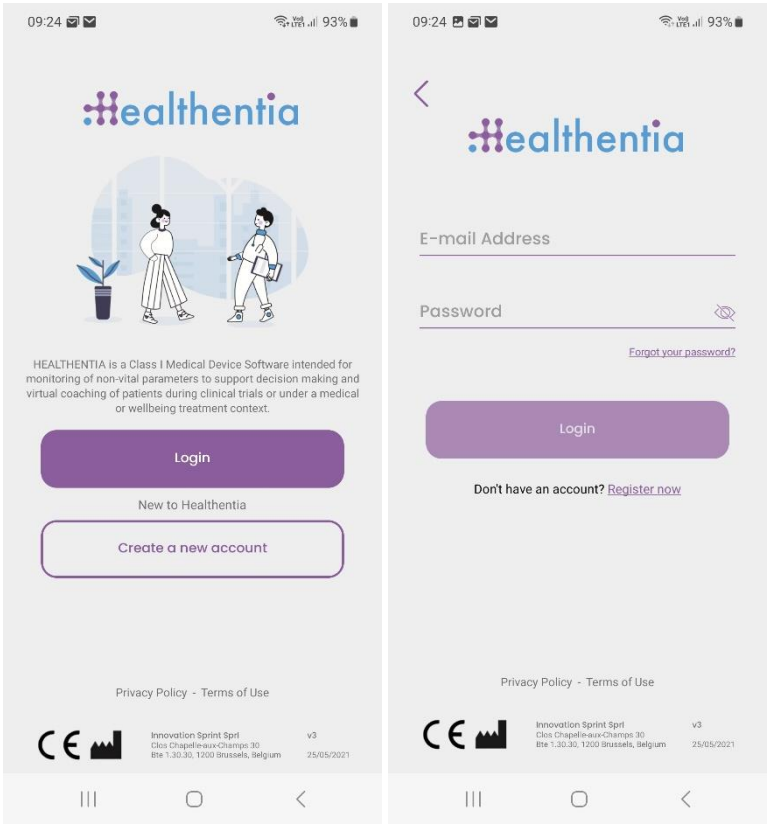


Figure 3: Process for pre-enrolled subjects. Initial and login pages.

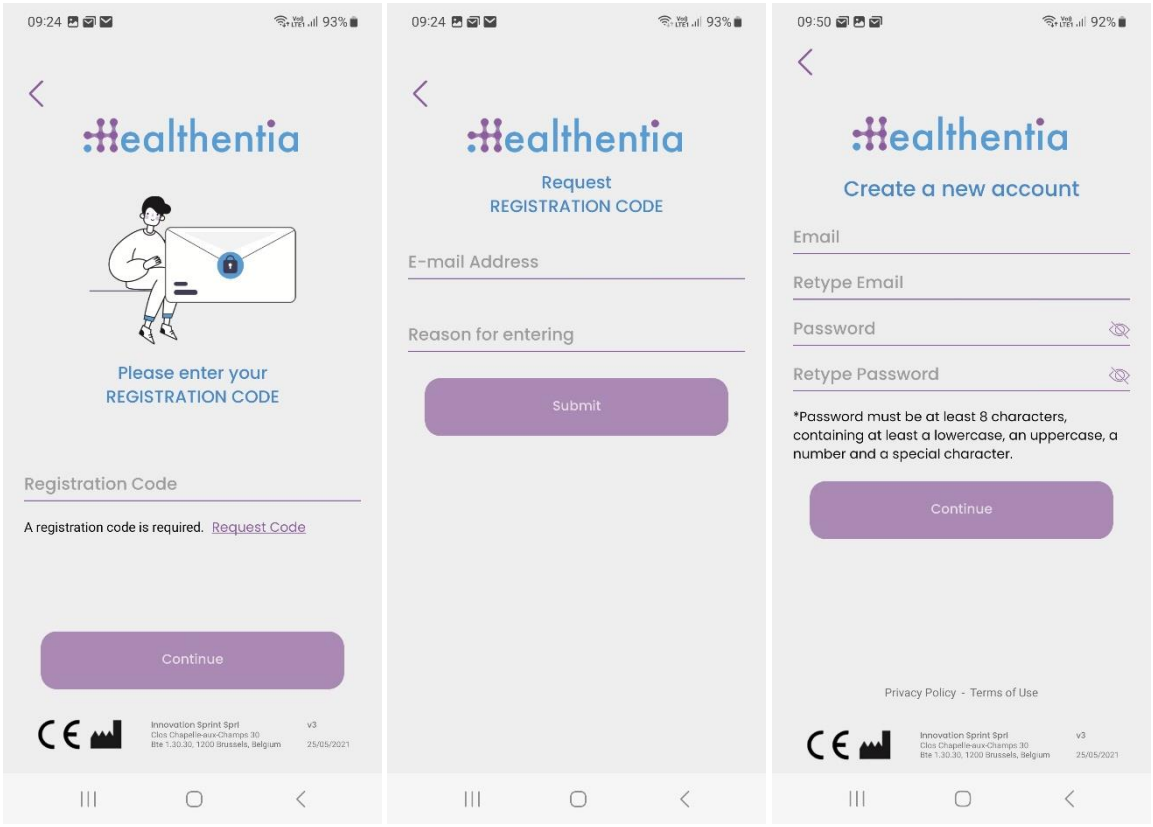


Figure 4: Process employing registration codes. Registration code, code request and account creation pages.

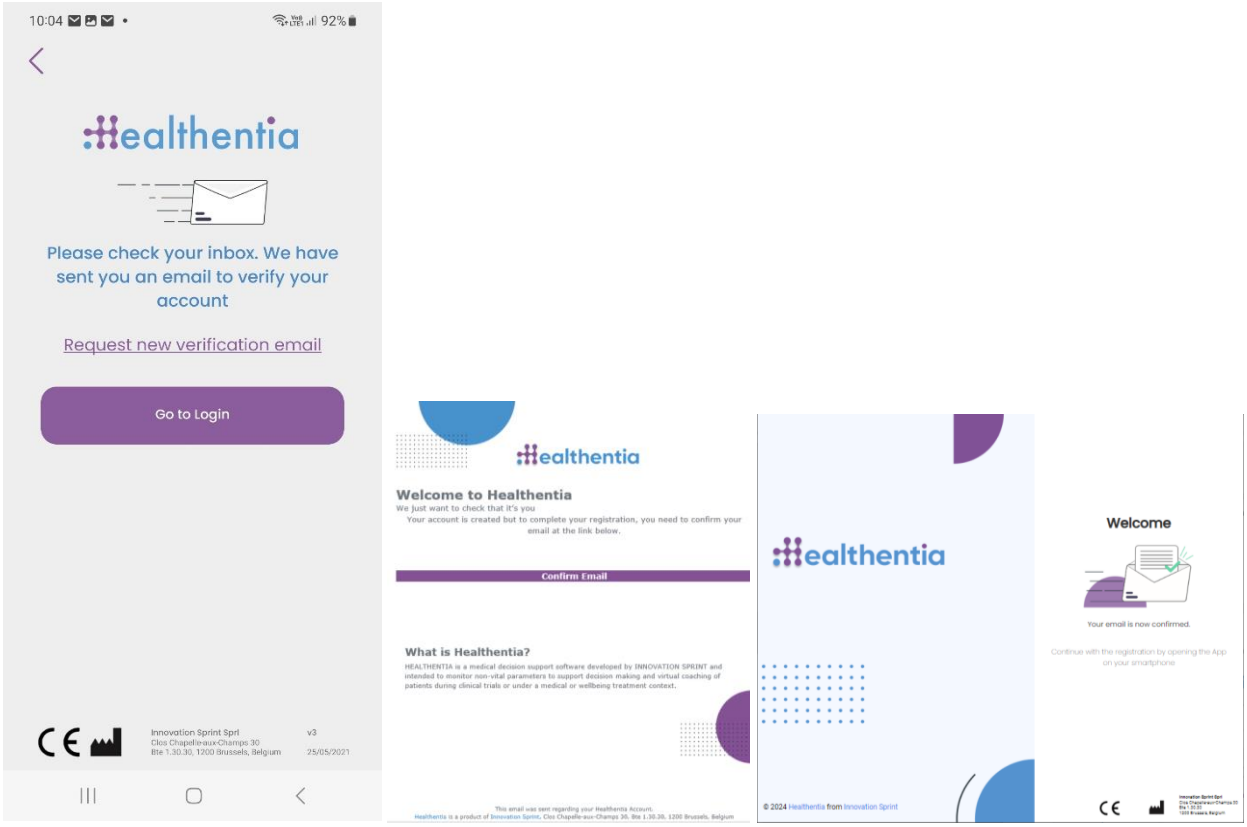


Figure 5: Email verification process. Verification page in the app, verification email and successful verification page in the browser.

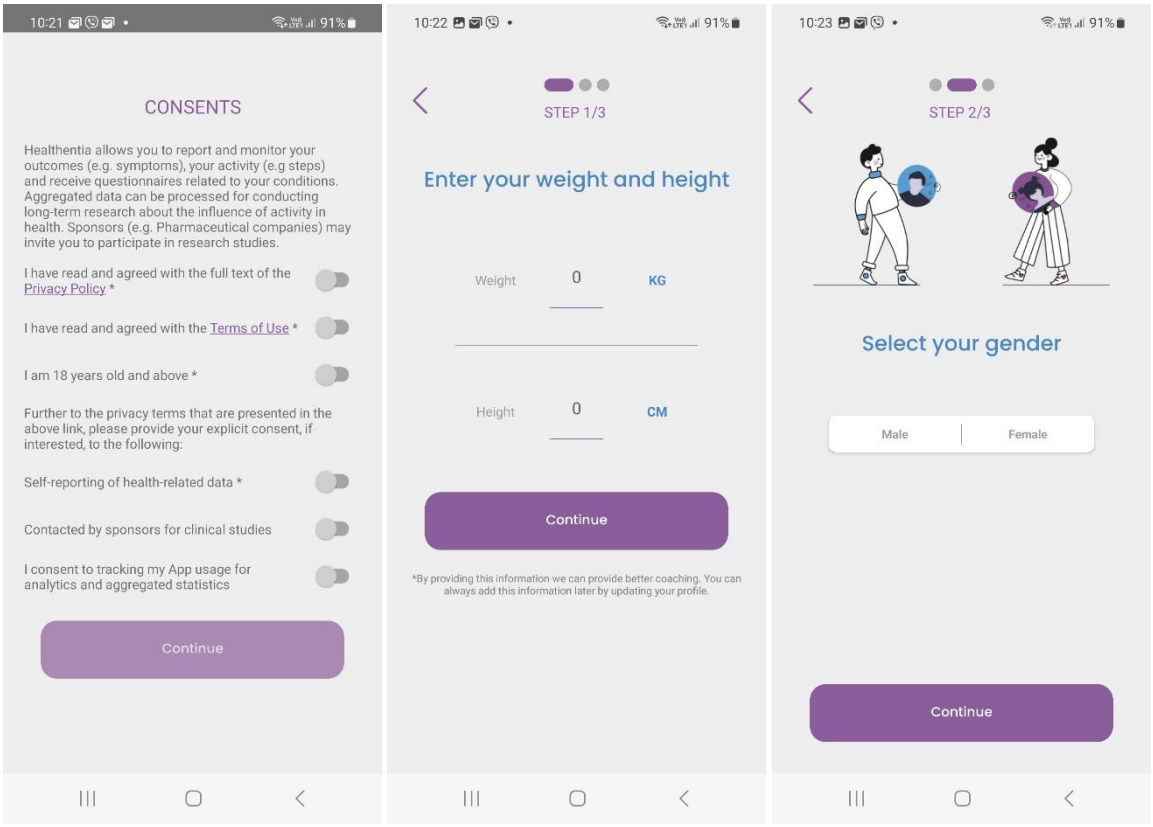


Figure 6: First time login process. Consent, body metrics and gender pages.

Up to this point the mobile app is agnostic to ONCOSCREEN. After login, it knows its user is an ONCO-CAWA user participating in some ONCOSCREEN study, hence the app is transformed into the ONCO-CAWA one, following the branding of ONCOSCREEN. The branding currently affects the primary and secondary colours, an ONCOSCREEN banner and the ONCOSCREEN logo. In the next version, more branding will be applied as explained in the following paragraphs.

No matter the registration process followed, after the first login the subject follows through the first login process shown in Figure 6. Initially they land in the consents page. The consents are dynamically built, so the page will be branded with ONCOSCREEN specific information in the next version. The subject needs to accept every item marked with an asterisk. A very important optional consent is that of tracking the app usage. If it is not accepted, then ONCO-CAWA will not be collecting data to build an app usage model for the subject, and offer personalized timing of the notifications.

The following three pages of the first-time login process are about entering the subject's weight and height (body metrics page), gender (gender page) and date of birth (date of birth page). The body metrics and date of birth pages will also be made dynamic in the future to allow for less detailed information if the ONCOSCREEN study participants decide so. E.g., the weight and height could be rounded, or more importantly the date of birth could be replaced by just year of birth.

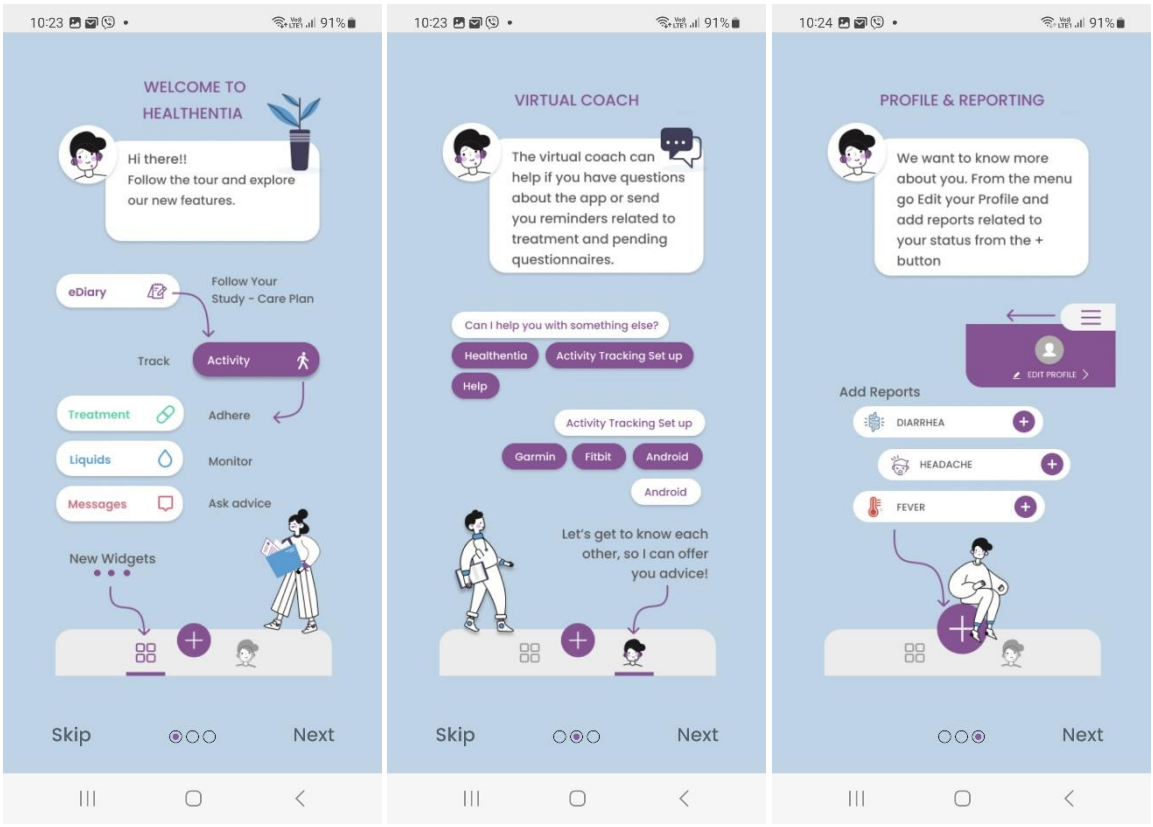


Figure 7: Mini manual pages.

Finally, the subject goes through some screens of a mini manual as shown in Figure 7, that will also be branded in the next version.

4.1.3.2 Home page

After login, the subject lands in the home page, shown in Figure 8. This is where ONCO-CAWA diverges from the parent Healthentia application and its ONCOSCREEN features are experienced. The page consists of different widgets, serving to either collect data from the subject, inform them, or do both.

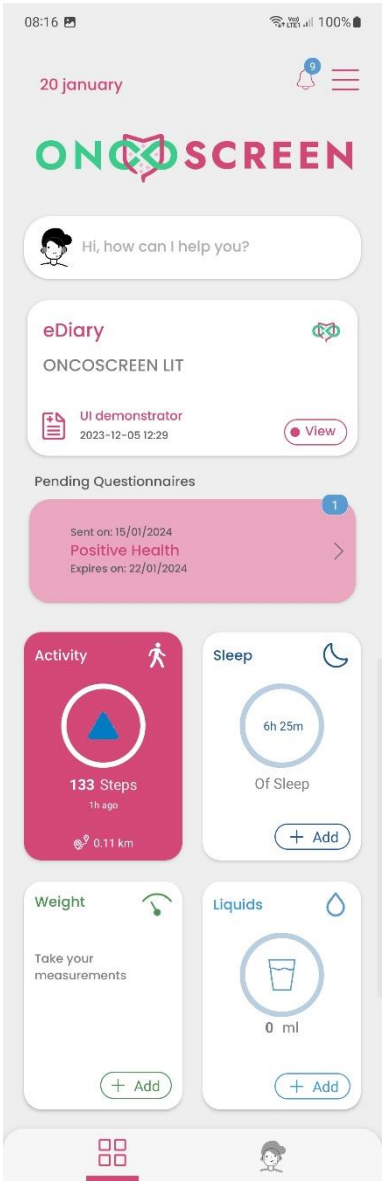
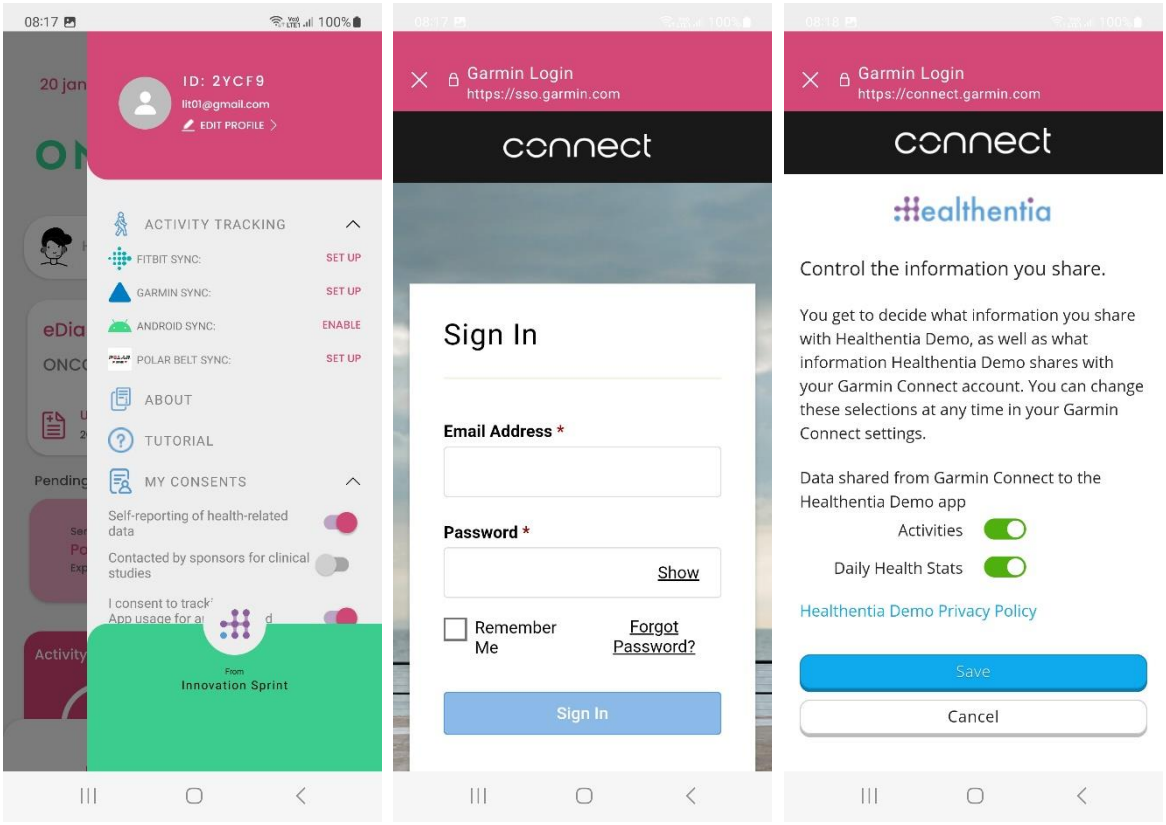


Figure 8: Home page.

4.1.3.3 Connection of activity trackers

After first login, the subject should connect their activity tracker. To do so, they access the sidebar via the top right drawer icon, see Figure 9, first screenshot. There are currently four options. Three of them are for the Fitbit, Garmin, and Polar activity trackers. The fourth option depends on the OS of the phone. For Android there is Android Sync, which uses the phone's sensors, while for iOS there is Apple Health, which uses whatever source connected to it.

Selecting e.g. Garmin, the subject goes along Garmin's process for connecting their Garmin account to ONCO-CAWA, including entering their credentials, accepting what is to be shared with ONCO-CAWA and agreeing to terms, as shown in the 2nd to fourth screenshots of Figure 9. After the completion of the process, the subject can verify that Garmin is connected, see final screenshot of Figure 9. Taping again will disconnect Garmin.



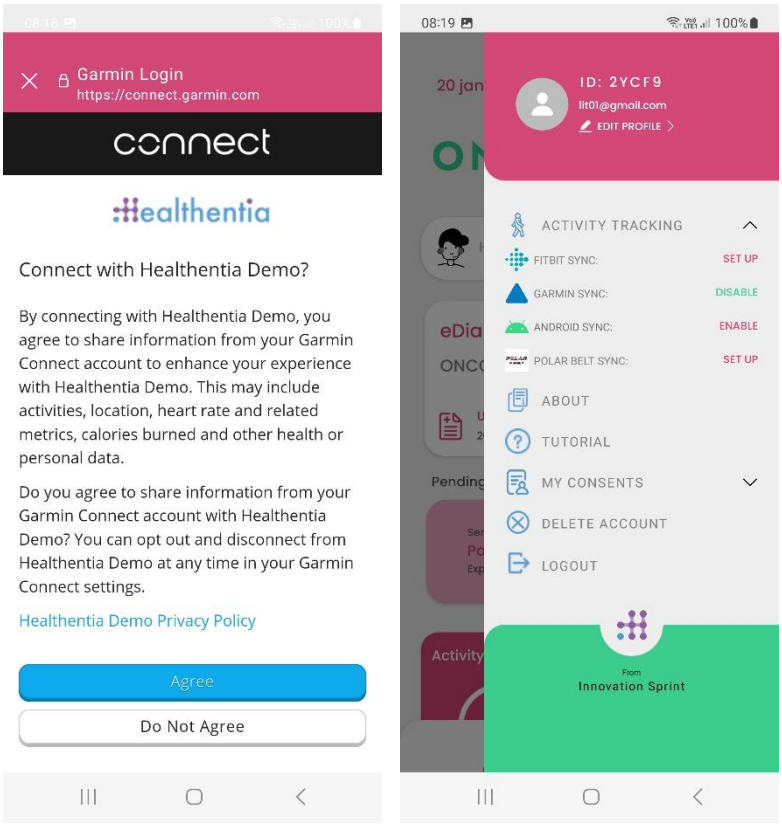


Figure 9: Activity tracker connection - Garmin.

4.1.3.4 Profile

Next the subject can visit their profile page, by tapping on the user icon on top of the sidebar. Initially the profile has just the mandatory information collected at the registration process, as shown in Figure 10 (left). The subject can opt to enter any other data they want, resulting to more complete info as shown in Figure 10 (right). Such non-mandatory info can be used to personalize dialogues offered by ONCO-CAWA or allow ONCOSCREEN study investigators contact the subject.

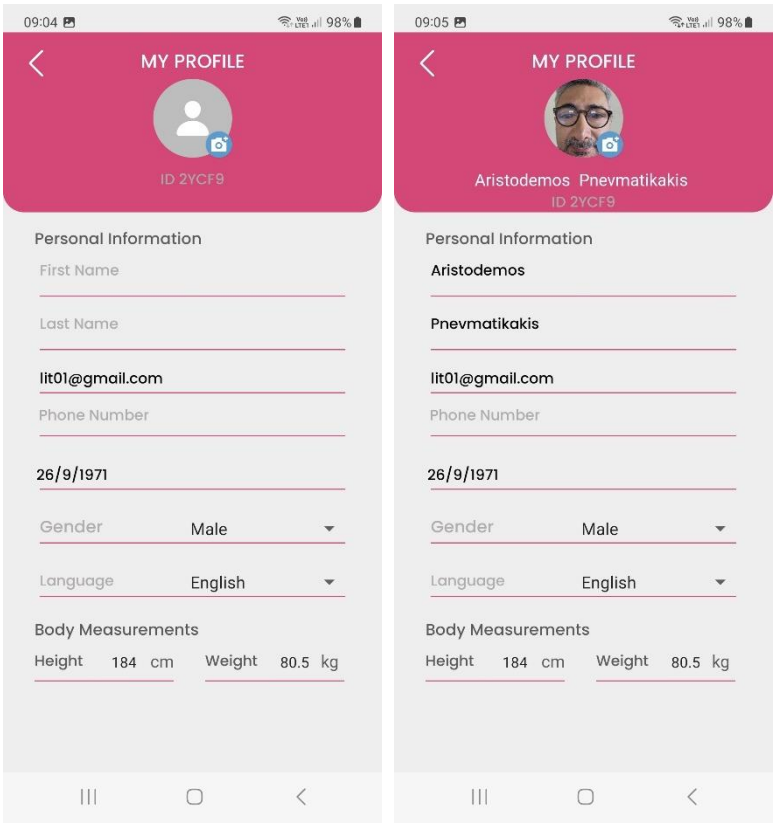


Figure 10: Profile page.

4.1.3.5 Data collection widgets

There are several data collection widgets on the home page, detailed below.

The physical activity widget primarily displays steps walked, but also floors climbed, calories burned and minutes of moderate or intense activity. All this info comes from the activity tracker. There are daily, weekly, and monthly views, as shown in Figure 11.

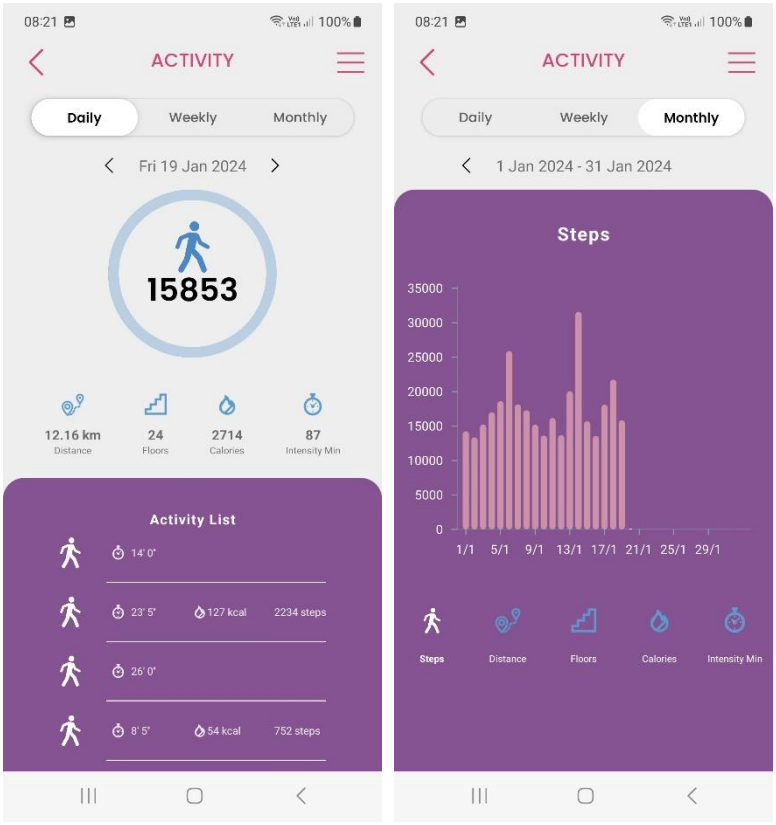


Figure 11: Physical activity widget.

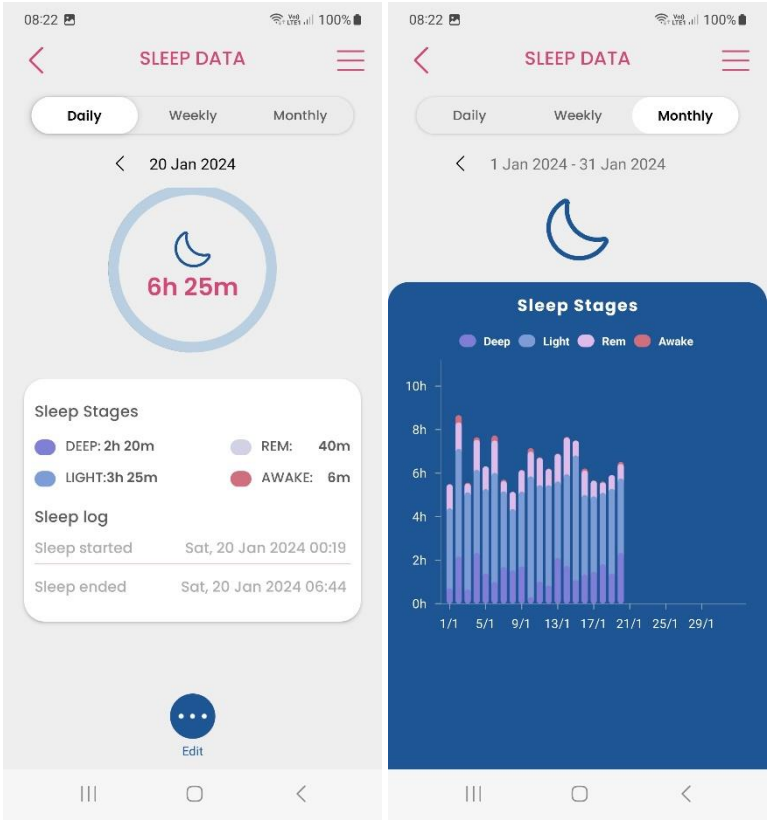


Figure 12: Sleep widget.

The sleep widget displays time spent sleeping, dividing it to the light, deep, REM and awake phases. The division is only possible if the information source is an activity tracker. Manual input of time to bed and time awaken is also possible. There are daily, weekly, and monthly views, as shown in Figure 12.

The heart rate widget also shows information from the activity tracker. The daily, weekly, and monthly views display the resting and the maximum heart rate of the selected period, as shown in Figure 13.

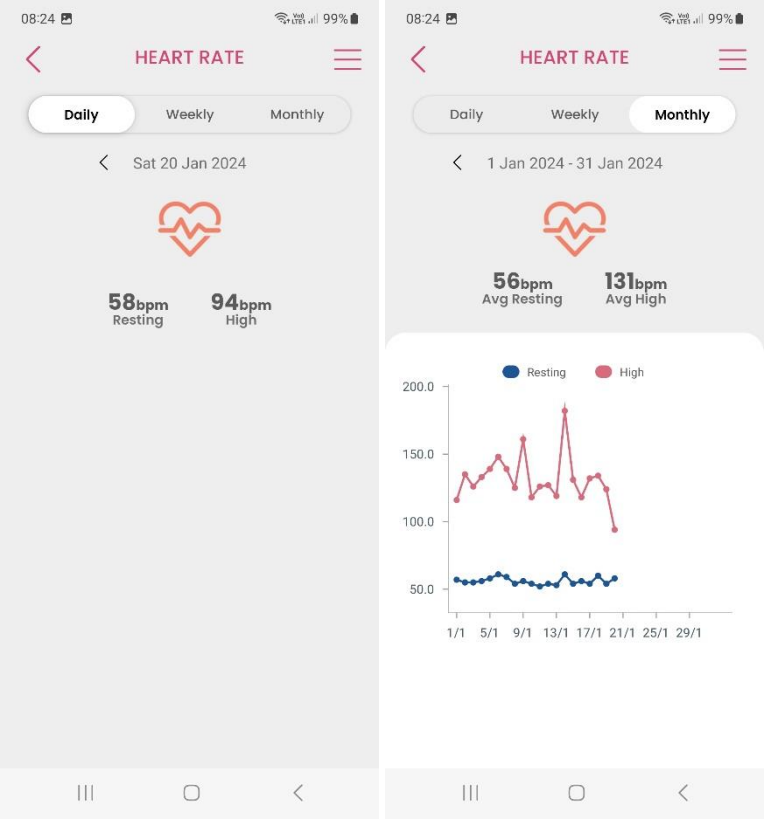


Figure 13: Heart rate widget.

The weight widget allows manual entry of weight, but also supports some Bluetooth enabled scales. Weight entry is shown in Figure 14. The entered information can be viewed in the daily, weekly, and monthly views.

The liquids widget allows manual entry of liquid consumption. Currently the primary category is water, followed by alcohol, beverages, coffee, and tea, see Figure 15. The liquids of interest might become selectable in a next version. The information is shown per day, and the subject can select any date to add any missing logs of liquids consumption.

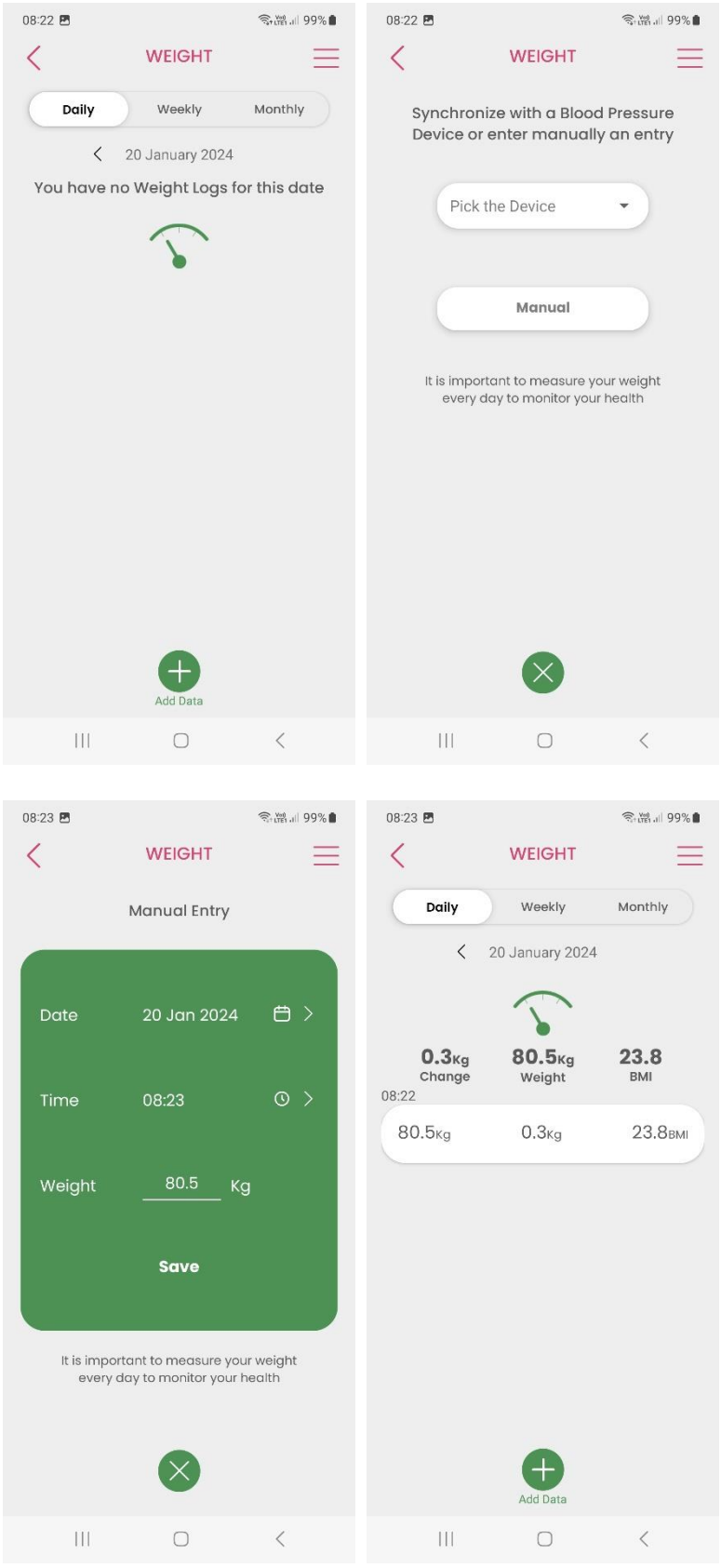


Figure 14: Weight widget.

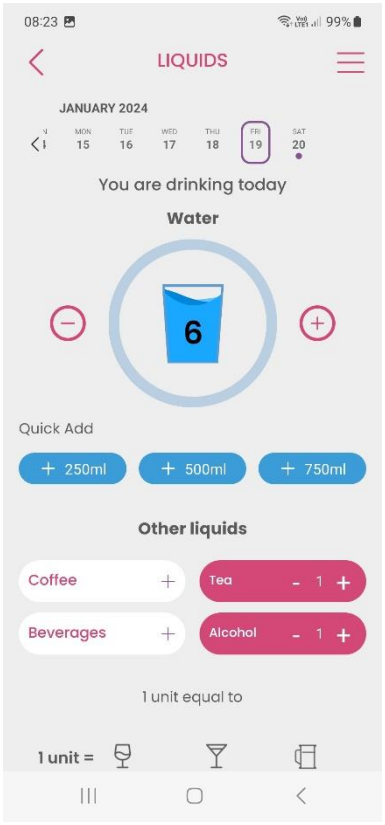


Figure 15: Liquids widget.

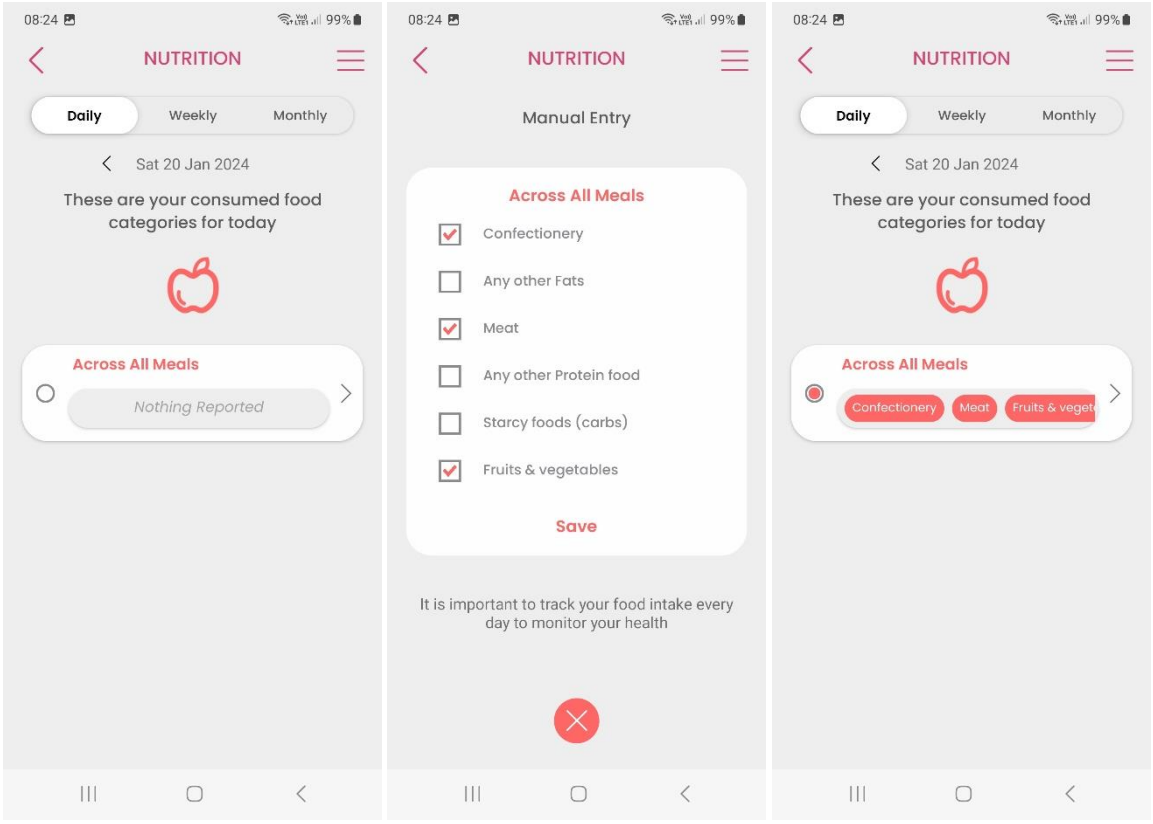


Figure 16: Nutrition widget.

The nutrition widget allows manual entry of food consumption. The food categories presented to the subject are configurable by the study investigator, so the categories shown in Figure 16 are just an example. The information is shown per day, and the subject can select any date to add any missing logs of food consumption.

More widgets might be implemented as the needs of ONCOSCREEN are consolidated. One widget already under design is the clinical widget, where the subject will be able to enter selected results of the clinical tests. The selection will be carried out by the study investigator, but the different screening and risk prediction tools of ONCOSCREEN will certainly be selected. The data entry will be either manual, or automatic, ingesting data from ONCOSCREEN's Kafka topics. An initial such ingestion mechanism, without any user interface, for risk prediction is shown in Figure 17.

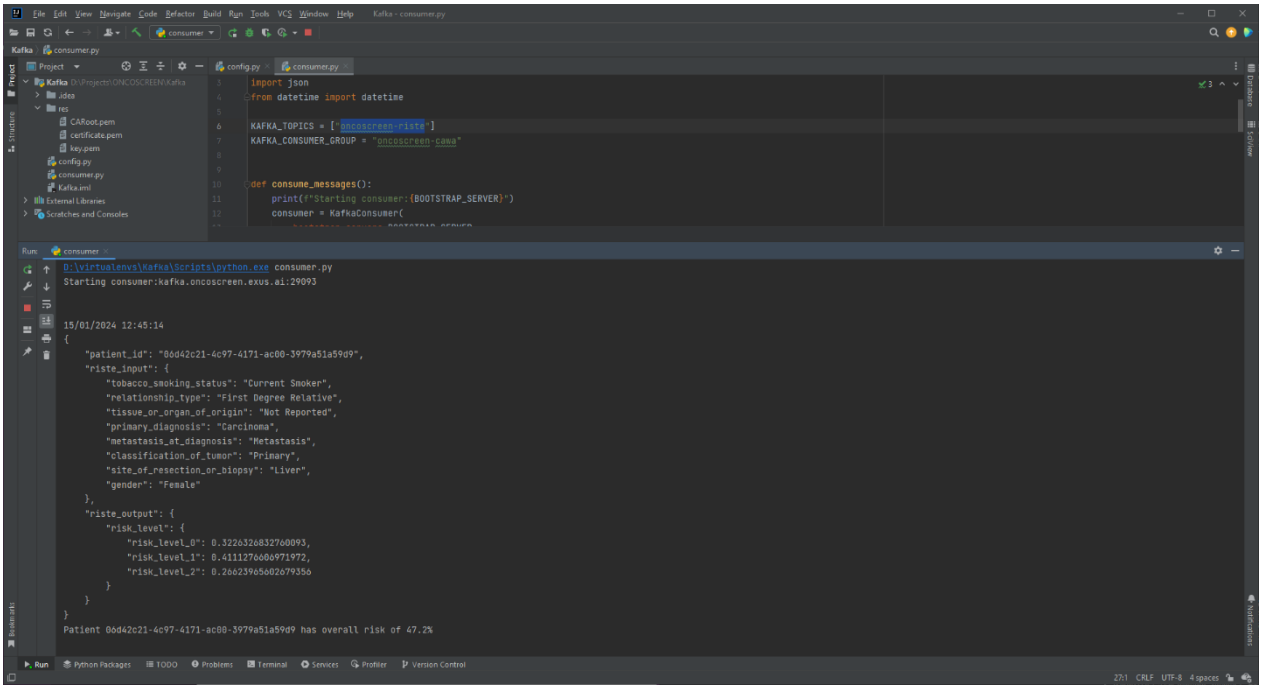


Figure 17: Ingesting Kafka messages for automatic data inclusion in ONCO-CAWA.

4.1.3.6 Questionnaire answering

Questionnaires are pushed to the subjects following the schedule selected by the study investigator. When new questionnaires are waiting for answer, the questionnaire widget appears on the home page, see Figure 18.

Scrolling to any of the pending questionnaires and tapping on it opens the questionnaire for answering. The process is shown in Figure 18. There are many different UI controls to facilitate answering, ranging from simple ones like option selection, numeric and text entry, to elaborate ones like uploading images, or selecting parts in body maps.

Answered questionnaires can be viewed via the eDiary widget, where all the questionnaires of a selected week are listed. Tapping on a list item displays the questions and provided answers, as shown in Figure 19.

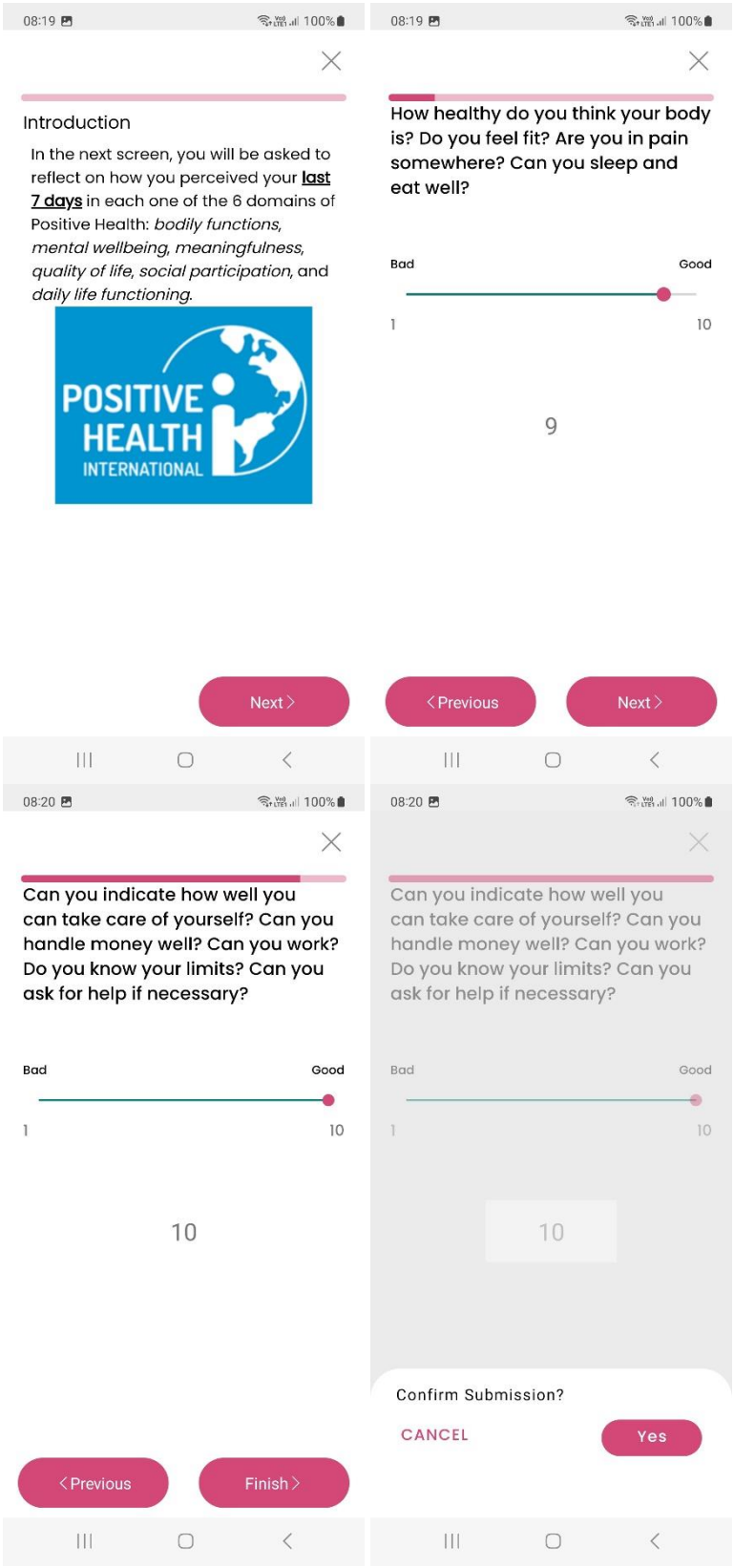


Figure 18: Answering a questionnaire.

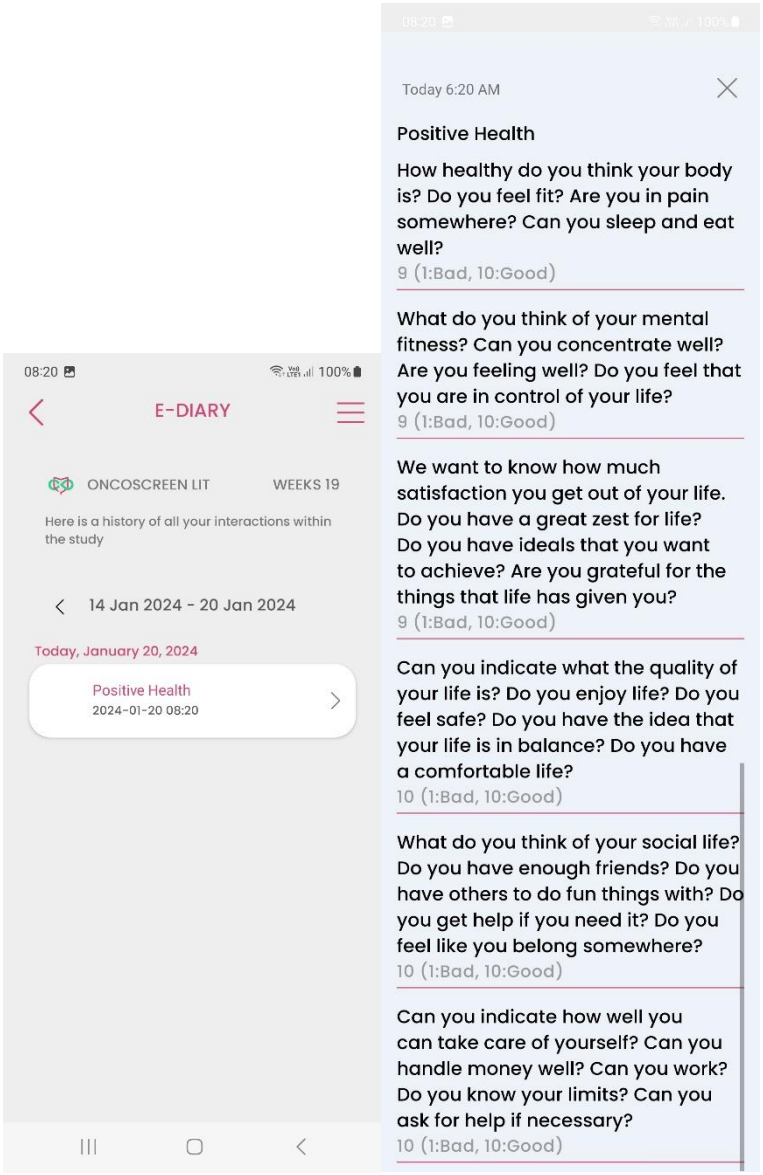


Figure 19: Browsing answers in the eDiary widget.

4.1.3.7 Virtual coach

The virtual coach is the ONCO-CAWA element for delivering advice from ONCOSCREEN to the subject. The coach can be accessed via the widget in the home screen to embark in the pre-defined dialogue on providing help with the app, or the user can select the dialog they have just been notified for. Notification warnings appear as numeric bobbles in the bell icon of the home screen. Tappin on the bell, the notification page opens, shown in Figure 20 (left). Selecting an unread notification for a dialogue (designated by the coach face-like icon), the subject runs through the dialogue, reading the messages and selecting one of the possible responses, as shown in Figure 20 (centre and right).

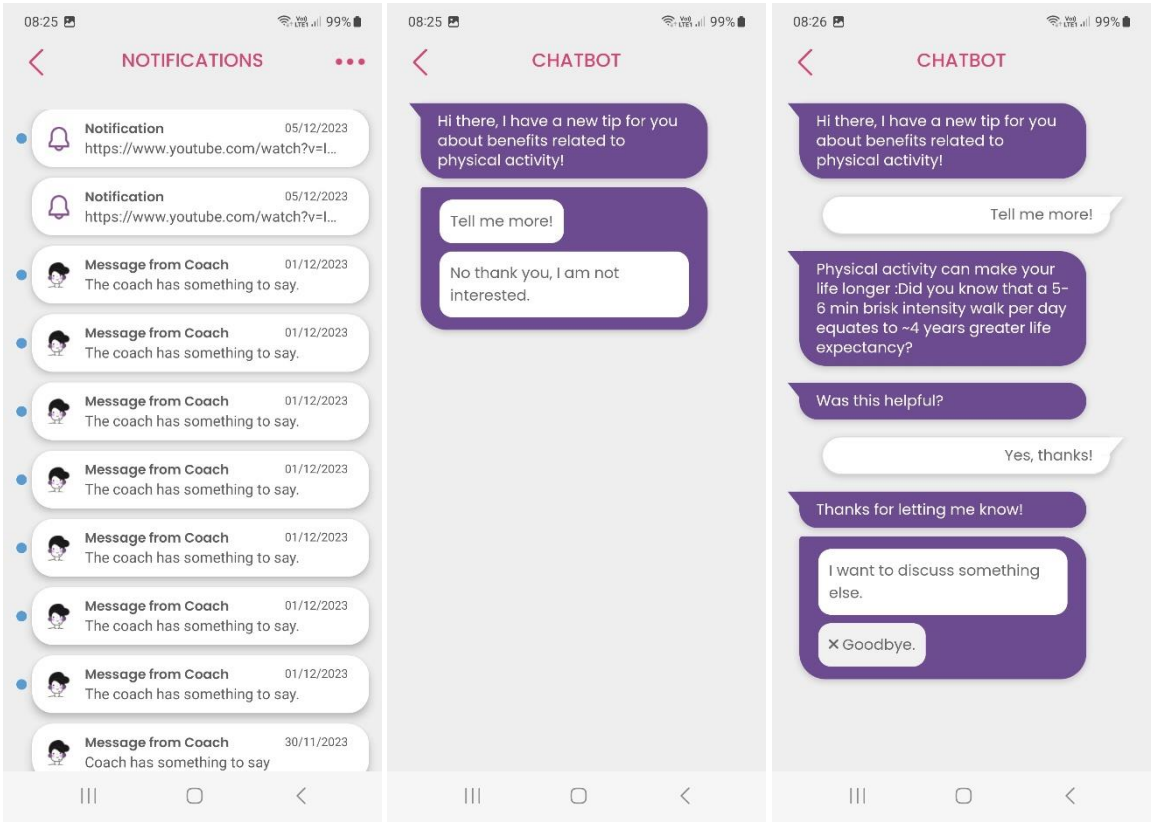


Figure 20: ONCO-CAWA notifications and running a dialogue via the virtual coach.

4.1.4 ONCO-CAWA Data Flows

ONCO-CAWA is a mobile app, backed by the backend of Innovation Sprint's Healthentia. The backend provides the API to communicate both with the mobile app, but also with the rest of ONCOSCREEN, to offer data. This is facilitated by a data ingestion service. Data is collected by ONCO-CAWA employing different mechanisms:

- The subject manually enters data in ONCO-CAWA.
- Devices have their data ingested by ONCO-CAWA when integrated via an SDK.
- Devices have their data ingested by Healthentia backend when integrated via an API.
- ONCOSCREEN data (from screening tools and risk predictors) is ingested by Healthentia backend from the Kafka topics via a Kafka ingestion service (1st option)
- ONCOSCREEN data are centrally sent to Healthentia backend (2nd option)

The API of Healthentia that facilitates all these data transfers is quite mature, but changes are expected to support the ONCOSCREEN-specific data. It is currently organised in different controllers (see part of its Swagger UI in Figure 21) as follows:

- Account: Authentication mechanism for the SW entities
- Aggregations: Get individual attributes for all in a study
- Vitals: Similar, only for vital signs
- Data provider: Get answers, scores, exercise sessions, and custom information fields

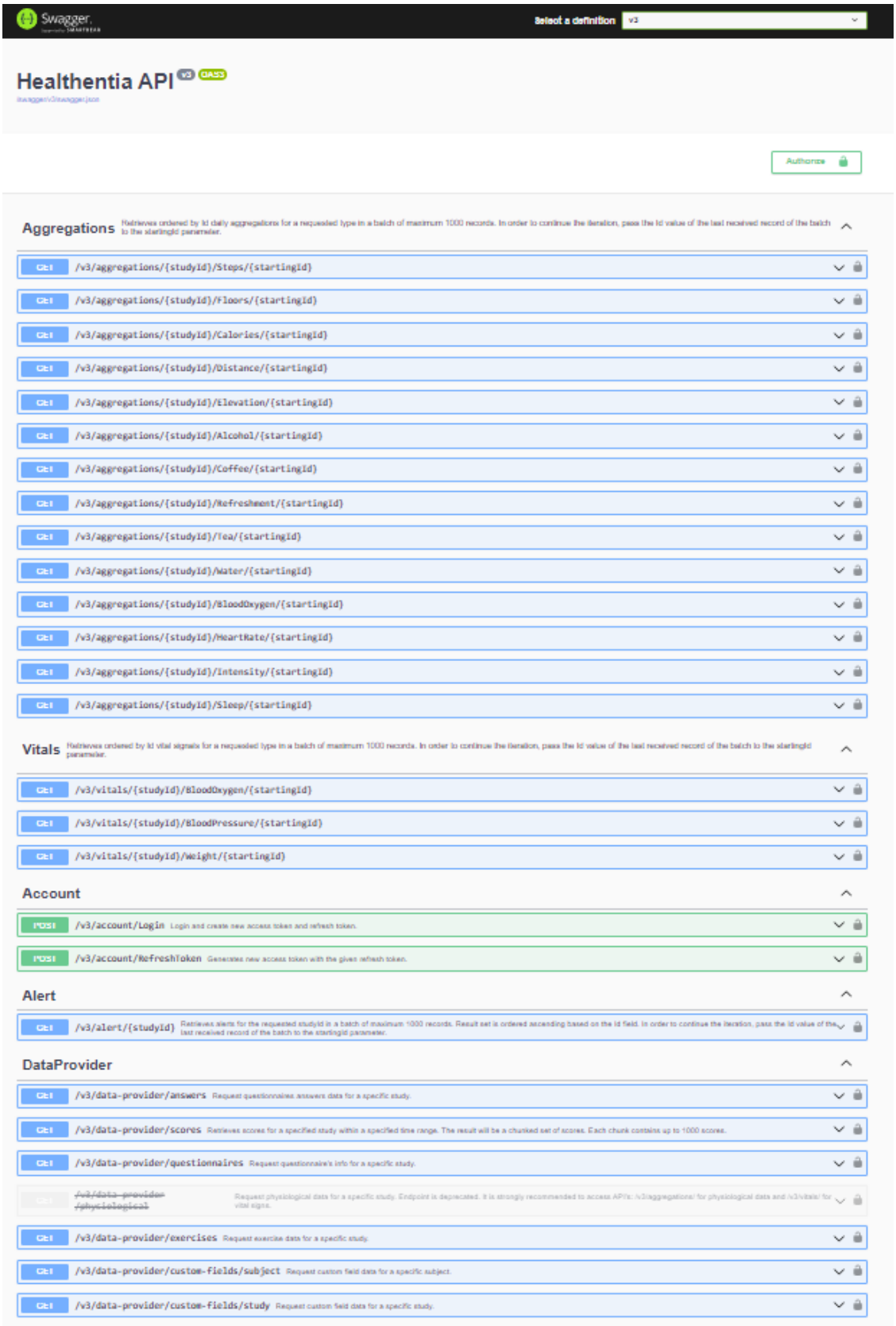


Figure 21: API endpoints offered by ONCO-CAWA for the provision of the collected information to ONCOSCREEN.

Risks and Mitigation Measures

In the following subsections the risks related to the design, implementation, and usage on ONCO-CAWA are detailed. The first pair are universal across ONCOSCREEN, while the second pair are specific to ONCO-CAWA.

4.1.5 Risk 1 – Lack of interaction between technical/clinician/end-users

Table 7: ONCO-CAWA Risk 1 - Lack of end-user engagement.

Risk Title	**Lack of interaction between technical/clinician/end-users**	**Status**	**Open**
Risk Category	Stakeholder Engagement	Related Task	T1.3
Risk Description			
The technical and non-technical sides fail to interact to get well-defined requirements.			
Responsible Partner	ICCS	Period Identified	Q1|2023
Probability		Impact	
Low		High	
Mitigation Plan (Indicative)			
Early interaction since kick-off for requirements. Involvement of the end-users, as discusses in section 2.2.			
Comments			
N/A			

4.1.6 Risk 2 – Lack of end-user engagement

Table 8: ONCO-CAWA Risk 2 - Lack of end-user engagement.

Risk Title	**Lack of end-user engagement**	**Status**	**Open**

Risk Category	Stakeholder Engagement	Related Task	T2.4
Risk Description			
End-users fail to properly specify the data collection needs, and/or do not give actionable feedback on ONCO-CAWA.			
Responsible Partner	SERVTECH	Period Identified	Q1|2023
Probability		Impact	
Low		High	
Mitigation Plan (Indicative)			
Early involvement of the end-users, as discusses in section 2.2.			
Comments			
N/A			

4.1.7 Risk 3 – Inefficient tools

Table 9: ONCO-CAWA Risk 3 - Inefficient tools.

Risk Title	**Inefficient tools**	**Status**	**Open**
Risk Category	Technical	Related Task	T4.4
Risk Description			
Difficulties in delivering suitable and efficient tools (mobile app).			
Responsible Partner	iSPRINT	Period Identified	Q2|2023
Probability		Impact	
Low		High	

Mitigation Plan (Indicative)
Adherence to requirements. Close monitoring of scientific and technological progress through the Technology Steering task. More tests before delivery to the end users will be performed
Comments
N/A

4.1.8 Risk 4 – End-User engagement with the mobile app

Table 10: ONCO-CAWA Risk 4 - End User engagement with the mobile app.

Risk Title	**End User engagement with the mobile app**	**Status**	**Open**
Risk Category	Stakeholder Engagement	Related Task	T4.4
Risk Description			
Low adherence to data collection protocol in trials.			
Responsible Partner	iSPRINT	Period Identified	Q2|2023
Probability		Impact	
Low		High	
Mitigation Plan (Indicative)			
End-user education: Demonstration of ONCO-CAWA to the end-users; hands-on training sessions. Gamification features in the app to make it more appealing.			
Comments			
N/A			

5 Clinical Decision Support System Tool analysis and architecture (ONCO-CLIDE)

The ONCOSCREEN AI-based Clinical Decision Support System (cDSS), named ONCOCLIDE, is being developed in collaboration with medical partners as part of the T4.5 initiative, is specifically tailored to advance the detection of colorectal cancer (CRC). This system is designed to support clinicians and medical experts, primarily in CRC screening processes. It does so by employing smart suggestions, utilizing the wealth of data from the ONCOSCREEN data lake and integrating various innovative methods and modules from Work Package 3 (WP3). ONCO-CLIDE focuses on enhancing the accuracy and efficiency of CRC detection through these intelligent recommendations. Its main goal is to aid in more effective and informed clinical decision-making during the screening phase, contributing to the early and precise identification of CRC.

Addressing End-user and Technical Requirements

5.1.1 Addressing End User Requirements

ONCO-CLIDE
CLIDE_UR1: Need to have precise diagnosis of presence or absence of malignancy
CLIDE_UR2: positive/negative predictive value (high sensitivity and specifity)
CLIDE_UR3: detect precancerous lesions and degree of dysplasia
CLIDE_UR4: distinguish the stage of malignacy as well the molecular subtype / mutational profile of the tumor.

Figure 22 ONCO-CLIDE End User Requirements.

5.1.2 Addressing Technical Requirements

ONCO-CLIDE	PRIORITY
CLIDE_TR1 (F): Integration with the various diagnostic solutions proposed in this project via the virtual data lake	High-Mandatory
CLIDE_TR2 (F): Integration/communication with the clinical center's systems to receive data from colonoscopies, CT scans, biopsies etc. via the virtual data lake	High-Mandatory
CLIDE_TR3 (F): The system should be able to consider the individual's risk level provided from ONCO-RISTE (T4.3) when formulating recommendations.	High-Mandatory
CLIDE_TR4 (F): The system must include a Virtual Tumor board that facilitates communication between doctors. This will be included in the web interface.	High-Mandatory
CLIDE_TR5 (F): The system should be able to perform correlation-based classification of citizens/cases across various adenoma-carcinoma sequence levels (0, I, II, III, IV) based on the European CRC guidelines	High-Mandatory
CLIDE_TR6 (F): The system should generate a set of recommendations based on the diagnostic results, classification level (staging) and individual risk level.	High-Mandatory
CLIDE_TR7 (F): Interaction between doctors and ONCO-CLIDE needs to happen through a dedicated web interface	High-Mandatory
CLIDE_TR8 (F): There needs to be an authentication system for the web platform with different roles.	High-Mandatory

Figure 23 ONCO-CLIDE Technical Requirements.

You may find more details of the above user and technical requirements on the deliverable D4.1 chapter 6.12.

ONCO-CLIDE Innovation Aspects

As the proposal also highlights there is a significant gap in current clinical Decision Support Systems: the lack of tools specifically designed for CRC. This gap is addressed by our team's innovative approach. Central to our innovation is the integration of new non-invasive diagnostic methods that are also being tested during the course of the project. These methods will be

incorporated into our system not only for display but also for making intelligent, data-driven recommendations. Furthermore, the incorporation of ONCO-RISTE enhances the system's capability. It allows for the generation of personalized suggestions, taking into account the user's lifestyle choices and a range of other risk factors. This personalized approach marks a significant advancement in the field of CRC detection.

ONCO-CLIDE Functionalities and Design

5.1.3 Functionalities

ONCO-CLIDE can be thought of as being divided into three parts/modules. The first one is the web interface that displays the data from the various project sources. The second one is the suggestion models that will provide assistance during the diagnostic process and the classification model that will output the stage of cancer. Finally, the third part is the virtual tumour board where doctors will be able to communicate on a specific case.

The fiunctionalities below are still in a prototype stage and changes might be made during the course of the project.

5.1.3.1 Web Interface

The web interface of ONCO-CLIDE is designed for ease of use, ensuring that all patient data and analytics are readily accessible to clinicians. It features a user-friendly dashboard that aggregates data from electronic health records, imaging systems, and other relevant sources, presenting a comprehensive patient profile. Secure login mechanisms ensure patient confidentiality, while intuitive navigation allows for quick access to various modules such as the suggestion engine and virtual tumour board.

5.1.3.2 Suggestion and classification models

Central to ONCO-CLIDE's functionality are its suggestion and classification models. Leveraging machine learning algorithms, these models analyse patient data to offer diagnostic suggestions and predict cancer stages. The suggestion model acts as a second opinion, offering diagnostic recommendations based on current guidelines and training data. The classification model employs advanced analytics to accurately stage the cancer, considering factors like tumour size, lymph node involvement, and metastasis, which are crucial for determining the appropriate treatment protocol.

5.1.3.3 Virtual tumour board

The Virtual Tumour Board is an integral component of ONCO-CLIDE, providing a straightforward platform for medical professionals to discuss patient cases. It is a space where doctors can gather online to exchange insights and opinions, ensuring a collaborative approach to cancer care. This feature supports real-time sharing of patient data and expert recommendations, streamlining the decision-making process for effective treatment planning. Its user-friendly interface is

designed to be accessible, promoting regular use and enhancing collective expertise in patient management.

5.1.4 Design

A high-level technical design of ONCO-CLIDE is depicted in Figure 5.1. It incorporates a central backend developed with Django. Django is a high-level Python web framework that encourages rapid development and clean, pragmatic design. It's known for its robustness and scalability, making it a suitable choice for handling the complex data interactions required by ONCO-CLIDE. This backend is responsible for managing interactions with the frontend elements of the system. Given the distinct computational requirements of the machine learning models, during the inference phase, a separate backend is designated for these tasks. This separation ensures that the processing demands of the machine learning operations do not interfere with the overall system performance.

On the frontend, ONCO-CLIDE is currently divided into two main sections. The first is the diagnostics interface, which is linked to a Kafka data stream. This setup is chosen for its efficiency in handling real-time data, which is essential for the diagnostic processes within the system. The second section is the virtual tumour board, a platform for medical professionals to discuss patient cases.

Both frontend components are built using JavaScript, with the React framework. React is selected for its effectiveness in developing user interfaces that are both functional and adaptable. This decision aligns with the requirement for a responsive and interactive user experience in medical diagnostic settings. While the two frontend components are separate as of now, there is consideration for their integration later in the project development.

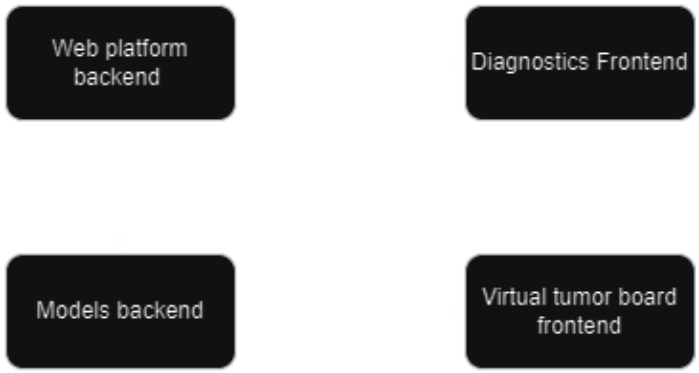


Figure 24 ONCO-CLIDE General design.

During this phase of the project, our efforts have been concentrated on developing the diagnostics frontend and the central backend. This stage has been crucial in laying the foundational structure for the ONCO-CLIDE. Below, we have provided some preliminary visuals

of the user interface. Please note that these are from an early stage of development and the interface is still being actively worked on.

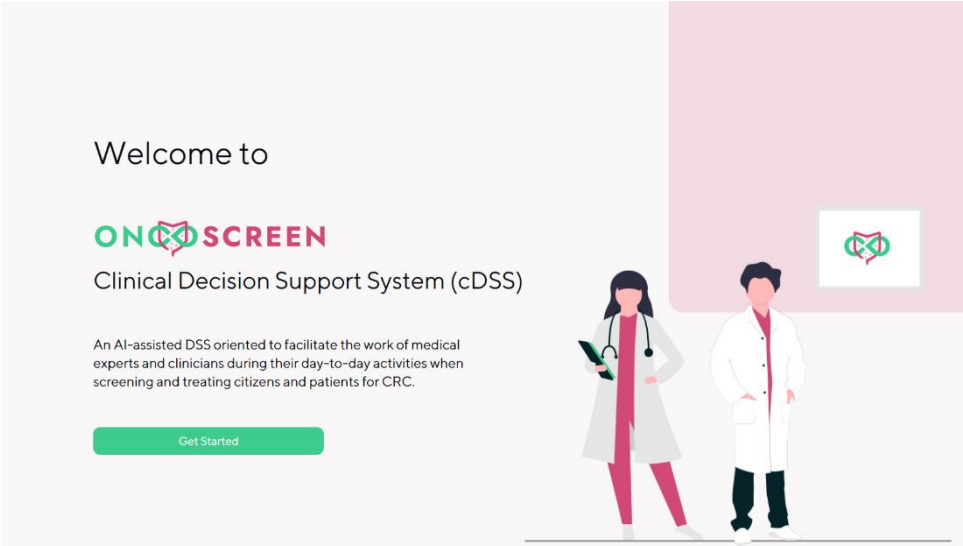


Figure 25 Welcome screen.

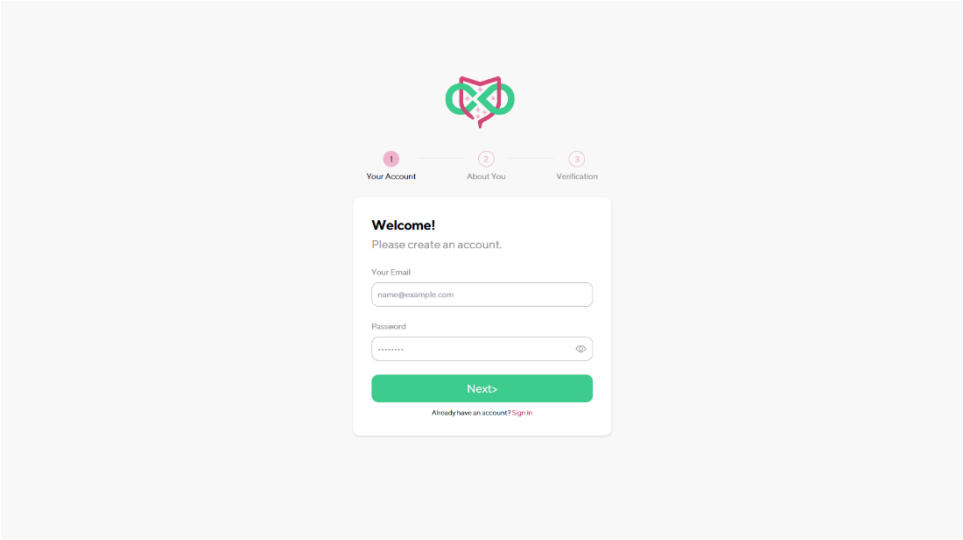


Figure 26 Signup page.

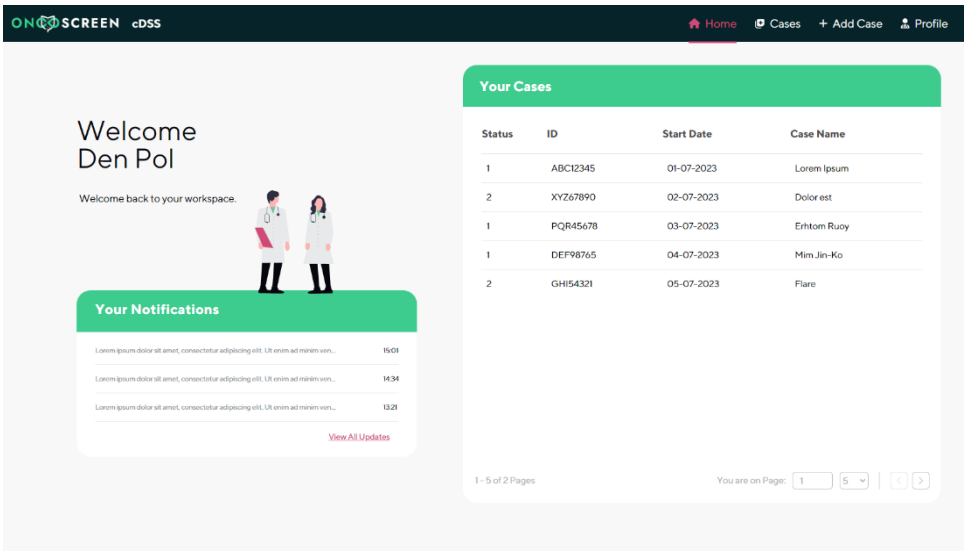


Figure 28 Main dashboard.

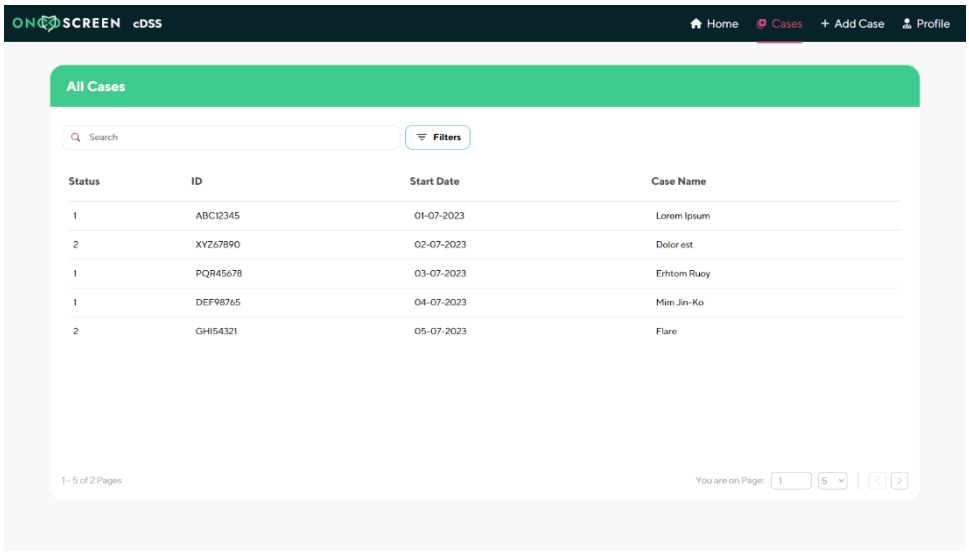


Figure 27 Cases page.

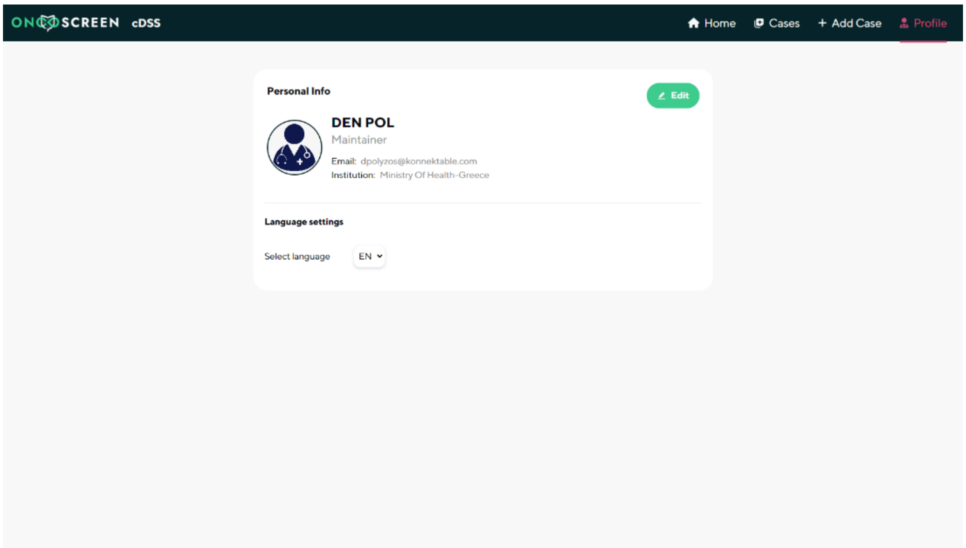


Figure 30 Patient page.

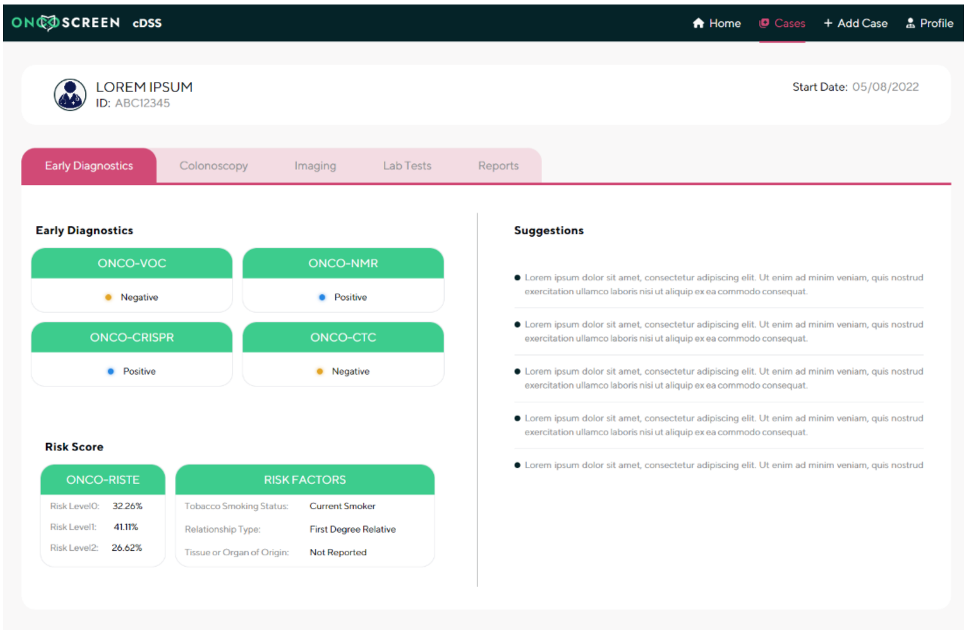


Figure 29 Profile screen.

5.1.5 ONCO-CLIDE Data Flows

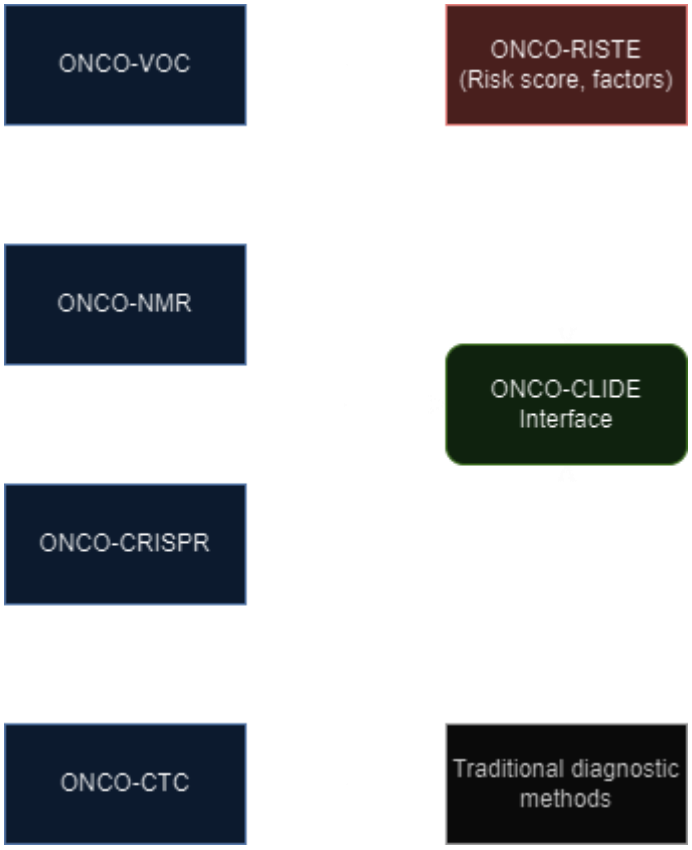


Figure 31 ONCO-CLIDE Data Flows.

The analytical data flows of ONCOCLIDE are described thoroughly in D4.1 chapter 3.2.13.3 the general data flow looks as in Figure 31

ONCO-CLIDE Trustworthiness

There has been some preliminary work on the selection of the predictive algorithms, though due to the lack of data at this stage of the project no final decisions have been made on this part. In selecting an algorithm for the Clinical Decision Support System, emphasis will be placed on performance metrics such as accuracy, sensitivity, and specificity, ensuring effective colorectal cancer identification. The algorithm will be rigorously evaluated for biases to ensure equitable decision-making across different groups. Additionally, robustness and generalizability will be key criteria, requiring the algorithm to perform reliably under various conditions and patient populations. This comprehensive approach ensures the algorithm is not only technically proficient but also clinically relevant.

Risks and Mitigation Measures

5.1.6 Risk 1 - Availability of early diagnostic data

Table 6 ONCO-CLIDE Risk 1.

Risk Title	Availability of early diagnostic data	**Status**	Open
Risk Category	Overall	**Related Task**	4.5
Risk Description			
The data from tools of WP3 will be available very late in the project affecting development of AI tools			
Responsible Partner	KONN	**Period Identified**	Q2|2023
Probability		**Impact**	
Medium		High	
Mitigation Plan (Indicative)			
Continuous monitoring of the progress of data collection together with the project coordination and possible use of only initially collected data as a placeholder until the total amount is collected.			
Comments			
N/A			

5.1.7 Risk 2 – Quality of early diagnostic data

Table 7 ONCO-CLIDE Risk 2.

Risk Title	Quality of early diagnostic data	**Status**	Open
Risk Category	Overall	**Related Task**	4.5
Risk Description			
The data collected from the early diagnostic methods of WP3 might be of low quality			
Responsible Partner	KONN	**Period Identified**	Q2|2023

Probability	Impact
Low	Low
Mitigation Plan (Indicative)	
Selection of specific algorithms that take into consideration any data quality limitations.	
Comments	
N/A	

5.1.8 Risk 3 - Availability of traditional diagnostic data

Table 8 ONCO-CLIDE Risk 3.

Risk Title	Availability of early diagnostic data	Status	Open
Risk Category	Overall	**Related Task**	4.5
Risk Description			
Some data need to be acquired from traditional diagnostic methods, and the clinical partners might be unwilling or unable to provide them due to the sensitive nature of them.			
Responsible Partner	KONN	**Period Identified**	Q1|2024
Probability		**Impact**	
High		High	
Mitigation Plan (Indicative)			
There have been two approaches considered as a resolution to this. The first is to avoid any complicated integration and have manual input fields on the cDSS platform. The second is to run a different local server on each clinical center so no data leaves the hospital			
Comments			
N/A			

5.1.9 Risk 4 - Participation of clinical partners

Table 9 ONCO-CLIDE Risk 4.

Risk Title	Availability of early diagnostic data	**Status**	Open
Risk Category	Overall	**Related Task**	4.5
Risk Description			
The intelligent suggestions that we are going to implement, are partially based on empirical observations of the clinical staff, so their participation is crucial to achieve this.			
Responsible Partner	KONN	**Period Identified**	Q1|2024
Probability		**Impact**	
Low		High	
Mitigation Plan (Indicative)			
Constant communication from the project coordination and constant channels of communication			
Comments			
N/A			

6 ONCO-EVIDA Tool analysis and architecture

According to the GA, ONCO-EVIDA is to be a "*powerful instrument for regional and national policymakers to enable them to combine and interpret results using user-friendly dashboards for more efficient and evidence-based decisions*".

As an intelligent analytics dashboard, it will integrate data from diverse sources to generate user-friendly visualizations and recommendations, with a focus on CRC risk factors and screening participation factors.

Users can upload and access their data, generate charts and customize (and save) their own dashboards while the tool generates concrete recommendations based on their queries.

A more comprehensive conceptual description can be found in D4.1 (section 3.2.14.1).

Addressing End-user and Technical Requirements

In total 20 ONCO-EVIDA end-user requirements are listed in detail in D4.1 (section 6.14.1) and 23 technical requirements for ONCO-EVIDA are listed in detail in D4.1 (section 6.14.2). In this iteration ONCO-EVIDA proceed in the advancement of various technical (mainly) and end user requirements as independent sub-systems focusing more on the user interface and in the chart capability. Due to some initial delays in tool development (caused due to problems in hiring a PhD researcher from VITO), most of the requirements are not fulfilled considering the ONCOEVIDA system as a whole, requiring further validation and development. Delays will be compensated over the next period with a tail heavy plan in regards to development and a closer end user and technical partners collaboration. All final updated requirements will be addressed in the second iteration of the deliverable.

ONCO-EVIDA Innovation Aspects

Compared to existing tools that visualize cancer data (e.g. European Cancer Pulse, ECIR Data Tool), ONCO-EVIDA will offer additional functionality in at least three respects: (1) the user's ability to customize and save the output of their queries for easy side-by-side comparisons, (2) the generation of policy suggestions specific to the query and (3) CRC screening (coverage and/or participation) can be treated as both an independent and a dependent variable.

ONCO-EVIDA Functionalities and Design

This sub-section will focus on the basic dashboard functionalities mentioned during LIT2 that focus on URs 1b-d, 4 and 11 and TRs 2-4, 6 and 13.

6.1.1 Functionalities

6.1.1.1 Datasets Access and Analysis

Users can access their custom and pre-curated datasets (based on available data sources accessible to them) within ONCO-EVIDA and organize relevant data for analysis. *(UR1b, TR2)*

6.1.1.2 Generating Charts

ONCO-EVIDA allows users to generate various charts directly from their datasets to visualize their data to gain insights and knowledge. *(UR4, TR6)*

6.1.1.3 Building Dashboards

Users can combine multiple charts representing at the same time a visualization of 'risk factors content' and of 'screening/health data content' into customized 'combined' graphical views. *(UR4, TR6)*

6.1.1.4 Customizing Layout

The tool will also provide options for users to customize the layout and design of their dashboards. *(UR4, TR6)*

6.1.1.5 Saving Dashboards

Users can save the dashboards they create (including the settings chosen) for future reference and export them for further analysis or sharing with colleagues and stakeholders (reproducibility and transparency). *(UR4, TR6)*

6.1.1.6 Evidence-Based Recommendations

Queries generate suggestions based on the analysis of the association between the independent and dependent variables of interest. *(UR1c-d, TR3-4)*

6.1.1.7 Access Control and User Authentication

Creating a password-secured user account is necessary to access ONCO-EVIDA's features. *(UR11, TR13)*

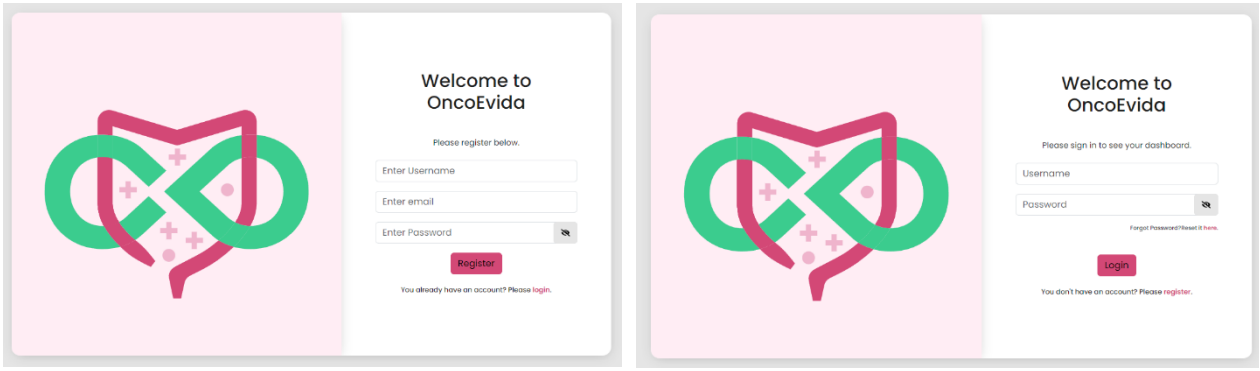
6.1.1.8 Additional Functionality

Some functionalities that are to be implemented in the future:

- Data integration from multiple sources
- Real-time data processing and display
- Data integration based on various metrics and dimensions
- Alerts and notifications based on predefined criteria
- Integration with machine learning algorithms

6.1.2 Design

This sub-section presents a curated selection of screenshots (Figure 33-Figure 39) that offer a comprehensive overview of ONCO-EVIDA's current design and functionalities.



(a) (b)

Figure 32 Registration (a) and Login (b) Page of ONCO-EVIDA.

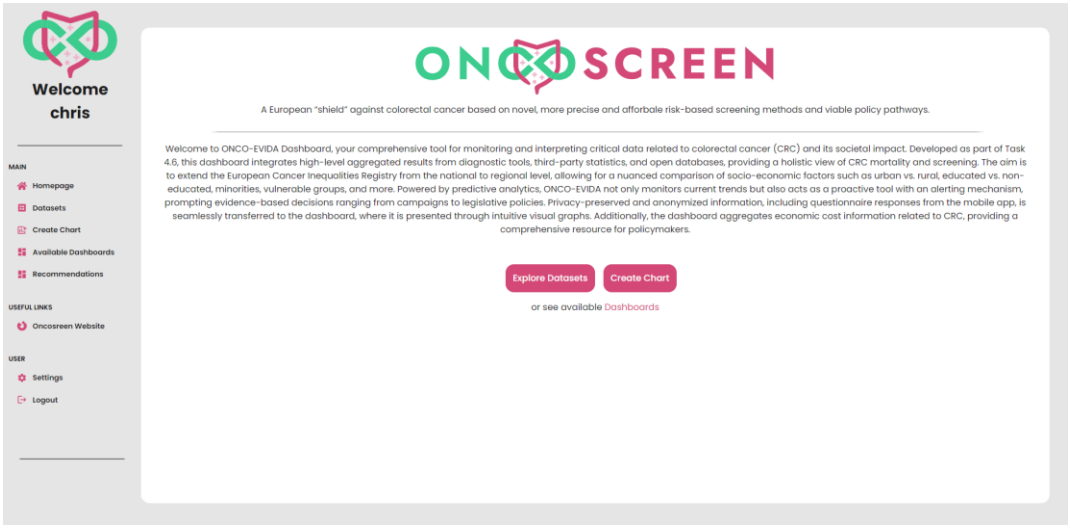


Figure 33 ONCO-EVIDA's home page.

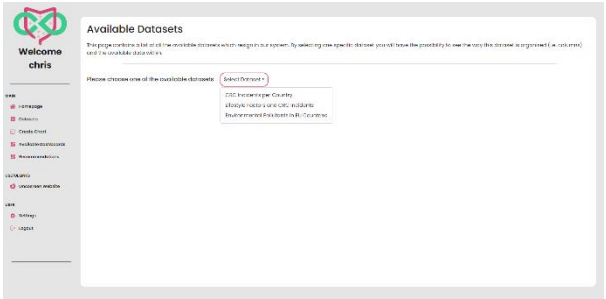


Figure 34 ONCO-EVIDA's dataset page, where the user can see an overview of the available datasets.

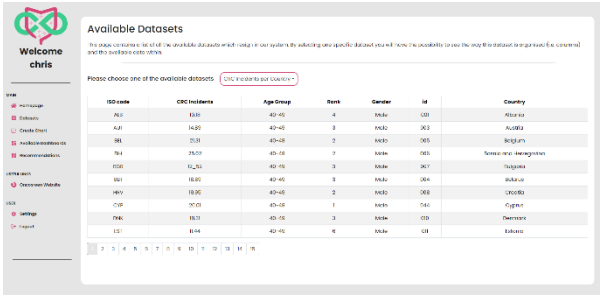


Figure 35 Information for each dataset is given in a table format.

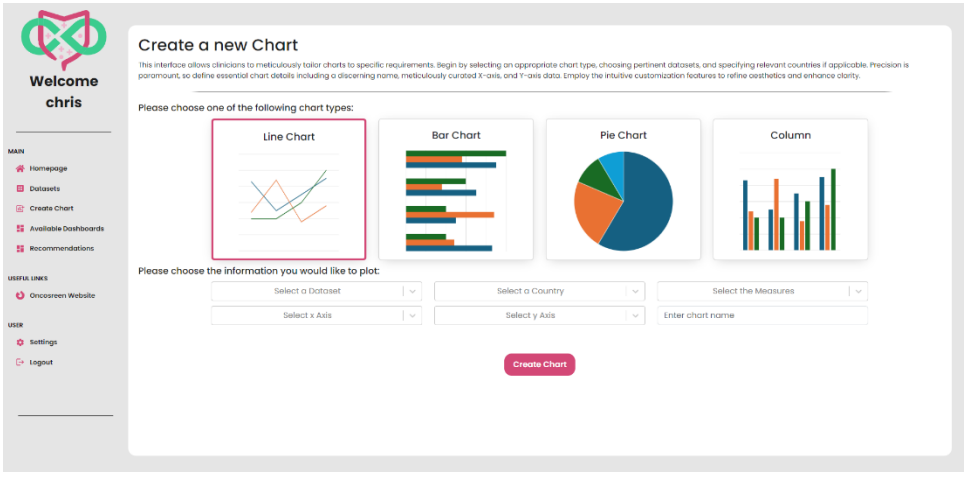


Figure 38 In this page the user can create a chart by selecting the desired chart type, the dataset, the countries of interest and the information they want on the x- and y-axes.

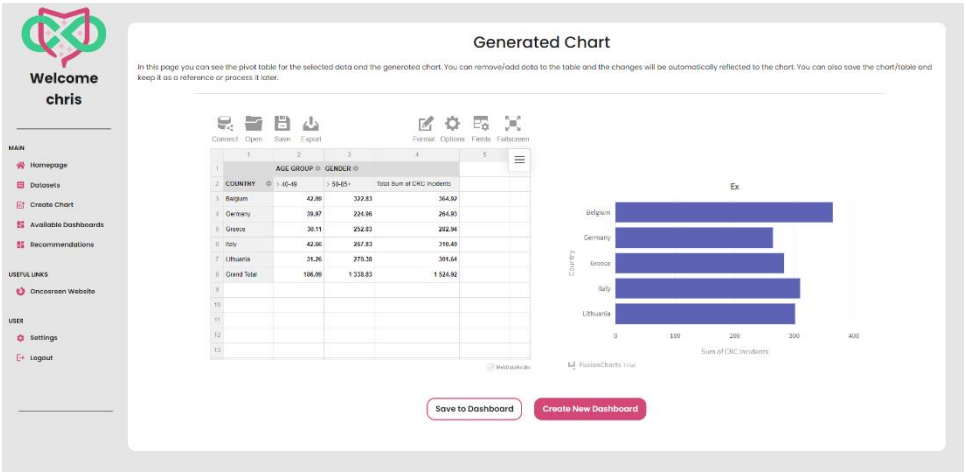


Figure 37 The chart page depicts the produced chart alongside a pivot table which contains the relevant information. The user has the possibility to save the chart on an existing dashboard or create a new dashboard and insert it there.

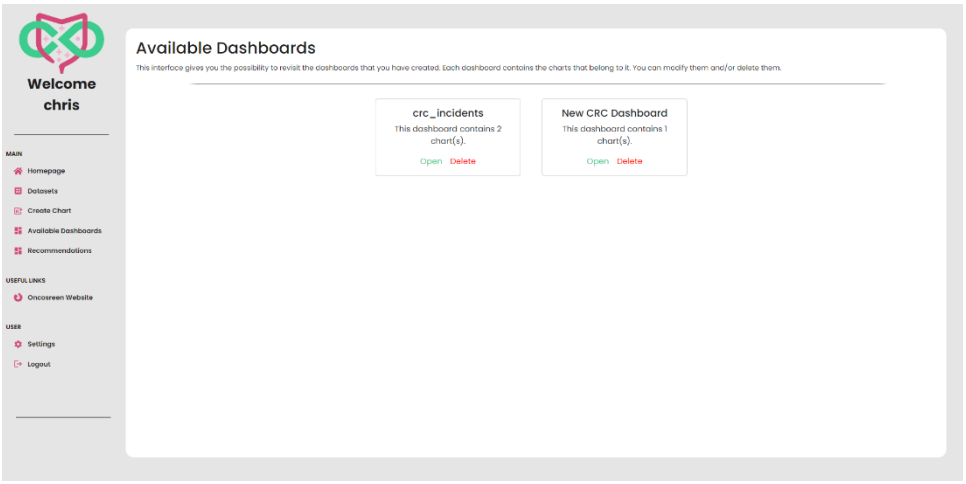


Figure 36 In this page the user is given an overview of the dashboards that have been created. Each dashboard consists of several charts. The user can either open/edit the dashboard or delete it. Upon deletion the charts belonging to this dashboard are also deleted.

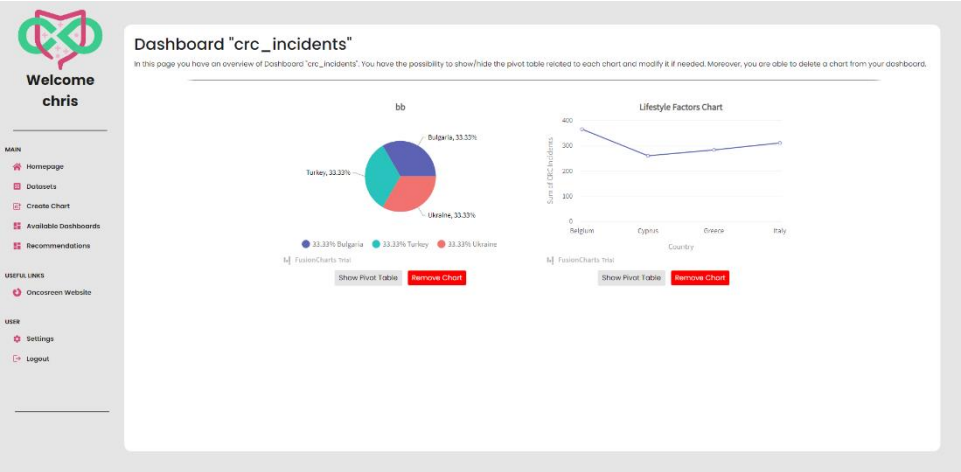


Figure 39 The individual's dashboard page depicts all the charts belonging to this dashboard. The user has the possibility to delete a chart from the dashboard or access and edit its pivot table.

The ONCO-EVIDA system architecture has been described in detail in D4.1 (section 3.2.14.3).

6.1.3 ONCO-EVIDA Data Flows

ONCO-EVIDA is primarily an output tool. It will receive most of its data from the Data Fusion Tool but will also be able to analyse data that came directly from third-party sources. For more information regarding the data flow to and from ONCO-EVIDA, please consult the overall system architecture in D4.1.

ONCO-EVIDA Trustworthiness

Non-aggregated cancer data will only be accessible in a federated structure; it will not leave or be copied from the data owners' repositories. ONCO-EVIDA output cannot be downloaded without authentication.

Risks and Mitigation Measures

6.1.4 Risk 23 – Lack of partner engagement

Table 10 ONCO_EVIDA Risk 23.

Risk Title	Lack of partner engagement	**Status**	Open
Risk Category	Stakeholder Engagement	**Related Task**	4.1
Risk Description			
Lack of interaction between technical partners and end-users.			

Responsible Partner	ICCS (*)	Period Identified	Q1|2023
Probability		**Impact**	
Low		Medium	
Mitigation Plan (Indicative)			
Composition of mixed working groups. Agile development methodology with frequent iterations and incremental releases. Regular planning of feedback moments. Bilateral outreach to (frequent) non-responders.			
Comments			
(*) Though ICCS is listed responsible partner in the Risk Registry, it falls to all ONCO-EVIDA partners to interact with end-users to gauge progress towards the URs.			

6.1.5 Risk 46 – High-resolution data not supported

Table 11 ONCO_EVIDA Risk 46.

Risk Title	High-resolution data not supported	Status	Open
Risk Category	Technical	**Related Task**	4.6
Risk Description			
High-resolution maps are not supported by the platform. Low geospatial levels (town or below) lead to cluttered visualization attempts due to a large number of units.			
Responsible Partner	Catalink VITO	**Period Identified**	Q1|2024
Probability		**Impact**	
High		Medium	
Mitigation Plan (Indicative)			
Increased emphasis on analysis results rather than visual representations for queries going below municipal level. Recommendation engine limits itself to mentioning only "worst performers".			

Comments
N/A

7 DATA FUSION Analysis and Architecture

The role of this tool is the multi-modal fusion of heterogeneous sources of data for mining and discovering new knowledge and correlations for CRC early detection, risk analysis and stratification of patients. The ONCOSCREEN's fused data lake is responsible for correlating ONCOSCREEN Data with 3rd party information while maintaining privacy by ensuring citizen anonymity.

Iterative Co-Design Activities with End Users

As stated in section 2.5 the data fusion task has no end user requirements. Therefore, at this stage of the project there are no co-design activities to be reported.

Addressing End-user and Technical Requirements

Data Fusion as a pure technical back-end tool that support the other tool doesn't have user requirements in its first iteration of the tool. Depending on the feedback rounds any end user requirements that may be presented shall be included in the second iteration of the deliverable. Nine Data Fusion technical requirements listed in detail in D4.1, although a particular number of them has been advanced, there are particular until their finalisation dealing with the overall 'system as a whole' response, having dependency with other tools' output. All requirements will be fulfilled in the second iteration of the deliverable.

DATA FUSION Analysis

At this stage of the project, only the analytic aspects of the data fusion component have been studied. By merging datasets from various open sources, the relation between environmental, lifestyle (nutrition, smoking, alcohol habits), comorbidities (obesity) and socioeconomic factors and CRC incidences were studied on country-level data from Europe. A correlation analysis and a linear regression analysis were performed using a time-lag and a moving average techniques, to study the effect of the exposure to the risk factors.

DATA FUSION Innovation Aspects

Metrics like partial correlations, following the principles of population methods and epidemiology, will be exploited to create a framework that will incorporate risk factors from open sources.

DATA FUSION Functionalities and Design Analysis

The data fusion comprises of two different tools; the data lake tool and the data fusion tool.

7.1.1 Analysis of Functionalities

7.1.1.1 Data lake

The data lake serves as a centralized physical repository dealing with clinical data from patients, video data from colonoscopies as well as image tissue data. All this types of data are gathered in a centralized repository and are shared to the respective ONCOSCREEN tools in order to be further analysed.

7.1.1.2 Data fusion

The role of the data fusion component is to reveal correlations between environmental, lifestyle (nutrition, smoking, alcohol habits), comorbidities (obesity) and socioeconomic risk factors and CRC incidences from open-source data on country level.

7.1.2 Design Analysis

Since the data fusion is at an early stage there are no specific design aspects to be reported in this deliverable.

7.1.3 DATA FUSION Data Flows

As described in detail in D4.1 subsection 3.2.11.4, the data lake component is connected to the ONCOAICO, ONCOAITI and ONCOBIOBA via REST APIs, while it is also connected via KAFKA SSL with the rest of the ONCOSCREEN tools. The data fusion tool is connected via KAFKA SSL with the rest of the ONCOSCREEN tools.

Data Fusion Tool Trustworthiness

The data lake will be designed based on SNOMED CT and the FHIR protocols and also will use SSL encryption, data anonymization and restricted access to data by certain tools to ensure secure exchange of the data.

Risks and Mitigation Measures

During all periods of the project, it is important to show that you have been aware of potential risks and that you have implemented mitigation measures to address them.

In this section, please align the risks with the ONCOSCREEN Risk Registry.

7.1.4 Risk 1 - Unsuitable cancer models

Table 3 DATA FUSION Risk 1.

Risk Title	**Unsuitable cancer models**	**Status**	**Open**
Risk Category	Technical	Related Task	4.2
Risk Description			

Difficulties in delivering suitable cancer-related data models			
Responsible Partner	CERTH	Period Identified	Q2|2023
Probability		Impact	
Medium		Medium	
Mitigation Plan (Indicative)			
Conduct extensive literature study and extensive surveys with the ONCOSCREEN policy-makers. Emphasise further the interdisciplinary dimension and consider additional correlating factors to improve predictions. Collect more datasets and iterate based on synthetic data.			
Comments			
N/A			

Risk 2 – Issues with System Integration

Table 4DATA FUSION Risk 2.

Risk Title	**Issues with System Integration**	**Status**	**Open**
Risk Category	Technical	Related Task	4.2
Risk Description			
System level integration or interoperability difficulties due to unforeseen complexity			
Responsible Partner	CERTH	Period Identified	Q2|2023
Probability		Impact	
Medium		High	
Mitigation Plan (Indicative)			

Clear description of system architecture, data models and interfaces following a standard modelling language. Modular software approach and well-defined interfaces between components from the very beginning. Structured and systematic approach to integration and verification activities, allowing sufficient time for tests. Close monitoring by the WP leaders and the Technical Coordinator to ensure smooth integration and interoperability.
Comments
N/A

8 Performance in Lab Tests/Clinical Trials

ONCO-RISTE in Laboratory Test 1 (LT1)

As of 1st LIT, no integration with other tools has been implemented, and the laboratory test focused solely on presenting the distinct functionalities of the ONCO-RISTE tool and the internal integration of the ONCO-RISTE back end with the fuzzy module.

In the ONCO-RISTE backend architecture, the functionality of fuzzy modules Python wheel library was utilized inside the backend to enhance the efficiency and capabilities of our system. This library was crafted and optimized for risk calculation prediction and provides a collection of

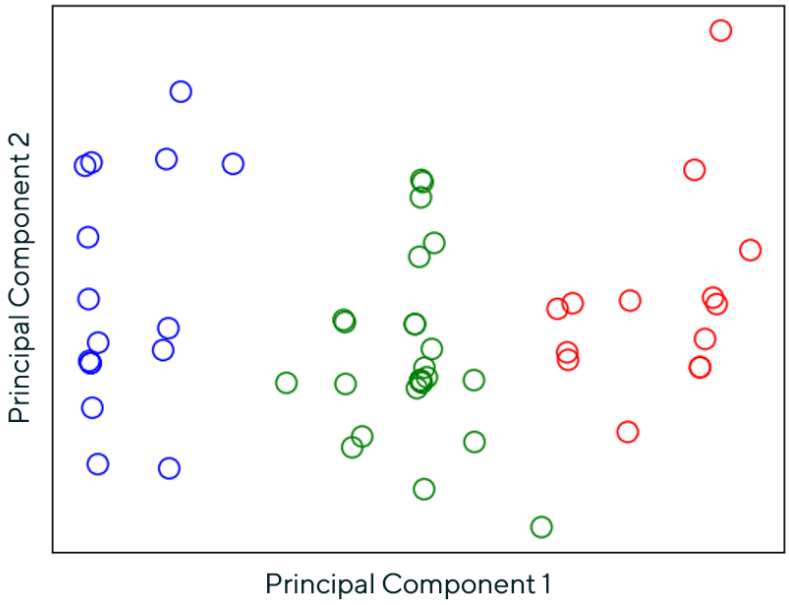


Figure 40 Result of Fuzzy C-means in LIT1 for 3-level risk. Principal Component technique was used to visualize these clusters.

pre-built functions that integrate with our backend processes.

Furthermore, for the need of the LITs in our project, we implemented an Integration Registry Excel as a centralized tracking system to monitor the connectivity between various tools. This registry serves as a dynamic record, detailing the relationships and interconnections between different components within our ecosystem. By documenting each tool's integration points, we ensure transparency and maintain a comprehensive overview of the entire system. The Integration Registry Excel not only aids in troubleshooting and debugging by swiftly identifying dependencies but also facilitates effective communication among team members. This proactive approach to monitoring tool connections enhances the project's overall efficiency and allows partners to respond promptly to any changes or issues in the integrated environment.

ONCO-RISTE introduced two key endpoints, namely "Soft Prediction" and "Fuzzy Prediction", offering distinct approaches for calculating risk scores based on user input. These endpoints

serve as pivotal components of ONCO-RISTE's functionality, providing users with valuable insights through straightforward and fuzzy assessments.

The "Soft Prediction" endpoint is designed to offer a straightforward risk assessment based on user input. By utilizing established algorithms, this endpoint delivers a risk score that represents a clear and direct evaluation of the provided information. In contrast, the "Fuzzy Prediction" endpoint incorporates fuzzy logic as part of the algorithm to process user input. This approach allows for a more context-aware assessment, considering the uncertainties and impreciseness inherent in healthcare data. The output is a risk score that reflects a more comprehensive understanding of the user's health status.

User Input Processing

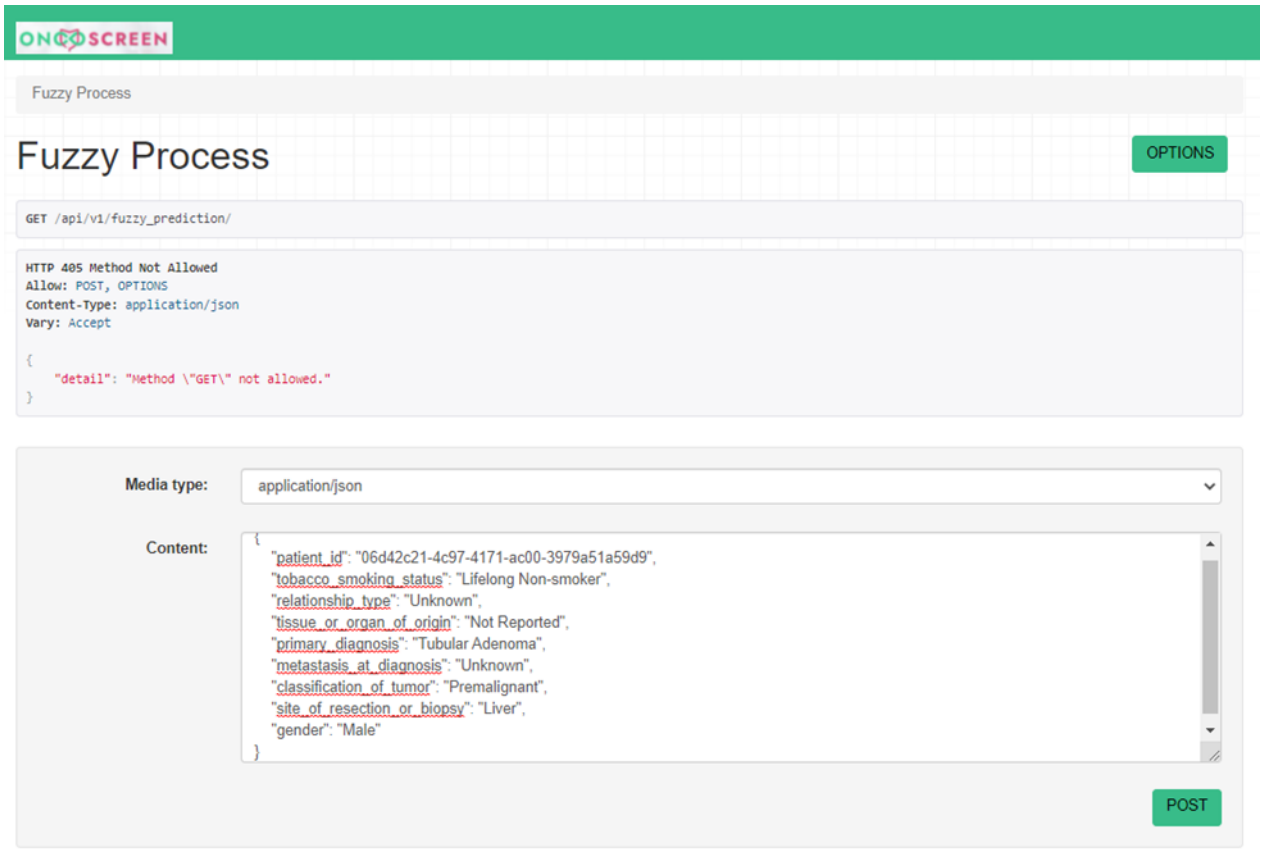


Figure 41 User input interface.

The user provides input through a designated interface, supplying relevant information for risk assessment. ONCO-RISTE processes this input according to the specific logic implemented in each endpoint. Included in this deliverable are detailed print screens capturing the user interface for input submission, the response interface displaying the risk scores, and console logs illustrating the internal processes and calculations performed by ONCO-RISTE during the prediction requests.

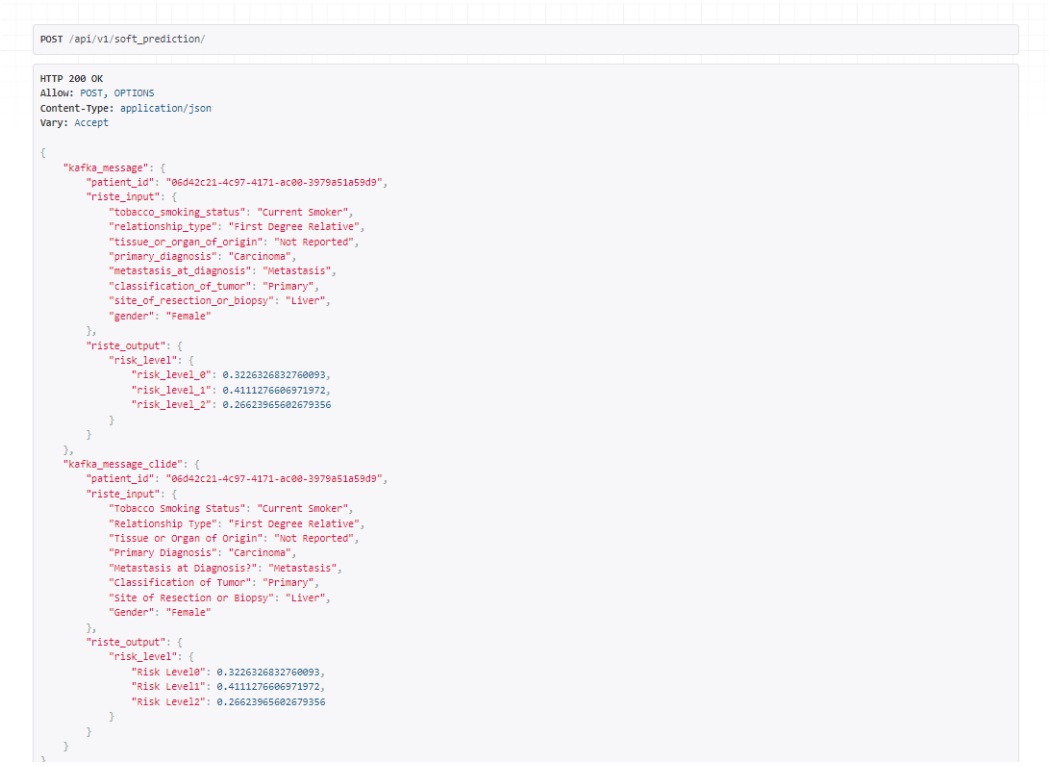


Figure 42: ONCO-RISTE 1st LIT output screen.

```
oncoscreen-oncoriste-backend-backend-1    | 2024-01-15 14:44:16,753 [INFO] app.settings: Fetching input from POST request
oncoscreen-oncoriste-backend-backend-1    | 2024-01-15 14:44:16,753 [INFO] app.settings: {'patient_id': '06d42c21-4c97-41
71-ac00-3979a51a59d9', 'tobacco_smoking_status': 'Current Smoker', 'relationship_type': 'First Degree Relative', 'tissue
_or_organ_of_origin': 'Not Reported', 'primary_diagnosis': 'Carcinoma', 'metastasis_at_diagnosis': 'Metastasis', 'classi
fication_of_tumor': 'Primary', 'site_of_resection_or_biopsy': 'Liver', 'gender': 'Female'}
oncoscreen-oncoriste-backend-backend-1    | 2024-01-15 14:44:16,764 [INFO] app.settings: Fuzzy result: {'06d42c21-4c97-41
71-ac00-3979a51a59d9': {'risk_level_0': 0.3226326832760093, 'risk_level_1': 0.4111276606971972, 'risk_level_2': 0.266239
65602679356}}
```

Figure 43: ONCO-RISTE console output during 1st LIT

ONCO-CAWA in Laboratory Test 1 (LT1)

The early version of ONCO-CAWA was presented in LT1, focusing on the subject registration process and the ONCOSCREEN branding. The registration included the different consents, notices, and disclaimers. Some widgets were also covered. Analytical screenshots were presented in previous. For the avoidance of duplication please see the reference to ONCO-CAWA tool previously in this document.

ONCO-CLIDE in Laboratory Test 1 (LT1)

During our inaugural lab integration test, we highlighted two principal advancements in our project's development. The initial segment of our presentation was dedicated to the unveiling of the system's high-level architecture, with a particular focus on the logical flow of information between components. This overview was critical in illustrating the theoretical underpinnings of our system's design, showcasing how data is expected to traverse through the various layers and

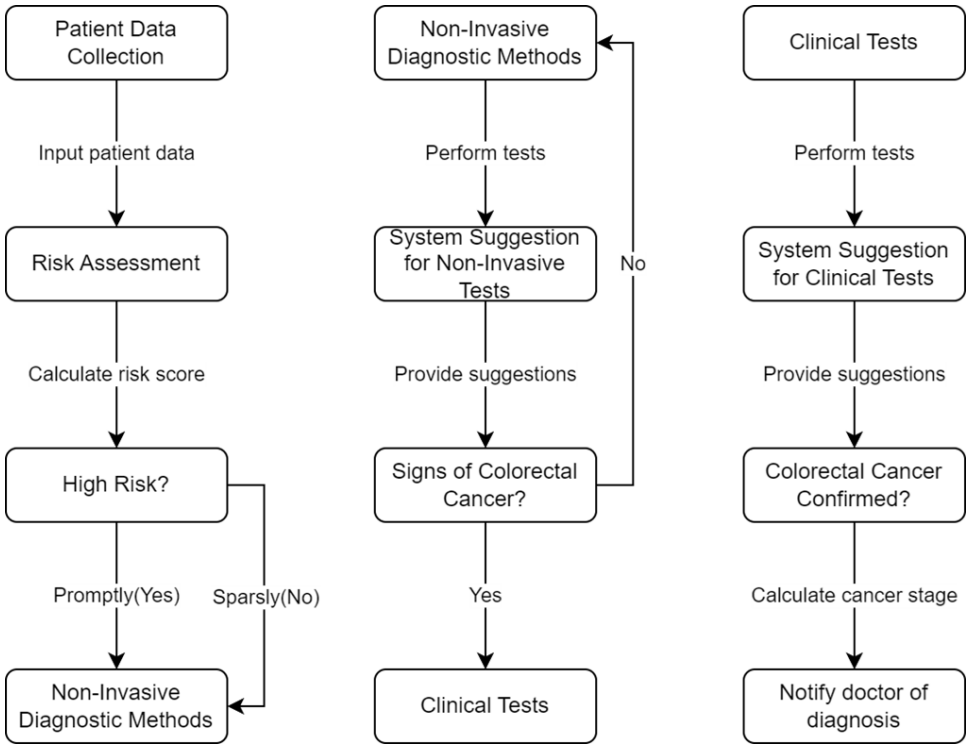


Figure 44 Logical flow chart of ONCO-CLIDE.

modules. By outlining the logical flow, we aimed to provide participants with a clear yet simple picture of the system's operational blueprint. This foundational architecture is crucial not only for understanding the interactions within the system but also for guiding the subsequent development phases, ensuring that the project progresses on a well-defined and coherent path.

The second part of our presentation shifted towards a more tangible aspect of our project, with a live demonstration of the initial design and implementation of the web interface. This hands-on showcase allowed us to bring the theoretical aspects of our project to life, presenting how users will interact with the system through a carefully crafted interface. The demonstration highlighted the practical application of our high-level architectural principles in the design and functionality of the web interface, emphasizing user experience and interaction. By bridging the gap between the system's logical flow and its user interface, this live demo not only underscored our commitment to a user-centric design approach but also provided an invaluable opportunity for gathering direct feedback, setting the stage for ongoing refinement and improvement.

Images from the web interface can be seen in section 5.1.4.

ONCO-EVIDA in Laboratory Test 1 (LT1)

For the first LIT meeting, the technical partners working on ONCO-EVIDA introduced the tool and its main functionalities, placing emphasis on the visualization of data, the customizability of resulting dashboards, and the future ability of ONCO-EVIDA to provide recommendations based on the end-user's specific query. A short live demo of the front-end was given, after which we reached out to end-users for feedback.

DATA FUSION in Laboratory Test 1 (LT1)

The role of data fusion is to reveal interrelations between different risk factors and colorectal cancer (CRC) metrics. These metrics can include CRC rate per population, deaths due to CRC or Disability-adjusted life years (DALYs) which is a metric reflecting the complete picture of the burden of the disease measuring the years of life lost due to ill-health, disability or early death.

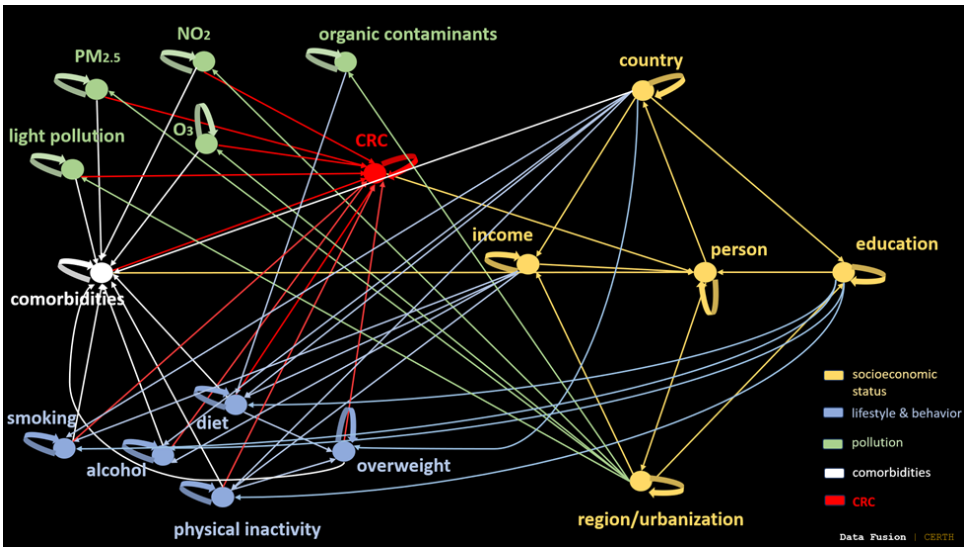


Figure 45 Interrelations between risk factors and CRC incidences.

The study of risk factors affecting CRC metrics leads to multidimensional data, including socioeconomic, environmental, lifestyle and behavioural features, which are strongly related to each other. For example, information about the region where a patient lives is highly related to environmental features, such as pollution metrics. These relations can be revealed by designing a complex architecture network, as can be seen in the Figure Fusion_1. In the figure, the direct and indirect correlations between the different risk factors, as well as the CRC incidences, can be seen.

ONCO-RISTE in Laboratory Test 2 (LT2)

The central focus of our efforts during the 2nd Lab Test resided in ensuring the successful integration of the various tools within the designated environment recognizing the critical importance of a harmonious integration for the overall success of the project.

For the 2nd LIT ONCO-RISTE fetched user input from ONCO-CAWA user questionnaires that were answered real time.

Integration with ONCO-CAWA ensures the utilization of accurate and up-to-date questionnaire answers, improving the overall quality of data processed by ONCO-RISTE. The integration established a seamless communication channel between ONCO-RISTE and ONCO-CAWA, allowing for the efficient exchange of essential health-related information. ONCO-RISTE in order to fetch questionnaire answers used ONCO-CAWA Rest APIs.

Subsequent to the successful integration efforts, our focus shifted towards the computation of risk scores. This involved the selection of the most important factors according to previous results and end-user's information, the formatting of the data that is received by the algorithm to be compliance with the aforementioned factors, and improvements in the automatic calculation of the number of clusters.

Following the comprehensive calculation process, the obtained risk scores were systematically transmitted through the Kafka messaging system. This strategic choice of utilizing Kafka ensured a reliable and scalable means of disseminating the calculated risk scores to the intended recipients, further solidifying the robustness and efficiency of the integrated system. In the 2nd LIT the calculated risk score was send to ONCO-CAWA, ONCO-CLIDE, DATA FUSION tools resulting to a complete integration test.

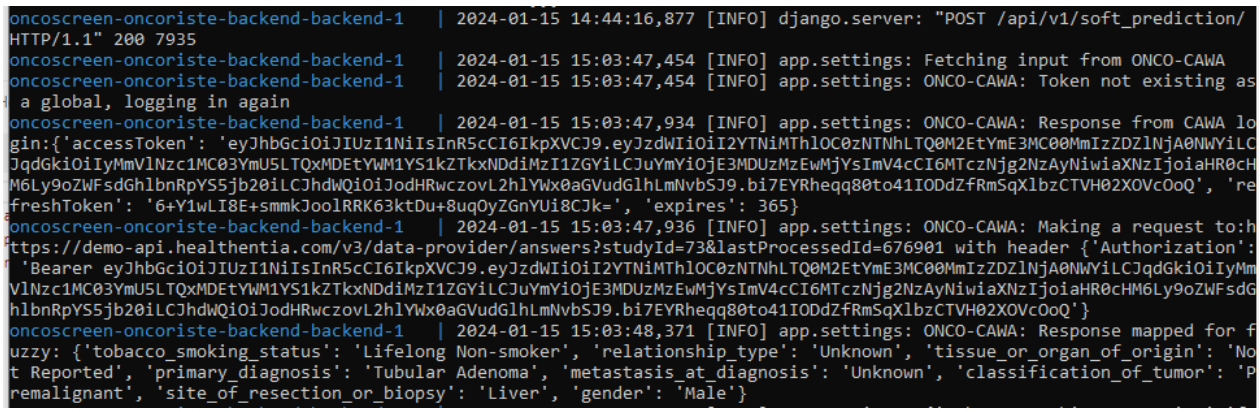


Figure 47: ONCO-RISTE automatically fetching user input data from ONCO-CAWA after authenticating itself.

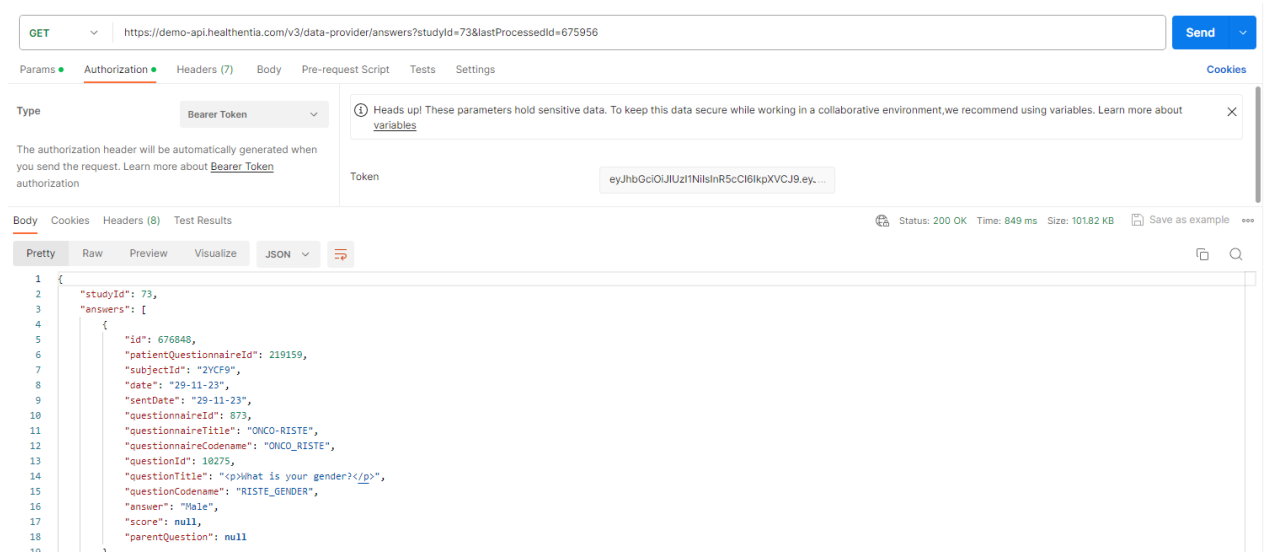


Figure 46: Test to fetch ONCO-CAWA user input data.

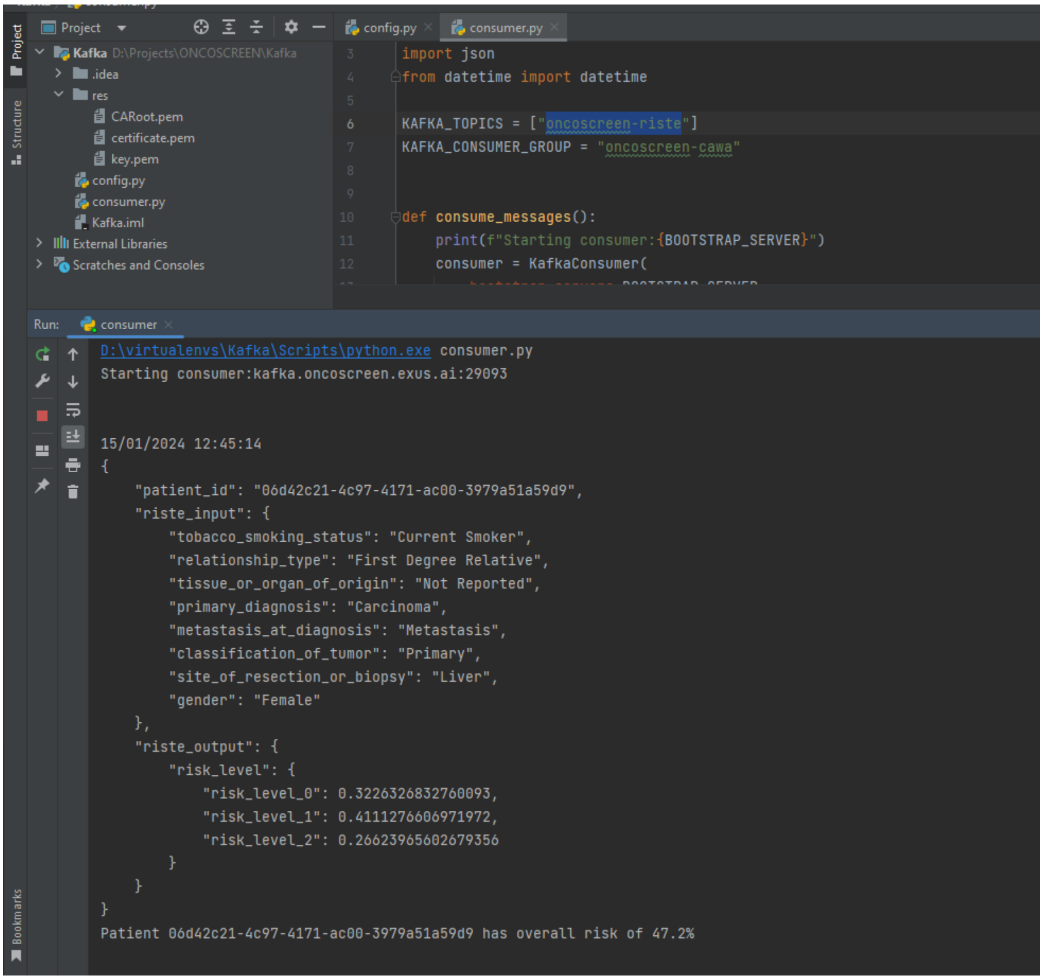


Figure 48: ONCO-CAWA receiving ONCO-RISTE output.

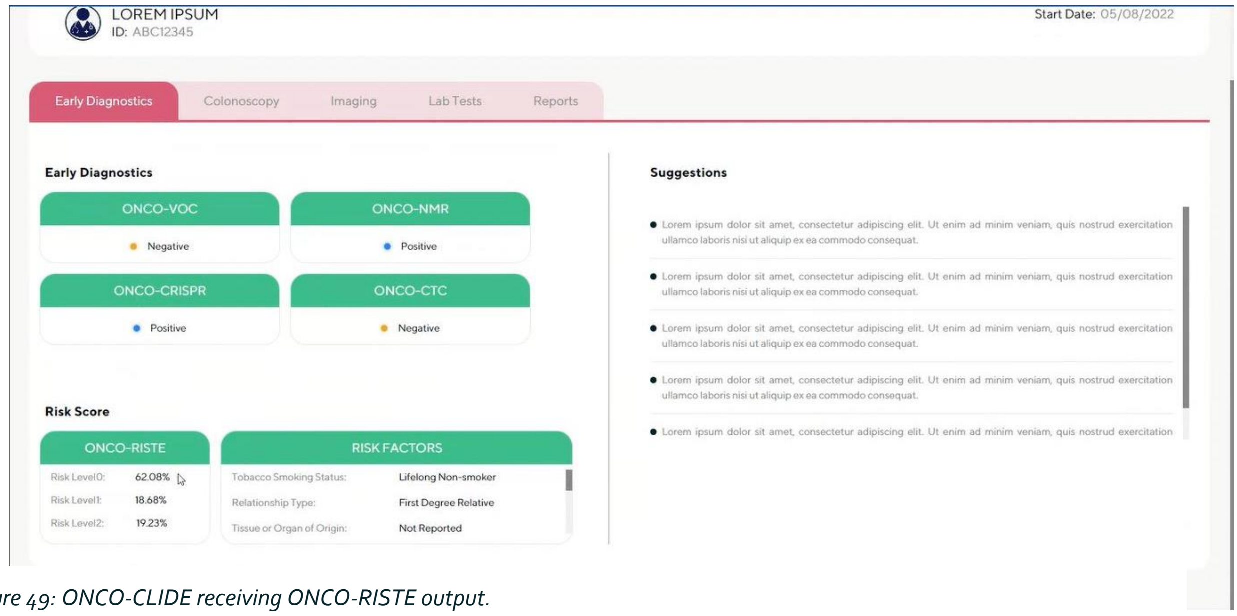


Figure 49: ONCO-CLIDE receiving ONCO-RISTE output.

ONCO-CAWA in Laboratory Test 2 (LT2)

In LIT2 ONCO-CAWA was demonstrated to end-user partners, placing the focus on data collection. Especially questionnaires were used, demonstrating the different UI capabilities in answering them.

The API of ONCO-CAWA was used to ingest the answers to the questionnaires, and feed ONCO-RISTE with the necessary attributes to perform risk assessment. The assessment was captured by a first version of the Kafka ingestion service of ONCO-CAWA. The demonstration is already described in section 8.6.

ONCO-CLIDE in Laboratory Test 2 (LT2)

The second lab integration test (LIT) marked a significant milestone in our project development, characterized by several key advancements. Primarily, our team successfully integrated user roles and user management features into the web platform. This enhancement is crucial as it enables distinct access levels and functionalities for different types of users, ensuring a tailored and secure user experience.

In addition to this, we showcased a preliminary example of integration between the web interface and two specialized tools: ONCO-RISTE and ONCO-VOC. This integration was facilitated through the Kafka setup implemented by EXUS, demonstrating the potential for seamless communication and data exchange between our web platform and these critical diagnostic tools. This integration is particularly noteworthy as it represents a significant step towards complete integration.

Furthermore, we have also undertaken various design changes in both the web interface and the backend. These modifications are not just aesthetic but are aimed at enhancing user interaction and the overall functionality of the system. By improving the user interface, we aim to make the platform more intuitive and user-friendly, while backend changes are focused on increasing the system's reliability, speed, and efficiency.

Overall, the second lab integration test was an opportunity to demonstrate the substantial progress we have made in developing a more integrated, user-centric, and efficient web platform.

ONCO-EVIDA in Laboratory Test 2 (LT2)

ONCO-EVIDA's showcase during LIT 2 built upon the presentation for the previous LIT meeting in order to better convey the progress made since. The presentation consisted of a quick recap of ONCO-EVIDA's nature as a user-friendly tool for policymakers and four parts of new content, one for each technical partner:

1) Introduction to and evolution of the back-end API (CERTH)
2) Live demo of the front-end interface (Catalink)

3) Wireframe of the recommendation engine (VITO)
4) Processing of air quality data (ICCS)

End-users (and other partners) were then once again encouraged to give feedback.

DATA FUSION in Laboratory Test 2 (LT2)

Even though the complex network described in Figure Fusion_1 is promising in correlating CRC metrics with socioeconomic and behavioural features, the lack of publicly available datasets that include such information in patient level make it impossible to be trained without a proper data collection plan. Therefore, the efforts regarding the data fusion tool were focused on the study of publicly available data in a country level. Various databases, including the Global Burden of Disease Study 2019 (GBD 2019) Data Resources (https://ghdx.healthdata.org/gbd-2019) and the Our World in Data online repository (https://ourworldindata.org/), were studied in order to retrieve country level data about socioeconomic, lifestyle, behavioural, health facilities related and environmental risk factors for CRC. This study resulted in a large dataset including food and alcohol consumption habits, smoking rates, share of obese people, antibiotic consumption and CO_2 and greenhouse gas emissions per European country from 2000 to 2020, and CRC metrics. Correlation and regression analysis were performed between risk factors and CRC metrics following two different approaches to study the exposure to each risk factor; namely a 1 and 2-year time-lag analysis and a 10-year moving average analysis.

This dataset will be enriched adding more data about other risk factors like socioeconomic features, such as education levels, human development index and gross national income, health facilities related data, and other lifestyle factors. In addition, such data will be added for more years and countries and age-standardized data will be taken into account for more accurate results. Environmental data analysis will not be performed by the data fusion tool since ICCS is responsible for analysing such data.

9 Next Steps

ONCO-RISTE

Next steps of ONCO-RISTE tool will focus on refining the computation of risk levels, with an expansion of the number of groups from the current tested ones to a total of five levels. The formulation of a rule base, provided by clinicians and the list of risk factors and groups described in D2.1, will further improve the quality of the classification. This will result in two distinct, yet comparable, approaches: one unsupervised, devoid of any rules, and the other incorporating semi-empirical rules and groups delineated by expert knowledge. The reliability of the tool will be ensured by disclosing the rules that are activated to generate the final output.

ONCO-CAWA

ONCO-CAWA is already mature. The next step is to invite ONCOSCREEN partners to use it, to gather feedback. In parallel, the data collection needs of ONCOSCREEN will be specified, and the implementation of the necessary mechanisms will be performed. Then, the clinical info widget will be designed and implemented. Finally, some gamification features will be designed and implemented based on the request of end users.

ONCO-CLIDE

Due to the scope of the task, there are multiple next steps for ONCO-CLIDE.

After preliminary data from the clinical trials are available and subsequent discussions with our partners, the system's architecture will be finalized and suitable algorithms will be selected.

Furthermore, considering the role of our tool as a final data consumer, upon further development of the project's other tools, we will be able to complete our integration.

Finally, a key decision awaits post-consultation with our clinical partners: determining whether to synchronize ONCO-CLIDE with existing health records systems or to opt for a standalone input method via our web platform. This decision will be guided by factors such as data sensitivity, ease of use for healthcare professionals, and overall system integrity.

ONCO-EVIDA

VITO will create template sentences for recommendations that are filled in as the user specifies their query. The strength of the final recommendation will depend on the strength of the correlation of the specified variables and other aspects of reliability (e.g. multivariable analyses are more reliable than univariable analyses).

VITO aims to procure Belgian CRC (screening) data to be used as a use case.

DATA FUSION

In the context of data fusion task, CERTH will design a data lake to serve as a centralized physical data repository for the ONCOSCREEN platform. The data lake will be responsible for handling the clinical data received from the virtual data harmonization layer. The rest of the ONCOSCREEN components will receive all necessary data from the data lake.

Regarding the data fusion tool, other methods for correlation analysis will also be studied and also more data will be collected. Also, CERTH will study the possibility of data fusion to integrate results from different ONCOSCREEN tools.

Conclusion

In conclusion, this deliverable marks a significant juncture in the project's trajectory, offering a comprehensive exploration of the initial development phase of the ONCO-RISTE, ONCO-CAWA, ONCO-CLIDE, ONCO-EVIDA, and Data Fusion Tool. The document details the iterative Co-Designed System activities, underscoring the integral involvement of end-users in the analysis and architecture phases. The project's commitment to meeting end-user requirements is evident, with workshops playing a crucial role in early requirement gathering and fulfilment.

The LITs conducted during the deliverable's development proved instrumental, with end-user feedback contributing to the enhancement of tool quality and guiding subsequent developmental steps.

Obviously, the presented tools differ at the level of maturity. In this iteration, ONCO-CAWA, ONCORISTE and ONCO-CLIDE were at an adequate level of maturity fulfilling required data according to their plan. Data fusion due to its dependency with other tool's output is quite normally expected to have a significant advancement over the second iteration. ONCOEVIDA had some delays in its initial development (due to a delay in hiring PhD student), that is however back on track. A tail-heavy plan has been created for ONCOEVIDA to fulfil its requirements, conducting necessary updates were appropriate in collaboration with the end users. It should be noted that intelligent analytics dashboard for healthcare policy makers are not common requiring additional discussions and collaboration needs among end user and technical partners.

In essence, D4.3 functions as the first stepping-stone that besides reporting on the progress and achievements until now, it also provides a clear roadmap for future iterations of the user-oriented tools that will be exploited in ONCOSCREEN.

Annex A: Code for Data Flow Messages

ONCO-RISTE – ONCO-CAWA -DATA FUSION Integration exchange message


```
{
        "patient_id": "06d42c21-4c97-4171-ac00-3979a51a59d9",
        "riste_input": {
            "tobacco_smoking_status": "Lifelong Non-smoker",
            "relationship_type": "Unknown",
            "tissue_or_organ_of_origin": "Not Reported",
            "primary_diagnosis": "Tubular Adenoma",
            "metastasis_at_diagnosis": "Unknown",
            "classification_of_tumor": "Premalignant",
            "site_of_resection_or_biopsy": "Liver",
            "gender": "Male"
        },
        "riste_output": {
            "risk_level": {
                "risk_level_0": 0.20525000686340086,
                "risk_level_1": 0.5978878025165388,
                "risk_level_2": 0.19686219062006027
            }
        }
    }
```


ONCO-RISTE – ONCO-CLIDE Integration exchange message


```
{
        "patient_id": "06d42c21-4c97-4171-ac00-3979a51a59d9",
        "riste_input": {
            "Tobacco Smoking Status": "Lifelong Non-smoker",
            "Relationship Type": "Unknown",
            "Tissue or Organ of Origin": "Not Reported",
            "Primary Diagnosis": "Tubular Adenoma",
            "Metastasis at Diagnosis?": "Unknown",
            "Classification of Tumor": "Premalignant",
            "Site of Resection or Biopsy": "Liver",
            "Gender": "Male"
        },
        "riste_output": {
            "risk_level": {
                "Risk Level0": 0.20525000686340086,
                "Risk Level1": 0.5978878025165388,
                "Risk Level2": 0.19686219062006027
            }
        }
    }
```