Grant agreement no. 101016112

Project ESSENCE

Empathic platform to personally monitor, Stimulate, enrich, and aSsist Elders aNd Children in their Environment

INNOVATION ACTION

Medical technologies, Digital tools and Artificial Intelligence (AI) analytics to improve surveillance and care at high Technology Readiness Levels (TRL)
SC1-PHE-CORONAVIRUS-2020-2B

Deliverable reference number and title:	**Data Management Plan**
Due date of deliverable:	30th April 2023
Actual submission date:	15th May 2023
Start date of project:	1st November 2020
End date of the project:	30th April 2023
Organisation name of lead contractor for this deliverable	UMIL
Other organizations involved	All

Version 1.0

Horizon 2020 Framework Programme (2014-2020)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

History chart

ISSUE	DATE	CHANGED PAGE(S)	CAUSE OF CHANGE	IMPLEMENTED BY
1.0	20.04.2023		Initial draft	UMIL
2.0	29.04.2023		Enhanced draft	UMIL
2.2	05.05.2023		Cybersecurity inserted	UMIL
3.0	09.05.2023		Comments by UHI, POLIMI and SXT	UMIL
3.1	10.05.2023		Section 14 data description and Section 16 data models	UMIL
4.1	11.05.2023		Integration of comments by POLIMI and SCOM, Section 14 revised	UMIL
4.2	12.05.2023		Appendix A revised	UMIL
5.0	13.05.2023		Contributions of POLIMI and SCOM integrated inside Appendix A	UMIL
5.2	15.05.2023		Contributions of ESE, POLIMI, SCOM, SXT, UMIL integrated in Sections 7, 14 and 16.	UMIL

This project has received funding from the *European Union's Horizon 2020 research and innovation programme under grant agreement No 101016112*

TABLE OF CONTENTS


essence

1 EXECUTIVE SUMMARY

This document provides the final plan for managing the data generated and collected during the project. It covers: a) the handling of research data during and after the project, b) what data have been collected, processed or generated, c) what methodology and standards have been applied, d) whether data are shared/made open and how and e) how data are curated and preserved.

As such, this is the final version of a living document that started before the signature of the Grant Agreement and it has been updated throughout the project through the delivery of an initial Data Management Plan (D1.1, M6) and the Final Data Management Plan (D1.3, the current document, M30).

It is a public document that illustrates the procedures, discussions, ethical issues tackled throughput the project to make the Public Data compliant with EC regulations on one side, and easy to use by the researchers on the other.

D1.3 contains also the list of the used data-sets, their description and model, and how the relevant data of the project are handled and stored.

Data sets will be preserved after the end of the project in an OpenAccess repository for further use by the research community. We have identified Zenodo as most suitable OpenAccess repository (https://zenodo.org/). This is a repository supported by CERN, that has already been used to keep the Monitoring OpenData of the MOVECARE project. To contain an OpenSource repository of ESSENCE data the community **https://zenodo.org/communities/essence2020** has been created in Zenodo

This deliverable integrates and extends the Data Protection Impact Assessment (DPIA) that was uploaded, as a separate document, together with D1.1 and validated by the Daata Protection Officer of the Politecnico di Milano.

2 DEFINITIONS

Dataset: Digital information created in the course of a research project but which is not a published research output. Research data excludes purely administrative records. The highest priority research data is that which underpins a research output. Research data do not include publications, articles, lectures or presentations.

Data Management Plan: A formal working document, which outlines how datasets are handled both during the active research phase and after the project is completed. DMPs in some form are now a requirement of a research grant proposals and therefore must be addressed at the earliest phase of the research lifecycle.

Metadata: Information about datasets stored in a repository/database template. For example, an image may require metadata that describe how large the picture is, the colour depth, the image resolution, when the image was created, and other data. A text document's metadata may contain information about how long the document is, who the author is, when the document was written, and a short summary of the document.

Repository: A digital repository is a mechanism for managing and storing digital content. Repositories can be subject or institutional in their focus

Secondary data: Sources that contain commentary on or a discussion about a primary source.

3 LIST OF ACRONYMS

AES – Advanced Encryption Standard

CA – Consortium Agreement

CBAC – Community Based Activity Center

CIA - Confidentiality, integrity and accessibility (of the data)

D1.1 – Deliverable 1.1 (Data Management Plan – preliminary)

D1.3 – Deliverable 1.3 (Final Data Management Plan)

D3.3 – Deliverable 3.3 (Report and prototype of the ESSENCE engineered components completed)

DDoS – Distributed Denial of Service

DMP – Data Management Plan

DoS – Denial of Service

DPIA – Data Protection Impact Assessment

DPO – Data Protection Officer

DSA – Disturbi Specifici Apprendimento

EAB – Ethical Advisory Board

EC – European Commission

FAIR - Findable, accessible, interoperable and re-usable (data)

GA – Grant Agreement

GDPR – General Data Protection Regulation

GUI – Graphical User Interface

ICT - Information & Communication Technology

JSON - JavaScript Object Notation

JWT – Java Web Token

MDCG - Medical Device Coordination Group

MFA – Multi Factorial Authentication

NDD - NeuroDevelopmental Disabilities

PSC – Project Steeering Committee

SD – Secondary Data

SLD - Specific Learning Disability

SPSS - Statistical Product and Service Solutions (software for statistics)

TLS – Transport Layer Security

TSR – Terminate and Stay Resident

WP – WorkPackage

4 Introduction to the document

Research data is as important as the publications they support. They allow other researchers to verify hypotheses and to build new research upon results without having to start again from scratch. This has been recognized as fundamental by the research community[1] and several OpenData initiatives have been proposed. Hence, the importance for ESSENCE to define a data management policy according to the European Commission Guidelines.

In fact, according to the EC, all project proposals submitted to "Research and Innovation actions" and "Innovation actions" have to include a section on research data management which is evaluated under the criterion 'Impact'. Projects participating in the pilot action on open access to research data have to develop a data management plan (DMP) to specify what data will be open.

This document introduces the final version of the project Data Management Plan (DMP). The ESSENCE DMP primarily lists the different datasets that have been produced by the project, the main exploitation perspectives for each of those datasets, and the major management principles the project implement to handle those datasets.

4.1 Data and Data Management in Horizon 2020

The DMPs have been introduced in the Horizon 2020 Work Programme since 2014:

"… Horizon 2020 … use of Data Management Plans (DMPs) detailing what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. The use of a Data Management Plan is required for projects participating in the Open Research Data Pilot. Other projects are invited to submit a Data Management Plan if relevant for their planned research."

In particular the following definition apply:

What is research data?	Research data refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion or calculation.
What is open research data?	Openly accessible research data can typically be accessed, mined, exploited, reproduced and disseminated, free of charge for the user.

Table 1: Research Data and Open Data as defined by the EC

The purpose of a DMP is to provide an analysis of the main elements of the data management policy that is used by the applicants with regard to all the datasets that are generated by the project.

It specifies the functional aspects: how the principles "FAIR" (findable, accessible, interoperable and re-usable) have been implemented inside the project.

It also specifies the structural aspects in particular with respect to security: how "CIA" (confidentiality, integrity and accessibility) has been implemented.

4.2 Data Management Plan Lifecycle

The data management plan covers the whole data life cycle.

[1] *Kitchin, Rob (2014). The Data Revolution. London: Sage. p. 49. ISBN 978-1-4462-8748-4*

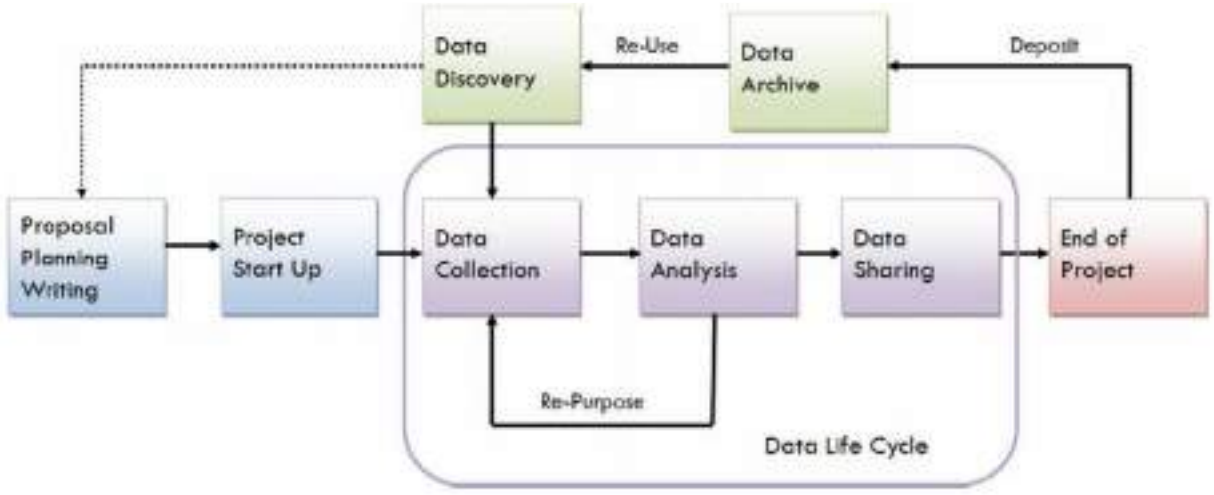


Figure 1 Source: Steps in the data life cycle. Source: University of Virginia, Research *Data Services*

Specifically, the DMP describes the data management life cycle for all datasets to be collected, processed and/or generated by a research project. It covers:

- The handling of research data during and after the project
- What data are collected, processed or generated
- What methodology and standards are applied
- Whether data are shared/made open and how
- How data are curated and preserved

4.3 Purpose of the document

Deliverable 1.3 Final Data management plan is the reference document for gathering, sorting, protecting and sharing all the datasets produced during the operation of the ESSENCE platform and its functioning inside the pilot. It also regulates the sharing of the data after project's end.

The purpose of the DMP is to provide an analysis of the main elements of the data management policy that has been used by the consortium with regard to all the datasets that are generated by the project.

The most important part in the document is the description of the datasets and the explanation of how the ESSENCE project has managed them (generate, store and transmit) and fulfil their adjustment to the requirements provided by the European Commission regarding H2020 projects.

This deliverable builds mainly on the results coming from:

T2.2 (M1-M6) Definition of the users, scenarios and functionalities,

T2.4 (M3-M12) Technical specifications of the components at the engineered prototype level

T2.5 (M3-M12) Design ESSENCE architecture and data model

T3.1 (M3-M24) Activity Center Architecture

T3.2 (M3-M24) Activities provided

T3.3(M3-M24) Invitation mechanisms

T3.8 (M3-M12) Implementation of cloud based data center

T5.1 (M13-30) Field testing of the older adults scenario

Monitoring data forms the basis of the ESSENCE project. They play a crucial role and should be effectively managed to ensure the verification and reuse of research results, and the sustainable storage of the datasets.

This Data Management plan aims at providing a timely insight into facilities and expertise necessary for data management both during and after the ESSENCE project, to be used by all ESSENCE partners and their environment.

The most important reasons for setting up this Data Management plan are:

- Embedding the ESSENCE project in the EU policy on data management, which is increasingly geared towards providing open access to data that is gathered with funds from the EU. The rationale is that the Horizon 2020 grant consists of public money and therefore the data should be accessible to other researchers.
- Enabling verification of the research results of the ESSENCE project.
- Stimulating the reuse of ESSENCE data by other researchers.
- Enabling the sustainable and secure storage of ESSENCE data in the consortium web based repositories.
- Support scientific dissemination.

The DMP considers all the data sets that have been collected, processed and/or generated within the project.

4.4 Application Area

The Data Management Plan provides clear guidelines on how each partner responsible of a dataset needs to take care of the preservation and sharing of the information. These guidelines have been shared and agreed by all ESSENCE partners. They provide a clear picture on each partner's responsibilities regarding the management of the data in the project.

4.5 Document Evolution Procedure

The DMP is not a fixed document but it is a living document outlining how research data have been handled during a research project and therefore it has evolved during the lifespan of the project. It gains more precision and substance during the project lifecycle.

As such, this is the final version of a living document that started before the signature of the Grant Agreement and it has been updated throughout the project through the delivery of an initial Data Management Plan (D1.1, M6) and the Final Data Management Plan (D1.3, the current document, M30).

This first version of the DMP included an overview of the datasets that were foreseen to be produced by the project, and the specific conditions that are attached to them. This was the starting point for developing a sound DMP and it was largely based on the Guidelines on Data Management in Horizon 2020.

All the sections have been updated in this document with respect to Deliverable D1.1. Deliverable D1.3 contains the final view of Data Management. Most significant changes with respect to D1.1 are related to Risk analysis and management that contains also experiments on Cybersecurity (Section 12) and the DataSet description and exchange (Section 14). The new Appendix A (Section 16) contains most relevant data models of ESSENCE.

The content of this deliverable will be considered final at the time of its submission to the European Commission.

4.6 ESSENCE methodology

The **methodology** the consortium follows to create the DMP is constituted of four main steps. These are the following:

1. Create a data management policy. To this end, we describe:
    a. The elements that the EU proposes to address for each dataset.
    b. The strategy that the consortium has used to address each dataset.
2. The elements that are used to create a data management plan are a set of templates, which are distributed to the partners of the consortium in order to fill them with information for each relative data set.
3. Analyze the completed data management plan templates filled by the project's partners. Particular care on data models has been put.
4. Provide a method to store and share the resulting identified datasets and the access policies to them.

These guidelines have been taken into account at all moments of the project.

5 GLOSSARY

The glossary reported in Deliverable D3.3, is reported here for sake of clarity.

Term	Definition
ECO-SYSTEM	Set of users and functionalities devoted to a specific vulnerable population. In ESSENCE, there are two eco-systems, one for children and one for seniors.
TARGET USER	Individuals for whom the ESSENCE system is designed and who use ESSENCE at home. They could be children of the first and second year of the primary school and non- or pre-frail seniors (65+ years).
PROFESSIONAL USER	Clinicians (general practitioners, psychologists, and child neuropsychiatrists) and/or teachers (only for the children eco-system), who interact with the target user at the point of need through ESSENCE providing remote tele-assistance.
EXTERNAL USER	Individuals who subscribe to the ESSENCE system and have access to some content of ESSENCE to interact with the target user. The external user might include the target users' peers, friends, or family members to foster between-generation exchange.
APPLICATION DOMAIN	Sub-set of functionalities which share a common aim. Functionalities are organized in four application domains, which are: - Tele-assistance - Remote monitoring - Stimulation - Social inclusion
MODULE	Aggregate of heterogeneous technical components which represents one of the building blocks of the ESSENCE system. ESSENCE is composed by the following modules: - *Community-based activity center* (CBAC), a holistic platform, based on virtual rooms, which provides the target and external user with a series of diverse activities with declared recreational, assistive, educational, and socialization purposes. - *AI-based monitoring system*, a collection of heterogeneous components, which gathers information from the activities mediated by the CBAC, from a smart ink pen, and from diverse applications to extract cognitive, physical, and emotion-related indicators. This module exploits Artificial Intelligence (AI) to provide alerts for deviations from physiological behaviours. - *ESSENCE manager*, the system control unit which manages user authentication and profiling, system configuration, possible

	system malfunctioning, and delivery of notifications to the users about the status of the ESSENCE system. - *Light Health Monitoring Module* (LHM), a web-interface for the professional users to plan activities of the target users, to schedule and access tele-consultation, to receive alerts from the AI-monitoring module about users' status, to monitor the usage of the ESSENCE system.
COMPONENT	Hardware and software tools used by the modules to achieve their specific objectives. The components can be devices, sensors, user interfaces, and applications.
FUNCTIONALITY	A specific application use case defining all the interactions between the users and the components, which are organized in four application domains. Examples of functionalities are tele-consultation, handwriting, voice analysis, virtual gym, cognitive games, etc. For a complete list of ESSENCE functionalities, please refer to Deliverable D2.2 (Month 12).
MEASUREMENT	Raw readings from a specific component associated to a single functionality.
INDICATOR	Features derived from measurements in order to assess the target user's status in the following domains: cognitive, physical and emotional valence and arousal.
NOTIFICATION	Type of feedback provided by the ESSENCE system (e.g., ESSENCE manager) to the target users. Examples of notifications are information about ESSENCE's status or reminders for the users, such as "recharge the smart pen", "remember the tele-consultation scheduled for tomorrow", etc.
REPORT	Feedback from the AI-based monitoring module to the professional users about the status of the target user, which is visualized in the LHM module. If a change from the normal behaviour is identified, the report is labelled as ALERT and specifically highlighted in the LHM module.
ACTIVITY	A single interaction task between a user and the ESSENCE system. A single functionality may include multiple activities. For example, the cognitive games functionality comprises several activities such as "playing puzzle", "playing cards", etc. Each activity is described by a model.
DATA CENTER	The main data repository that saves all the data (measurements, indicators, alerts, notifications, activities) managed by the ESSENCE system.

MEDICAL DEVICE (MDR 2017/745)	‘medical device’ means any instrument, apparatus, appliance, software, implant, reagent, material or other article intended by the manufacturer to be used, alone or in combination, for human beings for one or more of the following specific medical purposes: — diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease, — diagnosis, monitoring, treatment, alleviation of, or compensation for, an injury or disability, — investigation, replacement or modification of the anatomy or of a physiological or pathological process or state, — providing information by means of in vitro examination of specimens derived from the human body, including organ, blood and tissue donations, and which does not achieve its principal intended action by pharmacological, immunological or metabolic means, in or on the human body, but which may be assisted in its function by such means.
SOFTWARE AS A MEDICAL DEVICE (SaMD)	software intended to be used for one or more medical purposes that performs these purposes without being part of a hardware medical device.
HEALTH SOFTWARE	software intended to be used specifically for maintaining or improving health of individual persons, or the delivery of care
DAILY SCHEDULE	The activities that the user has to perform inside the day: these activities can be mandatory or suggested.
TELECONSULTATION	an activity in which the child/senior performs a cognitive test under the supervision of a clinician through a video call integrated in the test window. It produces a clinical report of the test.
USER PROFILE	all the personal information of a user required by the platform. It will be saved in a separate data base with respect to all other types of data

6 ESSENCE project and data

The ESSENCE project[2] is an Innovation Action project funded by the European Union Horizon 2020 research and Innovation programme under the grant agreement number 101016112.

It involves six countries and is coordinated by POLIMI. ESSENCE aims at using the data and information gathered during the Covid-19 pandemic period to open new opportunities for services targeting vulnerable populations:

- non- or pre-frail seniors, ages ≥ 65 years
- and children of the first years of primary school (5-7 years old).

6.1 ESSENCE functionalities

To contribute to the public health response in the context of the ongoing epidemic, and preparedness for future emergencies, ESSENCE aims at boosting the creation of a new model of home-based care that relies on:

- Stimulation
- Remote monitoring
- Tele-assistance
- Social inclusion, favouring the connection between users, families, and professionals.

Figure 2 ESSENCE users and application domains.

The main aim of this project was the design of an innovative, multi-faceted platform to connect, stimulate, assist and monitor two categories of vulnerable target users: non-frail or pre-frail older adults independently living at home, and children - with particular attention to those with learning difficulties. ESSENCE has been buildt upon the results of MoveCare H2020 and adopted an iterative optimization process between two pillars - technology transfer of a subset of successful modules and feedback received from testing on target users – with the final target of achieving CE mark and promptly entering the market.

Two different **eco-systems** (senior eco-system and children eco-system) have been developed for the two vulnerable groups of target users. Each of the two eco-systems is characterized by different types of users:

- **Target users**, who are the primary users of ESSENCE;
- **Professional users**, who are the professionals (clinicians and/or teachers) supervising the use of ESSENCE; each target user is associated to a single professional user, while each professional user can have in charge several target users;

[2] https://www.essence2020.eu/

- **External users**, as friends or family members of primary users, who are other potential target users who received a reduced set of ESSENCE functionalities and are mainly involved in the multi-players games for social inclusion.

The ESSENCE platform consists of four main components described in details in Deliverable D3.2 Report and prototype of the first release of the ESSENCE components, released at M24.Essence is a modular system that is composed of 4 modules:

A) Community-based activity center (CBAC), a holistic platform, based on virtual rooms, which provides the target and external user with a series of diverse activities with declared recreational, assistive, educational, and socialization purposes;

B) AI-based monitoring system, a collection of heterogeneous components, which gathers information from the activities mediated by the CBAC, from a smart ink pen, and from diverse applications to extract cognitive, physical, and emotion-related indicators. This module exploits Artificial Intelligence (AI) to provide alerts for deviations from physiological behaviours;

C) ESSENCE manager, the system control unit which manages user authentication and profiling, system configuration, possible system malfunctioning, and delivery of notifications to the users about the status of the ESSENCE system;

D) Light Health Module (LHM), a web-interface for the professional users that allows them to:

- plan activities of the target users,
- schedule and access tele-consultation,
- receive alerts from the AI-monitoring module about users' status,
- monitor the usage of the ESSENCE system.

ESSENCE adopted the continuous integration and testing approach, starting these activities early on (M5) and continuously performing them in parallel with the component and system prototyping. This process was facilitated with regular technical workshops organised biweekly until the release of final ESSENCE prototype. The technical workshops were mainly carried out in a remote/hybrid setup primarily due to COVID restrictions, with two face-2-face workshops organised in M18 and M24. In the beginning these workshops were more focused on integration, with the focus gradually shifting to testing activities as components and the system grew more mature. Similarly, testing activities gradually shifted from component, to integration, system, and early functional testing with target users, in line with the testing approach presented in D4.1 Protocols and metrics for system technical and functional testing at engineered level.

Components testing was performed by technical partners responsible for each component, and it involved testing the internal logic of software components, data application of stubs/drivers for testing in isolation the component's external interfaces, **data confidentiality, integrity and accessibility**, as well as hardware specific testing where relevant, most notably for voice monitoring and smart pen components.

Integration testing was concerned with proper interactions between components, once they were mature enough and (gradually) integrated with other components. Integration tests were performed as a joint activity among two or more involved partners responsible for different components. These tests involved stepwise test checklists derived from activity diagrams, which describe different ESSENCE functional workflows in a graphical manner, with a focus on high-priority functionalities.

The system was then tested in **real environments**: final tests of the whole system were performed to check the proper functionality and interoperation of various components in supporting various functional scenarios. The system tests involved describing a full platform use scenario with all the

steps involved, from initialisation steps to various activity planning and execution, involving all ESSENCE modules/components and high priority functionality. Given two separate pilots focused on somewhat different sets of ESSENCE functionalities for different end users, two such system test scenarios were used, i.e. system test scenario for seniors in Servymayor facilities and system test scenario for children in Gavirate school in Varese.

Finally, the early functional testing involved a small number of target users (both, end users and professional users/clinicians), who tested the system in an ecological environment to provide timely and valuable feedback from users' perspective, which helped improve the system in technical terms and, more importantly, in usability terms.

A longitudinal field testing of 12 months has been carried on both target populations with the deployment of the ESSENCE system at home.

6.2 The ESSENCE components

The Essence components can be subdivided between Home components to be deployed at users' home and Cloud components, allocated in the cloud. From the functional point of view, the following components are identified:

	Description
Smart pen	Sensorized pen to capture measurements that characterize the elder's frailty and physical abilities as well as the child's handwriting skill/progress. It is used as a standalone device storing data on-board and also to perform activities through the Activities Center's interface, e.g. controlled tests during teleconsultation.
Smart phone for voice monitoring	Smart phone of the user that runs a mobile application as a background service. It is used by monitoring module that combines voice analysis and profiling of phone conversations with the final goal of detecting early signs of cognitive decline (in elders), social interaction changes, and emotional valence and arousal. All indicators are stored in the cloud.
Tablet computer	Tablet is used as one of the interface modalities for users to access functionalities of the Activity Center, such as cognitive tests, serious/cognitive games, social interaction, tele-assistance, etc.
Home Station	The Home station comprises a n all-in-one PC and an RGBD camera with built-in microphone. The all-in-one computer is a compact desktop PC integrated with a display, which provides necessary processing power to run the CBAC component in user's home. The integrated display is used as the main interface for interacting with the CBAC. The camera is used to monitor and provide real-time tracking of user motion to animate its avatar inside the exerg-games.
PC	A standard PC for Health professionals for Light Health monitoring.
Data center	It stores the data collected from the use of ESSENCE system and the configuration data. It contains also all the cloud components functional to ESSENCE.

These components are connected with the users, from the functional point of view through the scheme reported below that outlines also the data exchanged by the different modules as shown

below.

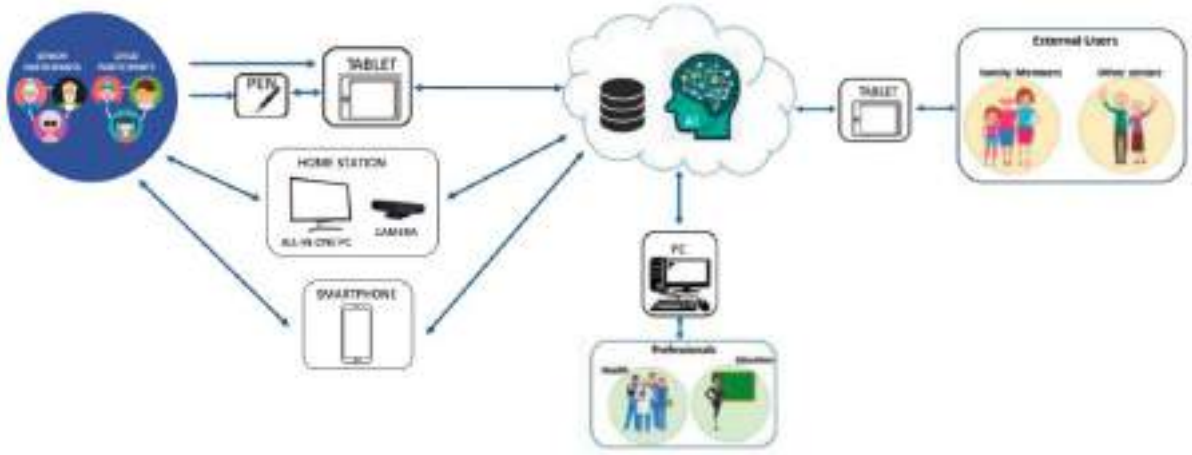


Figure 3. ESSENCE overview showing components provided *to the users.*

These components are connected according to the following Figure.

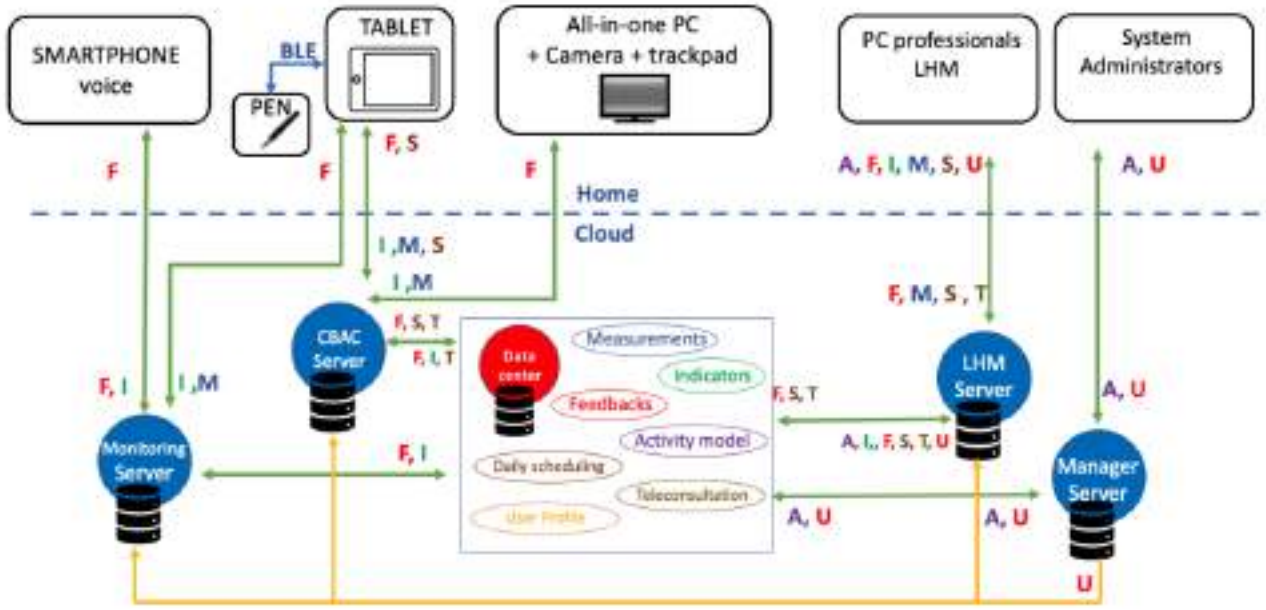


Figure 4. ESSENCE components and the data exchanged between them: Measurements, Indicators, Feedbacks, Activity models, daily Scheduling, Teleconsultation and User Profile are distinguished (cf. Glossary in Section 5).

7 Data Summary

In ESSENCE data are collected in three phases of the research process: Phase 1) Co-design of the engineered prototype (WP2), Phase 2) Integration and testing (WP4) and Phase 3) Field testing in the two relevant scenarios (WP5).

Phase 1: CO-DESING /

In the first months of the project, we administered questionnaires about COVID impact on daily life habits, humor, depression and general behavior of seniors and children in 4 of the countries involved in the consortium: Spain, France and Israel. The questionnaires were performed in the form of surveys under the supervision of FS in Spain, UIN in Italy, ESE in France and UH in Israel. In the same phase, focus groups with stakeholders have been organized to refine the requirements of the ESSENCE system.

The survey data acquisition procedure for the project purposes is in the form of qualitative data in the case of questionnaires and focus groups and quantitative data (cardinal, ordinal and nominal responses) to questions relating to user's socio-economic status, living arrangements and behavior.

All research methods adopted are well known and broadly utilized and have measures established to ensure ethical practice. All work has been undertaken in compatibility with national and EU law or with Israel (see Section 5.1.3 non-EU country) law, and none of the proposed research is foreseen to face any legal obstacles or objections.

Phase 2: EARLY TESTING

Starting at Month 18, 36 children and 9 seniors have be requested to participate in testing the ESSENCE system at the primary school of Gallarate (Varese, Italy), and in protected environments (protected apartments of Servimayor in Extremadura).

In particular, the system's early testing carried out in the children ecosystem allowed to stress test the CBAC operation with numbers of users that were larger than those typically tested in the lab (up to about 40 concurrent users compared to at most 5 in the laboratory for internal testing).

Phase 3: FIELD TESTING

In the last year of the project, seniors and children have been asked to participate in testing the ESSENCE system for a 1-year period.

For the children ecosystem, 66 children were recruited from Gavirate and Voltorre schools. Also 8 teachers and 4 clinicians took part in the field testing. The tablets were delivered to the participants in December 2021, but the structured activities only started in February 2022, with weekly meetings between groups of children, led by a teacher. Starting from June 2022, there were no more planned activities, and the use of the tablet took place on the child's voluntary initiative. The field testing phase ended in December 2022.

For seniors 66 people were recruited in different municipalities in Extremadura. Clinician of the CNC (third party of FS) participated in the field testing. Deployment started in September 2022 and the testing ended at the end of the project (30 April 2023). 6 seniors were also recruited in France to test the smart pen between March and May 2023. Data sets included information collected by the pen, subject's gender and age and MMSE test scores.

For both Phase 2 and 3, which include testing of the ESSENCE system on users (WP4/WP5), the national legislation of France, Italy and Spain has been the legal and ethical framework (cf. all deliverables associated to WP8).

In conducting research and deliberations, the national legislation of countries in which activity is implemented has been guided by ethics. The whole list of informed consents, information sheets and ethical committee procedures which are needed for data collection in all 3 phases is reported in Deliverable D8.1.

7.1 Types of data

The Processing under consideration is the management of the users' personal information and data in the framework of the ESSENCE project.

Complex set of data, including user's lifestyle and health related parameters, and environmental data, have been collected, stored and managed during the project lifecycle. All the data are processed according to GDPR[3] regulation.

The consortium has signed a Data Sharing Agreement (Appendix B) to ensure that there is in place proper arrangements relating to personal data transferred or shared between members of the ESSENCE consortium. In agreement with art. 26 of GDPR, it has been agreed that each of the parties is a joint data controller in relation to the data being transferred or shared under for the purpose described in the Grant Agreement. For scientific analysis and dissemination purposes, users' data are shared with all the partners of the consortium in a complete anonymized form.

Data which have been collected in the ESSENCE project are classified into two main categories:

- **General Personal Information (Private data)**

These data include mainly personal and identity information and contacts (phone number and email) and those data that are released directly by the users who give their explicit consent to the management conditions. These data are collected mainly in the written informed consent after detailed information ex art. 6 lect. a) of GDPR and in the form of questionnaires on user preferences.

This is in line with and respectful of the principle of privacy by design (art. 25 GDPR, privacy by default). The informed consent is provided in written form in accordance with deontological rules.

General personal information data are collected and stored at the recruitment site (FS for seniors and UIN for children). Contact data (email) are stored in the database under the responsibility of SCOM, the partner responsible for user authentication, but are encrypted and the access to data in the database is protected with authentication and authorisation mechanisms.

- **Behavioural Data and Health Data (Sensitive Data)**
    - *Behavioural Data* are acquired from the applications integrated in the CBAC module (e.g. exergames, serious games, social games, …), from the smartphone through the voice analysis application and from the smart ink pen. Then, data are gathered by the AI monitoring module which extracts relevant indicators in order to track the participant's cognitive state for the elder and to identify possible onset of DSA in children. These data are classified as personal data under the privacy regulation.

      These data are obtained through consent after detailed information ex art. 6 lect. a) of GDPR. This is in line with and respectful of the principle of privacy by design (art. 25 GDPR, privacy by default).
    - *Health Data* includes all the health data acquired through the Light Health Monitoring module and all the data acquired and stored during and after teleconsultation (e.g. professionals' reports of the tele-consultation on specific tests done during the consultation). These data mainly relate to physical and cognitive status of the user and are classified as special categories of personal data by article 9 of the GDPR.

[3] https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_it

- These data are obtained through consent after detailed information ex art. 6 lect. a) of GDPR. This is in line with and respectful of the principle of privacy by design (art. 25 GDPR, privacy by default).

  All partners can have accessed to behavioural and health data and data processors, but data are shared anonymised.

The raw data collected by the ESSENCE platform are called measurements (see Glossary, Section 0) and are stored into the ESSENCE cloud platform as a data source. Then, several applications on the ESSENCE servers included in the cloud architecture are processing the measurements to transform them in indicators. Such indicators are managed by the AI monitoring module, the intelligence of the system, which tries to detect preference and anomalies able to provide inputs to the ESSENCE Manager. Such inputs are called feedbacks and are suggested activities, alerts and reminders for the different users. The ESSENCE Manager distributes feedbacks to the user. The feedbacks are provided through notifications to the ESSENCE tablet or the smartphone, and/or via emails to professionals depending on the context and feedback meaning.

In Figure 3 all the main components of the ESSENCE project interacting with the users are shown. The target users at home are provided with a pen connected to a tablet, a home station for exergames, and some applications to be installed on their smartphones. The professional users instead are provided with web applications that can be run from their working stations. The external users can also connect to the ESSENCE platform with their own device (i.e., a tablet) to join social activities.

Personal data will be stored for five years and then destroyed. After five years, data will be kept in a complete anonymized form only for scientific purposes.

Anonymous data, in an aggregated form, are also shared in open access one month after data generation as declared in the Article 29.3 of the Grant Agreement.

In Section 13, a list of ESSENCE functionalities is reported: for each functionality, raw measurements and exemplary indicators are reported.

7.2 Purpose of the data

Data in ESSENCE are central at the beginning for defining user needs (phase 1), for testing the engineered prototype in real environment (phase 2) and then for monitoring and assisting the users at home and personalize their activities (phase 3).

The questionnaires used in Phase 1 include generic questions on the impact of COVID in the quality of life, on their habits and needs, on their use of technologies and so forth. All answers are pseudonymized. The data collected in Phase 1 are used to define the user needs, and thus refine the ESSENCE platform accordingly.

Concerning Phase 2 and 3, the data collected are used by the *ESSENCE manager* that communicates with all the users involved in the target user ecosystems (professionals, caregivers, the target user himself/herself and so forth).

In particular as reported in the DoA (pp 3-4 part B) the data are collected by the different modules as follows:

The monitoring module gathers heterogeneous information from the activities mediated by the CBAC, the smart ink pen, the smart phone and diverse applications. It extracts relevant indicators for both populations in order to track the participant's status and adapt the profile in the following domains:

- **Cognitive Status**: information collected through *Tele-consultation* (professionals' assessment from tele-consultation), Handwriting and Voice Analysis and cognitive digital tests with the goal of

monitoring seniors at risk of age-related cognitive decline and children at risk of Specific Learning Disability (SLD) and Neuro-Developmental Disabilities (NDD).

- **Physical Status**: information collected through the *Light Health Monitoring* Handwriting Analysis with the goal of monitoring eniors in terms of age-related changes in balance and tremor. .

The *Monitoring Module* exploits AI with a twofold aim: i) to enrich the users, providing personalized suggestions and feedback on their strengths, thus maximizing engagement and relieving stress; ii) to timely detect deviations from usual physiological behaviours and send alerts to others (family members, health and education professionals) to foster prevention and anticipation of care.

The ESSENCE Manager coordinates the platform: it is the system control unit that manages the user profile registrations, the system configuration, and possible system malfunctioning in a proactive way. In addition, it receives pre-processed data coming from the AI-based monitoring module and distributes notifications, alerts, or feedback accordingly to the user of interest."

All these behavioural data collected in Phase 2 and 3 cannot be considered strictly biometric data because from the time series it is not possible to identify univocally the person.

7.3 Data life cycle

The ESSENCE project conducts a prospective interventional study on the impact of novel technologies at keep or improve a good health and well-being status.

In particular, through the final field-testing study, the ESSENCE project is aiming at validating the ESSENCE platform in relation to three main factors:

- Feasibility, i.e. carrying out a pilot study on a small court (i.e. a not statistically significant sample size for assessing clinical outcomes, compose by 120 test users - 60 children and 60 elders)
- Usability
- Acceptance.

General Personal Information are collected by the two pilot sites of the ESSENCE project:

- FS in Spain, involved in the recruitment of seniors, helped by three third parties, as previously mentioned, which are:
  - Servimayor, involved in the early testing phase
  - CNC, involved in the on field testing phase
- UIN in Italy, involved in the recruitment of children, with the help of the Territorial School Office of Varese, as third party, for both testing phases.

These data are collected mainly in the informed consent on voluntary based participation.

Evaluation data have been collected by means of written questionnaires to have an evaluation by the user before and after testing the platforms. Data necessary for ESSENCE evaluation are inserted in a structured database according to the most common information technology standards (.xls, .CSV, json data format).

In Table 1 all the steps of the life cycle of data processing are reported and detailed from the creation of a user account towards the deletion of the whole data.

Table 1 Life cycle of data

PROCESS	DETAILED DESCRIPTION OF THE PROCESS
Create an account	The user provides identification data (email) and opens his/her new account. In the case of children, the parent or legal representative oversees the account creation.
Enter the initialisation data	The user chooses his/her preferences so that the configuration and initialisation data are entered on the device (tablet, smartphone, smart TV for the target user or PC for professionals). In the case of children, the parent or legal representative oversees data initialization.
Transfer data	Data are transferred to the cloud architecture
The ESSENCE platform collects data	The ESSENCE applications are used by the user and the collected raw data are stored to the cloud architecture.
Computation of indicators	The raw data are processed on the servers by means of automatic algorithms able to compute the daily indicators. The indicators are stored on the MongoDB data base in the ESSENCE cloud architecture.
AI reasoning	The monitoring AI processes all the indicators saved on the MongoDB data base on the cloud in order to derive feedbacks for the users in case of anomalies and/or behaviour changes. The feedback are suggested activities, alerts and reminders to the relevant user. The feedbacks are stored on the MongoDB data base.
Feedbacks sent to the home devices	The feedbacks are sent to the relevant user by means of notifications for the mobile, or the tablet or via email.
Share data	The generated data (indicators) are shared in open access one month after data generation as declared in the Article 29.3 of the Grant Agreement.
Delete data	Personal data are deleted 5 years after the end of the project.

7.4 Data collected characteristics

The data register delivers such information according to Annex 1 of the Horizon 2020 guidelines (2015) and the choice made by ESSENCE (in between brackets):

- **Data set reference and name**: Identifier for the data set to be produced (data sets do have a unique identified).
- **Data set description**: Descriptions of the data that are generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it is associated to a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse (an additional file explaining the data is stored with the data in the OpenData repository, information on the data is clearly available to partners who have shared data definition and design).
- **Standards and metadata**: Reference to existing suitable standards of the discipline. If these do not exist, an outline on the metadata and a data model is reported (standard format are used: .cvs and json).
- **Data sharing**: Description of how data are shared, including:
  - access procedures,
  - embargo periods (in ESSENCE there was none),
  - outline of technical mechanisms for dissemination (in ESSENCE Zenodo is the choice)

  - necessary software and other tools for enabling re-use, and definition of whether access can be widely open or restricted to specific groups (in ESSENCE there is no restriction and no additional tools or software are required).
  - Identification of the repository where data are stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc. – ESSENCE data center has been used throughout the project and Zenodo afterwards)
  - In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related, in ESSENCE all measurements that can be valuable for further research are shared, other data are for internal use).
- **Archiving and preservation (including storage and backup)**: a description of the procedures that have been put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered (see Section -).

Data sets description and details are reported in Section 16. In particular, data are complemented with a data model that contains the explanation of the data. A metadata file is assigned to datasets for effective and persistent citation when it is uploaded to the web repository. This metadata file can be used in any relevant publications to direct readers to the underlying dataset. The metadata file is stored in the ESSENCE data center and transferred in Zenodo for the data publicly accessed.

7.5 Origin of the data

Data in ESSENCE are heterogenous and provided by different devices.

Primary data are provided by the CBAC during the interactive activities and by the LHM. Other data are provided by the smart objects that belong to the smart network monitoring the subjects, that comprehends the ink pen and the smart phone. Additional data define the users s (e.g. – age, weight, MMSE score …).

From these primary data, indicators are computed by ESSENCE and stored in the Data Center. These are for instance the features automatically extracted on the fly from the voice signal over the smart phone, indexes of tremor intensity during handwriting computed from the pen raw data, speed with which an activity is performed, and so forth.

7.6 Size of the data

During the project we have collected the following amount of data (cf. Figure 1).

Essence manager server. The Essence Manager server + Data base was set for 25 GByte of disk usage. The data part, that contains all the user personal data, occupies 459 Mbyte of space.

Monitoring server. The monitoring application server + Data base was set for 56 GByte of disk usage. Inside this space both AI (26 GByte) and monitoring (30 GByte) data and applications are stored. In particular, inside the monitoring server the following data are contained.

- measurements 691.33 MB
- indicators 52.74 MB
- maintenance (it contains all the logs) 21.40 MB
- other (users' profile for the minigames) 114.59 KB

Most measurements are produced by the Smart pen: it samples at 50 Hz records constituted of 8 data (X, Y, Z acceleration; X, Y, Z angular velocity, pressure, timestamp) that produces 40 bytes 500 Byte / s. Hypothesizing a use of 10 minutes every day, we collected 2,1 Mbyte / month / user.

CBAC server. Two instances of the CBAC server have been deployed: one for the Spanish and one for the Italian pilot. One virtual machine is allocated for each instance; 20 GByte have been occupied for the Spanish instance and 17 GByte for the Italian one. In addition a third machine is allocated to store all the data. Inside this machine data from the Spanish CBAC and the Italian CBAC are stored separately. The amount of data space for the Spanish pilot is of 1,5 GByte and 0,7 GByte for the Italian pilot.

Most measurements are produced by the CBAC activities: it samples at 10 Hz, records of 3 data (x,y position, pointer activity (click, drag) that are 3 Bytes that produces 30 Byte /s. Hypothesizing a use of 30 minutes per day, we collect 378 Kbyte / month / user.

LHM Server. The Local LHM server + Dabase is set for 49,75 GByte of which 1,5 are reserved to Application Data and 80 Mbyte of logs.

8 Allocation of resources

There are no immediate costs anticipated to make the datasets produced in ESSENCE. The datasets are deposited in an OpenAccess website repository for at least 5 years after the conclusion of the project.

Each ESSENCE partner should respect the policies set out in this DMP. Datasets have been created, managed and stored appropriately and in line with European Commission and local legislation. Dataset validation and registration of metadata and backing up data for sharing through repositories is the responsibility of the partner that generates the data in the WPs.

The datasets in the ESSENCE web server project will be preserved in a UMIL storage archive, off-line, in line with the European Commission Data Deposit Policy, with no costs. The data will be preserved for 5 years and there are currently no costs for archiving data in this repository.

The data that will be made available to the scientific community will be moved to an OpenAccess repository like Zenodo at no costs.

Indeed, costs have been minimized as functionalities have been developed with inter-operability in mind: the data generated by all components will be re-used first into the project and therefore a clear data model has been set-up and maintained throughout the development.

In particular, data management has been supervised by UMIL and in particular by Prof. N. Alberto Borghese. His declared costs cover also these tasks and are associated to D1.1 and D1.3.

8.1 Data repositories

General personal information data include personal information and those data are collected mainly compiling and signing paper sheets.

For the questionnaires associated to phase 1 (co-design of ESSENCE platform), data from questionnaires collected in France, Israel and Spain have been pulled together in Israel. Data have been anonymized so that there is no personal data which enable to identify the individuals at the data file. The anonymized data are saved in a data base internal to the laboratory which is protected by University of Haifa.

For the questionnaires associated to pre-pilot and pilot, the data necessary to the ESSENCE platform are inserted, as pseudo-anonymized data, in a structured dataset or database according to the most common information technology standards (.xls, .CSV, .TXT, .JSON data format).

Instead, Behavioural and Health Data acquired from ESSENCE modules are organized in measurements (readings from a single sensor) and indicators (characteristics of a user which may be derived from measurements) and are represented as a JSON format for message exchange between ESSENCE modules, as well as for the storage in a NoSQL (Mongodb) database collection.

Personal data are encrypted and sent to the Data Center through ad hoc implemented API stored in a repository separated from that receiving all the other data. Measurements and indicators and the other data are sent to the Data Center without encryption.

Responsible for pre post test evaluartion of the pilot is FS for seniors and UIN for children. They used paper based questionnaires and then they inserted data in excel pseudonymised data set using the ESSENCE code of each user. The excel file is the one shared with the other partners for data processing

On the cloud, the AI monitoring module processes these data to provide the correct feedback to the user (e.g. suggested activities, required teleconsultation) or provide report/alert to the professionals on handwriting and voice analysis monitoring. The ESSENCE interface is the tablet CBAC application.

Informed consents will be kept on papers at recruitment sites and thus at UIN and FS for children and seniors respectively.

8.2 Data quality

All devices and apps run a thorough validation in terms of accuracy and reliability. An iterative testing approach of each single component and of the integrated prototype has been used (cf. Deliverable D4.2 – Integration and testing). This means that a prerequisite for the introduction in the field test of the technology is their approval by consortium on the basis of a test report.

The tests and their results are described in the corresponding deliverables:

- D2.3 - Definition of methodology and metrics for testing on users
- D4.1 - Protocols and metrics for system technical and functional testing at engineered level
- D 4.2 - Integration and testing of the complete ESSENCE prototype.

9 FAIR data (Findable, Accessible, Interoperable and Re-usable)

ESSENCE consortium is committed to make all the data produced FAIR: Findable, accessible, interoperable and re-usable.

All the public data of the project , once a data set has been completed, are made available in a public repository during the project by giving access rights to those third parties that need the research data to address the public health emergency.

9.1 Making the data Findable

Indeed, as declared in the Article 29.3 of the Grant Agreement, one month after collection anonymous data will be shared in open access Zenodo OpenAccess repository, using a specific community (Essence2020) at the link **https://zenodo.org/communities/essence2020**. Zenodo is a repository supported by CERN, that has already been used to keep the Monitoring OpenData of the MOVECARE project.

All data interesting for research have been made public in Zenodo, as soon as they are completed. Other data, for instance scheduling information, duration of activities, and so forth, are internally stored inside the data center.

9.2 Making the data Accessible

The ESSENCE project aims to collect and document the data in a standardized way to ensure that, the datasets can be understood, interpreted and shared in isolation alongside accompanying metadata and documentation (cf. Section 14).

A data model has been associated to data that fully describes how they were acquired, their format and their semantical meaning such that they can be used by other researchers fruitfully also long after the data have been produced. A reduced number of data models has been adopted to maximize re-use and standardization.

To this aim **metadata** have been extensively used to categorize the data and fully support semantic query in the cloud database for use inside the project and also outside the project as well as for re-use of the data.

Data have been acquired sequentially in time, and each new recording has its own time stamp that clearly distinguishes from the other data of the same user and of the same type, both in different days and inside the same day.

9.2.1 Open Access

The consortium strongly believes in the concepts of open science, and in the benefit that the European innovation ecosystem and economy can draw from allowing reusing data at a larger scale.

The datasets relevant to research have been made available for re-use through uploads during the project through the Zenodo repository. A specific community of Zenodo has been created to this purpose: **https://zenodo.org/communities/essence2020**.

Nevertheless, the consortium has transferred the ESSENCE shared data produced throughout the project to the Zenodo data repository, as soon as possible and in some cases soon after publications were accepted

Data produced by the ESSENCE platform are stored in the project cloud repository (Figure 2). They are made available to partners upon request, including in the context of checks, reviews, audits or investigations. Data are made accessible and available also for re-use and secondary analysis.

All the research data are of the highest quality, have long-term validity and are well documented in order other researchers to be able to get access and understand them after 5 years.

Data objects are deposited in the cloud repository under the following conditions:

- Partners access to data files and metadata and data files provided over standard protocols such as HTTPS.
- Use and reuse of data permitted.
- Privacy of its users protected.

If datasets are updated, the partner that possesses the data has the responsibility to manage the different versions and to make sure that the latest version is available in the case of publicly available data. Quality control of the data is the responsibility of the relevant responsible partner generating the data.

The definition of a data model allows to access the data easily. Data are stored with most common standard and formats like JSON format to fully support inter-operability. As such, simple APIs that access the data in the OpenAccess repository have been reported to access and download the data. For this reason, no particular documentation on Software for retrieving the data is required.

We aim to provide the right of use of the data only for research purposes and with the request to acknowledge the project name, ESSENCE, that has provided the data. This has been clearly stated in a companion document that specifies the license under which the data are provided and inserted inside the Zenodo repository. Moreover, access to the data has been monitored through the log of the accesses and actions operated by the Zenodo website.

ESSENCE partners also upload in Zenodo all the data and information related to scientific publications produced by the project, also after project's end.

In the last semester of the project the following data sets have been uploaded up to the 30th April 2023 (M30):

- Technology use characteristics among older adults during the COVID-19 pandemic: A cross-cultural survey.
- Deep Learning and Procrustes Analysis for Early Dysgraphia Risk Detection with a Tablet Application.
- Digital Tools for Handwriting Proficiency Evaluation in Children
- Can Free Drawing Anticipate Handwriting Difficulties? A Longitudinal Study
- Identification and characterization of learning weakness from drawing analysis at the pre-literacy stage
- Investigating the effects of COVID-19 lockdown on Italian children and adolescents with and without neurodevelopmental disorders: a cross-sectional study.

All these data sets correspond to journal publications.

The following additional data, acquired during the field testing phase, will be added in Zenodo after the end of the project:

For the Seniors ecosystem:

- Cognitive Tests (TMT, Bells)
- Handwriting data
- Voice analysis data
- CBAC activity

For the Children ecosystem:

- Serious Games

- Teleconsultation

The possibility of having restricted data was ruled. Restricted data would be agreed amongst all partners. If a restriction on open access to data was deemed necessary, attempts to make data available under controlled conditions to other individual researchers would have been put in place. However, this condition did not happen during the project.

9.3 Making the data Interoperable

Most used standards have been used to store data. Most used format to store the data is JSON format with a full description of the data record through a complete data model, this is the most used format for internal data.

The SPSS format has been used when statistical analysis is foreseen, for instance for the data set "Technology use characteristics among older adults during the COVID-19 pandemic: A cross-cultural survey)". Alternatively, a combination of Excel and Matlab files with their documentation is used to facilitate the processing by other researchers.

We provide a semantic description of the data that support semantic search both in the applicative and methodological domains.

Keywords in the data model have been chosen according to the best practices of the field of interest such that data can be easily identified and used in different disciplines. We resort to partners knowledge to use the best data description as possible.

9.4 Making the data Re-usable

Data are licensed under wide OpenAccess license, under Creative Commons, of the type: cc by-nc-nd 4.0, that allows full use of the data for non-commercial purposes [4]

Data quality is fundamental for the development of the project itself as all interventions are based on these data. Data quality has been checked first by the consortium partner that produces the data, and the partners who consume or read these data have double 3checked them.

9.4.1 IPR management and Security

Project partners obviously have Intellectual Property Rights (IPR) on their technologies and data, on which their economic sustainability relies. As a legitimate result, the ESSENCE project consortium protects these data and has consulted the concerned partner(s) before publishing data. This might result in a processing of the datasets to be made available to the public.

Another effect of IPR management is that – with the data collected through ESSENCE being of high value – all measures have been taken to prevent them to leak or being hacked as shown in Section 12. Hence, all data repositories used by the project include a secure protection of sensitive data.

9.4.2 Personal Data Protection

For some of the activities to be carried out by the project, it may be necessary to collect basic personal data (e.g. full name, contact details, background), even though the project has avoided collecting such data unless deemed necessary.

National legislations applicable to the project are also be strictly followed, such as the Italian Personal Data Protection Code2 or the Spanish LOPD[5] or see below.

[4] https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode
[5] https://www.boe.es/buscar/act.php?id=BOE-A-2018-16673

The General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679) is a regulation by which the European Parliament, the Council of the European Union and the European Commission intend to strengthen and unify data protection for all individuals within the European Union (EU). It also addresses the export of personal data outside the EU. The GDPR aims primarily to give control back to citizens and residents over their personal data and to simplify the regulatory environment for international business by unifying the regulation within the EU.[1] The GDPR replaced the data protection directive (officially Directive 95/46/EC)[2] of 1995. The regulation was adopted on 27 April 2016. It became enforceable from 25 May 2018 after a two-year transition period and, unlike a directive, it does not require national governments to pass any enabling legislation, and is thus directly binding and applicable.

10 CIA of the data (Confidentiality, Integrity and Accessibility)

Full CIA (Confidentiality, integrity and accessibility) of the data is considered according to documents reported in Section 15 and ESSENCE has taken into account the GDPR data requirements regulations.

All data collected by the project are considered compliant after giving data subjects full details on the experiments to be conducted, and after obtaining signed informed consent forms.

Data have been stored in a cloud data center, that guarantees redundancy in the storage and thus data integrity over time. The details on the cloud architecture are reported in Deliverable D3.1.

Following completion of the project, all the responsibility concerning data recovery and secure storage goes to the OpenAccess repository.

Accessibility has been granted to all partners of the project for the entire duration of the project. At the end of the project, partners have dumped the selected data before moving the data on permanent storage and on the OpenAccess data repository.

10.1 Data Confidentiality

The ESSENCE consortium is composed of 9 partners from the following countries, namely: Italy (4), Spain (1), France (1), Slovenia (1), Cyprus (1) and the Israel (1) covering all key research fields addressed in ESSENCE.

The coordinator is Politecnico di Milano, an Italian technical university. Being the coordinator, this is the entity entitled to have the final data management.

The persons in charge of the different roles and related responsibilities are the following ones:

- Scientific Coordinator: Prof. Ferrante Simona (Department of Electronics, Information and Bioengineering, DEIB, POLIMI), email: simona.ferrante@polimi.it
- Data Protection Officer: Dr. Vincenzo Del Core (Data Protection Officer. POLIMI), email: privacy@polimi.it
- Legal representative: Prof. Stefano Savaresi (Director of the DEIB Dept., POLIMI, as delegate of the Rector), email: stefano.savaresi@polimi.it
- All partners of the ESSENCE consortium have appointed an internal responsible for privacy compliance, who points of contact are reported in Table 1.

- Table 1 List of responsible for privacy compliance for all partner of the ESSENCE consortium.

Partner	Address	Point of contact for the management of personal data
Politecnico di Milano (POLIMI)	Piazza Leonardo da Vinci 32, Milano, 20133, Italy	Dr. Vincenzo Del Core (Data Protection Officer) Email: privacy@polimi.it
Università degli Studi di Milano (UMIL)	Via Festa Del Perdono 7, Milano, 20122, Italy	Data Protecion Officer Email: dpo@unimi.it
Università degli Studi dell'Insubria (UIN)	Via Ravasi 2, Varese 21100, Italy	Data Protecion Officer Email: privacy@uninsubria.it
Fundación para la Formación e Investigación de los Profesionales	Calle Pio Baroja 10, Merida 06800, Spain	Jonathan Gómez-Raja

de la Salud de Extremadura Fundesalud (FS)		Email: jonathan.gomez@fundesalud.es
University of Haifa (UH)	Abba Khushy Blvd Mount Carmel, Haifa 31905, Israel	Dr. Nadav Azoulay Email: nazoulay@univ.haifa.ac.il
SXT srl - Sistemi per telemedicina (SXT)	Via Torquato Tasso 29, Pogliano Milanese 20010, Italy	Luca Piccini Email: lpiccini@sxt-telemed.it
Smart Com d.o.o. Informacijski in Komunikacijski Sistemi (SCOM)	Ulica Brnciceva 45 Crnuce, Ljubljana 1231, Slovenia	Marko Žnidaršič Email: marko.znidarsic@smart-com.si
Signalgenerix Limited (SG)	Grigori Afxentiou 23c Mesa Geitonia, Limassol 4003, Cyprus	Marios Milis Email: marios.milis@signalgenerix.com
Initiation des Seniors aux NTIC Association (ESE)	Cite Phalsbourg 19, Paris 75011, France	Monique Epstein Email: monique.epstein@gmail.com

The main partners dealing with users data during the pilots are:

- SCOM: cloud and IT infrastructure provider (responsible for data storage)
- POLIMI, UMIL, SXT, SG: other technical partners responsible for the other modules to make the essence platform fully working
- UIN: pilot site (responsible for data collection for children)
- FS: pilot site (responsible for data collection for seniors)
- All other partners: responsible for data processor for research purposes.

The consortium has signed a Data Sharing Agreement (Appendix B) at the beginning of the project that contains proper arrangements relating to personal data transferred or shared between members of the ESSENCE consortium. In agreement with art. 26 of GDPR, it has been agreed that each of the parties is a joint data controller in relation to the data being transferred or shared under for the purpose described in the Grant Agreement. For scientific analysis and dissemination purposes, users data are shared with all the partners of the consortium in a complete anonymized form.

Third parties have also been involved in the management of sensitive data and different procedures have been adopted depending on their different roles, according to art. 28 of GDPR. In particular, three different categories of third parties are foreseen:

Institutions involved in the recruitment of the users, which are:

- Servimayor, a nursing home based on a non-profit association and third party of FS, involved in the early testing phase on Seniors;
- CNC, a Professional Association of Neuropsychologists in Spain and third party of FS, supporting in the recruitment of seniors in the field testing phase;
- Territorial School Office of Varese, a Third Party of UIN, supporting in the recruitment of children both for the early testing and the field testing phases.

In this case, since only specific persons within these institutions have access to the data, a written authorization letter to data processing has been signed by those persons.

Companies for Technical Support Services:

- Xtrem company has been hired by FS up to the Grant Agreement Amendment signature, following contracting rule of Spain and Extremadura region. It has been involved to provide technological support to the users during the field testing phase on seniors. Since it accesses the data for possible technical issues, it has appointed a responsible of the data, Mr. Raul Vadillo. After Amendment signature, Xtrem has been hired directly by SCOM with the same aims, because of administrative difficulties by FS in renewing the contract. The contracts stipulated between FS and Xtrem and between SCOM and Xtrem have set out the relevant aspects of data management: subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations.
- B link company has been hired by UIN following contracting rule of Italy. It has been involved to provide technological support to the users during the field testing phase on children. The contracts stipulated between UIN and B-link have set out the relevant aspects of data management: subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations.
- *Cloud platform* (AWS – Amazon Web Services as detailed in the following sections) has been contracted by SCOM. In AWS all data are uploaded: data storage has been limited geographically to EU. By applying for AWS Services SCOM agreed with the AWS Service Terms[6] which effectively represent a contract between SCOM and Amazon Web Services. The AWS Service Terms include the AWS GDPR Data Processing Addendum[7] and are thus GDRP compliant.

Third parties have also been involved in the management of sensitive data and different procedures have been adopted depending on their different roles, according to art. 28 of GDPR. In particular, three different categories of third parties are foreseen:

Institutions involved in the recruitment of the users, which are:

- Servimayor, a nursing home based on a non-profit association and third party of FS, involved in the early testing phase on Seniors;
- CNC, a Professional Association of Neuropsychologists in Spain and third party of FS, supporting in the recruitment of seniors in the field testing phase;
- Territorial School Office of Varese, a Third Party of UIN, supporting in the recruitment of children both for the early testing and the field testing phases.

In this case, since only specific persons within these institutions have access to the data, a written authorization letter to data processing has been signed by those persons.

Companies for Technical Support Services:

- Xtrem company has been hired by FS up to the Grant Agreement Amendment signature, following contracting rule of Spain and Extremadura region. It has been involved to provide technological support to the users during the field testing phase on seniors. Since it accesses the data for possible technical issues, it has appointed a responsible of the data, Mr. Raul Vadillo. After Amendment signature, Xtrem has been hired directly by SCOM with the same aims, because of administrative difficulties by FS in renewing the contract. The contracts stipulated between FS and Xtrem and between SCOM and Xtrem have set out the relevant

[6] AWS Service Terms. https://aws.amazon.com/service-terms/

[7] AWS GDPR Data Processing Addendum. https://aws.amazon.com/blogs/security/aws-gdpr-data-processing-addendum/

aspects of data management: subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations.

- B link company has been hired by UIN following contracting rule of Italy. It has been involved to provide technological support to the users during the field testing phase on children. The contracts stipulated between UIN and B-link have set out the relevant aspects of data management: subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations.
- *Cloud platform* (AWS – Amazon Web Services as detailed in the following sections) has been contracted by SCOM. In AWS all data are uploaded: data storage has been limited geographically to EU. By applying for AWS Services SCOM agreed with the AWS Service Terms[8] which effectively represent a contract between SCOM and Amazon Web Services. The AWS Service Terms include the AWS GDPR Data Processing Addendum[9] and are thus GDRP compliant.

Third parties have also been involved in the management of sensitive data and different procedures have been adopted depending on their different roles, according to art. 28 of GDPR. In particular, three different categories of third parties are foreseen:

Institutions involved in the recruitment of the users, which are:

- Servimayor, a nursing home based on a non-profit association and third party of FS, involved in the early testing phase on Seniors;
- CNC, a Professional Association of Neuropsychologists in Spain and third party of FS, supporting in the recruitment of seniors in the field testing phase;
- Territorial School Office of Varese, a Third Party of UIN, supporting in the recruitment of children both for the early testing and the field testing phases.

In this case, since only specific persons within these institutions have access to the data, a written authorization letter to data processing has been signed by those persons.

Companies for Technical Support Services:

- Xtrem company has been hired by FS up to the Grant Agreement Amendment signature, following contracting rule of Spain and Extremadura region. It has been involved to provide technological support to the users during the field testing phase on seniors. Since it accesses the data for possible technical issues, it has appointed a responsible of the data, Mr. Raul Vadillo. After Amendment signature, Xtrem has been hired directly by SCOM with the same aims, because of administrative difficulties by FS in renewing the contract. The contracts stipulated between FS and Xtrem and between SCOM and Xtrem have set out the relevant aspects of data management: subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations.
- B link company has been hired by UIN following contracting rule of Italy. It has been involved to provide technological support to the users during the field testing phase on children. The contracts stipulated between UIN and B-link have set out the relevant aspects of data management: subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations.

[8] AWS Service Terms. https://aws.amazon.com/service-terms/

[9] AWS GDPR Data Processing Addendum. https://aws.amazon.com/blogs/security/aws-gdpr-data-processing-addendum/

- *Cloud platform* (AWS – Amazon Web Services as detailed in the following sections) has been contracted by SCOM. In AWS all data are uploaded: data storage has been limited geographically to EU. By applying for AWS Services SCOM agreed with the AWS Service Terms[10] which effectively represent a contract between SCOM and Amazon Web Services. The AWS Service Terms include the AWS GDPR Data Processing Addendum[11] and are thus GDRP compliant.

Third parties have also been involved in the management of sensitive data and different procedures have been adopted depending on their different roles, according to art. 28 of GDPR. In particular, three different categories of third parties are foreseen:

Institutions involved in the recruitment of the users, which are:

- Servimayor, a nursing home based on a non-profit association and third party of FS, involved in the early testing phase on Seniors;
- CNC, a Professional Association of Neuropsychologists in Spain and third party of FS, supporting in the recruitment of seniors in the field testing phase;
- Territorial School Office of Varese, a Third Party of UIN, supporting in the recruitment of children both for the early testing and the field testing phases.

In this case, since only specific persons within these institutions have access to the data, a written authorization letter to data processing has been signed by those persons.

Companies for Technical Support Services:

- Xtrem company has been hired by FS up to the Grant Agreement Amendment signature, following contracting rule of Spain and Extremadura region. It has been involved to provide technological support to the users during the field testing phase on seniors. Since it accesses the data for possible technical issues, it has appointed a responsible of the data, Mr. Raul Vadillo. After Amendment signature, Xtrem has been hired directly by SCOM with the same aims, because of administrative difficulties by FS in renewing the contract. The contracts stipulated between FS and Xtrem and between SCOM and Xtrem have set out the relevant aspects of data management: subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations.
- B link company has been hired by UIN following contracting rule of Italy. It has been involved to provide technological support to the users during the field testing phase on children. The contracts stipulated between UIN and B-link have set out the relevant aspects of data management: subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations.

Cloud platform (AWS – Amazon Web Services as detailed in the following sections) has been contracted by SCOM. In AWS all data are uploaded: data storage has been limited geographically to EU. By applying for AWS Services SCOM agreed with the AWS Service Terms[12] which effectively represent a contract between SCOM and Amazon Web Services. The AWS Service Terms include the AWS GDPR Data Processing Addendum[13] and are thus GDRP compliant.The ESSENCE project consortium (all the partners and not only the coordinator) signed with the European Union a Grant

[10] AWS Service Terms. https://aws.amazon.com/service-terms/

[11] AWS GDPR Data Processing Addendum. https://aws.amazon.com/blogs/security/aws-gdpr-data-processing-addendum/

[12] AWS Service Terms. https://aws.amazon.com/service-terms/

[13] AWS GDPR Data Processing Addendum. https://aws.amazon.com/blogs/security/aws-gdpr-data-processing-addendum/

Agreement that defines all the obligations with respect to the data processing. In particular, Article 39 - Processing of Personal Data – establishes the rights and duties of the Commission and of the beneficiaries for what concerns the processing of personal data and the consequences for non-compliance.

The Data Sharing Agreement, which has been signed by all ESSENCE partners, defines specific rules and procedures intra-consortium. In it, the ESSENCE consortium establishes how third parties are foreseen to be given access to the Data.

In case any processing activity would be assigned to another entity, institution or person outside the ESSENCE consortium, a processing contract is signed with it or him, setting out all of the aspects stipulated in Art. 28 of the GDPR: duration, scope, purpose, documented processing instructions, prior authorisation where a processor is engaged, provision of any documentation providing evidence of compliance with the GDPR, prompt notification of any data breach, etc.

The parties in any case ensure that these third parties which are permitted by all Parties, undertake in writing the same obligations as agreed in the Data Sharing Agreement.

As far as data transfer of data outside Europe, we remark that the project consortium includes a partner from Israel. However, Israel is among the few non-EU countries which have received an 'adequacy determination' from the European Commission indicating that they have a data protection framework offering a level of protection equivalent to that provided under EU law.[14]

Furthermore, to minimize the risk of data transfer of data to non-EU countries, only pseudonymized data are transferred to and from Israel. Personal data are collected and stored by the pilot sites responsible at the recruitment of the users.

Data transfers with non-EU countries would be in accordance with Chapter V of the GDPR.

Moreover, the following is clearly expressed in the clauses of the Data Sharing Agreement:

3.2. Data sharing with Partners of the Consortium in Switzerland is covered by the 2000/518/EC: Commission Decision of 26 July 2000 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequate protection of personal data provided in Switzerland (notified under document number C (2000) 2304).

3.3. Exceptional events can bring to redefine the role of one or more consortium partners with respect to the belonging to the EU area: in this case this contract adopt the standard clauses defined by the COMMISSION DECISION of 15 June 2001 on standard contractual clauses for the transfer of personal data to third countries, under Directive 95/46/EC, and the COMMISSION DECISION of 27 December 2004 amending Decision 2001/497/EC as regards the introduction of an alternative set of standard contractual clauses for the transfer of personal data to third countries (notified under document number C(2004) 5271). Except these cases, the project does not foresee to transfer data outside the European Union.

From the operational point of view, contacts are maintained completely separated with respect to the other data Other sensitive data are uploaded in the ESSENCE platform in a pseudonymized form and different measures are taken to assure data protection according to the principle of privacy by design (art. 25 GDPR, privacy by default).All the Behavioural and Health data collected are stored into the ESSENCE cloud architecture as *measurements* that are then processed first to become *indicators* relevant to the single functionalities. Then, the *indicators* are managed by the monitoring

[14] The list of countries covered by a Commission adequacy determination is available at: https://ec.europa.eu/info/law/law-topic/data-protection/data-transfers-outside-eu/adequacy-protection-personal-data-non-eu-countries_en

AI to generate *feedbacks* in the form of suggested activities, alerts and reminders to the relevant user. All these *feedbacks* are orchestrated by the ESSENCE Manager.

All subjects have read and signed before starting using the platform an information sheet in which specific information on data treatment and security is described. Specific attention is given to the permission for open data. A full reference to article 25 of GDPR has been included.

All ESSENCE datasets have been pseudo-anonymized at the time of their publication in the OpenAccess repository, in order to assure privacy regarding the origins of the data.

10.1.1 National legislation

We report here additional documentation provided by single EC states involved in the project.

Italian national legislation

The Legislative Decree no. 196 of 30 June 2003 (the "Data Protection Code"), as amended by the Legislative Decree no. 101 of 10 August 2018, adapts Italian data protection laws to the new provisions of the GDPR. The Legislative Decree no. 101 entered into force on 19 September 2018.

https://www.gazzettaufficiale.it/eli/id/2018/09/04/18G00129/sg

Spanish national legislation

Data protection in Spain is ruled by organic law 3/2018 (Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales).

https://www.boe.es/buscar/doc.php?id=BOE-A-2018-16673

French national legislation

French laws apply the GDPR principles. In French legislation, this was transcribed by the law 2018-493 of 20 June 2018 on the protection of personal data.

https://www.legifrance.gouv.fr/loda/id/JORFTEXT000037085952/

The CNIL (Commission *nationale de l'informatique et des libertés* is the French Data Protection Authority. https://www.cnil.fr/

Cypriot national legislation

On 31 July 2018 the national law providing for the protection of natural persons with regard to the processing of personal data and for the free movement of such data (Law 125(I)/2018), was published in the official gazette of the Cyprus Republic (see Unofficial Translation in English:
http://www.dataprotection.gov.cy/dataprotection/dataprotection.nsf/2B53605103DCE4A4C225826300362211/$file/Law%20125(I)%20of%202018%20ENG%20final.pdf

Israeli national legislation

The legal framework for data protection in Israel is reported at the following link:

https://www.gov.il/en/departments/the_privacy_protection_authority

Slovenian national legislation

Slovenia has yet to implement in its legal system the Regulation (EU) 2016/679 of the European Parliament and the Council, of April 27, 2016, regarding the protection of natural persons with regard to the processing of their personal data and freedom of circulation of these data (GDPR), which became effective as of May 25 2018. For this purpose, a revised Protection of Personal Data Act (ZVOP-2) is currently under preparation in Slovenia. In the meantime, the data protection has been regulated since May 25 2018 by the direct implementation of the GDPR, which has precedence over the (old) Protection of Personal Data Act (ZVOP-1; Official Gazette of Republic of Slovenia No. 86/04, 51/07, 67/07, 94/07, 177/20). The latter is still valid, until superseded by the revised Act (ZVOP-2), but only applied to a limited extent.

The Information Commissioner (IPRS) is the Slovenian national data protection authority:

https://www.ip-rs.si/

10.2 Data Integrity

Data are managed and stored through a structured Mongo database. Specific measures have been implemented, e.g. pseudonymization, cryptography (physical encryption), and proprietary data format.

Specific measures have been implemented in data repositories, e.g. pseudo anonymization, cryptography (physical encryption), proprietary data format, and access control through authentication and authorisation.

The field testing is structured with a continuous monitoring of users' data to check their integrity and coherence. In addition, it is foreseen even in the protocol that users have been contacted on a regular basis to keep on track the pilot. At intermediate points, a check of data protection with users and on the system has been carried out. From this perspective we can say that the ESSENCE consortium performs a continuous update and monitoring of the DMP, assessment of the data quality and integrity and possible update of data risks.

From the implementation point of view, all sensitive personal data are encrypted inside the data centers of ESSENCE and secure transfer through HTTPS protocol has been implemented to guarantee protection.

Essential security mechanisms have been implemented, building on strong Advanced Encryption Standard (AES). Data in transfer has been secured with the TLS protocol. Data at rest are protected with relevant authentication and authorisation mechanisms. Authentication is carried out using cryptographically signed tokens (Java Web Token - JWT). After successful authentication, authorisation for access to services/data is enforced checking user assigned roles against access rights (privileges) assigned to different roles. The platform therefore provides protection at database and communication levels.

The cloud environment has been deployed on the Amazon Web Services platform (limiting geographically to EU the data storage), with all modules running on separate machines (each machine being under the responsibility of a lead technical partner – POLIMI, UMIL, SCOM, SXT) interfacing through secure service oriented (RESTful web services) and message queuing (MQTT) architectures. This modular architecture allows separate teams to build an integrated platform incrementally.

10.3 Data Accessibility

ESSENCE data can be accessed by all partners through the data center. Having a central single node to access the data makes accessibility easier.

A clear definition of the data models used to store the data has been defined with a tight collaboration of all partners and in particular of those partners who generate and possibly use the data. Such models

are maintained and updated throughout the project. Updates have been introduced to accommodate additional fields that can have potential interest (e.g. new indicators suggested by data analysis through AI). The use of NoSQL data bases (MondoDB) with a flexible data format like “json” format, as well as a modular design of the interfaces of the different modules, has facilitated this task.

11 Ethical Aspects

ESSENCE partners are to comply with the ethical principles as set out in Article 34 of the Grant Agreement, which states that all activities must be carried out in compliance with:

a. Ethical principles (including the highest standards of research integrity — as set out, for instance, in the European Code of Conduct for Research Integrity (European Science Foundation, 2011) — and including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct) and
b. Applicable international, EU and national law.

11.1 Purposes specification, making them explicit and legitimate

Before recruitment, each participant receive an information sheet including the information on the research ongoing, the information on the treatment of personal data in accordance with the EU Regulation 679/2016 (GDPR) and the informative note for the treatment of personal data.

In the information sheet is clearly specified the use of data and it has been asked permission also for sharing de identified data in open access repository for further research after the project, by other research teams. In case of denial of this clause, the data associated to that user are not transferred to the OpenAccess repository.

When research participants are children, the informed consent is signed not only by them but also by their legally authorised representative and it ensures that they have sufficient information to enable them to provide this on behalf and in the best interests of the participants.

The informed consent explicitly declares and informs subjects that they participate in a research project on a voluntary basis and of the risks related to their data processing.

Personal data are collected only for the specified, explicit and legitimate purposes of the ESSENCE pilot test and not further processed in a manner that is incompatible with those purposes, in accordance with the Art. 5.1 b) of [GDPR].

In case of dissemination of results data are presented in an anonymous format as in the standard scientific publication policy. This is also said to and approved by the users in the informed consent.

11.1.1 Data consent properties

Prior to the participation all subjects are informed about the data processing procedure and outcomes and their right about data management (access, rectification, opposition, erasure, portability and automated decision making) through the informed consent they sign to enter into the study.

The information is provided to subjects by verbal and written means, in the mother tongue of each participant in their own living country.

A checklist is provided to assure the user has the full comprehension and understanding of participation and data treatment during and after the trial.

After this time, during data collection we can distinguish two cases:

- data processing in the pilot;

- data processing outside the pilot.

During the pilot, the users receive feedback and alerts from the AI monitoring module which suggest activities and inform about any eventual change in behaviour (e.g., an anomaly detected in voice analysis features or handwriting features with respect to their normal patterns). Other data which are visible to the user are summaries on game scores performed with the CBAC.

After the pilot, data are stored in a safe and encrypted way on the project servers, and anytime anywhere the subject is entitled to ask for retrieve these data and processed information.

The consent is obtained by signing the form prepared by the consortium after he/she has read it and had sufficient time to ask questions to the responsible of the pilot: after this he/she can freely and aware decide to sign or not and be recruited or not for the trial.

For exercising the rights regarding access and data portability, the subject should contact the responsible of the pilot in his/her country which is the entitled person, and/or the DPO of the involved institution and ask for access or data portability.

The contact mean is the writing of an e-mail to the above-mentioned responsible people (whose contact details are in the informed consent sheet in the hands of the user) asking for the required action.

In accordance with Art. 20 of GDPR, by virtue of the right to request data portability, the users have a right to receive a copy of their personal data in a structured, commonly used, machine-readable format. The users may also request that the responsible of the pilot in his/her country transfer your data to another data controller indicated by him or her.

For exercising the rights regarding rectification and erasure, the subject should contact the responsible of the pilot in his/her country which is the entitled person, and/or the DPO of the involved institution and ask for rectification and/or erasure.

The contact mean is the writing of an e-mail to the above mentioned responsible peoples (whose contact details are in the informed consent sheet in the hands of the user) asking for the required action. The required operation should be done without undue delay.

In accordance with Art. 17 of GDPR, by virtue of the right to request data rectification and erasure, the users have a right to receive a notification of the rectification and erasure of their personal data without undue delay. The users may also request that the responsible of the pilot in his/her country transfer your data to another data controller indicated by him or her.

All data have been treated only in accordance with the ESSENCE objectives stated in the informed consent.

However, if the subject considers out of scope some data processing, for exercising the rights to restriction and objection, the subject should contact the responsible of the pilot in his/her country which is the entitled person, and/or the DPO of the involved institution and ask for them.

The contact mean is the writing of an e-mail to the above-mentioned responsible peoples (whose contact details are in the informed consent sheet in the hands of the user) asking for the required action.

In accordance with Art. 21 of GDPR, by virtue of the right to request data restriction or objection, the users have the right to receive a notification of the conclusion of the procedure in a structured, commonly used, machine-readable format.

The following CA clauses are relevant.

39.2 Processing of personal data by the beneficiaries

The beneficiaries must process personal data under the Agreement in compliance with applicable EU and national law on data protection (including authorisations or notification requirements).

The beneficiaries may grant their personnel access only to data that is strictly necessary for implementing, managing and monitoring the Agreement.

The beneficiaries must inform the personnel whose personal data are collected and processed by the Commission. For this purpose, they must provide them with the privacy statement(s) (see above), before transmitting their data to the Commission.

39.3 Consequences of non-compliance

If a beneficiary breaches any of its obligations under Article 39.2, the Commission may apply any of the measures described in Chapter 6 of the CA.

11.2 Confidentiality

ESSENCE partners must retain any data, documents or other material as confidential during the implementation for the project. Further details on confidentiality can be found in Article 36 of the Grant Agreement along with the obligation to protect results in Article 27.

11.3 Legal basis on data collection

The legal basis is represented by the GDPR regulation and all the national laws regarding data protection and research with human beings.

ESSENCE data are obtained after the subject has signed the consent after detailed information ex art. 6 lect. a) of GDPR.

This is in line with and respectful of the principle of privacy by design (art. 25 GDPR, privacy by default). Only personal data strictly necessary for data processing, e.g. contacts, have been collected and kept separately from data collected by the ESSENCE platform and analysed by the AI module. The informed consent is provided in written form in accordance with deontological rules.

Furthermore, and first of all, the driving principle that ESSENCE consortium is adopting is not to affect dignity and safety of people. All measures and procedures have been designed and put in place with these two pillars.

12 Risks on data management

Full risk analysis has been carried out.

According to the Medical Devices Coordination Group, MDCG 2019-16 rev. 1:

- Manufacturers should demonstrate state-of-the-art within their decisions (based on applicable standards, guidance, their own proprietary knowledge and publicly available scientific / technical information) while demonstrating appropriateness to proportionally address security risk.
- Provide a reference to technical documentation

These recommendations are relevant, if ESSENCE platform would be sold/given to some user organisation and they would then have to take care of proper security. In our case this is not relevant, since ESSENC is provided as a cloud-based service (mainstream approach nowadays), managed by the platform vendor(s),, i.e. ESSENCE technical partner(s).The ANNEX I of the MDCG also adds further information.

The Risk Analysis clearly indicates all the countermeasures that have been implemented according to the state of the art. Further, threat & vulnerability analysis as well as penetration tests have been performed in Q1 of 2023 on the infrastructure

A preliminary analysis was reported in D1.1, Preliminary Data Management Plan, to indicate the planned measures. This analysis has evolved throughout project development and the final view of risks is summarized in this section, that reports also details for cybersecurity.

It is worth note that in ESSENCE is important to prevent unwanted or unauthorised access to the patients' data. A potential attack to the ESSENCE system, with a consequent temporary interruption of the CBAC, LHM or other components, is not critical. ESSENCE is not a life support system, nor a crucial system used for direct diagnosis or to provide therapies. The activities can be postponed and rescheduled as soon as the system has been restored. For this reason, we report the measures adopted to create a safe system.

In the following table we are reporting all the measures implemented in the early phase of the project and used to augment data security. Some additional measures have been implemented based on the results of threat/vulnerability analysis and penetration testing, these additional measures are reported in Section **Error! Reference source not found.** and Section **Error! Reference source not found.**, respectively.

ID	Measures	Application
1	Encryption	Essential security mechanisms were implemented, building on strong AES encryption. Authentication is done using cryptographically signed tokens (JWT). After successful authentication, authorisation for access to services/data is enforced checking user assigned roles against access rights (privileges) assigned to different roles. The platform therefore provides protection at database and communication levels. Data in transfer were secured with the TLS protocol. Data at rest have been protected with encryption and relevant authentication and authorisation mechanisms. The authentication data include sensitive data (email, name surname) and are stored on a secured Database separated from the Data Center and they are encrypted. All other data are stored pseudonymized. Encryption keys are generated using a random source with high entropy (OpenSSL on an AMD Ryzen host). The encryption key is not stored on the

		same machine as the encrypted database, but on a remote virtual machine in a key vault. Change in the case of key compromise includes manual database decryption and re-encryption with a newly generated key.
2	Partitioning data	Data are partitioned on different servers by splitting over identification data from sensitive data. Even sensitive data are partitioned to assure the best privacy level. In particular, the authentication data include all personal sensitive data (email name surname mobile number) and an encrypted data storage is used for them. These data are stored in separate servers with respect to all the other data. The information stored in each module's local data server (e.g., the LHM local data server, **Figure 5** does allow the association between personal data and behavioral/health data. This is achieved by retrieving personal data on a need basis by exploiting specific API endpoints provided by the authentication module.
3	Logical access control	Only internal personnel have access to ESSENCE repository. A dedicated data repository has been set up for data from the users in the pilot. Only a subset of personnel (the responsible persons of the pilot sites and specially identified operators, and the DPO) are allowed to access the user data (both personal and sensitive). Access is granted with a userID/password mechanism.
4	Traceability (logging)	Access to data is granted with a userID/password mechanism. A log file traces accesses and operations.
5	Archiving	The generated data (indicators) are shared in open access one month after data generation as declared in the Article 29.3 of the Grant Agreement. The shared data are anonymized and are shared in Zenodo.
6	Paper document security	Paper documents are stored in a secure and locked placed with access limited only to the personnel involved in the recruitment under the responsibility of the partner that collected the data.
7	Operating security	Operations security is aligned with the ISO 27002 standard code of practice, i.e. the best practice recommendations for implementing and maintaining an information security management system. The recommendations cover procedures and responsibilities, malware protection, backup, logging and monitoring, control of operational software, technical vulnerability management and information systems audit coordination.
8	Clamping down on malicious software	ESSENCE is a closed infrastructure running on AWS servers (limiting geographically to EU the data storage) and using the Amazon Layer 4 Virtual Private Cloud Firewall for protecting communication among components and perimeter towards public internet. In addition, an Application Layer Firewall (Layer 7) is set up to assure the minimization of the access risk.
9	Managing workstations	The partner responsible for the infrastructure and integration (SCOM) is ISO 27001 certified and compliant. Moreover, all partners feature IT departments, which centrally implement relevant technical measures and security policies, including automatic workstation locking, regular updates, configuration, physical security, etc.
10	Website security	The website communication has been secured according to relevant recommandations, e.g. the ANSSI (Agence nationale de la sécurité des systèmes d'information). Access to website and servers from public internet is protected by the TLS protocol. Authentication and authorisation is done using cryptographically signed tokens (JWT), more details on JWT and

		encryption key management are provided above in the table item 1. Note that JWT token has validity limited to 5 minutes to limit misuse in case of token compromise.
11	Backup	The data backup is developed and provided according to the needs and policies defined in the experimental protocol of the study. The backup of sensitive data is done by dumping the database (which has already encrypted values), compressing it with a password and stored off-site at the SCOM.
12	Maintenance	By using the AWS cloud services hardware maintenance is transferred (outsourced) to Amazon. Remote monitoring and maintenance of components/applications is featured, where each partner has remote administrative access and responsibility for the maintenance of their respective component/app running in the cloud infrastructure. Maintenance of cloud infrastructure is provided by the ESSENCE partner SCOM. They are a qualified and ISO 9001 and ISO 27001 quality certified IT entity.
13	Network security	Network security management is based on the Amazon Virtual Private Cloud (VPC), most notably the Security Groups for VPC and Network Access Control Lists (ACL). A Security group acts as a virtual firewall for a virtual machine (VM) instance to control inbound and outbound traffic. When an instance is launched in a VPC, up to five security groups can be assigned to the instance. Security groups act at the instance level, not the subnet level. Therefore, each instance in a subnet in a VPC can be assigned to a different set of security groups. A network access control list (ACL) is an optional layer of security for a VPC that acts as a firewall for controlling traffic in and out of one or more subnets. One might set up network ACLs with rules similar to one's security groups in order to add an additional layer of security to one's VPC. Each partner has its own Security Group, internally these groups can access each other's networks. External access to these security groups is protected through firewalls (more details are provided in the table item 9 above).
14	Monitoring network activity	ESSENCE services are running in the AWS Cloud infrastructure, network level protection is provided by the inherent AWS security services, such as network security groups and ACL's, as described in table item 13 above. We additionally need to monitor the application layer (Layer 7) access, which is monitored and controlled by the Application layer firewall.
15	Personnel management	Personnel participates in pilot procedures definition and both technical and clinical operators have been properly trained before the pilot start. Dry run tests are done and confirm the commitment and preparation of the personnel for the trial in technical, clinical, legal and ethical issues.
16	Pseudonymization	The personal data management procedure implements pseudonymization, i.e., processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person. Hash technique is implemented.
17	Personnel training	The first adopted measure is the training of the project staff and their need to be aware of the risks involved in processing personal data and how to mitigate those risks though proper measures and countermeasures.

Additional risks can be grouped under illegitimate access to data, unwanted modification of the data, and data disappearance.

12.1 Threat and Vulnerability Analysis

12.1.1 Scope limitation

Risk assessment covered the Essence project, including all parts of the application and backend systems and the smart pen.

Exclusions:

- As infrastructure is hosted in Amazon cloud as Infrastructure as a Service, physical security and hardware issues are out of scope of the assessment.
- End user devices, like phones, laptops and tablets are owned by the user and under their control, therefore also out of scope.

Assumption: The environment used in pilot would also be used in production.

12.1.2 Methodology

The methodology used was NIST Risk Management Framework, as described in NIST SP 800-53 and NIST-30r1. Assessment was done using qualitative scale. For easier understanding the tables describing the scale used are added below. Assessment was based on interviews with developers.

The following tables are of interest in ESSENCE for classification of risks:

TABLE D-3: ASSESSMENT SCALE – CHARACTERISTICS OF ADVERSARY CAPABILITY

Qualitative Values	Semi-Quantitative Values		Description
Very High	96-100	10	The adversary has a very sophisticated level of expertise, is well-resourced, and can generate opportunities to support multiple successful, continuous, and coordinated attacks.
High	80-95	8	The adversary has a sophisticated level of expertise, with significant resources and opportunities to support multiple successful coordinated attacks.
Moderate	21-79	5	The adversary has moderate resources, expertise, and opportunities to support multiple successful attacks.
Low	5-20	2	The adversary has limited resources, expertise, and opportunities to support a successful attack.
Very Low	0-4	0	The adversary has very limited resources, expertise, and opportunities to support a successful attack.

TABLE D-4: ASSESSMENT SCALE – CHARACTERISTICS OF ADVERSARY INTENT

Qualitative Values	Semi-Quantitative Values		Description
Very High	96-100	10	The adversary seeks to undermine, severely impede, or destroy a core mission or business function, program, or enterprise by exploiting a presence in the organization's information systems or infrastructure. The adversary is concerned about disclosure of tradecraft only to the extent that it would impede its ability to complete stated goals.
High	80-95	8	The adversary seeks to undermine/impede critical aspects of a core mission or business function, program, or enterprise, or place itself in a position to do so in the future, by maintaining a presence in the organization's information systems or infrastructure. The adversary is very concerned about minimizing attack detection/disclosure of tradecraft, particularly while preparing for future attacks.
Moderate	21-79	5	The adversary seeks to obtain or modify specific critical or sensitive information or usurp/disrupt the organization's cyber resources by establishing a foothold in the organization's information systems or infrastructure. The adversary is concerned about minimizing attack detection/disclosure of tradecraft, particularly when carrying out attacks over long time periods. The adversary is willing to impede aspects of the organization's missions/business functions to achieve these ends.
Low	5-20	2	The adversary actively seeks to obtain critical or sensitive information or to usurp/disrupt the organization's cyber resources, and does so without concern about attack detection/disclosure of tradecraft.
Very Low	0-4	0	The adversary seeks to usurp, disrupt, or deface the organization's cyber resources, and does so without concern about attack detection/disclosure of tradecraft.

TABLE D-5: ASSESSMENT SCALE – CHARACTERISTICS OF ADVERSARY TARGETING

Qualitative Values	Semi-Quantitative Values		Description
Very High	96-100	10	The adversary analyzes information obtained via reconnaissance and attacks to target persistently a specific organization, enterprise, program, mission or business function, focusing on specific high-value or mission-critical information, resources, supply flows, or functions; specific employees or positions; supporting infrastructure providers/suppliers; or partnering organizations.
High	80-95	8	The adversary analyzes information obtained via reconnaissance to target persistently a specific organization, enterprise, program, mission or business function, focusing on specific high-value or mission-critical information, resources, supply flows, or functions; specific employees supporting those functions, or key positions.
Moderate	21-79	5	The adversary analyzes publicly available information to target persistently specific high-value organizations (and key positions, such as Chief Information Officer), programs, or information.
Low	5-20	2	The adversary uses publicly available information to target a class of high-value organizations or information, and seeks targets of opportunity within that class.
Very Low	0-4	0	The adversary may or may not target any specific organizations or classes of organizations.

TABLE D-5: ASSESSMENT SCALE – CHARACTERISTICS OF ADVERSARY TARGETING

Qualitative Values	Semi-Quantitative Values		Description
Very High	96-100	10	The adversary analyzes information obtained via reconnaissance and attacks to target persistently a specific organization, enterprise, program, mission or business function, focusing on specific high-value or mission-critical information, resources, supply flows, or functions; specific employees or positions; supporting infrastructure providers/suppliers; or partnering organizations.
High	80-95	8	The adversary analyzes information obtained via reconnaissance to target persistently a specific organization, enterprise, program, mission or business function, focusing on specific high-value or mission-critical information, resources, supply flows, or functions; specific employees supporting those functions, or key positions.
Moderate	21-79	5	The adversary analyzes publicly available information to target persistently specific high-value organizations (and key positions, such as Chief Information Officer), programs, or information.
Low	5-20	2	The adversary uses publicly available information to target a class of high-value organizations or information, and seeks targets of opportunity within that class.
Very Low	0-4	0	The adversary may or may not target any specific organizations or classes of organizations.

TABLE F-2: ASSESSMENT SCALE – VULNERABILITY SEVERITY

Qualitative Values	Semi-Quantitative Values		Description
Very High	96-100	10	The vulnerability is exposed and exploitable, and its exploitation could result in severe impacts. Relevant security control or other remediation is not implemented and not planned; or no security measure can be identified to remediate the vulnerability.
High	80-95	8	The vulnerability is of high concern, based on the exposure of the vulnerability and ease of exploitation and/or on the severity of impacts that could result from its exploitation. Relevant security control or other remediation is planned but not implemented; compensating controls are in place and at least minimally effective.
Moderate	21-79	5	The vulnerability is of moderate concern, based on the exposure of the vulnerability and ease of exploitation and/or on the severity of impacts that could result from its exploitation. Relevant security control or other remediation is partially implemented and somewhat effective.
Low	5-20	2	The vulnerability is of minor concern, but effectiveness of remediation could be improved. Relevant security control or other remediation is fully implemented and somewhat effective.
Very Low	0-4	0	The vulnerability is not of concern. Relevant security control or other remediation is fully implemented, assessed, and effective.

TABLE F-5: ASSESSMENT SCALE – PERVASIVENESS OF PREDISPOSING CONDITIONS

Qualitative	Semi-Quantitative	Description

TABLE H-3: ASSESSMENT SCALE – IMPACT OF THREAT EVENTS

Qualitative Values	Semi-Quantitative Values		Description
Very High	96-100	10	The threat event could be expected to have **multiple severe or catastrophic** adverse effects on organizational operations, organizational assets, individuals, other organizations, or the Nation.
High	80-95	8	The threat event could be expected to have a **severe or catastrophic** adverse effect on organizational operations, organizational assets, individuals, other organizations, or the Nation. A severe or catastrophic adverse effect means that, for example, the threat event might: (i) cause a severe degradation in or loss of mission capability to an extent and duration that the organization is not able to perform one or more of its primary functions; (ii) result in major damage to organizational assets; (iii) result in major financial loss; or (iv) result in severe or catastrophic harm to individuals involving loss of life or serious life-threatening injuries.
Moderate	21-79	5	The threat event could be expected to have a **serious** adverse effect on organizational operations, organizational assets, individuals other organizations, or the Nation. A serious adverse effect means that, for example, the threat event might: (i) cause a significant degradation in mission capability to an extent and duration that the organization is able to perform its primary functions, but the effectiveness of the functions is significantly reduced; (ii) result in significant damage to organizational assets; (iii) result in significant financial loss; or (iv) result in significant harm to individuals that does not involve loss of life or serious life-threatening injuries.
Low	5-20	2	The threat event could be expected to have a **limited** adverse effect on organizational operations, organizational assets, individuals other organizations, or the Nation. A limited adverse effect means that, for example, the threat event might: (i) cause a degradation in mission capability to an extent and duration that the organization is able to perform its primary functions, but the effectiveness of the functions is noticeably reduced; (ii) result in minor damage to organizational assets; (iii) result in minor financial loss; or (iv) result in minor harm to individuals.
Very Low	0-4	0	The threat event could be expected to have a **negligible** adverse effect on organizational operations, organizational assets, individuals other organizations, or the Nation.

TABLE G-2: ASSESSMENT SCALE – LIKELIHOOD OF THREAT EVENT INITIATION (ADVERSARIAL)

Qualitative Values	Semi-Quantitative Values		Description
Very High	96-100	10	Adversary is **almost certain** to initiate the threat event.
High	80-95	8	Adversary is **highly likely** to initiate the threat event.
Moderate	21-79	5	Adversary is **somewhat likely** to initiate the treat event.
Low	5-20	2	Adversary is **unlikely** to initiate the threat event.
Very Low	0-4	0	Adversary is **highly unlikely** to initiate the threat event.

TABLE G-3: ASSESSMENT SCALE – LIKELIHOOD OF THREAT EVENT OCCURRENCE (NON-ADVERSARIAL)

Qualitative Values	Semi-Quantitative Values		Description
Very High	96-100	10	Error, accident, or act of nature is **almost certain** to occur; or occurs **more than 100 times a year**.
High	80-95	8	Error, accident, or act of nature is **highly likely** to occur; or occurs **between 10-100 times a year**.
Moderate	21-79	5	Error, accident, or act of nature is **somewhat likely** to occur; or occurs **between 1-10 times a year**.
Low	5-20	2	Error, accident, or act of nature is **unlikely** to occur; or occurs **less than once a year**, but **more than once every 10 years**.
Very Low	0-4	0	Error, accident, or act of nature is **highly unlikely** to occur; or occurs **less than once every 10 years**.

TABLE G-4: ASSESSMENT SCALE – LIKELIHOOD OF THREAT EVENT RESULTING IN ADVERSE IMPACTS

Qualitative Values	Semi-Quantitative Values		Description
Very High	96-100	10	If the threat event is initiated or occurs, it is **almost certain** to have adverse impacts.
High	80-95	8	If the threat event is initiated or occurs, it is **highly likely** to have adverse impacts.
Moderate	21-79	5	If the threat event is initiated or occurs, it is **somewhat likely** to have adverse impacts.
Low	5-20	2	If the threat event is initiated or occurs, it is **unlikely** to have adverse impacts.
Very Low	0-4	0	If the threat event is initiated or occurs, it is **highly unlikely** to have adverse impacts.

TABLE G-5: ASSESSMENT SCALE – OVERALL LIKELIHOOD

Likelihood of Threat Event Initiation or Occurrence	Likelihood Threat Events Result in Adverse Impacts				
	Very Low	**Low**	**Moderate**	**High**	**Very High**
Very High	Low	Moderate	High	Very High	Very High
High	Low	Moderate	Moderate	High	Very High
Moderate	Low	Low	Moderate	Moderate	High
Low	Very Low	Low	Low	Moderate	Moderate
Very Low	Very Low	Very Low	Low	Low	Low

TABLE I-2: ASSESSMENT SCALE – LEVEL OF RISK (COMBINATION OF LIKELIHOOD AND IMPACT)

Likelihood (Threat Event Occurs and Results in Adverse Impact)	Level of Impact				
	Very Low	**Low**	**Moderate**	**High**	**Very High**
Very High	Very Low	Low	Moderate	High	Very High
High	Very Low	Low	Moderate	High	Very High
Moderate	Very Low	Low	Moderate	Moderate	High
Low	Very Low	Low	Low	Low	Moderate
Very Low	Very Low	Very Low	Very Low	Low	Low

12.1.3 Analysis

We first identify the sources of risks and then define measure to mitigate the risks.

12.1.3.1Source identification

12.1.3.1.1Threat Sources

The following malicious threat sources were identified:

Type of Threat Source	Capability	Intent	Targeting
Malicious insider: compromised(blackmail, bribe), dissatisfied employees	Moderate	Low	High
Malicious Priviliged insider	High	Low	High
Criminal group	High	Moderate	Low
Nation-state	Very High	High	Very High
Competitor	High	Moderate	High
Malicious low-skilled outsider	Very Low	Very Low	Very Low
Malicious outsider	Low	Very Low	Low

The following non-malicious or accidental threat sources were identified:

Vulnerability		
Identifier	Vulnerability Source of Information	Vulnerability Severity
V1	No reducancy setup.	Moderate
V2	No MFA(multifactor authentication) for administrators	Very High
V3	End user devices are owned by users, malware could be installed. No controll what other applications are installed	Moderate
V4	Physical access to devices gives access to application and data, due to persistent login	Moderate
V5	No controll over end user network. Potential to sniff traffic	Moderate
V6	Lack of control over supply chain - pencil	Moderate
V7	No WAF(Web Application Firewall)	Very High
V8	No DDOS protection(Distributed Denial of Service)	High
V9	Users are vulnerable to social engineering attacks	Low
V10	Impersonating healthcare professional as a trusted person	High
V11	No alerting	High
V12	No logging	Very high
V13	No audit for login	Very high

Predisposing Conditions		
Identifier	Predisposing Condition Source of Information	Pervasiveness of Condition
P1	End users are not proficient users of technology	High
P2	Handling health data pseudonimised	Very High
P3	Handling PII(Personally Identifiable Information)	Very High

12.1.3.1.3 Adversarial risks

The following risks on using the data center that can produce adversarial effects are:

Threat event	Threat source	Range of effects	Relevance	Likelihood of event Occurring	Vulnerabilities and predisposing conditions	Severity and pervasiveness	Likelihood event results in Adverse Impact	Overall likelihood	Level of impact	Risk	Suggested Mitigation
Overuse of resources (hard disk, memory, power, etc	Accidental IT Administrator	Application is slow, some sessions drop	Expected	Moderate	Not using autoscaling, no redundancy(V1, V14)	Moderate	Moderate	Moderate	Low	Low	Check sizing, use scaling for production
Network outage	Failure	Cannot administer platform and application	Anticipated	Low		Low	Moderate	Low	Moderate	Low	Several ISP providers for administration
Incorrect privilege settings	Accidental IT Administrator	Increased risk of accidental data damage/system breaches	Expected	Low	No audit(V13)	Very High	Moderate	Low	Moderate	Low	Audit access rights, access, performed actions
Software based malfunction: OS level	Failure	Issue in OS or database software could cause the application to stop working	Expected	Low	No redundancy(V1)	Moderate	High	Moderate	High	Moderate	Use well tested distributions meant for production servers
Software based malfunction: application level	Failure	Application could malfunction due to bugs, it could affect availability or limit functionality	Expected	Moderate		High	High	Moderate	High	Moderate	Test code in development and test environments before deploying to production environment
Access to device, leak of personal data	Accidental user	Disclosure of personal data to relatives, friends visiting without consent	Predicted	Low	Persistent login(V4)	Low	Moderate	Low	Low	Low	Mask private information
Pen malfunction	Failure	Data for analysis can not be recorded	Expected	Low		Moderate	Moderate	Low	Moderate	Low	QA of pens, testing of any software updates

12.1.3.1.4 Non-adversarial risks

The following risks on using the data center that can produce non-adversarial effects are:

Threat Event	Threat Source	Capability	Intent	Targeting	Relevance	Likelihood of attack initiation	Vulnerabilities and Predisposing Conditions	Severity and Pervasiveness	Likelihood Initiated Attack Succeeds	Overall Likelihood	Adverse Impacts	Level of Impact	Risk	Suggested mitigation
Denial of Service (DoS) attack	Criminal group	High	Moderate	Low	Anticipated	Moderate	No DDOS protection(V6), no autoscaling(V14)	High	Moderate	Moderate	No access to application. Appointments with health care professional that were scheduled through platforms cannot be done	High	Moderate	Deploy DDOS protection
Brute force login attempts/password spraying attacks	Malicious low-skilled outsider	Very low	Very low	Very low	Expected	Very high	Lack of integrated logging and alerting(V11, V12, V13), No MFA(V2)	Very High	Low	Moderate	Access to platform, application. Stealing data from database. Disclosure of health data, loss of reputation, penalties from regulatory bodies. DOS. Ransomware	High	Moderate	Use MFA for platforms access, [illegible] MFA for admin access to application and database
Supply chain attack: inserting components to the [illegible], stealing	Nation-state	Very High	High	Very High	Possible	Very Low	No control over supply chain(V5)	Moderate	High	Low	Stealing data from users. Disclosure of health data, loss of reputation, penalties from regulatory	Moderate	Low	
Malware gets installed on end user devices, stealing data	Malicious outsider	Low	Very Low	Low	Expected	Very High	No control over end user devices (V7). Users not proficient in IT, not educated in security behaviour (V8, P1)	Low	High	Very High	Stealing individual data	Low	Low	[illegible]
Sniffing network traffic at end user	Malicious outsider	Low	Very Low	Low	Anticipated	Moderate	No control over end user network (V9)	Moderate	Moderate	Moderate	Stealing individual data	Low	Low	Encrypted traffic
Application level attack: code injection etc	Malicious outsider	Low	Very Low	Low	Anticipated	High	No WAF(V?)	Very High	Moderate	Moderate	Access to application. Stealing data from database. Disclosure of health data, loss of reputation, penalties from regulatory bodies	High	Moderate	Deploy WAF, Pentest
Social engineering attack on end users	Criminal group	High	Moderate	Low	Expected	High	Users not proficient in IT, not educated in security behaviour (V8, P1)	Low	High	High	Stealing individual data	Low	Low	
Social engineering attack on professional user	Criminal group	High	Moderate	Low	Expected	High	No MFA, healthcare worker is trusted person (V2, V10)	High	Moderate	Moderate	Access to database. Use access to scam end users	High	Moderate	[illegible]
Compromised employee	Malicious insider	Moderate	Low	High	Possible	Low	No logging and alerting(V11, V12), No audit(V13)	Very High	Moderate	Low	access to limited data	Low	Low	
Compromised admin	Malicious privileged insider	High	Low	High	Possible	Low	No logging and alerting(V11, V12), No audit(V13)	Very High	Very High	Moderate	Access to platform, application. Stealing data from database. Disclosure of health data, loss of reputation, penalties from regulatory bodies. DOS. Ransomware. Leaking data or research to competitor	Very High	High	[illegible]
Stealing admin credentials using phishing or compromising application. Stealing data and research	Competitor	High	Moderate	High	Anticipated	Low	No MFA(V2), no logging and alerting(V11, V12)	Very High	Moderate	Low	Loss of competitive advantage	Very High	Moderate	MFA, logging, alerting, auditing, [illegible]
Cause integrity loss by creating, deleting, and/or modifying data on information systems	Competitor	High	Moderate	High	Anticipated	Low	No logging and alerting(V11, V12)	High	Moderate	Low	Loss of competitive advantage	Very High	Moderate	Logging, alerting, auditing

12.1.3.2 Risks mitigation analysis

Risk tolerance is suggested to be accepting everything with Low score or below on impact and requiring mitigations for score Moderate and High Impact. Here are the mitigation measures identified by risk analysis.

Following are the risks with High and Moderate scores:

12.1.3.2.1 Mitigation of High Impact

- Compromised administrator account

Potential consequences: Someone could access the platform or application itself, download and share or sell data, disable the system, remove the traces behind themselves. As there is personal and health data in the database, this would have regulatory consequences and potential high fines (GDPR, 2%-4% of global revenue).

Suggested mitigation:

- Enable logging of all administrative account activity.
- Alerts for high risk actions.
- Have independent person audit administrative activities.
- Enable account lockdown in case of several unsuccessful logon attempts.
- Enable Multi-Factorial Authentication (MFA) for admins.
- Vetting of potential employees before hiring.

12.1.3.2.2 Mitigation of moderate impact risk.

- Denial of Service Attack

Potential consequences: Not possible to access application. Appointments between patients and healthcare professionals cannot be booked or conducted through platform.
Suggested mitigation:
- Deploy Distributed Denial of Service (DDoS) protection available in AWS.
- Alternative communication method through phone.

- Brute force login attempts

Potential consequences: Unauthorized access to platform, leak of personal and sensitive data, data modification. Regulatory consequences, due to personal and health data, potentially high fines (GDPR 2-4% of global revenue). Once in the system attacker could deploy further malware, for example ransomware.
Suggested mitigation:
- Require MFA for platform access and admin access to application and database.

- Application-level attack

Potential consequences: Unauthorized access to the application, database, leak of personal and health data. Regulatory consequences, due to personal and health data, potentially high fines (GDPR 2-4% of global revenue).
Suggested mitigation:
- Deploy Web Application Firewall (WAF),
- perform penetration tests.

- Social engineering attack on professional user

Potential consequences: Taking over professional user account, leak of personal and health data. Take advantage that healthcare worker is trusted by the end users to conduct scams.
Suggested mitigation:
- Limit access for healthcare professionals to only access their own patients.
- Logging and monitoring of activity.

- Stealing admin credentials by competitor

Potential consequences: Stolen data about patients and research. Loss of competitive advantage.
Suggested mitigation:
- Deploy MFA for admin accounts.
- Deploy WAF to protect application.
- Enable logging, alerting, review activity by independent person regularly.

- Unauthorized modification of data by competitor

Potential consequences: after gaining access competitor might modify the research data to devalue it, so it cannot be used for research, might produce false conclusions.
Suggested mitigation:
- Enable logging, alerting and regular review by independent person.

- Software malfunction: operating system

Potential consequences: There could be issues with software used in the infrastructure supporting the application, like operating system or database. It could cause application to stop working.
Suggested mitigation:

- Use well tested distributions meant for production workloads, test patches before deploying them.

- Software malfunction: application

Potential consequences: Application could malfunction due to bugs, it might affect availability or limit functionality.
Suggested mitigation:

- Test code in development and test environment before deploying to production.
- Extensive unit testing

12.1.3.2.3 Summary of mitigation measures suggested by risk analysis

Here below is a table that summarizes mitigation suggestions.

Threat Event	Risk	Suggested mitigation
Denial of Service (DoS) attack	Moderate	Deploy DDOS protection
Brute force login attempts/password guessing attacks	Moderate	Use MFA for platforma access. Rate limiting for login. MFA for admin access to application and database.
Application level attack: code injection etc	Moderate	Deploy WAF. Pentest.
Social engineering attack on professional user	Moderate	Limit access for healthcare professionals to only have it for their patients, not everyone. Monitor activity.
Compromised admin	High	Logging, alerting, auditing IT admins. Account lockdown. Vetting potential employees.
Stealing admin credentials using phishing or compromising application. Stealing data and research	Moderate	MFA, logging, alerting, auditing, WAF
Cause integrity loss by creating, deleting, and/or modifying data on information systems	Moderate	Logging, alerting, auditing

Threat event	Risk	Suggested Mitigation
Software based malfunction: OS level	Moderate	Use well tested distributions meant for production servers
Software based malfunction: application level	Moderate	Test code in development and test environments before deploying to production environment

The summary of most important suggested mitigations is:

- Enable MFA (Multi-Factor Authentication) for admins.
- Enable logging and alerting for admin activity.
- Deploy DDoS protection
- Deploy WAF (Web Application Firewall)
- Testing application in test environment before production
- Pen testing

12.1.4 Risk mitigation mesaures implemented in ESSENCE

The recommendations from the threat and vulnerability analysis have been considered in the following manner:

- In order to strengthen the protection against compromise of admin credentials, MFA has been enabled for the AWS cloud services administrators, while certificate based SSH access has been configured for the virtual machine administrators
- Logging, alerting, and auditing is a mitigation against admins turning rogue and is planned in the commercial deployment phase
- DDOS protection is automatically provided by AWS Shield Standard, this protection is offered on all AWS services at no additional charge, it is always on and pre-configured. This service is complemented by the NGINX traffic throttling capability
- WAF protection is provided with the NGINX (authentication component) in combination with ModSecurity open source web firewall
- The adopted approach for production systems is to test each component in test environment, then deploy and test it in staging environment before finally releasing it and deploying in the production environment
- Penetration test has been performed, see Section **Error! Reference source not found.**

12.2 Cybersecurity testing

The cybersecurity audit, i.e. the penetration testing of the running ESSENCE platform prototype was performed by the Nexpose tool from Rapid7 LLC.

The audit was performed on 7 systems running in the AWS cloud infrastructure, 7 of which were found to be active and were scanned.

The following table lists the systems discovered during the cybersecurity test with an associated assessment of the quantitative risk score according to temporal risk score evaluation methodology of

the Nexpose, as explained at https://help.rapid7.com/nexpose/en-us/Files/Risk_scoring_FAQ.html . Nexpose calculates risk scores for every asset and vulnerability that it finds during a scan. The scores indicate the potential danger that the vulnerability poses to network and business security based on impact and likelihood of exploit.

The Temporal risk model is a mathematical calculation of the following factors:

- Time-based likelihood (t) is the number of days since vulnerability publicly disclosed. The overall score increases with the number of days.
- Proximity-based impact is the sum of four variables:
- access vector (AV) or the likelihood of exploit, based on whether the target is locally accessible, is accessible from within the network, or must be accessed from outside the network; local access results in a higher score
- confidentiality impact (C) or disclosure to unauthorized individuals or systems
- integrity impact (I) or unauthorized data modification
- availability impact (A) or loss of access to data
- exploit difficulty is the sum of two variables:
- access complexity (AC) or the likelihood of exploit based on how much skill is required to perform the exploit; an easier exploit results in a higher score
- authentication (Au) or the likelihood of exploit based on authentication requirements; no authentication results in a higher score

The following formula is used to calculate the Temporal scoring model:

$$\text{Risk} = \text{time} \times \frac{\text{proximity-based impact}}{\text{exploit difficulty}}$$

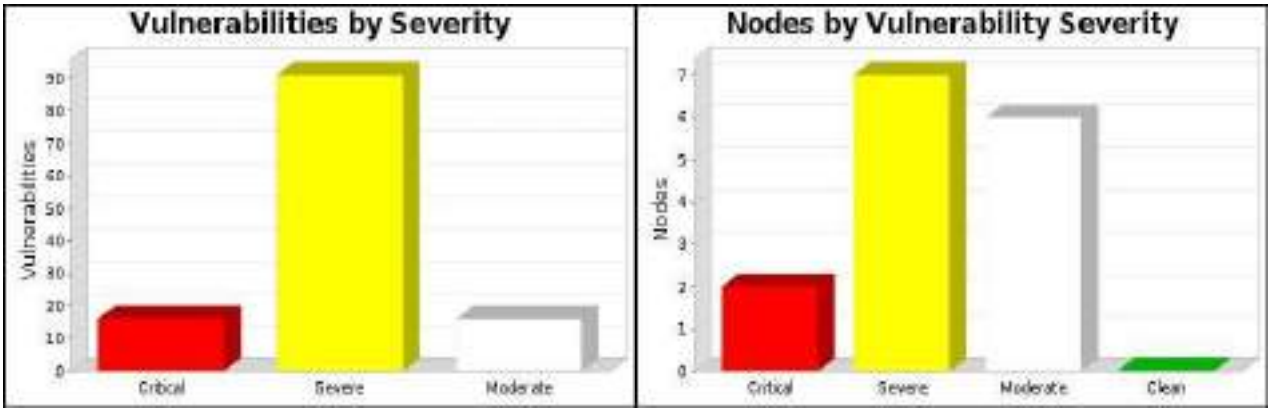
This formula can be broken down into its components as follows:

$$\text{Risk} = \sqrt{t} \times \frac{(AV + C + I + A)!}{(AC + Au)^2}$$

The score is expressed in high, whole numbers, ranging up to as many as six digits. There is no "highest" number. These numbers are relative to each other.

Node	Operating System	Risk	Aliases
18.195.78.58	Debian Linux 10.2	18,176	ec2-18-195-78-58.eu-central-1.compute.amazonaws.com
18.195.177.111	Debian Linux 10.2	18,176	ec2-18-195-177-111.eu-central-1.compute.amazonaws.com
18.159.6.189	Debian Linux 10.2	5,061	ec2-18-159-6-189.eu-central-1.compute.amazonaws.com

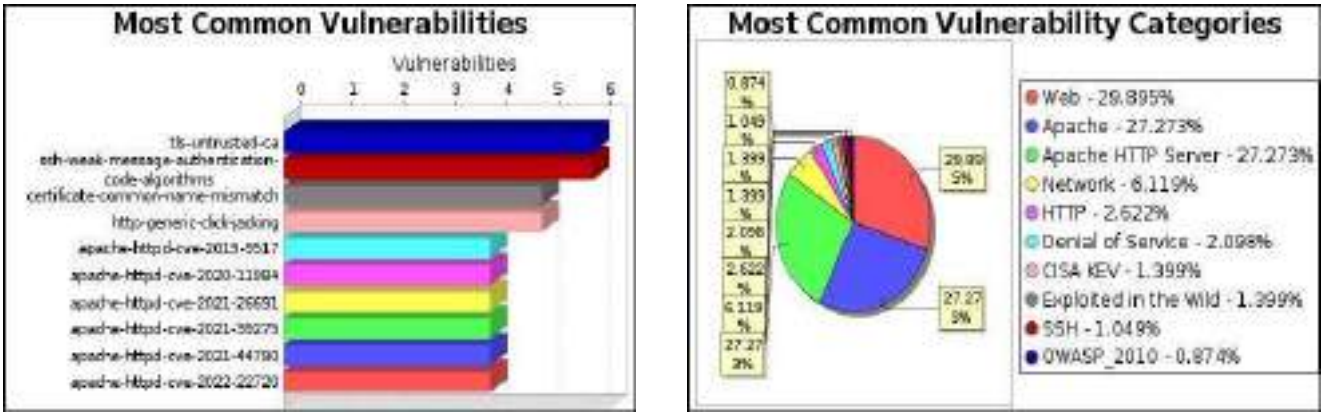
3.64.11.58	Microsoft Windows	4,651	ec2-3-64-11-58.eu-central-1.compute.amazonaws.com
18.185.215.209	Debian Linux 10.2	2,385	ec2-18-185-215-209.eu-central-1.compute.amazonaws.com
18.157.181.129	Debian Linux 10.2	599	ec2-18-157-181-129.eu-central-1.compute.amazonaws.com
3.71.245.143	Debian Linux 10.2	599	ec2-3-71-245-143.eu-central-1.compute.amazonaws.com



There were 123 vulnerabilities found during this scan. Of these, 16 were critical vulnerabilities. Critical vulnerabilities require immediate attention. They are relatively easy for attackers to exploit and may provide them with full control of the affected systems. 91 vulnerabilities were severe.

Severe vulnerabilities are often harder to exploit and may not provide the same access to affected systems. There were 16 moderate vulnerabilities discovered. These often provide information to attackers that may assist them in mounting subsequent attacks on your network. These should also be fixed in a timely manner, but are not as urgent as the other vulnerabilities.

Critical vulnerabilities were found to exist on 2 of the systems, making them most susceptible to attack. 7 systems were found to have severe vulnerabilities. Moderate vulnerabilities were found on 6 systems. No system was free of vulnerabilities.

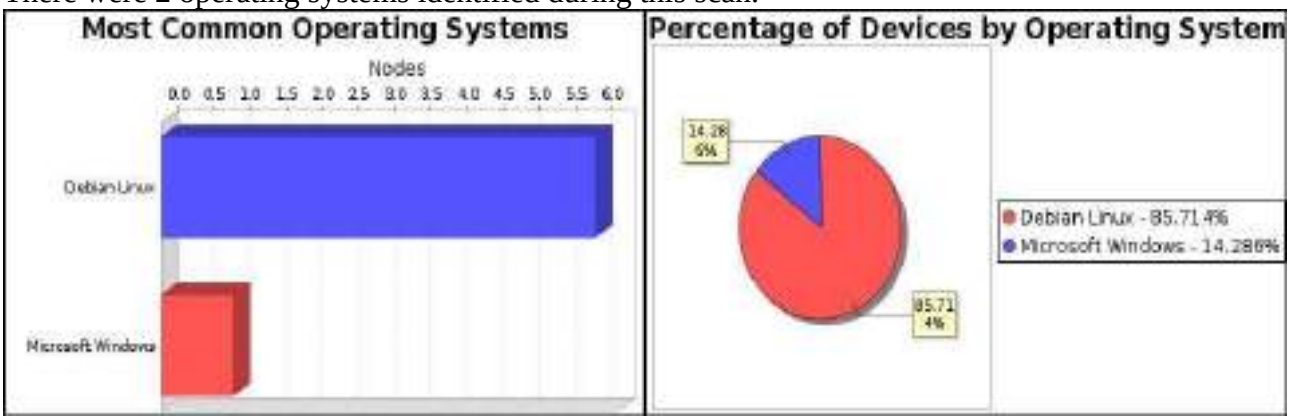


There were 6 occurrences of the tls-untrusted-ca and ssh-weak-message-authentication-code-algorithms vulnerabilities, making them the most common vulnerabilities.

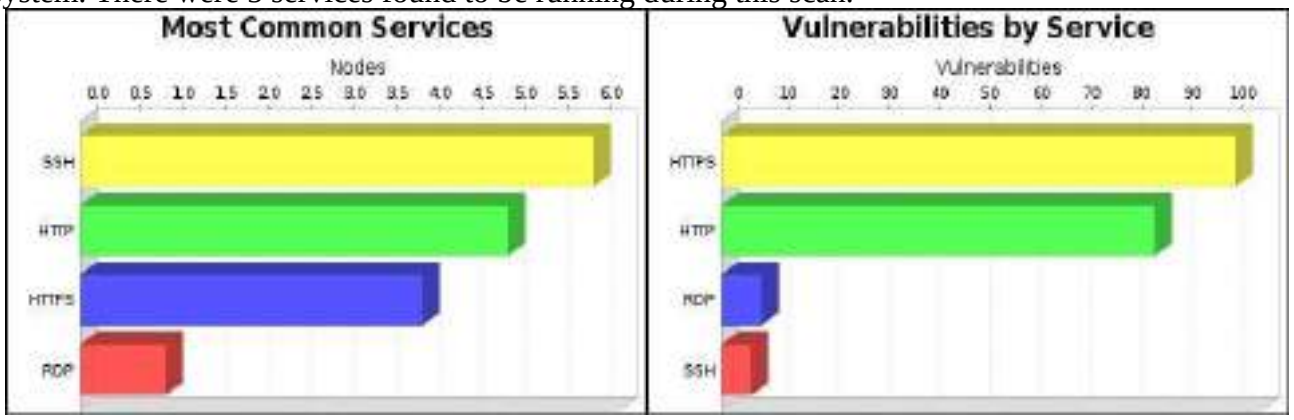
There were 171 vulnerability instances in the Web category, making it the most common vulnerability category.

The certificate-common-name-mismatch vulnerability poses the highest risk with a risk score of 4,192. Risk scores are based on the types and numbers of vulnerabilities on affected assets.

There were 2 operating systems identified during this scan.



The Debian Linux operating system was found on 6 systems, making it the most common operating system. There were 5 services found to be running during this scan.



The SSH service was found on 6 systems, making it the most common service. The HTTPS service was found to have the most vulnerabilities during this scan with 102 vulnerabilities.

12.2.1 Recommended remediation measures

Following this analysis, and implementing the following remediation measures, the risk was largely mitigated as reported in the following.

Remediation	Assets	Vulnerabilities		
1. Upgrade to the latest version of Apache HTTPD	2	78	4	0
2. Disable any MD5 or 96-bit HMAC algorithms within the SSH configuration	6	6	0	0
3. Fix the subject's Common Name (CN) field in the certificate	4	4	0	0
4. Obtain a new certificate from your CA and ensure the server configuration is correct	4	4	0	0
5. Disable insecure TLS/SSL protocol support	2	4	0	0
6. Disable HTTP OPTIONS method	3	3	0	0
7. Replace TLS/SSL server X.509 certificate	2	2	0	0
8. Disable any weak HMAC algorithms within the TLS configuration	2	2	0	0
9. Disable HTTP OPTIONS Method for Apache	2	2	0	0
10. Disable SSLv2, SSLv3, and TLS 1.0. The best solution is to only have TLS 1.2 enabled	2	2	2	0
11. Disable TLS/SSL support for static key cipher suites	2	2	0	0
12. Use HTTP X-Frame-Options	4	4	0	0
13. Remove the default page or stop/disable the IIS server	1	1	0	0
14. Disable HTTP OPTIONS Method for IIS	1	1	0	0
15. Disable TLS/SSL support for 3DES cipher suite	1	2	0	0
16. Generate random Diffie-Hellman parameters	2	2	0	0
17. Replace TLS/SSL self-signed certificate	1	1	0	0
18. Disable TCP timestamp responses on Linux	4	4	0	0
19. Enable TLS/SSL support for strong ciphers	2	2	0	0

A detailed analysis of the measures is hereafter reported.

12.2.1.1 Upgrade to the latest version of Apache HTTPD

Remediation Steps

Download and apply the upgrade from: http://archive.apache.org/dist/httpd/httpd-2.4.54.tar.gz

The latest version of Apache HTTPD is 2.4.54.

Many platforms and distributions provide pre-built binary packages for Apache HTTP server. These pre-built packages are usually customized and optimized for a particular distribution, therefore we recommend that you use the packages if they are available for your operating system. http://archive.apache.org/dist/httpd/httpd-2.4.54.tar.gz

Assets inolved

Name	IP Address	Site
ec2-18-195-177-111.eu-central-1.compute.amazonaws.com	18.195.177.111	Essence 2020 AWS Public IP - CBAC application Spain
ec2-18-195-78-58.eu-central-1.compute.amazonaws.com	18.195.78.58	Essence 2020 AWS Public IP - CBAC application Italy

12.2.1.2 Disable any MD5 or 96-bit HMAC algorithms ***within*** the SSH configuration

Remediation Steps

Consult the product documentation for instructions to disable any insecure MD5 or 96-bit HMAC algorithms within the SSH configuration.

Assets inolved

Name	IP Address	Site
Unknown	18.159.6.189	Essence 2020 AWS Public IP - Essence Manager + DB
ec2-18-157-181-129.eu-central-1.compute.amazonaws.com	18.157.181.129	Essence 2020 AWS Public IP - CBAC DB
ec2-18-185-215-209.eu-central-1.compute.amazonaws.com	18.185.215.209	Essence 2020 AWS Public IP - Monitoring Application - AI server
ec2-18-195-177-111.eu-central-1.compute.amazonaws.com	18.195.177.111	Essence 2020 AWS Public IP - CBAC application Spain
ec2-18-195-78-58.eu-central-1.compute.amazonaws.com	18.195.78.58	Essence 2020 AWS Public IP - CBAC application Italy
ec2-3-71-245-143.eu-central-1.compute.amazonaws.com	3.71.245.143	Essence 2020 AWS Public IP - Monitoring Application - server + DB

12.2.1.3Fix the subject's Common Name (CN) field in the certificate

Remediation Steps

The subject's common name (CN) field in the X.509 certificate should be fixed to reflect the name of the entity presenting the certificate (e.g., the hostname). This is done by generating a new certificate usually signed by a Certification Authority (CA) trusted by both the client and server.

Assets

Name	IP Address	Site
Unknown	18.159.6.189	Essence 2020 AWS Public IP - Essence Manager + DB
ec2-18-185-215-209.eu-central-1.compute.amazonaws.com	18.185.215.209	Essence 2020 AWS Public IP - Monitoring Application - AI server
ec2-18-195-177-111.eu-central-1.compute.amazonaws.com	18.195.177.111	Essence 2020 AWS Public IP - CBAC application Spain
ec2-18-195-78-58.eu-central-1.compute.amazonaws.com	18.195.78.58	Essence 2020 AWS Public IP - CBAC application Italy

12.2.1.4 Obtain a new certificate from your CA and ensure the server configuration is correct

Remediation Steps

Ensure the common name (CN) reflects the name of the entity presenting the certificate (e.g., the hostname). If the certificate(s) or any of the chain certificate(s) have expired or been revoked, obtain a new certificate from your Certificate Authority (CA) by following their documentation. If a self-signed certificate is being used, consider obtaining a signed certificate from a CA. References: Mozilla: Connection Untrusted Error SSLShopper: SSL Certificate Not Trusted Error Windows/IIS certificate chain config Apache SSL config Nginx SSL config What's My Chain Cert?

Assets involved

Name	IP Address	Site
Unknown	18.159.6.189	Essence 2020 AWS Public IP - Essence Manager + DB
ec2-18-185-215-209.eu-central-1.compute.amazonaws.com	18.185.215.209	Essence 2020 AWS Public IP - Monitoring Application - AI server
ec2-18-195-177-111.eu-central-1.compute.amazonaws.com	18.195.177.111	Essence 2020 AWS Public IP - CBAC application Spain
ec2-18-195-78-58.eu-central-1.compute.amazonaws.com	18.195.78.58	Essence 2020 AWS Public IP - CBAC application Italy

12.2.1.5 Disable insecure TLS/SSL protocol support

Remediation Steps

Configure the server to require clients to use TLS version 1.2 using Authenticated Encryption with Associated Data (AEAD) capable ciphers.

Assets inolved

Name	IP Address	Site
Unknown	18.159.6.189	Essence 2020 AWS Public IP - Essence Manager + DB
ec2-3-64-11-58.eu-central-1.compute.amazonaws.com	3.64.11.58	Essence 2020 AWS Public IP - Light Health Monitoring - app + DB

12.2.1.6 Disable HTTP OPTIONS method

Remediation Steps

Disable HTTP OPTIONS method on your web server. Refer to your web server's instruction manual on how to do this.

Web servers that respond to the OPTIONS HTTP method expose what other methods are supported by the web server, allowing attackers to narrow and intensify their efforts.

Assets inolved

Name	IP Address	Site
ec2-18-195-177-111.eu-central-1.compute.amazonaws.com	18.195.177.111	Essence 2020 AWS Public IP - CBAC application Spain
ec2-18-195-78-58.eu-central-1.compute.amazonaws.com	18.195.78.58	Essence 2020 AWS Public IP - CBAC application Italy
ec2-3-64-11-58.eu-central-1.compute.amazonaws.com	3.64.11.58	Essence 2020 AWS Public IP - Light Health Monitoring - app + DB

12.2.1.7 Replace TLS/SSL server X.509 certificate

Remediation Steps

Obtain a new certificate and install it on the server. The exact instructions for obtaining a new certificate depend on your organization's requirements. Generally, you need to generate a certificate request and save the request as a file. This file is then sent to a Certificate Authority (CA) for processing. Please ensure that the start date and the end date on the new certificate are valid.

Your organization may have its own internal Certificate Authority. If not, you may have to pay for a certificate from a trusted external Certificate Authority.

After you have received a new certificate file from the Certificate Authority, you have to install it on the TLS/SSL server. The exact instructions for installing a certificate differ for each product. Please follow their documentation.

Assets inolved

Name	IP Address	Site
ec2-18-195-177-111.eu-central-1.compute.amazonaws.com	18.195.177.111	Essence 2020 AWS Public IP - CBAC application Spain
ec2-18-195-78-58.eu-central-1.compute.amazonaws.com	18.195.78.58	Essence 2020 AWS Public IP - CBAC application Italy

12.2.1.8 Disable any weak HMAC algorithms within the TLS configuration

Remediation Steps

The following recommended configuration provides a higher level of security. This configuration is compatible with Firefox 27, Chrome 22, IE 11, Opera 17 and Safari 9. SSLv2, SSLv3, TLSv1 and TLSv1.1 protocols are not recommended in this configuration. Instead use TLSv1.2 protocol.

Refer to your server vendor documentation to apply the recommended cipher configuration:

ECDHE-ECDSA-AES256-GCM-SHA384:ECDHE-RSA-AES256-GCM-SHA384:ECDHE-ECDSA-CHACHA20POLY1305:ECDHE-RSA-CHACHA20-POLY1305:ECDHE-ECDSA-AES128-GCM-SHA256:ECDHE-RSA-AES128-GCMSHA256:ECDHE-ECDSA-AES256-SHA384:ECDHE-RSA-AES256-SHA384:ECDHE-ECDSA-AES128-SHA256:ECDHE-RSAAES128-SHA256:!aNULL:!eNULL:!EXPORT:!DES:!RC4:!3DES:!MD5:!PSK:!SHA1:!DSS

Assets inolved

Name	IP Address	Site
Unknown	18.159.6.189	Essence 2020 AWS Public IP - Essence Manager + DB
ec2-3-64-11-58.eu-central-1.compute.amazonaws.com	3.64.11.58	Essence 2020 AWS Public IP - Light Health Monitoring - app + DB

12.2.1.9Disable HTTP OPTIONS Method for Apache

Remediation Steps

Disable the OPTIONS method by including the following in the Apache configuration:

<Limit OPTIONS>
  Order deny,allow
  Deny from all
</Limit>

Assets inolved

Name	IP Address	Site
ec2-18-195-177-111.eu-central-1.compute.amazonaws.com	18.195.177.111	Essence 2020 AWS Public IP - CBAC application Spain
ec2-18-195-78-58.eu-central-1.compute.amazonaws.com	18.195.78.58	Essence 2020 AWS Public IP - CBAC application Italy

12.2.1.10 Disable SSLv2, SSLv3, and TLS 1.0. The best solution is to only have TLS 1.2 enabled

Remediation Steps

There is no server-side mitigation available against the BEAST attack. The only option is to disable the affected protocols (SSLv3 and TLS 1.0). The only fully safe configuration is to use Authenticated Encryption with Associated Data (AEAD), e.g. AES-GCM, AES-CCM in TLS 1.2.

Assets inolved

Name	IP Address	Site
Unknown	18.159.6.189	Essence 2020 AWS Public IP - Essence Manager + DB

ec2-3-64-11-58.eu-central-1.compute.amazonaws.com	3.64.11.58	Essence 2020 AWS Public IP - Light Health Monitoring - app + DB

12.2.1.11 Disable TLS/SSL support for static key cipher suites

Remediation Steps

Configure the server to disable support for static key cipher suites.
For Microsoft IIS web servers, see Microsoft Knowledgebase article 245030 for instructions on disabling static key cipher suites. The following recommended configuration provides a higher level of security. This configuration is compatible with Firefox 27, Chrome 22, IE 11, Opera 14 and Safari 7. SSLv2, SSLv3, and TLSv1 protocols are not recommended in this configuration.

Instead, use TLSv1.1 and TLSv1.2 protocols.

Refer to your server vendor documentation to apply the recommended cipher configuration:

ECDHE-RSA-AES128-GCM-SHA256:ECDHE-ECDSA-AES128-GCM-SHA256:ECDHE-RSA-AES256-GCM-SHA384:ECDHE-ECDSA-AES256-GCM-SHA384:DHE-RSA-AES128-GCM-SHA256:DHE-DSS-AES128-GCM-SHA256:kEDH+AESGCM:ECDHE-RSA-AES128-SHA256:ECDHE-ECDSA-AES128-SHA256:ECDHE-RSA-AES128-SHA:ECDHE-ECDSA-AES128-SHA:ECDHERSA-AES256-SHA384:ECDHE-ECDSA-AES256-SHA384:ECDHE-RSA-AES256-SHA:ECDHE-ECDSA-AES256-SHA:DHE-RSA-AES128-SHA256:DHE-RSA-AES128-SHA:DHE-DSS-AES128-SHA256:DHE-RSA-AES256-SHA256:DHE-DSS-AES256-SHA:DHE-RSA-AES256-SHA:!aNULL:!eNULL:!EXPORT:!DES:!RC4:!3DES:!MD5:!PSK

Assets involved

Name	IP Address	Site
Unknown	18.159.6.189	Essence 2020 AWS Public IP - Essence Manager + DB
ec2-3-64-11-58.eu-central-1.compute.amazonaws.com	3.64.11.58	Essence 2020 AWS Public IP - Light Health Monitoring - app + DB

12.2.1.12 Use HTTP X-Frame-Options

Remediation Steps

Send the HTTP response headers with X-Frame-Options that instruct the browser to restrict framing where it is not allowed.

Assets involved

Name	IP Address	Site
Unknown	18.159.6.189	Essence 2020 AWS Public IP - Essence Manager + DB
ec2-18-195-177-111.eu-central-1.compute.amazonaws.com	18.195.177.111	Essence 2020 AWS Public IP - CBAC application Spain

ec2-18-195-78-58.eu-central-1.compute.amazonaws.com	18.195.78.58	Essence 2020 AWS Public IP - CBAC application Italy
ec2-3-64-11-58.eu-central-1.compute.amazonaws.com	3.64.11.58	Essence 2020 AWS Public IP - Light Health Monitoring - app + DB

12.2.1.13 Remove the default page or stop/disable the IIS server

Remediation Steps

If this server is required to provide necessary functionality, then the default page should be replaced with relevant content.

Otherwise, this server should be removed from the network, following the security principle of minimum complexity.

If the server is not needed, it can be disabled in the following way: in the Services window of the Control Panel's Administrative Tools section, right-click on the 'World Wide Web Server' entry and select 'Stop'. Set its startup type to 'Manual' so that it does not restart if the machine is rebooted (this is done by selecting 'Properties' in the right-click menu).

Assets involved

Name	IP Address	Site
ec2-3-64-11-58.eu-central-1.compute.amazonaws.com	3.64.11.58	Essence 2020 AWS Public IP - Light Health Monitoring - app + DB

12.2.1.14 Disable HTTP OPTIONS Method for IIS

Remediation Steps

Disable the OPTIONS method by doing the following in the IIS manager

1. Select relevent site
2. Select Request filtering and change to HTTP verb tab
3. Select Deny Verb from the actions pane
4. Type OPTIONS into the provided text box and press OK

Assets inolved

Name	IP Address	Site
ec2-3-64-11-58.eu-central-1.compute.amazonaws.com	3.64.11.58	Essence 2020 AWS Public IP - Light Health Monitoring - app + DB

12.2.1.15 Disable TLS/SSL support for 3DES cipher suite

Remediation Steps

Configure the server to disable support for 3DES suite.

For Microsoft IIS web servers, see Microsoft Knowledgebase article 245030 for instructions on disabling 3DES cipher suite. The following recommended configuration provides a higher level of security. This configuration is compatible with Firefox 27, Chrome 22, IE 11, Opera 14 and Safari 7. SSLv2, SSLv3, and TLSv1 protocols are not recommended in this configuration.

Instead, use TLSv1.1 and TLSv1.2 protocols.

Refer to your server vendor documentation to apply the recommended cipher configuration:

ECDHE-RSA-AES128-GCM-SHA256:ECDHE-ECDSA-AES128-GCM-SHA256:ECDHE-RSA-AES256-GCM-SHA384:ECDHE-ECDSA-AES256-GCM-SHA384:DHE-RSA-AES128-GCM-SHA256:DHE-DSS-AES128-GCM-SHA256:kEDH+AESGCM:ECDHE-RSA-AES128-SHA256:ECDHE-ECDSA-AES128-SHA256:ECDHE-RSA-AES128-SHA:ECDHE-ECDSA-AES128-SHA:ECDHERSA-AES256-SHA384:ECDHE-ECDSA-AES256-SHA384:ECDHE-RSA-AES256-SHA:ECDHE-ECDSA-AES256-SHA:DHE-RSA-AES128-SHA256:DHE-RSA-AES128-SHA:DHE-DSS-AES128-SHA256:DHE-RSA-AES256-SHA256:DHE-DSS-AES256-SHA:DHE-RSA-AES256-SHA:!aNULL:!eNULL:!EXPORT:!DES:!RC4:!3DES:!MD5:!PSK

Assets inolved

Name	IP Address	Site
ec2-3-64-11-58.eu-central-1.compute.amazonaws.com	3.64.11.58	Essence 2020 AWS Public IP - Light Health Monitoring - app + DB

12.2.1.16 Generate random Diffie-Hellman parameters

Remediation Steps

Configure the server to use a randomly generated Diffie-Hellman group. It's recommend that you generate a 2048-bit group. The simplest way of generating a new group is to use OpenSSL:

openssl dhparam -out dhparams.pem 2048

Assets inolved

Name	IP Address	Site
ec2-18-195-177-111.eu-central-1.compute.amazonaws.com	18.195.177.111	Essence 2020 AWS Public IP - CBAC application Spain

12.2.1.17 Replace TLS/SSL self-signed certificate

Remediation Steps

Obtain a new TLS/SSL server certificate that is NOT self-signed and install it on the server. The exact instructions for obtaining a new certificate depend on your organization's requirements. Generally, you need to generate a certificate request and save the request as a file. This file is then

sent to a Certificate Authority (CA) for processing. Your organization may have its own internal Certificate Authority. If not, you may have to pay for a certificate from a trusted external Certificate Authority, such as Thawte or Verisign.

Assets inolved

Name	IP Address	Site
ec2-18-185-215-209.eu-central-1.compute.amazonaws.com	18.185.215.209	Essence 2020 AWS Public IP - Monitoring Application - AI server

12.2.1.18 Disable TCP timestamp responses on Linux

Remediation Steps

Set the value of net.ipv4.tcp_timestamps to 0 by running the following command:

sysctl -w net.ipv4.tcp_timestamps=0

Additionally, put the following value in the default sysctl configuration file, generally sysctl.conf:

net.ipv4.tcp_timestamps=0

Assets involved

Name	IP Address	Site
Unknown	18.159.6.189	Essence 2020 AWS Public IP - Essence Manager + DB
ec2-18-157-181-129.eu-central-1.compute.amazonaws.com	18.157.181.129	Essence 2020 AWS Public IP - CBAC DB
ec2-18-195-78-58.eu-central-1.compute.amazonaws.com	18.195.78.58	Essence 2020 AWS Public IP - CBAC application Italy
ec2-3-71-245-143.eu-central-1.compute.amazonaws.com	3.71.245.143	Essence 2020 AWS Public IP - Monitoring Application - server + DB

12.2.1.19 Enable TLS/SSL support for strong ciphers

Remediation Steps

Enable support for at least one of the ciphers listed below:

- TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256
- TLS_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256
- TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384
- TLS_ECDHE_ECDSA_WITH_AES_256_GCM_SHA384
- TLS_DHE_RSA_WITH_AES_128_GCM_SHA256
- TLS_DHE_DSS_WITH_AES_128_GCM_SHA256
- TLS_DHE_DSS_WITH_AES_256_GCM_SHA384
- TLS_DHE_RSA_WITH_AES_256_GCM_SHA384

Assets involved

Name	IP Address	Site
Unknown	18.159.6.189	Essence 2020 AWS Public IP - Essence Manager + DB
ec2-3-64-11-58.eu-central-1.compute.amazonaws.com	3.64.11.58	Essence 2020 AWS Public IP - Light Health Monitoring - app + DB

12.2.2 Implemented mesaures

Most of the vulnerabilities detected with the security/penetration testing are due to the fact that web services on virtual machines in AWS cloud are accessible directly via their public IP addresses, this access has been enabled for development purposes during the ESSENCE project.

This access has been closed after the penetration tests, thus mitigating majority of identified vulnerabilities. In order to allow admin access to virtual machines by responsible partners, only certificate-based SSH service is allowed via public IP addresses of virtual machines, with SSH tunnelling as an option for stronger protection of access to other services on these servers, if needed for further development/testing purposes. Moreover, weak hash algorithms (MD5, 96-bit HMAC) have been disabled in SSH configuration as per the remediation recommendation #2, see section 12.2.1.2.

Thus, for irregular development purposes, AWS servers can only be accessed via SSH, while regular access to ESSENCE AWS servers for production use is only possible through the NGNIX reverse proxy as part of the comprehensive ESSENCE security facilities, see section **Error! Reference source not found.**.

12.3 Analysis of Residual risks

12.3.1 Illegitimate access to the data

If an illegitimate access to data, we envisage here a moral feeling of invasion of privacy. This might happen for misuse of information from the qualified personnel, loss of personal device (e.g. tablet) or breach into Essence platform.

This risk is miminised as data protection in the platform has been enforced by design. The following control and design measures have been adopted for this aim:

- Encryption of sensitive data
- Partitioning data
- Traceability (logging)
- Clamping down on malicious software
- Backup
- Maintenance
- Personnel management and training
- Paper on document security
- Logical access control
- Pseudo-anonymization
- Managing workstations

- Website security
- Monitoring network activity
- Network security
- Operating security

This risk can be analyzed answering the following questions:

What could be the main impacts on the data subjects if the risk were to occur?
Moral feeling of invasion of privacy

What are the main threats that could lead to the risk?
Misuse of information from the qualified personnel

What are the risk sources?
Loss of personal device (tablet), ESSENCE platform breach

Which of the identified controls contribute to addressing the risk?
Encryption, Partitioning data, Traceability (logging), Clamping down on malicious software, Managing workstations, Website security, Backups, Maintenance, Personnel management, Paper document security, Logical access control, Pseudoanymisation, Training of the Personnel

How do you estimate the risk severity, especially according to potential impacts and planned controls?
Negligible.
The risk severity is "negligible" because the only result of a data breach could result in feeling an invasion of privacy without a real or objective harm

How do you estimate the likelihood of the risk, especially in respect of threats, sources of risk and planned controls?
Negligible.

Being a research project, the trial is conducted with a limited population (120 target users). Having set up all the proper countermeasures from the technical, legal and ethical perspective, the likelihood of the risk is estimated to be low.

12.3.2 Unwanted modification of the data

The loss of data integrity can let the users encounter inconvenience from unauthorised alteration of health data (signals and statistics), which could even make it difficult for them to receive appropriate feedbacks in the form of notifications and/or alerts and could result in unnecessary tele-consultations. This may potentially raise mental stress.

The loss of data availability could again hinder the timely and accurate feedbacks to subjects.

The main threats that can lead to this hazard are: Data Hacking from external sources, Data manipulation by internal personnel without proper qualification or training, Software bias or errors

This risk is miminized as data protection in the platform has been enforced by design. The following control and design measures have been adopted for this aim:

- Encryption of sensitive data

- Traceability (logging)
- Clamping down on malicious software
- Backup
- Maintenance
- Personnel management and training
- Paper on document security
- Logical access control
- Pseudonymization
- Managing workstations
- Website security
- Monitoring network activity
- Network security
- Operating security

This risk can be analyzed answering the following questions:

What could be the main impacts on the data subjects if the risk were to occur?

The loss of data integrity can let the users encounter inconvenience from unauthorised alteration of health data (signals and statistics), which could even make it difficult for them to receive appropriate feedback in the form of notifications and/or alerts and could result in unnecessary tele-consultations. This may potentially raise mental stress.

The loss of data availability could again hinder the timely and accurate feedback to subjects.

What are the main threats that could lead to the risk?

Data Hacking from external sources, Data manipulation by internal personnel, Software bias or errors

What are the risk sources?

Hackers, Personnel without proper qualification or training, Software bugs

Which of the identified controls contribute to addressing the risk?

Encryption, Traceability (logging), Logical access control, Backups, Maintenance, Pseudonymisation, Clamping down on malicious software, Managing workstations, Website security, Monitoring network activity, Network security, Operating security Partitioning data, Personnel management,

How do you estimate the risk severity, especially according to potential impacts and planned controls?

Limited.

Due to the implemented controls, if on one side the risk severity could be high if a person is receiving wrong feedback and alerts, on the other side the overall risk severity is estimated as limited. Indeed, no final decisions on health status are taken solely by the AI monitor but users have the possibility to do teleconsultation with clinicians and teachers to understand the potential anomalies highlighted by the AI model.

How do you estimate the likelihood of the risk, especially in respect of threats, sources of risk and planned controls?

Limited.

In respect of the identified threats and sources of risk, thanks to the planned controls, if on one side the risk severity could be high if a person is receiving wrong feedback and alerts, on the other side the overall risk severity is estimated as limited. The training of personnel and its participation to the pilot definition and conduction also assures a constant monitoring that is minimising the likelihood of the internal risk. The estimation is taken with overestimation to "limited", but it is also possible to assume it as "negligible" and due mainly to the external risk sources.

12.3.3 Data disappearance

Data disappearance can result in wrong alerts and feedbacks from the system to the user, because of missing data. This may be attributed to hardware or Software malfunctioning or to hacking. Main risk sources are identified in:

- External human resources
- Internal Human resources
- Hardware and software malfunctioning

This risk is miminized as data protection in the platform has been enforced by design. The following control and design measures have been adopted specifically for this aim:

- Backup,
- Clamping down on malicious software
- Maintenance
- Logical access control
- Personnel management

This risk can be analyzed answering the following questions:

What could be the main impacts on the data subjects if the risk were to occur?

Data disappearance can result in wrong alerts and feedback from the system to the user.

What are the main threats that could lead to the risk?

Hardware malfunctions, Software malfunctions, Hacking

What are the risk sources?

External human resources, Internal Human resources, Hardware malfunction and software malfunction

Which of the identified controls contribute to addressing the risk?

Backups, Clamping down on malicious software, Managing workstations, Website security, Monitoring network activity, Network security, Operating security, Maintenance, Logical access control, Personnel management

How do you estimate the risk severity, especially according to potential impacts and planned controls?

Negligible.

The ICT platform is implementing proper countermeasures at hardware and software level to prevent such eventuality. Also proper personnel training is minimising the risk for the internal personnel.

How do you estimate the likelihood of the risk, especially in respect of threats, sources of risk and planned controls?

Negligible.

The ICT platform is implementing proper countermeasures at hardware and software level to prevent such eventuality. Also proper personnel training is minimising the risk for the internal personnel. This risk can be therefore considered negligible.

12.4 Completeness of risk assessment

The risk assessment activity is based on the experience acquired by the manufacturer in consideration of the many types of devices, including similar ones, treated.
gure 4. Risk overview.

13 Functionalities that generate data

Functionality (D22)	Responsible	Brief Description	Target Population	Measurements	Example of Indicators
Handwriting Standalone	POLIMI	Daily unconstraint handwriting to monitor features correlated to the handwriting performance and tremor	Senior	3D Acceleration, 3D Gyroscopes, 3D magnetometer, tip force and timestamp raw data	handwriting and tremor indicators such as tilt, approximate entropy
Voice Analysis Standalone	SG	Voice acoustic features extracted on the fly during phone call to monitor cognitive decline, araousal and valence	Senior	-	acoustic features computed on the fly such as pitch...
Mini Games	POLIMI	Mini games in three domains of learning: writing reading and calculation; these games should be played by all children and if the child is identified as a child at risk the game should be planned by the teacher as reinforcement	Child	Raw data of the interaction of the user with the tablet during the game	game score, frequency of use, game level ..

Serious Games	POLIMI	Serious game used to detect potential signs of learning delays; 1. copy square; 2. copy sequence; 3. tunnel square; 4. tunnel ELE	Child	Raw data of the pen interaction of the user with the tablet during the game (x,y coordinate, pressure)	features correlated to handwriting production during the test
Teleconsultation	UMIL/POLIMI	The teleconsultation module is provided inside the CBAC. The module provides activities to be performed during teleconsultation (such as cognitive tests and cognitive exercises). Two users with different roles: consultant and consultee. The interface shows the videos of the two roles, an active area is shown and used to display tests or images or to let the consultee draw or write during the visit. In this latter case the drawing of the consultee is shown in real time to the consultant and the consultant can collect results and fill in reports.	Senior/Child with clinician	Digital Clinical Test raw data	Report

Digital Tests (TMT, Bells)	UMIL	Digital Tests to be performed by the senior alone with the automatic supervision of the SW	Senior	raw x,y inputs of the trace executed during the test	test scores
Postural Exergames for elders	UMIL	Postural exergames carried out with the RGB-D camera and skeleton analysis to extract the user movement.	Senior	-	game score, frequency of use, game level and features correlated to postural control
Video Tutorial gym for elders	UMIL	It is a virtual room in which elders can see a tutor carrying out a gym video lecture and the elder can replicate the movements / physical exercises. are following a tutor that is doing exercises	Senior	-	frequency of use
Video Tutorial for children	UIN	The same setting as virtual gym but finalized at teaching psychomotricity	Children	-	frequency of use
Multiplayer Cognitive Games	UMIL	Cards, Pictionary, Puzzle, Ruzzle, Bingo of rhyms and syllabus	Senior Child	-	frequency of use

14 Datasets in ESSENCE

The ESSENCE components exchange data between them as shown in the following **Figure 5**, in which the data flow between the components is highlighted. In particular, we distinguish with alphabet letters the data payloads in:
- Measurements (M): raw data acquired by a component.
- Indicators (I): processed raw data and outcomes of activities.
- Feed-backs (F): any type of data that returns to a module: feed-back for users, alerts and notifications.
- Activity model (A): defines an activity with the parameters and indicators associated to an activity.
- Daily scheduling (S): is produced by the professional through the LHM and it is constituted of a set of activities that are suggested or mandatory for a user in a given day. (The digital neuropsychological tests - TMT-A, TMT-B and Bells - are defined as mandatory activities).
- Teleconsultation (T): it contains the teleconsultation tests required to the user. It is initialized with the teleconsultation tests definition and it incorporates also the test clinical report, when the test has been completed. The Teleconsultation behaves differently with respect to activities in terms of data. This is because the Teleconsultation cannot be used freely by the users, but each session is scheduled and then carried out in a precise time slot. The scheduling procedure (performed by a psychologist through the LHM) produces a document, called Session, which is saved in the central datacenter. The Teleconsultation app uses this document to setup a Teleconsultation session. When the session is finished, the original Session document in the datacenter is updated by the Teleconsultation with the data of the performed tests.
- User profile (U): personal data of the user.

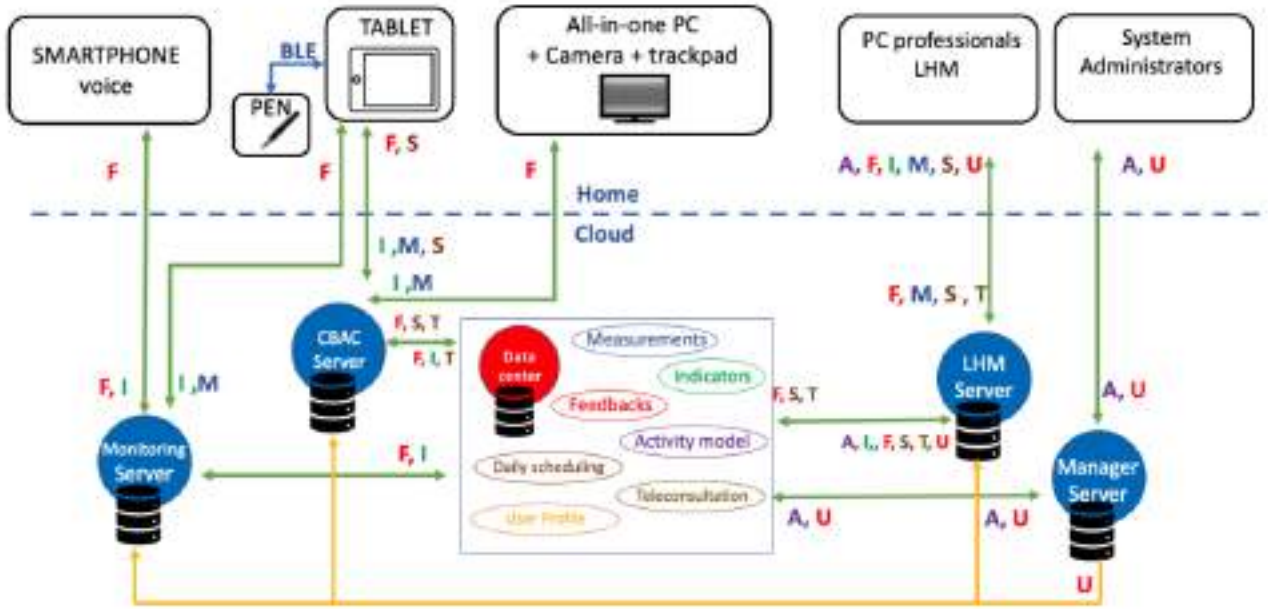


Figure 5. ESSENCE components and the data exchanged between them: Measurements, Indicators, Feedbacks, Activity models, daily Scheduling, Teleconsultation and User Profile are distinguished (cf. Glossary in Section 5). This is the same as Figure 4, duplicated here for sake of clarity.

14.1 Data flow

All components exchange data with the associated Local Data Server. This allows a high modularity of the data structure and facilitates maintenance and upgrading of the different modules. Each local server can automatically process data incoming from one or more components to produce secondary data to be transmitted back or transferred to the Data Center for final storage.

The Data Center works as a central node of the system through which all data go through. This choice allows maximum modularity of the system on one side and on the other to centralize all data security measures. Specific data models used for the different types of data are reported in Appendix A.

The exchange of data between modules is here after described. It is the final view of the data flow shown in Deliverable D3.3, that optimizes data transfers.

14.1.1 System Administration module

The starting point is the registration of users carried out through the System Administrators module, that is a Web application. This allows entering the personal data of the user (U), that are transferred to the local Manager Server. The same panel allows to enter the model of the different activities (A). These data are then transferred to the Data Center. At any time the structure of an activity (A) as well as users with their authorizations and properties (U) can also be queried in the Data Center through the Manager Server and displayed. The system administration, being a Web application, module could be used also by a PC or a tablet.

14.1.2 PC for professionals

The PC for professionals is used by the Clinicians to run the Light Health Monitoring module. It contains the GUI for a clinician/teacher to view user data, plan the activities for the user and review thoroughly the results obtained on the activities.

Through the LHM, clinicians produce a daily Scheduling (S) based on the activities available retrieved from the Data Center (A). It produces also a Teleconsultation session (T) for the user selected. Scheduling and Teleconsultation are sent to the LHM Server and in turns to the Data Server.

The clinicians can also input specific additional personal data for a particular user (M), such as the diagnosis based on the ICD-10 dictionary, the body weight... These data are kept inside the LHM server.

The LHM can also group his/her users (U) in homogenous classes (e.g. children of 1st grade, patients within a given region) to which the same daily scheduling (S) is applied. Teleconsultation instead is always personal, as it requires setting up a one-one tele-visit.

The LHM analyzes the outcomes contained inside the Indicators (I) of the different activities carried out by a user to compute the compliance to his/her activity plan (F).

The LHM displays:

- the outcomes (contained in the Indicators – I) computed by all functionalities including the daily outcome computed by the voice processing and pen data analysis visualizing them through dashboards.
- the clinical report of teleconsultation (T)
- the report/alert computed by the monitoring AI on the voice and handwrting data (F).
- specific additional personal data of his/her patients (U)

14.1.3 The smartphone

The smartphone extracts on the fly features (I), from the acoustic signal collected through smartphone calls. No acoustic signal is stored. These are used to detect early cognitive decline. These features are transferred to the Monitoring server and the Data Center.

14.1.4 The local monitoring server

The local monitoring server processes the pen raw data to extract features for further analysis (I). It also hosts the AI module that processes the indicators produced by the indicators produced by the smart phone and the smart pen to generate alerts and notifications that are sent to the Data Center (F). These are retrieved by the LHM to be displayed to the clinicians.

14.1.5 The tablet

The tablet is the most versatile component. It is used by the CBAC (that could also be served by a PC) to provide the following functionalities:

- Games for elders and children (A)
- Digital cognitive tests of elders (A)
- Screening tests for teleconsultation for children (T) with and without the use of the smart pen.

These components have been developed as Web applications under the model view control approach. In this approach all the logic stays in the CBAC server that sends to the client (the tablet) only the current view that has to be shown. The Tablet, in this view, returns only the raw data of single activities (M) and the indicators specific of each activity (I).
The CBAC supports also the serious games for children that have been developed as native applications for tablet. In this case, the CBAC server sends to the tablet the Schedule required and receives also measurements (M) and indicators (I) related to the minigames from the tablet.

Tablet displays also notifications and alerts (F). These are treated as push notifications that are displayed upon their arrival to the tablet. Some notifications come from the Data Center and are visualized on the tablet similarly to what is done with the Smart phone. Other notifications are generated inside the CBAC server, as a result of the analysis of the Teleconsultation (T) or Activity schedule (S), for instance when a mandatory activity is required or a teleconsultation is pending.

The tablet communicates with the CBAC server which returns to the Data Center:

- the indicators associated to each activity (I) including the outcome.
  the teleconsultation session, filled with the test report, for each teleconsultation (T).

14.1.6 The Smart pen

An additional data exchange takes place when the smart pen is used for screening tests in teleconsultation. In this case, the raw data from the pen (M) are routed to the Monitoring server (along with its indicators (I). From here, only indicators are transmitted to the Data Center. Besides these data the monitoring server compiles a report (F) that is also sent to the Data Center. Lastly, the Monitoring Server supervises the good operation of the smart pen and sends notifications to the tablet (F), like “pen low-battery”, to be displayed on the tablet.

The smart pen can be used also in stand-alone mode. In this case, its data are acquired by the tablet and sent to the monitoring server, in a complete transparent way as it is managed by an application running continuously in background. The data exchanged with the Monitoring Servers are the same as before.

14.1.7 All-in-one PC

The all-in-one PC + camera + trackpad is used by elders at home to play exer-games for physical exercising. It provides exergames, video tutorials and video chat upon elder’s request through a specific menu. It exchanges with the CBAC server the same data of the Tablet.

14.1.8 Technical management of the data flow

Beyond supporting data exchange, each local server produces, statistics, usage data and alerts that are read by local technical administrators in charge to manage the servers. Local monitoring server is managed by POLIMI, local CBAC server by UMIL, Local LHM server by SXT and local Manger server by SCOM.

Moreover, through the local Manager, the global technical administrator, SCOM, can supervise the use of the entire data infrastructure and analyze the general use of the structure as well as of the single local server

15 References

Reference website for updates about Data Protection in the EU:

https://ec.europa.eu/info/law/law-topic/data-protection

Reference website for updates about Cyber Security strategy in the EU:

https://ec.europa.eu/digital-single-market/en/cybersecurity

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) can be found at:

https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32016R0679

Ethics and Data Protection relevant doument:

https://ec.europa.eu/info/sites/info/files/5._h2020_ethics_and_data_protection_0.pdf

WHO guideline on digital health interventions:

https://apps.who.int/iris/bitstream/handle/10665/311941/9789241550505-eng.pdf?ua=1

Towards a framework for policy development in cybersecurity - Security and privacy considerations in autonomous agents:

https://www.enisa.europa.eu/publications/considerations-in-autonomous-agents

Recommendations on European Data Protection Certification:

https://www.enisa.europa.eu/publications/recommendations-on-european-data-protection-certification

Handbook on European data protection law - 2018 edition Council of Europe

https://www.coe.int/en/web/data-protection/documentation

Handbook on Security of Personal Data Processing, Enisa

https://www.enisa.europa.eu/publications/handbook-on-security-of-personal-data-processing

Guidelines on Data Management in Horizon 2020

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oapilot-guide_en.pdf

Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, Version 2.0, 30 October 2015:

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oapilot-guide_en.pdf

Annotated GA version 30 October 2015, p. 218

Fact Sheet: Open Access in Horizon 2020:

http://www.nks-swg.de/media/content/FactSheet_Open_Access.pdf

Webpage of European Commission regarding Open Access:

http://ec.europa.eu/research/science-society/open_access

European Commission (2016): Guidelines on Data Management in Horizon 2020, Version 2.1, 15 February 2016:

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European Science Foundation (2011), European Code of Conduct for Research Integrity of ALLEA (All European Academies) and ESF, March 2011:

http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-ofconduct_en.pdf

FAIR data principles (FORCE11 discussion forum):

https://www.force11.org/group/fairgroup/fairprinciples

OpenAIRE repository:

https://www.openaire.eu/opendatapilot-dmp

UK Data Service:

https://www.ukdataservice.ac.uk/manage-data/document

16 Appendix A – Data models

We report here the data models used for OpenAccess data as well as for data internal+ly used.

16.1 OpenAccess datasets

Most of these OpenAccess data sets are associated to journal publications and are stored in the Zenodo repository of the ESSENCE project: **https://zenodo.org/communities/essence2020.**

Datasets produced by the pilot have still to be completely analyzed to publish the results. The latter are inserted in the OpenData repository after project end.

The data sets in Zenodo associated to publications, uploaded at the 30th of April 2023,are:

- Elimelech, Ortal Cohen; Ferrante, Simona; Josman, Naomi; Meyer, Sonia; Lunardini, Francesca; Gomez-Raja, Jonathan; Galan, Carmen; Caceres, Pilar; Sciama, Piera; Gros, Marianne; Vurro, Clodia; Rosenblum, Sara. Technology use characteristics among older adults during the COVID-19 pandemic: A cross-cultural survey. *Technology in Society,* 71, 102080. *Data are stored in a .SAV file, that is an SPSS file that contains the answers to the questionnaire and their encoding*.
- Lomurno, E., Dui, L. G., Gatto, M., Bollettino, M., Matteucci, M., & Ferrante, S. (2023). Deep Learning and Procrustes Analysis for Early Dysgraphia Risk Detection with a Tablet Application. Life, 13(3), 598. *Description of the data has been inserted in the publication page. The data sets contain a description of the data: anonymized personal data are stored in Excel files while interaction data are stored in Matlab files.*
- L. G. Dui, E. Calogero, M. Malavolti, C. Termine, M. Matteucci and S. Ferrante, "Digital Tools for Handwriting Proficiency Evaluation in Children," 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Athens, Greece, 2021, pp. 1-4, doi: 10.1109/BHI50953.2021.9508539. *Description of the data has been inserted in the publication page. The data sets contain a description of the data: anonymized personal data are stored in Excel files while interaction data are stored in Matlab files.*
- Dui, L. G., Toffoli, S., Speziale, C., Termine, C., Matteucci, M., & Ferrante, S. (2022, September). Can Free Drawing Anticipate Handwriting Difficulties? A Longitudinal Study. In 2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI) (pp. 1-4). IEEE. *Description of the data has been inserted in the publication page. The data sets contain a description of the data: anonymized personal data are stored in Excel files and data in Matlab files.*
- Dui, L. G., Lomurno, E., Lunardini, F., Termine, C., Campi, A., Matteucci, M., & Ferrante, S. (2022). Identification and characterization of learning weakness from drawing analysis at the pre-literacy stage. Scientific Reports, 12(1), 21624. *Description of the data has been inserted in the publication page. The data sets contain a description of the data: anonymized personal data are stored in Excel files and data in Matlab files.*
- Termine, C., Dui, L.G., Borzaga, L. et al. Investigating the effects of COVID-19 lockdown on Italian children and adolescents with and without neurodevelopmental disorders: a cross-sectional study. Curr Psychol (2021). https://doi.org/10.1007/s12144-021-02321-2. *Description of the data has been inserted in the publication page. Data are stored in .csv format.*

16.2 Data sets internal to ESSENCE

Seven main types of data sets are exchanged inside ESSENCE (cf.

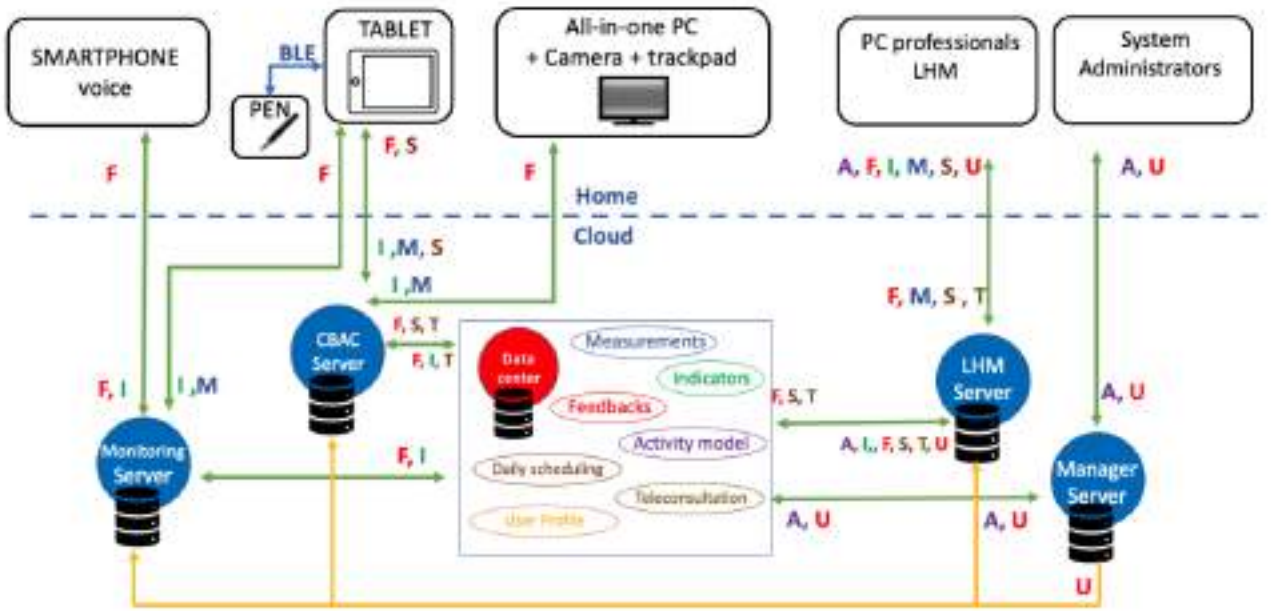


Figure 5)):

a) Measuremen
b) Indicators,
c) Feed-backs,
d) Activity models,
e) Daily scheduling
f) Teleconsultation
g) User profile

For each type, we report here the data models, the content of specific fields of the data model in the different functionalities, some meaningful examples of these data.

16.2.1 Measurements

Measurements are raw data specific for each activity and follow a model suitable to the data. However, two basic models have been adopted to describe three important types of measurements:

a. CBAC activities
b. CBAC Serious games
c. Smart pen raw data

These data model are documented in the following sections.

16.2.1.1Smart Pen Measurements

Raw data are acquired by the smart ink pen in three scenarios:

1. Unconstrained use. The user simply uses the smart ink pen to write whatever they want, without interacting with other ESSENCE components.
2. Controlled use. The user is guided in the performance of three handwriting tests with the smart ink pen by the ESSENCE smart ink pen application, installed on the user's tablet.
3. Teleconsultation. The acquisition is triggered by the clinician during the teleconusltation.

In all scenarios the same raw data are stored on the smart ink pen onboard memory, then transferred to the tablet by the ESSENCE smart ink pen application. The application is in charge of transmitting the data to the ESSENCE monitoring server. The following data are stored on the monitoring server:

- General information: this includes the unique identifier associated to the user, the date and hour in which the smart ink pen stored the data, the serial number of the pen which stored the data, information about the smart ink pen firmware.
- Activity: it represents the activity which generated the data. It can be “onboard” for the first and third scenario, “onboard; CU - pangram caps” or “onboard; CU - pangram lowers” or “onboard; CU - spiral” for the second scenario.
- Session: it represents the identifier of the teleconsultation session in which the data were generated.
- Image: it represents the encoding of the image of the handwriting tests performed by the user during the teleconsultation session.
- Raw Data: this includes the time series of the signals acquired by the smart ink pen.

16.2.1.2 Raw Data produced by the CBAC activities (through the tablet)

Each CBAC Activity, besides publishing a report inside the Indicator payload, logs a set of row data regarding all events that happened during the execution of the activity. These data are stored inside the local CBAC server for further analysis and their type and data format are specific to each individual activity. In this section, we provide a brief description of raw data acquired by each activity of the CBAC.

16.2.1.2.1 Card games Raw data

The data collected by cards game includes all data acquired during gameplay. This data is critical to understanding how players interact with the game, how they decided to think about the next move, and how they performed each move.

The following data are recorded:

- Game information: This includes the date and time of the game, the version of the game being played, the participants, ahd the final score.
- Player moves: this data include information about each move made by each player, such as which card they played, which card they discarded, and whether they took any other actions during their turn (e.g., making a point with a “Scopa”).
- Player actions: this data include information about any actions the player takes while playing the game, such as dragging a card to a new location, moving the card to rearrange the deck, or interacting with the game's user interface in any other way.
- Time taken: The time each player takes to think about their next move and to execute their turn are also recorded. This information can help identify patterns in how players approach the game and can be used to understand how much time is required by each player to think and to execute each move
- Game results: Finally, the results of each hand of the game are stored, including which player won the hand and any other relevant information about how the hand was played.

16.2.1.2.2 Pictionary Raw data

The data stored by Pictionary include information about the game session, the players, the game rounds, and the actions of each player during the game. Here are the main data points are recorded:

- Game session information: This includes the date and time of the game session, the version of the game being played, and any other relevant information about the game session, such as the number of players participating and the ID of all players. If users enter/exit a session, this event is recorded.

- Player information: Information about each are stored, including their username, display name, and any other relevant information about their account.
- Game round information: Each round of the game is recorded, including the drawing player, the guessing players, and the word that is being drawn or guessed.
- Drawing data: Information about the drawing created by the drawing player are recorded, including the timestamped event of a drawing, choice of a different color or of the eraser, and any other relevant information about the act of drawing made by the user.
- Guessing data: Information about the guesses made by the guessing player are also recorded, including the time taken to make the guess and the actual word guessed.
- Game progress data: Information about the progress of the game are recorded, including the number of rounds played, the score of each player, and any other relevant information about the game.

16.2.1.2.3 Ruzzle Raw data

The data stored the Ruzzle game include information about the game session, the players, the game board, and the words identified by each player. Here are the main data points that are recorded:

- Game session information: This includes the date and time of the game session, the version of the game being played, and any other relevant information about the game session, such as the number of players participating. If users enter/exit a session, this event is recorded.
- Player information: Information about each player are stored, including their username, display name, and any other relevant information about their account.
- Game board data: Information about the game board is recorded, including the letters in the 4x4 matrix, their location on the board, and any other relevant information about the board.
- Word data: Information about the words identified by each player are recorded, including the word itself, the time taken to identify the word, and the score earned for that word.
- Game progress data: Information about the progress of the game are recorded, including the number of rounds played, the score of each player, and any other relevant information about the game.

16.2.1.2.4 Puzzle Raw data

The data stored the Puzzle game include information about the game session, the players, the puzzle completed by each player, and the time taken by each player to complete the puzzle. Here are the main data points that are:

- Game session information: This includes the date and time of the game session, the version of the game being played, and any other relevant information about the game session, such as the number of players participating.
- Player information: Information about each player are stored, including their username, display name, and any other relevant information about their account.
- Puzzle data: Information about the puzzle completed by each player are also recorded, including the puzzle image or design, the puzzle difficulty level, and any other relevant information about the puzzle.
- Time data: Information about the time taken by each player to complete the puzzle are be recorded, including the start and end time of the puzzle, and the time taken to complete the puzzle.
- Player actions: this data include information about any actions the player takes while playing the game, such as dragging a tile to a new location, moving the tile, or interacting with the game's user interface in any other way.

- Game progress data: Information about the progress of the game are recorded, including the number of rounds played, the score of each player, and any other relevant information about the game.

16.2.1.2.5 Rubamazzetto Raw data

Data recoreded by the Rubamazzetto Activity are similar to those recoded by the other cards game. Additionally, we record the configuration parameters of the game.

16.2.1.2.6 Tessilatela Raw data

The data stored the Tessilatela game include information about the game session, the players, the puzzle completed by each player, and the time taken by each player to complete the puzzle. Here are the main data points that are:

- Game session information: This includes the date and time of the game session, the version of the game being played, and any other relevant information about the game session, such as the number of players participating.
- Player information: Information about each player are stored, including their username, display name, and any other relevant information about their account.
- Tessilatela data: Information about the Tessilatela completed by each player are recorded, including the words in the table, the entire table of the letters, board difficulty level, and any other relevant information about the game.
- Time data: Information about the time taken by each player to identify each word are also recorded, including the start and end time of the puzzle, and the time taken to complete the game.
- Player actions: this data include information about any actions the player takes while playing the game, such as guessing a word, or making a wrong guess.
- Game progress data: Information about the progress of the game are recorded, including the number of rounds played, the score of each player, and any other relevant information about the game.

16.2.1.2.7 Tombola Raw data

The data stored by Tombola include information about the game session, the players, the tiles used, the board of each player, and the game results. Here are the main data points that are recorded:

- Game session information: This includes the date and time of the game session, the version of the game being played, and any other relevant information about the game session, such as the number of players participating. We also record the configuration parameters of the game, as the difficulty.
- Player information: Information about each player are stored, including their username, display name, and any other relevant information about their account.
- Tombola board data: Information about the Tombola cards are recorded, including the card numbers, the location of each word on the card, and any other relevant information about the card.
- Extracted words data: Information about the words called during the game are recorded.
- Game results data: Information about the game results are recorded, including the winning patterns, the players who won each pattern, and any other relevant information about the game results.

16.2.1.2.8 Videotutorial Raw data

The data stored the Videotutorial include information about the users, their entering and exiting time of the platform, as well as the video they've watched during the session.

16.2.1.2.9 Videochat Raw data

The data stored the Videochat include information about the users, their entering and exiting time. Here are the main data points that are recorded:

- User information: Information about each user, including their username, display name, and any other relevant information about their account.
- Connection data: Information about the connections between users are be recorded, including the time of entering and exiting from the videochat session.

16.2.1.2.10 Exergames Raw data

The data stored by Exergames include information about the users, the start and end time of the exergame session.

For each game played, we record:

- The game played by the user,
- The duration of the game,
- The configuration parameters of the game, as the difficulty,
- The points made by the user while playing the game,
- A set of indicators that are specific for each game and that indicates the activity performed by the user and its score.

16.2.1.2.11 Digital Tests Raw data

The raw data stored by digital tests include information about the users, the start and end time of the digital session, the tests executed, and all actions performed on the board by the user. More precisely, we record:

TMT A/B:

- Time: time required to complete the test (main outcome of the test).
- Targets: total number of target-TMT events (when a user entered in a target with the draw line).
- Errors: number of error-TMT events, where the user drawn a line in a target in the wrong order
- Errors/targets the number of error-TMT events divided by the number of target-TMT events.
- ΔT: the average time between 2 successive target-TMT events.
- ΔT_n: the time between 2 successive target-TMT events, averaged over the last $21-n$ target-TMT events (with $1 \leq n \leq 20$).
- Enter-exit in a target: the number of times the user draw a line that enters/exits a target
- Draw data: the timestamped trajectory of the draw made by the user
- Pauses: the events where the user paused in drawing the line
- Line interruptions: the events where the user disconnects the pen from the screen, thus interrupting the line.

Bells:

- Targets: total number of target-Bells events (main outcome of the test). A target-bell event is when the user circle one of the bells.

- ΔT: the average time between 2 successive target-Bells events.
- ΔT_n: the time between 2 successive target-Bells events, averaged over the last 36−n target-Bells events (with 1≤n≤35).
- 1Bell: the location of the first target-Bells event, expressed in terms of subarea. The canvas of the test is divided into 9 rectangular subareas of equal sizes (3 rows [1, 2, and 3]×3 columns [A, B, and C]).
- Errors: total number of error-bells events, where the user circled an item that was not a bell.
- Omissions: total number of bells not identified by the user.

16.2.1.3Raw Data produced by the CBAC mini games for children (through the tablet)

The serious games were performed via the CBAC by the users in the children ecosystem when scheduled through the LHM. The raw data generated by this activity were stored in the ESSENCE monitoring server and included:

- General information: this includes the unique identifier associated to the user, the date and hour in which data were generated.
- Activity: it represents the specific serious game which generated the data. It can be "Copia_quadrato", "Copia_sequenza", "Tunnel_quadrato" or "ParolaTunnel_Parola".
- Session: it represents the identifier of the serious games session in which the data were generated.
- Games information: this includes the game level and its parameters, together with a flag which encodes for the level outcome (successfully executed or not)
- Raw Data: this includes the time series of the signals acquired by the tablet.

16.2.2 Indicators

Indicators are a very flexible structure that is shared among several components and can assume different schemas. They are produced by:

- Tablet:
  - Digital cognitive tests (through the CBAC)
  - Activities (through the CBAC)
  - Digital cognitive tests (through the CBAC)
  - Mini-games (through the CBAC)
- – All-in-one
  - Exer-games (through the CBAC)
- Monitoring server
  - Features extracted from smart pen data analysis
- Smart phone
  - Features extract on the fly from voice in phone conversations.

16.2.2.1CBAC indicators

Each activity performed through the CBAC results into **Indicators** that contain also the outcome of the activities. All activities but the mini-games are formatted according to the following schema:

```
{
        "UserId":{ "type":"string","required":true },
        "ActivityDate":{ "type":"string","required":true},
        "ActivityId":{"type":"string","required":true},
        "ActivityRevision":{"type":"number","required":true},
        "ActivityName":{"type":"array",
                "contains":{"type":"string"},"minItems":3,"required":true
                },
```


```
                "Parameters":[
                        {"ParameterId":{"type":"string","required":true},
                        "ParameterName":{"type":"array",
                                "contains":{"type":"string"},"required":true},
                        "ParameterType":{"type":"string","required":true,
                                "enum":["string","number","int","float","object","array","boolean",
                        "Null"]},
                                "ParameterValue":{"type":"object"
                                },
                        "ParameterOptions":{"type":"array",
                                "contains":{"type":"array","required":false}}
                                }
                        ],
                "Outcomes":[
                        {"OutcomeId":{"type":"string","required":true},
                        "OutcomeValue":{"type":"object","required":true},
                        "OutcomeName":{"type":"array",
                                "contains":{"type":"string"},"required":true
                                },
                        "OutcomeType":{"type":"string","required":true,
                                "enum":["string","number","int","float","object","array","boolean",
                        "null"]}}
                        ]
        }
```


Here we provide a brief description of the Indicators fields:

- UserID: an incremental number indicating uniquely a single user. E.g., “283”.
- ActivityDate: the date of the activity. E.g., "2023-05-04",
- ActivityID: the unique identifier of the activity. E.g., "UMIL_ruzzle", for the Ruzzle game.
- ActivityName: the name of the activity to be displayed, translated in three languages. E.g. "Scopa, Escoba, Cards",
- Parameters: a set of parameters required to configure the activity, before start. The parameters are optional and change from activity to activity. Each parameter has a field ParameterID containing its own unique name, and a field called ParameterName with the translation of the name of the parameter in three languages. The field ParameterValue indicates its value, formatted accordingly to the ParameterType. E.g., "ParameterId":"Difficulty", "ParameterName":["Difficulty", "Difficoltà","Dificultad"], "ParameterType":"array", "ParameterValue":"Medium" }

Indicators contains a set of **Outcomes** that are the results of the activity after it was completed. The Outcomes are mandatory and change from activity to activity. Each parameter has a field OutcomeID containing its own unique name, and a field called OutcomeName with the translation of the name of the parameter in three languages. The field OutcomeValue indicates its value, formatted accordingly to the OutcomeType. E.g.,

"OutcomeID": "StartTimestamp", "OutcomeName:["Start activity timestamp", "Timestamp inizio attività", "Timestamp Inicio actividad "], "OutcomeType":"int", "OutcomeValue":" 1683134533132 " }

This approach allows maximum flexibility In the description of the results of an activity.

The Outcome of the **CBAC activities** follow the Outcomes data model and contain the following elements:

- Timestamp of the start time of the activity,
- Timestamp of the end time of the activity,
- Duration of the activity, in seconds,
- Date of the activity.

These Outcomes are provided by all social, physical and cognitive activities of the CBAC:

- Cards
- Briscola
- Puzzle
- Pictionary
- Ruzzle
- Tombola
- Tessilatela
- Rubamazzetto
- Videotutorials
- Videochat

The Outcome of the **digital cognitive tests (TMT-A/B, Bells)** follow also the Outcomes data model and contain the following elements:

The indicators reported as Outcomes, for the TMT A/B tests, are:

- The duration of the TMT-A test
- The duration of the TMT-B test
- The number of errors performed in the TMT-A test
- The number of errors performed in the TMT-B test
- The number of missed items in the TMT-A test. A missed item is an item not included in the sequence by the user, e.g. in the sequence 1-2-3-5-6-…-20, the number 4 is the missed item.
- The number of missed items in the TMT-B test.
- The joint duration of the TMT-A and the TMT-B tests.

The indicators reported as Outcomes, for Bell test, are:

- The number of errors, i.e. the number of item wrongly identified as bells
- The number of omissions, i.e. the number of bells not found by the user
- The number of omissions in the left part of the picture
- The number of omissions in the right part of the picture
- The number of omissions in the central part of the picture

The Indicators of the **Exergames activities** describe the time where the activity was executed and its duration and indicates the score obtained in such. The description of the time where the activity was executed follows the same data schema of the other CBAC activities.

The Outcomes contain the score, and it varies from game to game. As an example, the outcome of the Exergame “Hay Collect” contains the fields “HitRatio” and “Points”.

Serious Games Indicators

The serious games indicators are computed from the serious games measurements. The indicators data model is reported hereafter:


```
{
 "icode": {
"type": "string",
"required": true
},
 "datetime": {
"type": "date",
"required": true
```

```
},
  "userid": {
"type": "string",
"required": true
},
"measureid": {
"type": "array",
"contains": {
"type": "string",
}
"required": true
},
"data": {
"sessionid": {
  "type": "string",
  "required": false
   },
 "activity": {
   "type": "string",
   "required": true
    },
 "indicators": {
   "type": "array",
   "contains": {
   "type": "object",
   "properties": {
   "name": {
   "type": "string",
   "required": true
    },
   "unit": {
   "type": "string",
   "required": true
   },
   "value": {
   "type": "number",
   "required": true
    }}}
 "required": true
 }}
```

In the following we provide a brief description of the fields contained in the schema:

- icode: strings that identifies the monitoring functionality from which the idicators are generated. For the smart ink pen its value is “SGD”.
- datetime: date in which the measurement used for the indicators computation was generated. The date format is “YYYY-MM-dd T HH:mm:ss.SSS Z.”
- userid: the unique identifier of the subject the indicators refer to, for example 1”.
- measureid: the unique identifier associated to the measurements from which the indicators were computed, for example “5knn1krgl9vbw”. The numner of identifiers can be mor than 1.
- data: this field contains several info related to the indicators instance.

- sessionid: unique identifier associated to the serious games session in which the measurement was generated.
- activity: string which specifies the type of activity from which the indicators were computed, for example “Copia_quadrato”.
- indicators: it is an array containing all the computed indicators. Each indicator is characterized by the following field
  - name: name of the indicator, for example “R2e”
  - unit: measurement unit of the indicator, for example “SAT”
  - value: numeric value of the indicato, for example 52.31

16.2.2.2 Voice Monitoring Indicators

The Indicators used for voice processing are the features extracted by the Smart phone app. These follow the here after schema.


```
{
"title":"Voice App Schema",
"description":"Schema for Voice App features",
"type":"object",
"properties":{
"userid":{
"description":"The User Id",
"type":"string"
},
"icode":{
"const":"VOICE_APP"
},
"callid":{
"description":"The id of the call: random number",
"type":"string"
},
"time":{
"type":"object",
"properties":{
"temporality":{
"const":"timeinterval"
},
"t0":{
"description":"The time that voice data acquisition starts",
"type":"string",
"format":"date-time"
},
"dt":{
"description":"The interval of voice data acquisition in seconds",
"type":"number"
}
},
"required":[
"temporality",
"t0",
"dt"
]
},
"data":{
"description":"An array with the values for all estimated features/keys",
"type":"array",
```

```
"items":{
"type":"object",
"properties":{
"key":{
"type":"string"
},
"value":{
"type":"number"
},
"unit":{
"type":"string"
}
},
"required":[
"key",
"value",
"unit"
]
}
}
},
"required":[
"userid",
"icode",
"callid",
"time",
"data"
],
"additionalProperties":false
}
```

16.2.2.3Smart Pen Indicators

The smart ink pen indicators are computed by an automatic routine starting from the smart ink pen measurements. The indicators data model is reported hereafter:

```
{
  "icode": {
"type": "string",
"required": true
},
 "datetime": {
"type": "date",
"required": true
},
   "userid": {
"type": "string",
"required": true
},
"measureid": {
"type": "array",
"contains": {
"type": "string",
}
"required": true
```


```
},
"data": {
"sessionid": {
  "type": "string",
  "required": false
   },
 "activity": {
   "type": "string",
   "required": true
    },
 "indicators": {
   "type": "array",
   "contains": {
   "type": "object",
   "properties": {
   "name": {
   "type": "string",
   "required": true
    },
   "unit": {
    "type": "string",
   "required": true
   },
   "value": {
   "type": "number",
   "required": true
    }}}
 "required": true
 }}
```


In the following we provide a brief description of the fields contained in the schema:

- icode: strings that identifies the monitoring functionality from which the idicators are generated. For the smart ink pen its value is "PEN".
- datetime: date in which the measurement used for the indicators computation was generated. The date format is "YYYY-MM-dd T HH:mm:ss.SSS Z."
- userid: the unique identifier of the subject the indicators refer to, for example "343".
- measureid: the unique identifier associated to the measurement from which the indicators were computed, for example "5knn1krgl9vbw".
- data: this field contains several info related to the indicators instance.
    - sessionid: unique identifier associated to the teleconsultation session in which the measurement was generated.
    - activity: string which specifies the type of activity from which the indicators were computed, for example "onboard; CU – pangram caps".
    - indicators: it is an array containing all the computed indicators. Each indicator is characterized by the following fields
        - name: name of the indicator, for example "ExecutionTime"
        - unit: measurement unit of the indicator, for example "s"
        - value: numeric value of the indicato, for example 52.31

16.2.3 Teleconsultation

The Teleconsultation app behaves differently from the activities in terms of data. This is because the Teleconsultation cannot be used freely by the users, but each session is scheduled and then carried out in a precise time slot.

The scheduling procedure (performed by a psychologist through the LHM) produces a document, named session document, which is written according to the teleconsultation schema, and saved in the central datacenter. The Teleconsultation app uses this document to setup a Teleconsultation session. When the session is finished, Teleconsultation data are updated with the report and uploaded inside the Data Center with the data of the performed tests.

The Initial Teleconsultation data contain the session document that follows this schema:

```
{
    "session_id": {"type": "string", "required": True},
    "date": {"type": "object",
        "properties": {"start_date_time": {"type": "string", "required": True},
            "end_date_time": {"type": "string},
            "start_timestamp": {"type": "number", "required": True},
            "end_timestamp": {"type": "number"}}
                },
    "participants": { "type": "object",
        "properties": {"professional_id": {"type":"number"},
            "user_id": {"type":"number}
            }
        },
    "ecosystem": {"type": "string", "required": True, "enum": ["Italy", "Spain", "Demo"]},
    "status": {"type": "string", "enum": ["pending", "completed", "canceled"]},
    "performed_tests": {"type":"array", "items": {type": "object",
                                "properties":{"test_id": {type": "string", "required": True},
                "test_name":{"type":"array", "required": True, "minItems":3,
                                                        items":{"type": "string"}},
                        "test_description": {"type": "array", "minItems": 3, "items":
                                                                        {"type":"string"}},
                 "raw_data": {type": "object"},
                 "test_report": {type": "object", "properties":{
                       "type": {"type":"string", "enum": ["pdf", image", "url"]},
                         "description":{type":"array", "minItems":3,
                             "items":{ "type":"string"}
                        "data":{"type":"string"}
                    }
                  }
                }
             }
    },
    "session_notes": {type": "string"}
  }
}
```

In the following, we provide a short description about the fields contained in the previous schema:

- session_id: the unique id of the session (generated by the system), e.g., “9a3d8da9-474c-4f9a-acb7-1c36dfcb9e18”
- date: a dictionary containing the starting and ending date of the session, both encoded as timestamp and as a readable string. It contains:
    - start_date_time: the starting time of the session expressed as a human readable string. E.g., “10/04/2023 09:00:00”
    - start_timestamp: the starting time of the session expressed as an integer (timestamp), e.g., “1681117200000”
    - end_date_time: the ending time of the session. For design purposes, it is automatically set as one hour after the starting date
    - end_timestamp: the ending time of the session expressed as an integer (timestamp)
- participants: a dictionary containing the id of the participant users. In particular, it contains:
    - professional_id: the id of the clinician, e.g., “351”

- user_id: the id of the patient, e.g., “318”

- ecosystem: a string to identify the ecosystem from which the session is carried out. The possible values are: “Italy”, “Spain”, and “Demo” for testing sessions
- status: this field give information about the status of the session, which can be “pending” (for a session not yet started), “completed” (when a session is completed and saved on the datacenter), or “cancelled” when the results have been discarded
- session_notes: general notes written by the clinician about the session
- performed_test: is a lists that contains the data of the performed tests. For each of them we save:
  - test_id: a unique id (automatically generated by the system) to identify each single test, e.g., “21c0o15lgalav9c”
  - test_name: an array which contains the name of the test in the Italian, Spanish, and English (in this order)
  - test_description: an array with the descriptions of the test the Italian, Spanish, and English (in this order)
  - test_report: an object which contains the clinical report of test and the metadata to interpret it. This dictionary is composed of:
    - type: the type of the report, which can be a “pdf”, an “image”, or an “url”. If this filed is not specified, the report is assumed to be save in pdf format
    - data: the data of the report encoded as a base64 string
  - raw_data: the low level data of the test which has been used to create the test report. The organization of this field is different for each test

16.2.4 Daily Scheduling

Daily scheduling is structured according to the following schema. It indicates the list of activities to be performed, their priority, as well as the parametrization of each activity (if required). Moreover, it contains the information about the type of user of the daily plan (field UserTypes), and the type of professional who created the daily plan (field SpecialistTypes). For each activity, an icon is also specified (as a link to an image) for visualization purposes.


```
{"UserId":{"type":"string","required":true},"ScheduledDate":{"type":"string,
"required":true},
"Activities":[{"Priority":{"type":"number","required":true},"ActivityModel":{"ActivityId":{"type":"string","requir
ed":true},"ActivityRevision":{"type":"number","required":true},
"ActivityClassification":{"type":"array","contains":{"type":"string"},"minItems":3,"required":true},
"ActivityName":{"type":"array","contains":{"type":"string"},"minItems":3,"required":true},
"SpecialistTypes":{"type":"array","contains":{"type":"string"},"enum":["Clinical","Teacher"],"required":true},
"EcoSystem":{"type":"array","contains":{"type":"string"},"enum":["Italy","Spain","Demo"],"required":true},
"Configurators":{"type":"array","contains":{"type":"string"},"required":true},
"Tags":{"type":"array","contains":{"type":"string"}},
"UserTypes":{"type":"array","contains":{"type":"string"},"enum":["Child","Senior","ExternalUser"],"required":tr
ue},
"Icon":{"type":"string","required":true},
"Parameters":[{"ParameterId":{"type":"string","required":true},
"ParameterName":{"type":"array","contains":{"type":"string"},
"required":true},"ParameterType":{"type":"string","required":true,"enum":["string","number","int","float","obj
ect","array","boolean","Null"]},
"ParameterValue":{"type":"object"},
"ParameterOptions":{"type":"array","contains":{"type":"array","required":false}}}],
"CbacParameters":[{}],
"Outcomes":[{"OutcomeId":{"type":"string","required":true},
"OutcomeName":{"type":"array","contains":{"type":"string"},"required":true},
```

"OutcomeType":{"type":"string","required":true,"enum":["string","number","int","float","object","array","boolean","null"]}}}}}}}

16.2.5 User Profile

The user profile describes the main information about the user. As it is customary for such data, it is supported by a relational data base spread across different tables (Figure 6).

The main table, **users,** contains the unique **id** of the user, his/her email, his/her name and surname, and other information about his/her personal data. It also stores the user (encrypted) password and the information about the status of the registration. The data in this table are inserted at registration time by the administrators and finalized by the clinicians. It can be updated by system administrators.

This data model is used by the authentication utility running on this server, to authenticate sessions of the users, providing a **token** associated to a user and the specific session running for that user.

Each user can have multiple **roles** associated to his/her profile. **Roles** are defined in the roles table. Examples of a roles is that a user can be a "Pilot user" for the Senior ecosystem, or a "caregiver" for the children ecosystem.

Roles are also associated to different **modules**, i.e. the main components of the platform. As an example, the LHM and the CBAC are defined as modules. This is used to provide to each user the adequate permission that they need to use the platform.

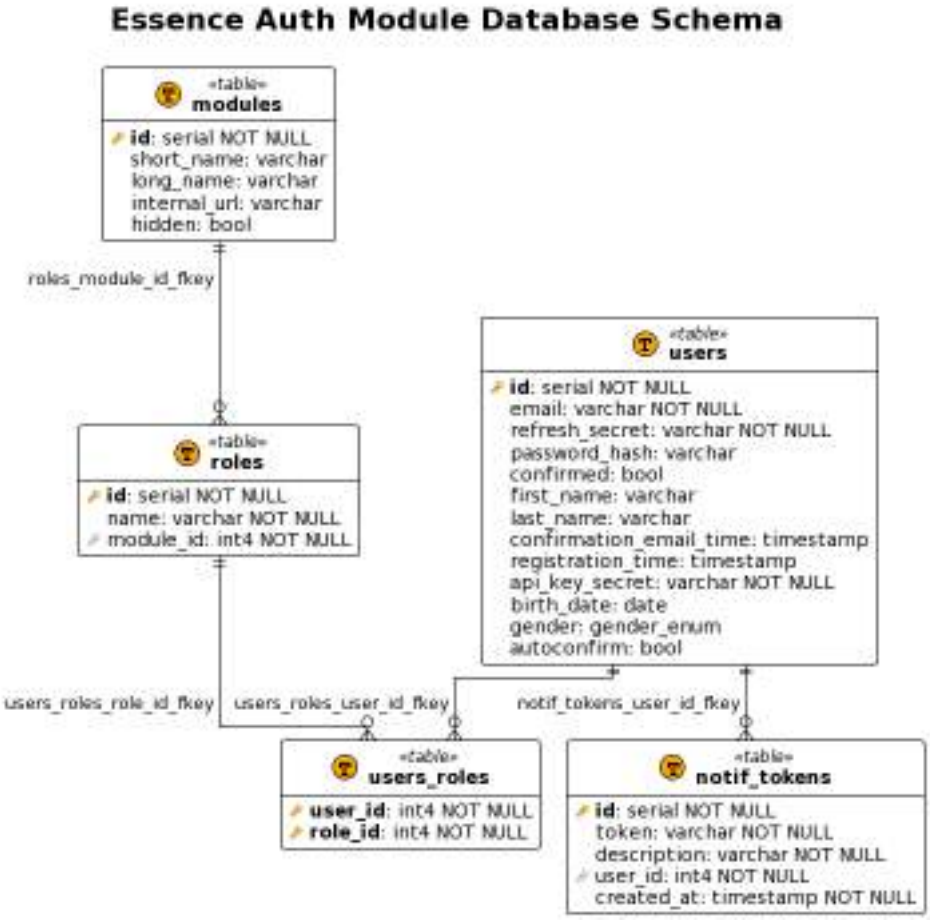


Figure 6. User profile data model.

16.2.6 Activity Model

The Activity Model describes each activity proposed by ESSENCE, This document is structured according to the following schema, and indicates the name of the activity, the type of user who can do that activity (field UserTypes), the type of professional who is interested in the Outcomes of that activity the daily plan (field SpecialistTypes), the ecosystem where the activity is deployed and used

(field EcoSystem). Moreover, it contains the list of all the parameters that are required to configure the activity (field Parameters) and the Indicators produced by the activity (field Outcomes).


```
{"ActivityId":{"type":"string","required":true},
"ActivityRevision":{"type":"number","required":true},
"ActivityClassification":{"type":"array","contains":{"type":"string"},"minItems":3,"required":true},
"ActivityName":{"type":"array","contains":{"type":"string"},"minItems":3,"required":true},
"SpecialistTypes":{"type":"array","contains":{"type":"string"},
"enum":["Clinical","Teacher"],
"required":true},
"EcoSystem":{"type":"array","contains":{"type":"string"},"enum":["Italy","Spain","Demo"],"required":true},
"Configurators":{"type":"array","contains":{t"type":"string"},"required":true},
"Tags":{"type":"array","contains":{"type":"string"}},
"UserTypes":{"type":"array","contains":{"type":"string"
},
"enum":["Child","Senior","ExternalUser"
],"required":true},
"Icon":{"type":"string","required":true
},
"Parameters":[{"ParameterId":{"type":"string","required":true
},"ParameterName":{"type":"array","contains":{"type":"string"
},"required":true
},"ParameterType":{"type":"string","required":true,"enum":["string","number,"int","float","object","array","boolean","Null"]},
"ParameterOptions":{"type":"array","contains":{"type":"array","required":false}}}],
"CbacParameters":[{}],
"Outcomes":[{"OutcomeId":{"type":"string","required":true},
"OutcomeName":{"type":"array","contains":{"type":"string"},"required":true},
"OutcomeType":{"type":"string","required":true,"enum":["string","number","int","float","object","array","boolean","null"]}}]}
```


16.2.7 Feedbacks

The feedback provided by the systems are of three types:

a. Notifications
b. Reports.
c. Alerts.

Notifications and Alerts share the same data model while reports adopt a different data model.

16.2.7.1 Notifications and Alerts Data Model

We report here a data schema of a notification produced by the LHM


```
{
"type":"object",
"properties":{
"module_id":{
"type":"string",
"enum":[
"LHM",
"CBAC",
"MONITORING",
"ESSENCE_MANAGER",
"AI_MODULE"
]
},
```


```
"user_id":{
"type":"number",
"description":"Recipient ID"
},
"timestamp":{
"type":"number"
},
"date_time":{
"type":"string"
},
"title":{
"type":"string"
},
"text":{
"type":"string"
},
"event_kind":{
"type":"string",
"description":"TDB"
},
"link":{
"type":"string",
"description":"TDB"
},
"notification_status":{
"type":"string"
},
"notification_id":{
"type":"string"
}
},
"required":[
"module_id",
"user_id",
"title",
"text"
]}
```


16.2.7.2 Report Data Model

We report here an example of report data schema produced by the LHM


```
{
"UserId":{
"type":"string",
"required":true,
"description":"Patient user ID"
},
"ReportDate":{
"type":"object",
"properties":{
"date_time":{
"type":"string",
"description":"Must be in ISO format",
"required":true
},
```


```
"timestamp":{
"type":"number",
"description":"UNIX timestamp in milliseconds for date_time",
"required":true
}
},
"required":true
},
"ReportDescription":{
"type":"array",
"contains":{
"type":"string"
},
"minItems":3,
"required":true,
"description":"Description of the report (eng, ita, esp)"
},
"ReportId":{
"type":"string",
"required":true,
"description":"Report ID"
},
"ReportRevision":{
"type":"number",
"required":true,
"description":"Version of the document"
},
"ReportName":{
"type":"array",
"contains":{
"type":"string"
},
"minItems":3,
"required":true,
"description":"Name of the report (eng, ita, esp)"
},
"Alert":{
"type":"boolean",
"required":true,
"default":false,
"description":"Flag indicating if this is an alert"
},
"Viewed":{
"type":"boolean",
"required":true,
"default":false,
"description":"Flag indicating if this report has been seen and closed by a professional"
},
"ViewedBy":{
"type":"string",
"required":false,
"description":"ID of the professional who first sees and closes the report"
},
"Rating":{
"type":"integer",
```


```
"required":false,
"description":"Rating of the report by the professional"
},
"Documents":{
"type":"array",
"contains":{
"type":"object",
"properties":{
"DocumentId":{
"type":"string",
"description":"ID of the document",
"required":true
},
"DocumentName":{
"type":"array",
"contains":{
"type":"string"
},
"minItems":3,
"required":true,
"description":"Name of the associated document (eng, ita, esp)"
},
"DocumentDescription":{
"type":"array",
"contains":{
"type":"string"
},
"minItems":3,
"required":true,
"description":"Description of the associated document (eng, ita, esp)"
}}}},
"minItems":1,
"description":"Lists of documents associated to this report"
}
```

```
{
"DocumentId":{
"type":"string",
"required":true,
"description":"Document ID"
},
"DocumentType":{
"type":"string",
"required":true,
"description":"Type of the document"
},
"DocumentContent":{
"type":"string",
"required":true,
"description":"Content of the document encoded in Base64"
}
}
```

17 Appendix B – Data Sharing Agreement

Data sharing agreement was signed by all partners in the first year of the project. It regulates all aspects related to data. It has been compiled starting from the DPIA and it is aligned to Deliverable D1.1 of ESSENCE.