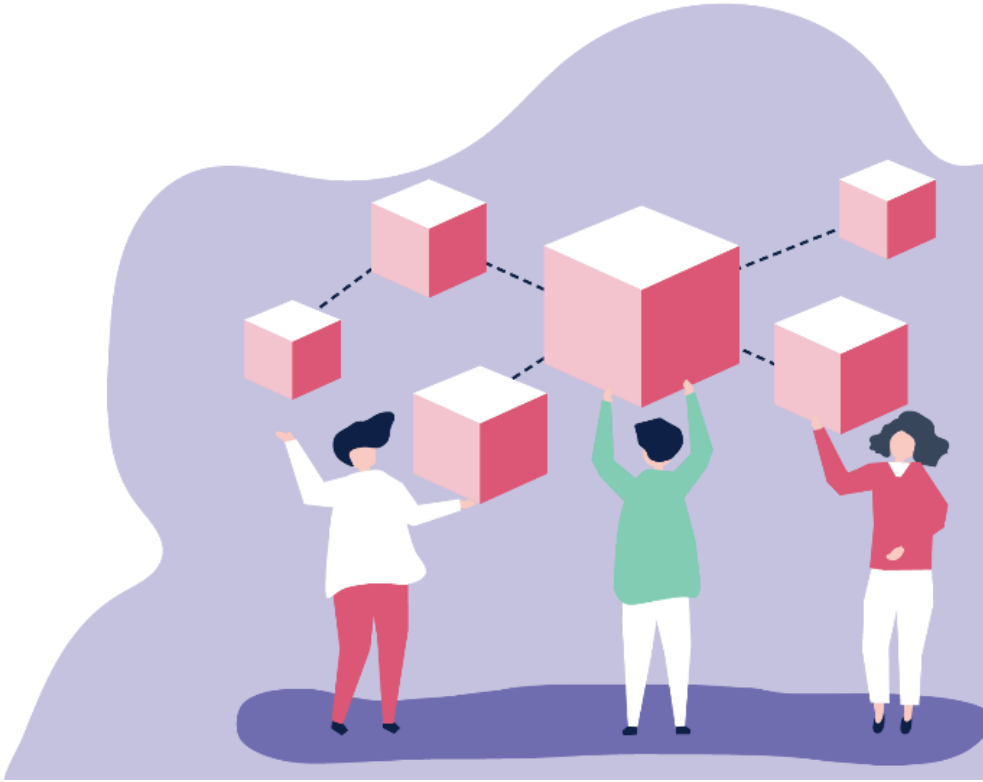
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 857223

GATE KEEPER

D1.9 D1.4.2 Data Management Plan

Deliverable No.	D1.9 D1.4.2	**Due Date**	31/12/2023
Description	This version of the Data Management Plan provides the final view of the project's activities regarding data management, knowledge, FAIR principles and data security. It reflects the data-related practices on a pilot level and the IPR management of the solutions that were developed throughout the project.		
Type	Report	**Dissemination Level**	PU
Work Package No.	WP1	**Work Package Title**	Project coordination, IPR and Ethics Issues
Version	1.0	**Status**	Final



Authors

Name and surname	Partner name	e-mail
Vasiliki Tsiompanidou	UDG	vtsiompanidou@udgalliance.org
Adrian Quesada Rodriguez	MI	aquesada@mandint.org
Alexandra Ivan	MI	aivan@mandint.org
Stea Miteva	UDG	smiteva@udgalliance.org

History

Date	Version	Change
25/08/2022	0.1	Creation of the Table of Content
14/09/2022	0.2	Additions to the Table of Content taking into account results from previous pilot consultations (D1.10, Ethical Assessments etc)
02/10/2023	0.3	Input added in sections 1 and 2
16/11/2023	0.4	Input added in sections 3, 4 and 6
11/12/2023	0.5	Final input
15/12/2023	0.6	Submission for peer review
08/02/2024	0.7	Peer review completed
09/02/2024	1.0	Final version ready for submission

Key data

Keywords	Data; Data Protection; Data Management; FAIR principles; Open Science; Datasets; Intellectual Property Rights
Lead Editor	Vasiliki Tsiompanidou (UDGA), Adrian Quesada Rodriguez (MI)
Internal Reviewer(s)	Frans Folkvord (PBY)

Abstract

This document describes the final version of the Data Management Plan and presents the relevant activities undertaken by the partners of the GATEKEEPER project.

The deliverable expands on the first version of the Data Management Plan, in which the framework for data management was defined both at the project and at the pilots' level, and reports on the data management practices of the GATEKEEPER partners both on a pilot level and infrastructure level. It is the final update of the Data Management Plan reflecting the work performed throughout the project's lifecycle and includes the information provided by pilots in their individual DMPs performed at a local level.

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Legal disclaimer

The information in this document is provided "as is" and as it has been collected according to the inputs provided by the different partners. The above referenced consortium members shall have no liability to third parties for damage of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law. This data management plan is a living document and will evolve with the advancement of the project.

Acronyms

AI Artificial Intelligence
DMP Data Management Plan
DoA Description of Action
DPIA Data Protection Impact Assessment
EEA European Economic Area
EHDS European Health Data Space
EU European Union
FAIR Finable, Accessible, Interoperable, Reusable
GDPR General Data Protection Regulation
GK GATEKEEPER
IDS Intrusion Detection System
IoT Internet of Things
IPR Intellectual Property Rights
KET Key enabling technologies
LSPs Large Scale Pilots
OU Open University
PROMs Patient Reported Outcome Measures
RUC Reference Use Case
WP Work Package

Table of contents

List of tables

List of figures

1 Executive summary

This document constitutes the final version of the data management plan performed within the GATEKEEPER project, including information relevant for the pilot sites, as well as for the GATEKEEPER infrastructure as a whole. As such, the document is structured as follows:

1) Section 2: the GATEKEEPER project and its achievements are briefly described, explaining the role of the present deliverable in the context of the project.
2) Section 3: A data summary is provided, including the categories of data collected, the purposes, types and formats, the overall data flow, as well as their retention period and storage details.
3) Section 4: The compliance of the project with the FAIR principles is presented, referring to the Findability, Accessibility, Interoperability and Reusability of the projects' datasets and further solutions developed.
4) Section 5: A summary of the Intellectual Property Rights (IPR) developed within the project is presented, along with their exploitation strategy.
5) Section 6: The technical and organisational measures to enhance security are detailed, both implemented by the pilots and referring to the overall GATEKEEPER infrastructure.
6) Section 7: The project's compliance with Ethics and Legal Aspects is explained.

In order to achieve the above, continuous monitoring of the project's activities and communication with partners was paramount. The individual Data Management Plans (DMPs) and the Data Protection Impact Assessment (DPIAs) performed at each pilot site and continuously updated throughout the project's lifecycle served as the baseline for the reporting performed in this Deliverable.

2 Introduction

2.1 The GATEKEEPER project

The GATEKEEPER project aimed from its conception at connecting healthcare providers, businesses, entrepreneurs, elderly citizens and the communities they live in, which has also been reflected at its Consortium composition. Envisioning to create an open, trust-based arena for matching ideas, technologies, user needs and processes, GATEKEEPER has established an open source, European, standard-based, interoperable and secure framework available to all developers, for creating combined digital solutions for personalised early detection and interventions that:

i. harness the next generation of healthcare and wellness innovations;

ii. cover the whole care continuum for elderly citizens, including primary, secondary and tertiary preventions, chronic diseases and co-morbidities;

iii. straightforwardly fit 'by design' with European regulations, on data protection, consumer protection and patient protection

iv. are subjected to trustable certification processes;

v. support value generation through the deployment of advanced business model based on the VBHC paradigm.

Focusing on chronic diseases, such as heart failure, Parkinson's, and diabetes, GATEKEEPER aims at empowering patients in managing diseases through the use of innovative digital tools, harnessing the power of data.

GATEKEEPER has, thus, established its plan based on Large-Scale pilots in eight regional communities, from seven member states of the European Union (EU). Overcoming the challenges imposed by the Covid-19 pandemic, it has managed to address the needs of over 40,000 EU citizens.

Based on the above, the GATEKEEPER ecosystem is comprised of various spaces dedicated to different purposes, including a healthcare space, a business space and a "consumer" space. As such, the Figure 1 below provides an overview of the GATEKEEPER ecosystem.

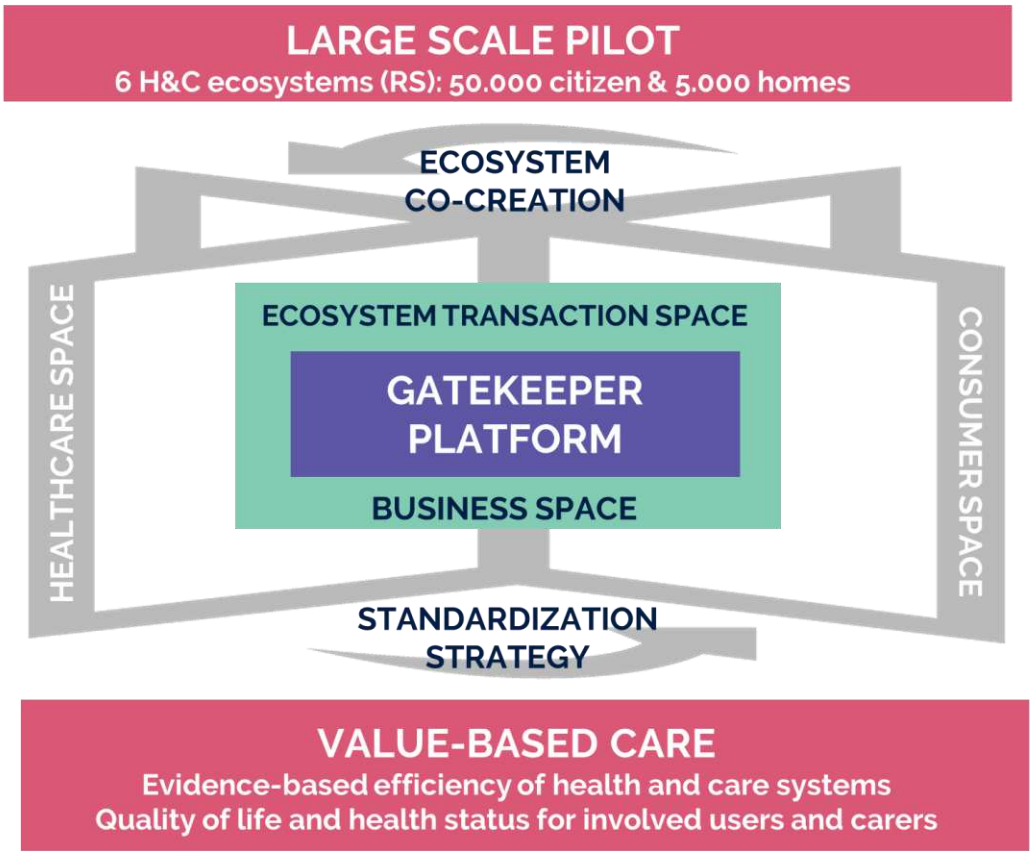


Figure 1 –The GATEKEEPER Project

2.2 Purpose of the document

This document constitutes the final version of the Data Management Plan, reflecting the practices adopted at the deployment sites and the GATEKEEPER ecosystem as a whole. The first iteration of the DMP and Deliverables D1.5 and D1.10 served as a baseline for the design of both pilots' activities and the GATEKEEPER platform and related solutions. The project-wide DMP is complemented by the DMP and the Data Protection Impact Assessment (DPIA) performed at an individual pilot level in order to encompass at the greatest degree possible the specific data management measures adopted along with the safeguards in place to ensure data protection.

As such, this final version of the DMP aims at summarising the data management practices that have been adopted project-wide and the ways the solutions and datasets generated comply with the FAIR principles. Focusing on data security, it also aims at presenting the project's privacy and security by design approach that was adopted throughout the project's lifecycle.

Table 1: Deliverable context

PROJECT ITEM	RELATIONSHIP
Objectives	The deliverables provides the overview of the data management practices adopted at a project level, encompassing in particular the different pilots' activities.
Exploitable results	The deliverable follows up on the model for data management that was created at its first iteration and explains the consortium's related practices.
Work plan	The deliverable serves as an update to the original reporting in accordance with the DoA.
Deliverables	The deliverable is to be read in conjunction with D 1.11.

3 Data Summary

The present section will provide a high-level description of the data that was generated, collected and/or processed within the GATEKEEPER project with respect to pilots' confidential information. By providing a brief description of the methodology behind the pilots' choices, this section will simultaneously present a summary of the privacy-forward considerations in each step of the process.

In particular, it will provide additional information on the following specific topics:

a) The personal data that was collected and/or processed by each pilot;

b) The overall project data flow;

c) The purposes for which personal data was collected and/or processed by each pilot;

d) The types and formats of the datasets;

e) The processing of existing datasets;

f) The storage location of the data;

g) The retention period for the data of each pilot.

3.1 Data Description

In the context of the GATEKEEPER project, personal data had a pivotal role in the design and implementation of the project's activities and the achievement of its objectives. As such, personal data related to health were collected in the course of the pilots' activity, either through medical devices provided for this purpose, patients' Electronic Medical Records, direct contact with patients or questionnaires.

In order to determine the types of personal data collected and processed and the method of collection, each pilot manager took into consideration the Reference Use Case (RUC) for which the data would be utilised, the envisioned goals and any intricacies present, collecting only the data that would be necessary to accomplish the desired result and abiding by data minimisation requirements. In addition to that, where applicable, pilots provided patients with an Informed Consent Form, including all required information, ensuring they acquired the freely-given, prior, explicit and informed consent of the patients participating in the pilot activities.

Given the sensitive nature of the data involved, being a special category of data according to Article 9 of the GDPR, all data processing performed within the project was held to high standards of protection. A privacy and ethics risk assessment was performed at a project level for each pilot (Deliverable D1.10), in addition to the Data Protection Impact Assessment (DPIA) and the Ethical Approval procedure that was completed at each pilot site.

In view of the above, the table below presents precisely the types of data collected and processed in correlation to the respective pilot and RUC. The table also reports on the level of complexity of each data processing activity, differentiating among Low, Moderate, and High and it identifies the data controller for each dataset.

Table 2 Description of Personal Data per RUC collected on a pilot level.

TASK	REFERENCE USE CASE(S) / COMPLEXITY LEVEL / PILOT SITE / DATA CONTROLLER	PERSONAL DATA PROCESSED
T6.3	*RUC 2, 5, 7 / Moderate & High / Aragon, Spain / SALUD*	○ **Medical device data:** Oxygen saturation, blood pressure, heart rate, respiratory rate, body temperature, ECG, dyspnoea degree, blood glucose ○ **EMR data:** e.g., demographics, medication, clinical activity, comorbidities ○ **Questionnaires:** Barthel, PAM, Barber, EQ5D
	RUC 3 / High / Basque Country, Spain / OSA	○ **Medical device data:** Intermittently scanned continuous glucose monitoring (isCGM), blood pressure, heart rate, ECG ○ **Non-medical-grade monitoring** of health (e.g., heart rate, blood pressure) and fitness (e.g., activity, sleep) ○ **EMR data:** e.g., chronic conditions, age, gender, therapy prescription
	RUC 3 / High / Central Greece, Greece / DCCG	○ **Medicall device data:** Real-time continuous glucose monitoring (rt-CGM), pulse rate, heart rate variability, electrodermal activity, skin temperture, blood pressure ○ **Additional data from e-CRF of the trial**, intergrated into Diabetes Management Platform by CERTH, regarding demographics, anthropometrics, biochemical tests and comorbidities. ○ **Questionnaires:** PAID, HFS-II, EQ-5D-3L / EQ VAS
	RUC 7 / High / Cyprus – AMEN / AMEN	○ **Non-medical-grade monitoring** of health and fitness, e.g., heart rate, step count, stress, respiration rate, SpO2, sleep

	○ **Questionnaires:** EuroQol- 5 Dimension (EQ-5D-3L), Global Deterioration Scale, Geriatric Anxiety Scale (GAS), Geriatric Depression Scale (GDS)
RUC 7 / High / Cyprus – PASYKAF / PASYKAF	○ **Non-medical-grade monitoring** of health and fitness, e.g., heart rate, step count, stress, respiration rate, SpO2, sleep ○ **Questionnaires:** EORTC Quality of Life Questionnaire Core-30 (EORTC QOL C30), Hospital Anxiety and Depression Scale (HADS), Integrated Palliative care Outcome Scale – Patient Version (IPOS)
RUC 7 / Moderate / Bangor, UK / OU	○ **Non-medical-grade monitoring** of health and fitness, e.g., heart rate, step count, blood pressure, sleep ○ **EMR data:** e.g., age, sex, BMI, cancer primary, TNM stage, concomitant diseases, concomitant medications ○ **Questionnaires:** Edmonton Symptom Assessment System (ESAS), UK Oncology Nursing Society (UKONS) Oncology/Haematology 24 Hours Triage Scale

3.2 Overall Data Flow in the GATEKEEPER Ecosystem

The GATEKEEPER Ecosystem is comprised of multiple components that are meant to facilitate its functionalities in a manner that is interoperable and privacy be design. As will be further detailed in Section 4 of the present, the GATEKEEPER ecosystem is based on a Web of Things architecture and is complemented by a Trust Authority to further promote privacy by design. Data analytics is performed in a federated learning manner, in order to ensure the original Data Holders (pilots) maintain control over the data entrusted to them by data subjects.

Considering the above, the figure below, as was also included in Deliverable D4.14, depicts the overall data flow within this ecosystem. Personal data collected by participants in the pilots following all necessary procedures, including consent, have been collected either directly or through the use of devices. Before entering the GATEKEEPER ecosystem, personal data has been anonymised and/or pseudonymised, depending on the needs of each pilot, and, where possible, used to develop the GATEKEEPER AI tools. Finally, the goal is to make knowledge, synthetic datasets and GATEKEEPER tools and solutions available through the GATEKEEPER Marketplace, even beyond the project's duration, under the

condition that no personal data is included in any of the solutions or otherwise jeopardised.

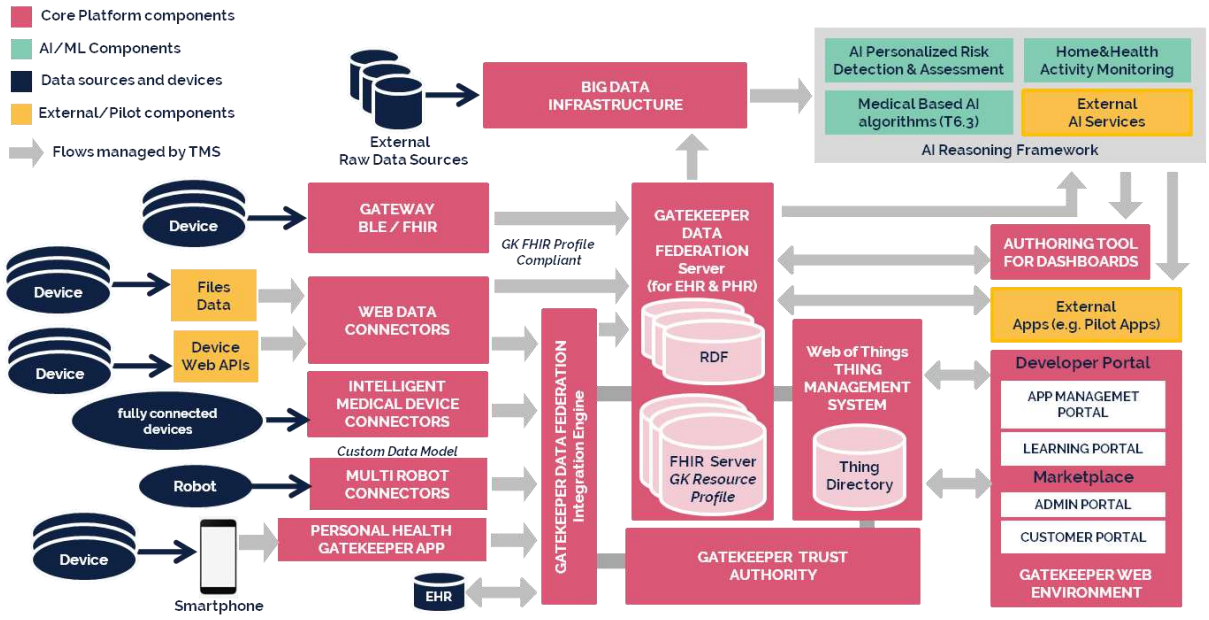


Figure 2 - GATEKEEPER Platform architecture demonstrating the overall data flows

3.3 Purpose of data collection/generation and relation to the project's objectives

The processing of personal data in the context of GATEKEEPER was performed in accordance with privacy requirements, aiming at protecting the confidentiality of any personal data involved and the adequate exercise of patients' rights. Respecting the autonomy, dignity and safety of participants to the project's various activities, while ensuring the lack of biased outcomes were key elements during this procedure.

Given the pilots' varying activities, the specific outcomes of each data processing envisioned differ depending on the pilot site, RUC, and the use of Artificial Intelligence (AI)/Machine Learning (ML) solutions. In order to better reflect the different outcomes, the table below provides a relevant overview per RUC/pilot.

Table 3 Description of the Nature of data processing and respective outcomes.

TASK	REFERENCE USE CASE(S) / COMPLEXITY LEVEL / PILOT	NATURE OF THE PROCESSING

	SITE / DATA CONTROLLER	
T6.3	*RUC 2, 5, 7 / Moderate & High / Aragon, Spain / SALUD*	**OUTCOME DESCRIPTION:** ○ Prediction of an acute exacerbation event of COPD, HF or multimorbid condition over a predefined time interval ○ Prediction of emergency room visit or hospitalization because of an exacerbation over a predefined time interval **AI/ML APPROACH:** ○ **Predictive Modelling (Inductive Reasoning):** Classification of the patients according to the risk of adverse outcomes such as exacerbations in COPD, HF, or multimorbid condition, or equivalently, prediction of the probability of a COPD exacerbation over a predefined time interval treated via classification algorithms (e.g. Class 0: No Event, Class I: Event, or Class 0: Low Risk, Class I: Intermediate Risk, Class II: High Risk) or time to event prediction modelling ○ **Pattern Mining and Clustering (Inductive Reasoning):** (a) Searching for novel patterns - Recognition of data patterns leading to/associated with an exacerbation, (b) Identification of high-risk groups.
	RUC 3 / High / Basque Country, Spain / OSA	**OUTCOME DESCRIPTION:** ○ Prediction of the probability of a cardiovascular event over a predefined time interval ○ Prediction of the probability of a short-term cardiac arrythmia event, induced by hypoglycaemia, based on ECG recording **AI/ML APPROACH:** ○ **Predictive Modelling (Inductive Reasoning):** Prediction of the probability of a cardiovascular event OR a short-term cardiac arrythmia event induced by hypoglycaemia over a predefined time interval treated via ML/DL classification methods (e.g. Class 0:

	No Event, Class I: Event) or time to event prediction modelling ○ **Pattern Mining and Clustering (Inductive Reasoning)**: Searching for novel data patterns associating long-term metabolic control with a subsequent cardiovascular event OR short-term glycaemic control, specifically hypoglycaemia, with cardiac arrythmias , or identification of high-risk groups
RUC 3 / High / Central Greece, Greece / DCCG	**OUTCOME DESCRIPTION:** ○ The estimated risk of a hypoglyceamic event based on the estimated sequence of interstitial glucose values over a 60-min (maximum) prediction horizon **AI/ML APPROACH:** ○ **Predictive Modelling (Inductive Reasoning):** Short-term prediction (up to 60 minutes) of hypoglycaemia defined considering the interstitial or capillary blood compartment of glucose measurement. A time series prediction problem addressed through adaptive linear time series or non-linear ML/DL approaches. ○ **Pattern Mining and Clustering (Inductive Reasoning):** Recognition of novel data patterns consistently leading to hypoglycaemic events
RUC 7 / High / Cyprus – AMEN / AMEN	**OUTCOME DESCRIPTION:** ○ Index of a patient's status and disease worsening. Specification of the Outcome based on: (i) the Geriatric Depression Scale (GDS) and the Geriatric Anxiety Scale (GAS) **AI/ML APPROACH:** ○ **Predictive Modelling (Inductive Reasoning):** A classification problem addressed through non-linear ML/DL methods. ○ **Pattern Mining and Clustering (Inductive Reasoning):** (a) Searching for novel patterns - Recognition of data patterns associated

	with depression and anxiety symptom levels, (b) Identification of high-risk groups.
RUC 7 / High / Cyprus – PASYKAF / PASYKAF	**OUTCOME DESCRIPTION:** o Index of a patient's status and disease worsening. Specification of the Outcome based on the Integrated Palliative care Outcome Scale - Patient Version (IPOS) **AI/ML APPROACH:** o **Predictive Modelling (Inductive Reasoning):** A classification problem addressed through non-linear ML/DL methods. o **Pattern Mining and Clustering (Inductive Reasoning):** (a) Index of patient's (daily) status and disease worsening (based on collected data) aiming at the prioritirization of daily actions and interventions, (b) Identification of high-risk groups based on interventions, disease progression, cancer type.
RUC 7 / Moderate / Bangor, UK / OU	**OUTCOME DESCRIPTION:** o Specification of the Outcome based on Edmonton Symptom Assessment System (ESAS), and UK Oncology Nursing Society (UKONS) Oncology/Haematology 24 hours triage scale **AI/ML APPROACH:** o **Pattern Mining and Clustering (Inductive Reasoning):** Searching for novel patterns - Recognition of data patterns associating EHR and Non-medical-grade PHR data with ESAS questionnaire and/or UKONS PROMs (i.e. symptoms, toxicities, problems).
	OUTCOME DESCRIPTION: o Predicting the probability of transition from State 1 to State 2 and from State 1/2 to State 3; State 1: Cancer in remission (partial or complete) & symptoms free (baseline state), State 2: Emergent symptoms, State 3: Cancer recurrence/progression, State 4: Death. **AI/ML APPROACH:**

	○ **Predictive Modelling (Inductive Reasoning):** Risk model - Predicting the probability of transition from State 1 to State 2 and from State 1/2 to State 3.

Similarly, and in order to ensure a greater level of transparency with regards to the use of AI for the processing of data, a separate table has been created to better detail the AI-based services developed and utilised within the project, as described below. It is worth highlighting that any and all AI-based activities and models developed and used within the project abide by high ethical standards and relevant legislation, both in force and upcoming.

In this regard, the AI services described below are focusing on the prediction of potential negative outcomes connected to the disease in question in order to provide preventive medicine services and related advice. As such, the design and/or use of AI models was completed in a way that would not entail negative outcomes for data subjects involved. Data anonymisation and/or pseudonymisation were essential for the further protection of the data in question.

Table 4 Description of AI services developed/used within GATEKEEPER

TASK	REFERENCE USE CASE(S) / COMPLEXITY LEVEL / PILOT SITE / DATA CONTROLLER	AI SERVICE(S)
T6.3	*RUC 2, 5, 7 / Moderate & High / Aragon, Spain / SALUD*	I. Prediction of exacerbations for people with COPD, or heart failure, or polymedicated people (Moderate Complexity) II. Prediction of exacerbations for people with COPD, or heart failure, or polymedicated people (High Complexity)
	RUC 3 / High / Basque Country, Spain / OSA	I. Long-term prediction of cardiovascular events in people with type 2 diabetes II. Short-term prediction of hypoglycaemia-related cardiac arrythmias in people with type 2 diabetes
	RUC 3 / High / Central Greece, Greece / DCCG	I. Hypoglycaemia predictive modelling

RUC 7 / High / Cyprus – AMEN / AMEN	I.	Explaining depression and anxiety levels in people with dementia
RUC 7 / High / Cyprus – PASYKAF / PASYKAF	I.	Recognition of data patterns associated with PROMs for advanced cancer patients
RUC 7 / Moderate / Bangor, UK / OU	I.	Emerging prognostic and diagnostic patterns connecting cancer symptoms
	II.	Risk prediction of cancer symptoms and recurrence

3.4 Types and formats of datasets

The datasets that have been generated, processed and/or stored within the GATEKEEPER project were aligned in accordance with a coordinated methodology in order to ensure interoperability and a homogenous approach. As such, the pilot datasets were analysed throughout their lifecycle, from the data acquisition, its characterisation to its transformation, as well as its use for the training of the AI/ML models, where applicable.

Through the creation of a comprehensive record that was continuously monitored and updated, a number of elements were being reported, including, among others, information on the data source, the data type and volume, the availability, the usability and accessibility, as well as the tools used. Though most of the original datasets were in the format of JSON, CSV and FHIR objects, ultimately the goal was to transform them so that they were aligned with HL7 FHIR (Fast Healthcare Interoperability Resources).

Table 5 Data Format per Pilot and RUC

Pilot	RUC	Data Format
Aragon	RUC 1	JSON, FHIR
Aragon	RUC 2	EHR, Medical Devices, Questionnaires, Vitalpatch health stream
Bangor	RUC 7	FHIR API
Basque Country	RUC 1	FHIR
Basque Country	RUC 3	FHIR API
Basque Country	RUC 7	CSV
Cyprus	RUC7 -AMEN	FHIR object, Garmin Venu SQ, Questionnaires, Data Federation

Cyprus	RUC7 - PASYKAF	FHIR object, Garmin Venu SQ, Questionnaires, Data Federation
Greece	RUC 1	CERTH platform Data from non-medical devices, Demographics, Anthropometrics, Biochemical tests and questionnaires
Greece	RUC 3	CERTH platform Data from non-medical devices, Demographics, Anthropometrics, Biochemical tests and questionnaires
Puglia	RUC 1	FHIR API
Puglia	RUC 3	FHIR API
Poland	RUC 7	JSON, FHIR API
Saxony	RUC 7	FHIR API

In addition to the above, the project has led to the generation of a number of deliverables and written reports. The respective WPs, as well as the formats of those documents are described in the table below.

Table 6 Documents produced in GATEKEEPER

Work-package	Deliverable-Asset	Format
WP 1	D1.1 Project Reference Manual & Quality Plan	.doc; .pdf
WP 1	D1.2 Periodic Management Report	.doc; .-pdf
WP 1	D1.3 Final Report	.doc; .pdf
WP 1	D1.4 Data Management Plan	.doc.; .pdf
WP 1	D 1.5 Legal, Ethics and Privacy Protection (LEPP) Management	.doc; .pdf
WP 1	D1.6 D1.2.2 Periodic Management Report	.doc; .pdf
WP 1	D1.7 D1.2.3 Periodic Management Report	.doc; .pdf
WP 1	D1.8 D1.2.4 Periodic Management Report	.doc; .pdf
WP 1	D1.9 D1.4.2 Data Management Plan	.doc; .pdf
WP 1	D1.10 D1.5.2 Legal, Ethics and Privacy Protection (LEPP) Management	.doc; .pdf
WP 1	D1.11 D1.5.3 Legal, Ethics, and Privacy Protection (LEPP) Management	.doc; .pdf
WP 2	D2.1 Initial Ecosystem Management Plan	.doc; .pdf
WP 2	D2.2 GATEKEEPER Trust Framework	.doc; .pdf
WP 2	D2.3 User Requirements and Taxonomy	.doc; .pdf
WP 2	D2.4 Open Innovation and co-creation workshop	.doc; .pdf
WP 2	D2.5 RRI approach for the ICT for AHA domain	.doc; .pdf
WP 2	D2.6 Open Calls Report	.doc; .pdf

WP 2	D2.7 Scaling up twinnings report	.doc; .pdf
WP 2	D2.8 GATEKEEPER Trust Authority Report	.doc; .pdf
WP 2	D2.9 D2.4.2 Open Innovation and co-creation workshop	.doc; .pdf
WP 2	D2.10 D2.4.3 Open Innovation and co-creation workshop	.doc; .pdf
WP 2	D2.11 D2.6.2 Open Calls report	.doc; .pdf
WP 2	D2.12 D2.6.3 Open Calls report	.doc; .pdf
WP 3	D3.1 Functional and technical requirements of GATEKEEPER platform [M12]	.doc; .pdf;
WP 3	D3.2 Overall GATEKEEPER architecture [M 10, M 18]	.doc; .pdf;
WP 3	D3.3 Interoperability within GATEKEEPER [M 06, M15]	.doc; .pdf;
WP 3	D3.4 Semantic Models, Vocabularies & Registry [M08, M 24]	.doc; .pdf
WP 3	D3.5 GATEKEEPER binary FHIR optimization for IoT [M16, M24]	.doc; .pdf, code
WP 3	D3.6 D3.2.2 Overall GATEKEEPER architecture	.doc; .pdf
WP 3	D3.7 D3.2.2 Interoperability within GATEKEEPER	.doc; .pdf
WP 3	D3.8 D3.4.2 Semantic Models, Vocabularies & Registry	.doc; .pdf
WP 3	D3.9 D3.5.2 GATEKEEPER binary FHIR optimization for IoT	.doc; .pdf, code
WP 4	D4.1 Microservices Containerization & Deployment [M18, M30, M 40	.doc; .pdf, code
WP 4	D4.2 Thing Management System [M 12, M 24]	.doc; .-pdf, code
WP4	D4.3 GATEKEEPER advanced Big Data services, Models and analytics for personalized risk detection & interventions [M24, M36, M40]	.doc, .pdf, code
WP 4	D4.4 Data federation and Integration and Health Semantic Data Lake [M15, M 27, M39]	.doc, .pdf, code
WP 4	D4.5 GATEKEEPER Trust Authority [M12, M24].	.doc, .pdf, code
WP 4	D4.6 GATEKEEPER Marketplace Services [M24, M36].	.doc, .pdf, code
WP 5	D5.1 Application Programming Interfaces for GATEKEEPER	.doc; .pdf, code
WP5	D5.2 Advancing and personalizing the analytic of Home Activity Monitoring and Health Activity Monitoring	.doc; .pdf, code
WP5	D5.3 AI-powered services for personalised early risk detection and risk assessment	.doc; .pdf, code

WP 5	D5.4 Intelligent Connected Care Services and IoT	.doc; .pdf, code
WP 5	D5.5 Design of authoring tool for adaptive and multimodal interfaces	.doc; .pdf, code
WP 5	D5.6 Robotic assistance in community care: general framework, requirements and evaluation	.doc; .pdf, code
WP 5	D5.7 Technical validation report	.doc; .pdf, code
WP 5	D5.8 D 5.1.2 Application Programming Interfaces for GATEKEEPER	.doc; .pdf, code
WP 5	D5.9 D5.2.2 Advancing and personalizing the analytic of Home Activity Monitoring and Health Activity Monitoring	.doc; .pdf, code
WP 5	D5.10 D 5.3.2 AI-powered services for Personalised early risk detection and risk assessment	.doc; .pdf, code
WP 5	D5.11 D5.4.2 Intelligent Connected Care Services and IoT	.doc; .pdf, code
WP 5	D5.12 D5.5.2 Design of authoring tool	.doc; .pdf, code
WP 5	D5.13 D5.6.2 Robotic assistance in community care: general framework, requirements and evaluation	.doc; .pdf, code
WP 5	D 5.14 D5.7.2 Technical validation report	.doc; .pdf
WP 6	Large scale pilots' datasets[11]	.csv[12]
WP 6	LSP scientific manuscripts	.doc; .pdf
WP 6	D6.1 Medical use cases specification and implementation guide	.doc; .pdf
WP 6	D6.2 Early detection and interventions operational planning	.doc; .pdf
WP 6	D6.3 GATEKEEPER Big Data and Data analytics strategies	.doc; .pdf
WP 6	D6.4 Clinical Study and CRF	.doc; .pdf
WP 6	D6.5 All Ethical approval package	.doc; .pdf
WP 6	D6.6 Report about the pilots' outcome	.doc; .pdf
WP 6	D6.7 D6.1.2 Medical use cases specification and implementation guide	.doc; .pdf
WP 6	D6.8 D6.1.3 Medical use cases specification and implementation guide	.doc; .pdf
WP6	D6.9 D2.2.2 Early detection and interventions operational planning	.doc; .pdf
WP 6	D6.10 D6.2.3 Early detection and interventions operational planning	.doc; .pdf
WP 6	D6.11 D6.3.2 GATEKEEPER Big Data and Data analytics strategies	.doc; .pdf
WP 6	D6.12 D6.3.3 GATEKEEPER Big Data and Data analytics strategies	.doc; .pdf
WP 6	D6.13 D6.6.2 Report about the pilots' outcome	.doc; .pdf
WP 6	D6.14 D6.4.2 Clinical Study and CRF	.doc; .pdf

WP 7	D7.1 Pilot Studies Use Case Definition and Key Performance Indicators (KPIs)	.doc; .pdf
WP 7	D7.2 Updated KPI Evolution Report (I to IX)	.doc; .pdf
WP 7	D7.3 New Use case demonstrations conclusion (I to IX)	.doc; .pdf
WP 7	D7.4 Pilot Studies Evaluation Results and sustainability plan	.doc; .pdf
WP 7	D7.5 D7.2.2 Updated KPI Evolution report (I to IX)	.doc; .pdf
WP 7	D7.6 D7.2.3 Updated KPI Evolution Report (I to IX)	.doc; .pdf
WP 7	D7.7 D7.2.4 Updated KPI Evolution Report (I to IX)	.doc; .pdf
WP 7	D7.8 D7.2.5 Updated KPI Evolution Report (I to IX)	.doc; .pdf
WP 7	D7.9 D7.2.6 Updated KPI Evolution Report (I to IX)	.doc; .pdf
WP 8	D8.1 Overview of relevant standards in smart living environments and gap analysis	.doc; .pdf
WP 8	D8.2 Initial standardization strategy	.doc; .pdf
WP 8	D8.3 -Certification scheme strategy and sustainability plan	.doc; .pdf
WP 8	D8.4 Standardization report and recommendations	.doc; .pdf
WP 8	D8.5 Initial Plan on the Overall Governance for Procurements	.doc;.pdf
WP 8	D8.6 Report on the overall governance for procurements	.doc; .pdf
WP 9	D9.1 Dissemination and communications plan	.doc; .pdf
WP 9	D9.3 Dissemination and communication activities and materials	.doc; .pdf
WP 9	D9.4 GATEKEEPER Socio-Economic assessment reports	.doc; .pdf
WP 9	D9.5 GATEKEEPER exploitation and sustainability	.doc; .pdf
WP 9	D9.6 D9.3.2 Dissemination and communications activities and materials	.doc; .pdf
WP 9	D9.7 D9.3.3 Dissemination and communications activities and materials	.doc; .pdf
WP 9	D9.8 D9.4.2 GATEKEEPER Socio-Economic assessment reports	.doc; .pdf
WP 9	D9.9 D9.4.3 Socio-Economic assessment reports	.doc; .pdf
WP 9	D9.10 D9.5.2 GATEKEEPER exploitation and sustainability	.doc; .pdf
WP 9	D9.11 D9.5.3 GATEKEEPER exploitation and sustainability	.doc; .pdf

WP 10	D10.1 HCT- Requirement No. 1	.doc; .pdf
WP 10	D10.2 H- Requirement No.2	.doc; .pdf
WP 10	D10.3 H-Requirement No. 3	.doc; .pdf
WP 10	D10.4 POPD-Requirement No.4	.doc; .pdf

3.5 Data Storage Management and Retention

The GATEKEEPER architecture, both on a project level and on dedicated pilot sites, has been designed to be privacy and security by design, ensuring that datasets generated and/or collected within the project remain protected throughout their lifecycle.

As such, pilots had already defined from the project's early stages a secure location for the storage of the data collected, as well as the technical and organisational measures that would be implemented, including, but not limited to anonymisation/pseudonymisation of the data, encryption and password protection, as well as access rights limitations according to the needs of each RUC.

Similarly, the pilots have defined the data retention periods in accordance with EU and national personal data protection requirements, taking into account, in particular, the purpose of the data collection and processing, the possibility to anonymise the data, as well as compliance with other legal obligations. Where data cannot be anonymised securely, the deletion of the data after the expiration of the retention period is envisioned.

In view of the above, the following table summarises where pilot data is stored within the GATEKEEPER project, as well as the predefined retention period, as was reported in the pilots' dedicated DMPs, DPIAs and/or Ethical Approval procedures.

Table 7 Pilot Data Storage and Retention Period

Pilot	Data Storage	Retention Period
Aragon	Any personal data has been pseudonymised and stored to the GATEKEEPER platform servers. Only SALUD authorised personnel will have access to personal data.	The retention period has been defined in accordance with the GDPR and national requirements.
Basque Country	Any personal data has been anonymised or pseudonymised data and stored at the headquarters of the project coordinator (Kronikgune).	Access to the anonymised and pseudonymised patient data will be maintained for 12 months after the end of the study. Access to evaluation results data that cannot lead to the identification of

		data subjects will be maintained for 10 years after the end of the study.
Cyprus	The data has been anonymised and stored in the HPE GATEKEEPER Infrastructure, protected by encryption at rest measures.	Data will be stored until the end of the study. After the end of the study, all personal information will be deleted.
Greece	Data has been pseudonymised or anonymised and safely kept in CERTH (Greece) and HPE (Italy) premises in accordance with the DPIA of the study.	The retention period has been defined in accordance with the GDPR and national requirements.
Poland	Any personal data has been pseudonymised and stored in a password - protected database at the coordinating researcher's office.	The retention period has been defined in accordance with the GDPR and national requirements.
Puglia	Dedicated cloud cluster with access controls in place.	15 years
Saxony	Any personal data has been pseudonymised and stored at a separate GK private space at the HPE platform, protected by encryption at rest measures. Any sharing of the data with the GATEKEEPER Consortium was performed only after the anonymisation of the data.	Personal data will be retained for 5 years after the completion of the project. After that, the data will be deleted. For non-personal data, they will be retained for at least 10 years and will be then deleted.
UK:		Personal data collected will be retained for the duration of the project and will then be deleted.
Milton Keynes	Data about the feasibility study and robot are hosted in the Open University servers, while data concerning the impact of the devices has been collected and processed by Samsung and provided in an anonymous aggregated form to the	

	University of Warwick for the pilot evaluation.	
Bangor	Personal data has been pseudonymised and stored in the HPE platform.	

Where medical devices were provided to patients, pilots had already determined the procedure that participants would have to follow to return the device to the respective pilot site and delete the relevant data stored in the devices, and had informed participants accordingly.

4 FAIR Principles

Within the GATEKEEPER project, the solutions developed have been designed to be as open as possible, while preserving the privacy of any personal data involved and the confidentiality of relevant information. This includes not only the data that has been generated, collected and/or processed within the project, but also the solutions that have been developed in the context of the project.

Taking the above into consideration, the GATEKEEPER platform was envisioned from the beginning to be based on the Web of Things (WoT) architecture and framed into a layered structure composed of the following layers:

1) Access;
2) Certify;
3) Find;
4) Share;
5) Compose.

Each of the above layers, as was in-depth reported in Deliverable D3.6, is intrinsically linked to the FAIR (Findability, Accessibility, Interoperability, Reusability) principles, as follows:

Layer 1: Access, provide Accessibility of FAIR principle: This layer is responsible for turning any Thing into a Web Thing that can be interacted with using HTTP requests just like any other resource on the Web. In GATEKEEPER this layer is provided by the Things Management System (TMS), one of the core components dedicated to the implementation of the functionalities associated with accessing and finding layers of WoT architecture. The Thing Description provided within the TMS enables:

(i) The management of multiple Things by a cloud service,

(ii) The simulation of devices/Things that have not yet been developed,

(iii) Common applications across devices from different manufacturers that share a common Thing model, and

(iv) Combining multiple models into a Thing.

Layer 2: Certify, improve the FAIR principles with Trustability: This layer is specific to the GATEKEEPER platform, with respect to the Web of Things layered reference architecture, dedicated at fostering trust in the GATEKEEPER platform through certification, as well as establishing a way to securely share data across services. Within the GATEKEEPER architecture the certify layer is enabled through the interaction between the TMS and the GATEKEEPER Trust Authority (GTA), which provides the Certify layer of the GATEKEEPER architecture, while the GATEKEEPER Marketplace service will be in charge of sharing the GATEKEEPER Things Things that have been certified by the GTA). The Trust creation is managed using Blockchain technology with the aim of having a decentralized trust system, allowing participants to verify data correctness and ensure its immutability. Things

then use Blockchains to register themselves and organise, store, and share streams of data effectively and reliably.

Layer 3: Find, provide Findability of FAIR principle: This layer is dedicated at providing ways for easy discovery and consumption of Things. In GATEKEEPER it will be implemented through a Marketplace that will provide Things offered through the *consumer space*, the *healthcare space* and the *business spac, e*ach oriented to a different type of market user. These core features are supported by the Networked Things architecture that provides the reference model in home and health-oriented devices forming the GATEKEEPER Platform's Business Space. The ecosystem is split into clear boundaries around 3 spaces, Business-to-Government (B2G), Business-to-Consumer (B2C) and Business-to-Business (B2B).

Layer 4: Share, provide Interoperability of FAIR principle: This layer provides the explanation of what a Thing is, what data or services it offers, and so on. Through these functionalities a Thing is not only easily used by other HTTP clients but is also findable and automatically usable by other WoT applications. Reusing Web semantic web standards to describe Things and their services enables searching for Things through search engines and other Web indexes, while allowing the automatic generation of user interfaces or tools to interact with Things. At this level, technologies such as JSON-LD (a language for semantically annotating JSON) are in use. In GATEKEEPER, all the Things will use as communication language the Web of Things standard with JSON-LD contexts, including FHIR standard and SAREF ontology.

Layer 5: Compose, provide Reusability of FAIR principle: This layer provides the integration of data and services from heterogeneous Things into an immense ecosystem of tools such as analytics software, mash-up platforms and developer platforms. Within GATEKEEPER the compose layer provides all the intelligent services for early detection and intervention and a developer platform where developers can compose GATEKEEPER Things in order to provide advanced services.

These early detection, prediction and proactive services for healthcare were validated in the pilot sites in order to populate the Consumer and Healthcare spaces within the GATEKEEPER Marketplace where these services are envisioned to be available as Things to third party users in order to compose more advanced services and share valuable knowledge.

4.1 Findability

As already explained, the GATEKEEPER ecosystem envisions the further dissemination of knowledge acquired and the tools developed beyond the project's duration through the creation of a Marketplace, which is fully aligned with the FAIR principles.

In particular, with regard to the Findability principle, all datasets made available through the Marketplace, where that is feasible in an anonymised format and without jeopardising data subjects' privacy, will be accompanied by relevant metadata, as was reported in Deliverable D4.14 and demonstrated below.

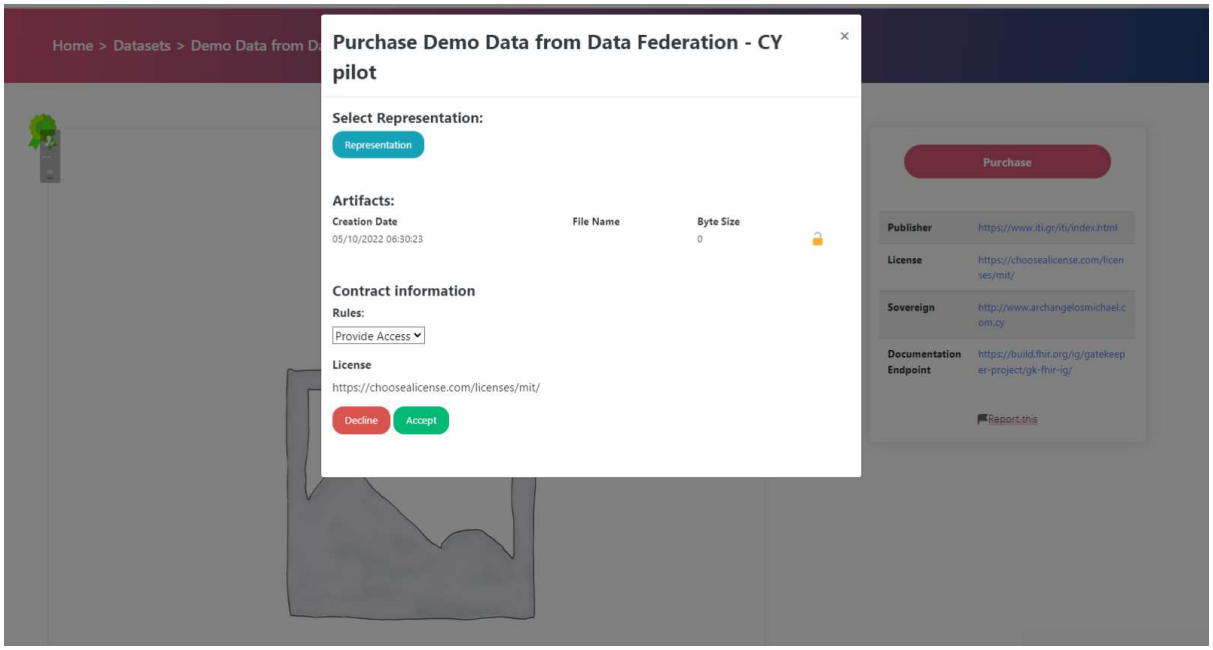


Figure 3 - Metadata available to the Consumer before making an agreement to receive data

In order to ensure homogeneity, a table explaining the attributes that need to be included in the data models in order to enable its findability by assisting the filtering of the relevant searches has been designed and reported in Deliverable D4.6, as follows:

Table 8: Attributes for discovery in data model

Attribute	Allowed Values	Applies to
Category	• App • Device • Thing • Dataset	All
Domain	• Health • Energy • Transport • Education • …	All
Disease / Medical condition	<type>	All
Medical Use Case/ Purpose	<type>	All
Developer	<username>	All
City, Country	<text>	All
Pilot	Greece, …	All
Price	<any number>	All
Audience	One or more from: Consumer, Business, Healthcare	All

Platform	…	Applications
Tools	Docker, NPM, Webpack, Gulp, Composer	Applications
Programming Language	….	APIs
Measurements	BP, HR, …	Device
Device type	Wearable, robot, ambient, …	Device
Size	Number (in MB)	Datasets
Licenses	CC-by-4.0, …	Datasets
Provider	<name>	Datasets
Data formats	CSV, XML, TXT, …	Datasets

In addition to the above, each pilot has adopted case-specific measures in order to ensure the findability of any datasets made available, which wil be detailed below.

4.1.1 UK Pilot

Outline the discoverability of data (metadata provision)
The definition of the metadata follows the OU guidelines for research data http://www.open.ac.uk/library-research-support/sites/www.open.ac.uk.library-research-support/files/files/RDM-Guidelines-for-creating-readme-style-metadata.pdf.
Specifically, records are described by:

- the date and location of the data collection;
- the person responsible for the data collection;
- the identifier of the data subject;
- the phase of data collection (entry or exit);

Datasets are described by features included in the ORDO repository and:

- Location and period of the data collection;
- The phase of the data collection (start or end);
- Version and last date of the change.

Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
Datasets are assigned a DOI and a standard description, as defined by the OU repository for research data (ORDO).

Outline naming conventions used
Metadata are described in the readme file attached to the datasets. The naming of data property is self-descriptive, e.g. LOCATION_OF_COLLECTION, DATE_OF_COLLECTION.
Name of datasets will include reference to the batch, phase, location and date of the data collection.

Outline the approach towards search keyword

Data is documented. The documentation and metadata is included in the project repository in ORDO. The documentation includes reference to used guidelines and formats concerning the data features and scales. Data will be mapped to existing ontologies to support their interoperability and reuse, and then made available as linked data on open.data.ac.uk

Outline the approach for clear versioning
Versioning is managed by ORDO and the OU cloud (OneDrive). These systems provide a versioning system including changelogs, date and person.
Versioning of source code is managed through Git.

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how
We refer to the DDI https://ddialliance.org/Specification/DDI-Lifecycle/3.2/#3.2schema.

4.1.2 Spain Pilot – Aragón

Specific actions for making data findable are not foreseen. At present, SALUD is undergoing a process of transformation of its Electronic Healthcare Record into a patient centred and modular concept in which the codification of information and use of structured data is of utmost importance.

4.1.3 Basque Country Pilot

Discoverability of data
Osakidetza has the Electronic Health Record with patients´ demographic and clinical information. This information, together with the metadata, can be extracted by the Osakidetza´s Oracle Business Intelligence tool. Extracted data comes from structured data, and it is codified, pseudonymized/anonymized and individualized/aggregated. The datasets that are generated in order to be used under research conditions have to be approved by the Euskadi Ethics Committee and in accordance with the European and local legal framework.
All the data are under standard description and with the international standardized homogenization.

Conventions used
All the datasets generated are in accordance with the needs of each study protocol.

Approach towards search keyword and for clear versioning
The approach depends on the needs of each study protocol and the strategy has to be developed in each study. Further information will be provided once the variables of each use case are defined

Standards for metadata creation
International standards.

4.1.4 Cyprus Pilot

What documentation and metadata will accompany the data?
Any information that is reported in the research papers in the write-up process with not be person-identifiable. If data will be shared on the plan of Open Research Data will also be no person-identifiable.

Metadata creation
As metadata, we will thus provide:

- Publication date;
- Title;
- Authors including contact information;
- Description;
- Version;
- Language;
- Keywords;
- Grant acknowledgment, and
- References to all publications referring to the dataset.

Discoverability
All data will be uploaded together with the relating metadata, including project context and lab book entries. These collections will be linked to scientific articles, conference proceedings, reports, and other sources to be published. For this, we will make use of persistent and unique Digital Object Identifiers (DOI) via the data storage facility. A description of available data collections will also be added to the Cyprus partners' websites.

Naming conventions, keywords, and versioning
We ascertain that the data will be easily recognized and correlated to experiments via the following naming conventions:

Raw data: YYMMDD_[experiment]_[technique]_XXX.*
Processed results: YYMMDD_[experiment]_[technique]_XXX_analysis_ZZZ.*
Herein, symbols represent the following:

- YYMMDD - the inverted date of the day the experiment was conducted [experiment] - a short title for the experimental series [technique] - a unique denominator for each technique
- XXX - a running number for individual measurements
- ZZZ - a running number for separate processes of analysis

The same naming formats will be used for other data. Other documentation will include the methodology and analytical procedure.

The data, metadata, and documentation are compliant to disciplinary standards, open file formats, and use controlled vocabularies and the standard metadata schema for easy interoperability and re-use.

4.1.5 Saxony Pilot

The Terminology used will be in line with H2020 projects and standard medical/psychological terminology.

4.1.6 Greece Pilot

Outline the discoverability of data (metadata provision)
The definition of the metadata follows the OU guidelines for research data http://www.open.ac.uk/library-research-support/sites/www.open.ac.uk.library-research-support/files/files/RDM-Guidelines-for-creating-readme-style-metadata.pdf.
Specifically, records are described by:

- the date and location of the data collection;
- the person responsible for the data collection;
- the identifier of the data subject;
- the phase of data collection (entry or exit);

Datasets are described by features included in the HPE repository and:

- Location and period of the data collection;
- The phase of the data collection (start or end);
- Version and last date of the change.

Outline the identifiability of data and refer to the standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
Datasets are assigned a DOI and standard description.

Outline naming conventions used
Metadata are described in the readme file attached to the datasets. The naming of data property is self-descriptive, e.g. LOCATION_OF_COLLECTION, DATE_OF_COLLECTION.
The name of datasets will include a reference to the batch, phase, location, and date of the data collection.

Outline the approach towards search keyword
Data is documented. The documentation and metadata are included in the project repository in HPE. The documentation includes reference to used guidelines and formats concerning the data features and scales.

Outline the approach for clear versioning
Versioning is managed by CERTH and the HPE infrastructure. These systems provide a versioning system including changelogs, date, and person.

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how
We refer to the DDI https://ddialliance.org/Specification/DDI-Lifecycle/3.2/#3.2schema

4.1.7 Poland Pilot

The Pilot has taken into consideration the guidelines provided in the first version of the DMP and has aligned its activities accordingly.

4.1.8 Puglia Pilot

Outline the discoverability of data (metadata provision)
Metadata for datasets that will be openly shared (see subsection below) will be defined according to policies and standards decided at the overall GATEKEEPER project level, when they will be available.

Identifiability of data and standard identification mechanisms
Mechanisms to assign persistent identifiers to datasets that will be openly shared (see subsection below) will be defined according to policies and standards decided at the overall GATEKEEPER project level, when they will be available.

Naming conventions
Naming conventions for datasets that will be openly shared (see subsection below) will be defined according to policies and standards decided at the overall GATEKEEPER project level, when they will be available.

Search keyword
Search keywords for datasets that will be openly shared (see subsection below) will be defined according to policies and standards decided at the overall GATEKEEPER project level, when they will be available.

Versioning
Versioning for datasets that will be openly shared (see subsection below) will be defined according to policies and standards decided at the overall GATEKEEPER project level, when they will be available.

Standards for metadata creation
Metadata creation standards for datasets that will be openly shared (see subsection below) will be defined according to policies and standards decided at the overall GATEKEEPER project level, when they will be available.

4.2 Accessibility

Intrinsically tied to the above, accessibility focuses on ensuring that metadata are indeed retrievable and accessible using a standardised communication protocol, which is open, free and universally implementable, allowing, where required, for the implementation of authentication and/or authorisation procedures. As such, more details regarding the pilots' possibilities to make datasets openly accessible and the measures towards that end will be provided below.

4.2.1 UK Pilot

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

Data concerning pilot participants will be not made available. These data include personal and sensitive information that pose risks for the pilot participants.
Aggregated data will be shared within the project and to the public at the end of the project.

Specify how the data will be made available
Data will be made available through ORDO repository (the OU repository for research data).

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
Data will be in .csv format and data description in .txt format.

Specify where the data and associated metadata, documentation and code are deposited
Data, metadata and documentation will be stored in ORDO repository. Source code will be uploaded in Git-Hub.
Working data and documentation will be stored in the OU cloud (OneDrive).

Specify how access will be provided in case there are any restrictions
Personal and sensitive data will be made available only to the project team through a specific OU repository for sensitive data. Access will be managed and monitored by project team.

4.2.2 Spain Pilot – Aragòn

Partial accessibility of data will be based on the specific purpose of this potential access. Anonymized data might be made accessible for building predictive models if a solid hypothesis and a scientific plan are well justified and considered as relevant for the organisation. The same rule applies for pseudo-anonymized data. If some external entity would like to have access to pseudo-anonymized data in the context of GATEKEEPER they must justify not only the purpose but also the necessity of pseudo-anonymizing data.

4.2.3 Basque Country Pilot

Data protection framework regarding data collection, storage, access, protection and sharing will be ensured. This framework will follow the procedures indicated in the new general data protection regulation No 2016/679 that has been recently applied in Europe and will seek local Ethics Committees approval ensuring data access, process and storage.
Then, no data related to personal information will be transferred, only anonymized or pseudonymized data will be transferred (depending on the use case). This information will not be openly accessible.

It will be agreed with the consortium how long the data will be stored for this study, what data can be archived and what safeguards will be setup.

Once the results of the study are validated, they will be disseminated in scientific and social forums and the databases can be placed with all the legal requirements in open repositories.

How the data will be made available
Data related to personal data will not be openly accessible due to data protection implications.

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
There are no methods or software tools needed, because no data related to personal data will be shared.

Specify where the data and associated metadata, documentation and code are deposited
The databases that contains information related to patient diagnoses, prescriptions, and laboratory test results, together with referrals and use of resources (visits, hospital admissions, emergency department visits etc) are supported by the EHR.

Specify how access will be provided in case there are any restrictions
During the project, it is required to keep data safe and secure. Data security is needed to prevent unauthorized access. Otherwise the data could be disclosed, changed or deleted. Personal data is available by the EHR-Osabide Global-only to the professionals enabling accessing and collecting all relevant data concerning each patient to guide in the decision-making.

Access to data will be analyzed in each use case based on the preferences and requirements of the Basque Health Service.

4.2.4 Cyprus Pilot

How will you share the data?
All data will be made available. However, there will be different access levels. Anonymized data will be made openly available. Sensitive data will not be publicly available, according to data protection law. Access can be granted onsite at the repository (visiting scientist) or - with sufficient clearance - through controlled remote data processing.
Anonymised interview data will be shared through academic publications and conference presentations. Data will only be shared through dissemination activities and will not be shared in a raw form with other parties. The dataset does not have significant long-term value and therefore will not be held for longer than 4 years.

Are any restrictions on data sharing required?
The raw data will be kept for the project duration and destroyed thereafter. The existence, range, and nature of the project's original data will be publicised via

references in published outputs by including relevant dataset DOIs, as well as via conference presentations and materials produced during the project.

4.2.5 Saxony Pilot

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

Raw data concerning pilot participants will not be made available. These data include personal and sensitive information that pose risks for the pilot participants. Aggregated data will be shared within the project and to the public at the end of the project (e.g. summaries).

Specify how the data will be made available

Aggregated data will be made available through summaries and reports will be shared via various media and community channels, including: mailing lists, the GATEKEEPER-project website, other relevant online platforms. Analyzed data will also be used for academic publications. Data will be pseudonymized for all publication reasons.

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

It is planned to use redcap data software. If possible, data from questionnaires will be transformed in .fhir format. Other data formats need to be further discussed with Samsung.

Specify where the data and associated metadata, documentation and code are deposited

The raw data collected by TUD (questionnaires) will be saved in the network of the Faculty of Medicine at the Technische Universität Dresden using individually defined access authorizations. Data are stored in pseudonymized form. Access to the pseudonymized data is restricted to the scientists of the research group "GATEKEEPER". Device-, sensors-related data and speech recognition-related data will be saved by Samsung. An appropriate data security concept needs to be worked out with Samsung.

Specify how access will be provided in case there are any restrictions

Access to the pseudonymized data is restricted to the scientists within the pilot. Selective data can be shared with relevant partners through secured data management within GATEKEEPER.

4.2.6 Greece Pilot

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

Data concerning pilot participants will be not made available. These data include personal and sensitive information that poses risks for the pilot participants. Aggregated data will be shared within the project and to the public at the end of the project.

Specify how the data will be made available
Data will be made available through the GK Marketplace as a Thing.

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
Data will be in various formats such as FHIR/xml, .csv format and data description in .txt format.

Specify where the data and associated metadata, documentation and code are deposited
Data will be stored in the HPE infrastructure. Metadata and documentation will be stored in the GK Marketplace repository. The digital application will be also exposed as a thing to the GK Marketplace.

Specify how access will be provided in case there are any restrictions
Personal and sensitive data will be made available only to the project team through a specific HPE repository for sensitive data. Access will be managed and monitored by the project team.

4.2.7 Poland Pilot

The Pilot has taken into consideration the guidelines provided in the first version of the DMP and has aligned its activities accordingly.

4.2.8 Puglia Pilot

Openly available data

As a baseline, it must be considered that, as previously illustrated, the primary objective for collecting data in the GATEKEEPER Puglia Pilot experiment is to improve care for elderly citizens in the Puglia Region, and allow them active and health aging, including by providing cost-effectiveness information to healthcare administrators.

Sharing data with the research community is a secondary objective, that can improve the prospects of achieving the primary objective in the future.

Based on such premises, and in consideration of the need to fully meet applicable privacy protection and ethics regulation, the criteria to be used for deciding which datasets will be made openly available are as follows (see also sections below):

- Datasets should be necessary and useful to improve research on areas linked to active and healthy aging
- Dataset should fully respect privacy protection and ethics limitations mentioned in relevant sections below
- Datasets should be fully anonymizable before sharing, without losing their value as per first bullet

Data that will not respect such criteria will not be shared.

Relevant determinations will be made during the course of the Pilot experiment, in coordination with other project Pilots and with the overall GATEKEEPER management.

It has to be noted that - because of the involvement of administrative, law-regulated approval cycles that are outside the scope of the GATEKEEPER project and that cannot be fully determined in the frame of the project itself - in principle, data directed to healthcare administrators will be shared only with administrators of the Puglia Regional healthcare system. It will be up to such healthcare administrators to decide about possible sharing of such data with other actors, outside the scope of this Data Management Plan.

How data will be made available

Data to be openly shared will be made available through relevant open-access repositories, such as Zenodo.

Relevant determinations will be made during the course of the Pilot experiment, in coordination with other project Pilots and with the overall GATEKEEPER management.

Methods and software tools needed to access the data

Data to be openly shared will be made available through standard formats as recommended by the GATEKEEPER project, and will thus be accessible through correspondingly standard methods and software tools.

Data and associated metadata, documentation and code

Data to be openly shared, as well as and associated metadata and documentation, will be deposited in relevant open-access repositories, such as Zenodo.

Relevant determinations will be made during the course of the Pilot experiment, in coordination with other project Pilots and with the overall GATEKEEPER management.

No code is expected to be shared by the GATEKEEPER Puglia Pilot experiment.

Restriction to access

Restriction to access will be applied, in particular, in order to address relevant ethics issues (see section below). Restrictions to usage will be specified and managed through the formulation of relevant, legally binding "terms of use".

4.3 Interoperability

As has already been reported throughout the project, including in Deliverable D3.3.2, Interoperability of both datasets and developed solutions within the project is a focal point of the activities designed. The GATEKEEPER platform builds on top of the Web of Things to create that uniform framework. In GATEKEEPER, all resources are represented with a Thing Description that leverages JSON-LD and ontologies to expose resource semantics in a machine-interpretable way.

As such, interoperability within the GATEKEEPER platform has been pursued on the following levels:

a) Technical: Ensuring that the infrastructure links without issues the systems and services developed within the project through common secure communication protocols.

b) Syntactic: Permitting the communication and exchange of data between each system contained in the project.

c) Semantic: Enabling the "comprehension" of the data exchanged between the systems.

d) Organisational: Aligning organisational practices among the various project partners to ensure the homogeneity of the solutions developed.

e) Cross-Layer technologies: Using Web of Things, HL7-FIR and Data federation/Data space connectors to enable the interoperability of solutions.

In addition to the above, the following sections provide more information on the pilots' practices in order to achieve interoperability within the project.

4.3.1 UK Pilot

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

The project team will follow the OU guidelines for describing data http://www.open.ac.uk/library-research-support/research-data-management/describing-data

Concerning the data repository, we will adopt the description of data repository used in ORDO while, concerning the data we consider as reference the DDI schema https://ddialliance.org/Specification/

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

We will use standard metadata vocabulary, adopted in ORDO repository. Concerning data features, we cannot at this stage commit to a specific ontology. We will provide a mapping as soon the questionnaires for the data collection are defined.

4.3.2 Spain Pilot – Aragòn

SALUD technical infrastructure works with the HL7 standard so as to allow the interoperability among the different modules that make up the EHR. The specific modules also use standards in order to encode information as LOINC or DICOM. Companies providing KETs in the context of the Aragón pilot will be requested to use also standard formats.

4.3.3 Basque Country Pilot

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

Diagnostics: ICD-9 and ICD-10;

Lab results: Local code RICs for Specialized Care and DBP for Primary care;

Pathological Anatomy: SnomedCT;

Medical images: DICOM (Digital Imaging and Communications in Medicine);

Medications: ONPP (Official Nomenclature of the Pharmaceutical provision of the National Health System in Spain). International ATC in HL7-CDA in the Summary Clinical Record Report;

Allergies: local code;

Vital signs: DBP (Primary Care); RIC (Specialized Care);

Encounter visits: local code;

Patient Reported Outcome Measurements: Local code RICs for Specialized Care.

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

Standard vocabulary for data will be used. For the non-standardized data a dictionary will be needed in order to provide a mapping to the local vocabulary.

4.3.4 Cyprus Pilot

Regarding data interoperability, standard data vocabulary will be used

4.3.5 Saxony Pilot

For data interoperability keywords will be used common within ageing and technology studies and innovations.

4.3.6 Greece Pilot

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards, or methodologies you will follow to facilitate interoperability.

The data and metadata will be stored in HL7/FHIR formats. For data exchange, established interoperability standards such as XDS, CDA, etc. will be explored.

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

The pilot will use the following standard vocabularies:

Socio-demographics, Clinical data, sensor data and patient-reported: FHIR resources;

Medication: ATC codes;

Medical record: SNOMED, ICD-10.

4.3.7 Poland Pilot

The Pilot has taken into consideration the guidelines provided in the first version of the DMP and has aligned its activities accordingly.

4.3.8 Puglia Pilot

Data interoperability

Metadata vocabularies, standards and methodologies to be followed for facilitating interoperability of data that will be openly shared will be defined according to policies and standards decided at the overall GATEKEEPER project level, when they will be available.

Standard vocabulary

The usage of vocabularies for specific datasets, as well as the possibility and opportunity to link such vocabularies with interdisciplinary ontologies, will be defined according to policies and standards decided at the overall GATEKEEPER project level, when they will be available.

4.4 Reusability

Reusability of the datasets is the ultimate goal throught the establishment of the GATEKEEPER Marketplace. As such, it has been of utmost importance to pilots to ensure that data can be safely and securely be shared with other parties aiming at performing research on the conditions/diseases central to the project, without compromising privacy. The following information per pilot describes how the pilots intend to make their datasets reusable.

4.4.1 UK Pilot

Specify how the data will be licenced to permit the widest reuse possible

Data will be released in open access. The licensing will be defined with the management of the large-scale pilot.

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

Data will made available (aggregated data) as soon as evaluation of the UK Pilot is completed and this has been cleared by the management of the large-scale pilot. This extra step will assure a last quality check of data, considering the overall quality of data collected across the GATEKEEPER pilots.

Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

The collected data include personal and sensitive data required to assess the personal state of pilot participants. This data is necessary for the evaluation of the UK pilot and of the GATEKEEPER large-scale pilot. The value of collected data is strictly related to the evaluation of the technology-driven interventions. The reuse of this data is on the one hand limited to the specific type of intervention and, on the other hand, to the risks for the data subjects. Aggregated data will be made

available at pilot scale and at European scale, but data concerning individual participants will not be shared for reuse but destroyed.

The value of the collected data in the evaluation of the GATEKEEPER large-scale pilot. Data will be made available for re-use at European scale, to support further studies and the replicability of interventions in other countries.

Describe data quality assurance processes

The project team will follow the OU guidelines on quality of research data http://www.open.ac.uk/library-research-support/research-data-management/data-quality

Data will be collected through questionnaires in one-to-one sessions supported by a trained project team member or external researcher. Each batch (20-30 questionnaires) will be controlled by the project team.

Data will be collected through digital web-based supports. Data will be stored in ORDO, this will provide a versioning system and a standard set of metadata. Furthermore, the research team will follow the OU guidelines concerning naming and organising research data http://www.open.ac.uk/library-research-support/research-data-management/organising-your-files

The data collected, scale and formats, will be defined in collaboration with the large scale management and with a project partner with expertise in data-driven evaluation. The engagement with the large-scale pilot management will ensure that the quality of data is consistent with the rest of the project pilots and that good practices will be shared across the project.

Specify the length of time for which the data will remain re-usable

Aggregated data concerning the pilot evaluation will be stored and made available for at least 10 years in the ORDO repository.

Personal and sensitive data will be destroyed at the end of the project or right after the data analysis.

4.4.2 Spain Pilot – Aragón

Data is not intended to be made re-usable by default. Only specific requests based on scientific hypothesis may pave the way to a potential re-usability of the data generated in GATEKEEPER.

4.4.3 Basque Country Pilot

Specify how the data will be licensed to permit the widest reuse possible

When possible, the data set will be licensed under an Open Access license. However, this will depend on the level of privacy, and the Intellectual Property Right (IPR) involved in the data set. Datasets to be available will be decided by owners/ partners of them.

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

Data will be accessible considering the legal, contractual or ethical issues of the Basque Country PS. Data classified as confidential will as default not be reusable due to privacy concerns.

Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

The Basque Country PS aims to make as much data as possible re-usable for the third parties. Restriction will only apply when privacy, IPR or other exploitation ground are in play. However, the use by third parties have to be agreed beforehand.

Describe data quality assurance processes

A quality control is needed at the local level in the collection process. During data collection, the local data manager is the main responsible for quality control, who must ensure that the data reflect the actual facts, responses, observations and events, and that the current regulation is respected.

Throughout the data collection process, several measures will be undertaken to ensure a high quality and standardised data set. Two main measures will facilitate a rigorous and uniform data collection: a) a unique codebook for the five pilot sites and b) local field guides for collecting the data.

The codebook (to be agreed with the GATEKEEPER Evaluation team) will be based on the operationalized list of indicators. It provides the structure and layout of the data files and will include:

definitions of indicators;

source of information;

response codes for each indicator;

measures to indicate non-response and missing data;

Specify the length of time for which the data will remain re-usable

The decision about long-term provision will be taken as the data are stored following the directives of Osakidetza regarding research studies. Data will be probably stored at the Osakidetza's servers, and will be kept maintained, at least, for 5 years after the end of the project (with a possibility of further prolongation for extra years).

4.4.4 Cyprus Pilot

Given the inclusion of sensitive personal data, reusability will be minimised and only anonymised data will be shared within the project.

4.4.5 Saxony Pilot

Specify how the data will be licensed to permit the widest reuse possible

The data will remain available for future analysis after the end of the project term. Data will remain available in a secured archive for at least 10 years, in accordance with the relevant regulations

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

In case of re-use, the regulations of the consortium contract will be followed. Permission should be given by the local data management, and the GATEKEEPER project managers and coordinators.

Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

The collected data include personal and sensitive data required to assess the personal state of pilot participants. This data is necessary for analysis of the Saxony Pilot and of the GATEKEEPER large-scale pilot. The reuse of this data is on the one hand limited to the specific type of intervention and, on the other hand, to the risks for the data subjects. Aggregated data will be made available at pilot scale and at European scale, but data concerning individual participants is not intended to be reused.

Describe data quality assurance processes

Data will be collected using academic based methods. Data collected will be thoroughly analysed by members of the team. The engagement with the large-scale pilot management will ensure that the quality of data is consistent with the rest of the project pilots and that good practices will be shared across the project.

Specify the length of time for which the data will remain re-usable

The explicitly derived, and pseudonymized data sets and calculation bases are stored on an access-restricted, clinic-internal server area for at least 10 years.

4.4.6 Greece Pilot

Specify how the data will be licensed to permit the widest reuse possible

Data will be released in open access. The data will be extracted as an HL7/FHIR repository and registered in the GATEKEEPER Marketplace. The licensing will be defined with the management of the large-scale pilot.

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

Data will be made available (aggregated data) as soon as the evaluation of the GR Pilot is completed and this has been cleared by the management of the large-scale pilot. This extra step will assure a last quality check of data, considering the overall quality of data collected across the GATEKEEPER pilots.

Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

The collected data include personal and sensitive data required to assess the personal state of pilot participants. This data is necessary for the evaluation of the GR pilot and of the GATEKEEPER large-scale pilot. The value of the collected data is strictly related to the evaluation of technology-driven interventions. The reuse of this data is on the one hand limited to the specific type of intervention and, on

the other hand, to the risks for the data subjects. Aggregated data will be made available at the pilot scale and at the European scale, but data concerning individual participants will not be shared for reuse but destroyed.

The value of the collected data in the evaluation of the GATEKEEPER large-scale pilot. Data will be made available for re-use at the European scale, to support further studies and the replicability of interventions in other countries.

Describe data quality assurance processes

The project team will follow the OU guidelines on quality of research data http://www.open.ac.uk/library-research-support/research-data-management/data-quality

Data will be collected through the digital solutions supported by a trained project team member or external researcher. Data will be stored in CERTH secure space and then in HPE infrastructure (as soon as this is ready), this will provide a versioning system and a standard set of metadata. Furthermore, the research team will follow the OU guidelines concerning naming and organizing research data http://www.open.ac.uk/library-research-support/research-data-management/organising-your-files

The data collected, scale, and formats will be defined in collaboration with the large scale management and with a project partner with expertise in data-driven evaluation. The engagement with the large-scale pilot management will ensure that the quality of data is consistent with the rest of the project pilots and that good practices will be shared across the project.

Specify the length of time for which the data will remain re-usable

Aggregated data concerning the pilot evaluation will be stored and made available for at least 10 years in the GATEKEEPER Marketplace. Personal and sensitive data will be destroyed at the end of the project or right after the data analysis.

4.4.7 Poland Pilot

The Pilot has taken into consideration the guidelines provided in the first version of the DMP and has aligned its activities accordingly.

4.4.8 Puglia Pilot

Data licensing for reuse

In general, reuse of data will be permitted to researchers for the purpose of furthering scientific research on active and healthy aging, as mentioned previously.

As previously mentioned, data licensing provisions - which have not yet been defined at the time of this writing - will be specified in relevant, legally binding "terms of use". These will include consideration of IPRs linked to the "copyrightable layer" of the shared datasets (i.e. database schemata, ontologies developed in GATEKEEPER, other proprietary metadata), that may be owned by relevant GATEKEEPER partners or other third parties.

Timing for re-use

In general, datasets that will be openly shared, will be made available as soon as they will be available (after accounting the time for their anonymization and preparation). Normally, this means after the completion of the Puglia Pilot experiment, i.e. in the final phases of the GATEKEEPER project.

Embargo periods will be considered, in case they are needed to allow partners of the Puglia Pilot team - possibly in cooperation and coordination with other GATEKEEPER partners, according to project's internal agreements - to develop relevant scientific publications based on the shared data.

Usability by third parties

As illustrated before, during the project data from the Puglia Pilot will be used by partners involved in the Pilot team. After the end of the project, third parties will be allowed to reuse data if they are research institutions conducting research in the area of active and healthy aging. Restriction to this category is linked to ethics issue (see section below).

Data quality assurance processes

Data quality assurance processes will be defined according to policies and standards decided at the overall GATEKEEPER project level, when they will be available.

Time limits for re-usability

In general, specific time limits for re-usability are not established, apart from the overall consideration of the decreasing value of datasets as time passes. Although a definite determination on this issue has yet to be made, at the time of this writing a period of 10 years before data are deleted due to loss of value, seems reasonable.

5 Intellectual Property Rights

The Intellectual Property Rights (IPR) Strategy has been designed in a manner that promotes Open Science while protecting the Consortium's and third-parties' rights with regard to IPR-protected work. As per the Consortium Agreement, the results of the experiments performed within GATEKEEPER are owned by the beneficiaries or third parties that generate them. Similarly, where two or more partners have contributed to the generation of results, they are joint owners and commonly decide the terms of protection.

In order to ensure a common understanding of the IPR in question, it is worth noting that such IPR involved may include the following:

- **Copyright**, the right of authors/creators over their work, including software, code and publications;
- **Patents**, a set of exclusive rights granted to inventors in exchange for a public disclosure of the invention;
- **Trademarks**, particularly useful for the differentiation and recognition of products or services;
- **Database rights**, a sui generis right protecting the content of a database in view of the substantial investment required for its creation;
- **Trade Secrets,** meant to protect confidential information of an organisation. Within the GATEKEEPER project, trade secrets have been maintained to a minimum, in order to promote the dissemination of knowledge and to further innovative actions beyond the project.

5.1 Ownership of Background Information and Results

During the signing of the Consortium Agreement, all partners reported on the background that would be included in the project. Such background remains under the ownership of the partner contributing it, providing access to the other partners under certain conditions. In particular, such access rights were:

a. Free of any administrative transfer costsl
b. Granted on a non-exclusive basis;
c. Used only for the purposes defined in the Consortium Agreement;
d. Sub-licensed to affiliated entities, as long as they worked in the implementation of the project.

Similarly, access rights to results are:

a. Granted on fair and reasonable conditions;
b. Granted on a royalty-free basis for internal research activities;
c. Granted to third parties only under the agreement of the partners.

A request for Access Rights may be made up to twelve months after the end of the Project.

5.2 Intellectual Property Rights within GATEKEEPER

Given the project's multifaceted activity, encompassing all aspects of digital healthcare, including healthcare professionals, businesses and citizens, the IPR protection of the solutions developed is of utmost importance. At the same time, and in order to ensure the continuity of the solutions developed, alignment with the task responsible for exploitation has been a crucial part of the IPR strategy.

The following table describes the components that were developed in the context of GATEKEEPER, as well as the exploitation strategy that was defined in Deliverable D9.10.

Table 6 - Component list overview

Component Name	Related Task	Exploitation strategy (Open source, License, Reference, Undecided)
Things Management System	4.2	Open source
Things Directory	4.2	Open source (that could scale with some adaptation on other platform that want to be compliant with Web of Things. An exploitation strategy of this component could be a commercial product based on this open source that may include support and customization depending on the client needs)
GATEKEEPER Advanced Big Data services	4.3	License (even if all the containers use open-source software and components, the infrastructure is based on a licensed platform)
GK-Integration Engine	4.4	Reference
Data federation and Integration	4.4	Reference
RDF Semantic Data Lake	4.4	Open source / Reference
Trust Authority	4.5	License / Reference
GATEKEEPER Marketplace Services	4.6	License
Advancing and personalizing the analytic of Home Activity Monitoring and Health Activity Monitoring	5.2	License (Tiered approach based on business type)
AI-powered services for Personalized	5.3	License (Tiered approach based on business type)

early risk detection and risk assessment		
Intelligent Connected Care Services and IoT	5.4	License (Tiered approach based on business type)
Design of authoring tool for adaptive and multimodal interfaces	5.5	Open source
Robotic assistance in community care: General framework, requirements and evaluation	5.6	Open source / Reference
Models and Analytics for personalized risk detection & Interventions	6.3	License (Tiered approach based on business type)

Considering the above, Deliverable D9.11 will provide more clarity with regard to the IPR status of the solutions described above, in order to better reflect the most appropriate approach from an exploitation perspective, as was agreed in the respective task.

Finally, with regards to copyright, and in order to maximise the project's impact and knowledge sharing, the project has worked on a number of publications prioritising open access.

6 Data Security

6.1 Technical and Organisational Measures

Given the sensitive nature of the data involved in the project, security has been at the forefront of the efforts since the beginning. As such, the solutions developed and the infrastructure have been designed to be privacy and security by design. In this regard, as was also reported in Deliverable D3.6, the GATEKEEPER Data Centre infrastructure, managed by HPE, has adopted a number of technologies, products and services to that end, including:

- VPN access, by means of OpenVPN open-source software, tailored to the partners needs and providing two kinds of VPN profiles:
  - *Road warrior,* for GATEKEEPER partners users, supporting on-demand connections from PC clients
  - *Site-to-site,* for GATEKEEPER pilots, supporting always-on connections from Pilot sites
- Support for different VPN access authentication types:
  - Two Factor Authentication (2FA using password + One-Time-Password, OTP), used by partners participating to the GATEKEEPER project
  - Digital Certificate-based Authentication (Client Certificate + Password), used for Site-to-site connections
- Firewall devices and policies, used to determine whether a given user/pilot can access a network or a GATEKEEPER service
- Security services, including:
  - Identity Management, centrally managing user identities to access services (e.g. VPN, servers, VMs, container platform, big data platform)
  - Public Key Infrastructure (PKI), in particular an internal private Certification Authority that releases digital certificates (e.g. for VPN user access or internal web sites/services) and manages their lifecycle (e.g. expiration, revocation, renewal)
  - Intrusion Detection System (IDS), a service to block malicious attacks based on security rule-sets
  - Proxy Server based on HTTP
  - Log Management: all devices (e.g. operating system, backup, switches, firewalls, etc.) are traced, and logs are kept in a Log Management system for security purposes
- Infrastructure services, including:
  - Backup of all devices and services for supporting high resiliency
  - Monitoring of physical devices for hardware fault.

What is more, as previously discussed, the GATEKEEPER Trust Authority has added functionalities promoting privacy and security by design, in alignment with the security and privacy by design principles of the Open Web Application Security Project (OWASP). Based on the top 10 web application security risks made public by the OWASP project, the GATEKEEPER processes offered by the Trust Authority have been mapped against the corresponding OWASP security concerns for web applications and the measures implemented to address them, as reported in Deliverable D4.14 and presented in the table below.

Table 9: GTA processes and OWASP security risks mapping

GTA process	GTA subcomponent	OWASP web application security risk	GTA solution
Authentication	User Management Module	A2:2017-Broken Authentication	Best practices, such as the ones for Forgot Password functionality [10], can be configured by clients
Authorisation / Access Control	Trusted Things Sharing	A5:2017-Broken Access Control	Access control through IDS Connectors. Access is denied by default, unless signed DAT is received. Owing to decentralization, only the Data Owner can update/delete the data.
Auditing / Logging / Digital evidence	Things Action Tracking	A10:2017-Insufficient Logging and Monitoring	High-value transactions have an audit trail in Blockchain, which ensures immutability and non-repudiation.

Similarly, and in order to increase security of the data, the GATEKEEPER Trust Authority incorporated, at the request of the pilots, to introduce a "Things Anonymisation" component in order to de-identify the FHIR data, whether referring to the pilot dataset per se or regarding datasets potentially donated by the pilots for secondary analyses at the end of the project.

In order to achieve that, a variety of techniques are used for the de-identification of data, including encryption, masking, and generalisation. These have been applied to identifiers such as the name and to a selection of quasi-identifiers, such as the address, telephone number, email address, and device id, combinations of which can lead to identification. The deletion of the association table available to

the data controller so that natural persons cannot be identified by them is handled by the respective data controller for each dataset. The Table below, as was originally reported in Deliverable D4.14, presents the (quasi-) identifiers edited and the method used.

Table 10: Identifiers and quasi-identifiers edited by the current implementation

(Quasi-) Identifier	FHIR Resource(s)	Method applied
name	Patient, Practitioner	▪ encryption (for pseudonymisation) ▪ masking with pre-defined value (for further de-identification)
name	FamilyMemberHistory	▪ encryption (for pseudonymisation) ▪ masking with pre-defined value (for further de-identification)
telecom	Patient, Practitioner	encryption
birthDate	Patient, Practitioner	generalisation (year of birth instead of date of birth)
address	Patient, Practitioner	masking with pre-defined value
identifier	Device	encryption
deviceIdentifier	Device	encryption

Moreover, and in order to further increase the security of the data involved, each pilot has introduced and implemented their own additional security measures, which are described in brief in the table below. Additionally, pilots performed individual Data Management Plans and their own DPIAs in order to ensure that a thorough analysis of risks and mitigation measures has been performed.

Table 11: Pilots' Technical and Organisational Measures

Pilot	RUC	Technical and Organisational Measures
Aragon	RUC 1 & 2	• Patients' informed consent • Replacement of the participants' personal information with randomised codes

		• Sharing of exclusively aggregated data, containing no personal information • Data anonymisation and pseudonymisation • Access controls
Basque Country	RUC 1	• Data anonymisation and pseudonymisation techniques • Access controls
Basque Country	RUC 3	• Assignment of alphanumeric identification codes to patients • Secure storage locally of the list linking the identification codes to patients, accessible only by the principal investigator • Data anonymisation and pseudonymisation techniques • Access controls
Basque Country	RUC 7	• Data anonymisation and pseudonymisation techniques • Access controls
Cyprus	RUC7 -AMEN	• Encryption at rest • Data anonymisation • Password protection • Access controls
Cyprus	RUC7 - PASYKAF	• Encryption at rest • Data anonymisation • Password protection • Access controls
Greece	RUC 1	• Data anonymisation and pseudonymisation techniques • Encryption of stored data • Weekly/Bi-weekly monitoring of the systems to ensure its proper usability and functionality • Regular monitoring of data quality

		• Access controls
Greece	RUC 3	• Data anonymisation and pseudonymisation techniques • Encryption of stored data
Puglia	RUC 1	• Data anonymisation and pseudonymisation techniques • Only sharing anonymised datasets • Access controls
Puglia	RUC 3 - CSS	• Data pseudonymisation • Cryptographic protection, using TLS/SSL-based HTTPS protocol • Password protection • Access controls
Puglia	RUC 1-8 - AReSS	• Data pseudonymisation • Separate and secure storage of re-identification key • Cryptographic protection • IDS protection of the storage platform provided by HPE • VPN Access • TLS/SSL encryption mechanism • Regular backups • Regular maintenance • Physical measures including dual NIC, dual power supply, stacked switch etc • Access controls and access records
Poland	RUC 7	• Data pseudonymisation • Only pseudonymised data will be shared with the research teams • Password protection of stored data • Storage of personal data locally at the Coordinating Researcher's Office • Access controls

Saxony	RUC 1 & 7	• Data anonymisation and pseudonymisation techniques • Only sharing anonymised datasets • Encryption at rest • Regular backups • Access controls
UK	RUC 7 - Bangor	• Data pseudonymisation • A password-encrypted MS Excel file on the CI's MS One-Drive linked to the NHS email will be the only place where the participants' identifiers will be linked to the study case number • Data will be available in a pseudonymized or aggregated form only and within the frame of the GATEKEEPER project • Password protection • Access controls
UK	Milton Keynes	• Data anonymisation • Data will be shared only in aggregated form, revealing no personal data • Access controls

7 Ethical and Legal Aspects

Given the sensitive nature of the data involved in the GATEKEEPER project and the involvement of human participants in the pilots, compliance with ethical and legal requirements at an international, European and national level have played a major role when designing and implementing the project's pilots and its solutions.

To that end, pilots have performed individual Data Protection Impact Assessments which they have maintained updated throughout the project's lifecycle in order to ensure continuous compliance with the relevant requirements. Said DPIAs served as the baseline for a project-wide ethical and privacy impact assessment which was reported accordingly in Deliverable D1.5 and updated in D1.10. Similarly, mitigation measures were proposed in order to address any ethical and legal risks identified.

At the same time, the project has established a Policy, Legal and Gender Board, led by the project's Ethical, Legal and Gender Manager, in charge of monitoring and providing assistance in handling the following matters:

a. Legal aspects: the legal issues associated to the deployment of GATEKEEPER tools and actions (e.g. IPR, data protection and access, privacy issues, ethical aspects, etc.),
b. Policy issues: how new policies could help innovative smart living technologies get users acceptance and market uptake,
c. Gender issues: the implementation the gender equality policy of the project, and
d. Ethical, security and data management concerns in relation to data management.

Deliverable D1.11 provides a more in-depth analysis of the ethical and legal aspects that are relevant for the project, analysing the evolving regulatory framework in this regard and providing an overview of the project's compliance actions.

8 Conclusion

The current deliverable includes the best available information on data management procedures at the pilot level and at the GATEKEEPER project as a whole. Both pilots and technical partners have been considering data management and ethics requirements from the start of the project and have developed strategies in order to ensure the integrity of the solutions developed and the security, privacy and confidentiality of the data involved.

As is evident, data, both personal and non-personal, play an increasingly important role in the medical research domain and e-health and the GATEKEEPER project is no exception. From this perspective, the project's efforts have concentrated in the two-fold strategy regarding data, as follows:

a. Ensuring non-personal data can be used in order to share knowledge, useful insights and innovative solutions with the community so as to enhance citizens' quality of life, without compromising the Consortium's Intellectual Property Rights;

b. Ensuring that personal data remains protected and secure. In this regard, anonymisation and pseudonymisation techniques have been at the forefront of partners' efforts.

In this framework, the FAIR principles have been taken into consideration during the design and implementation of the projects' activities, with regards to both pilot datasets and the GATEKEEPER infrastructure.