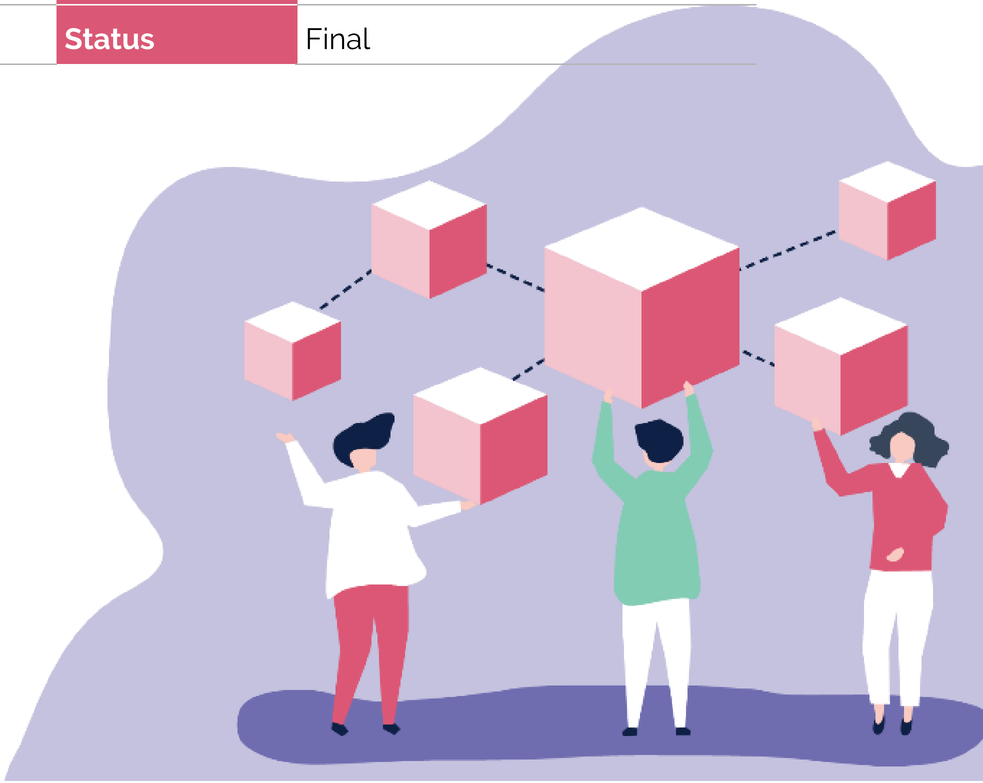
This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement N° 857223

D1.4Data Management Plan

Deliverable No.	D1.4.	**Due Date**	31/08/2020
Description	The Data Management Plan provides the strategy and principles to be adopted within the project regarding the data management, knowledge and IPR issues. It also includes the data management plans of the different pilot zones. This is a living document that will be constantly updated during the duration of the project according to the changes that might happen at a project level and at the pilots level.		
Type	Report	**Dissemination Level**	Public
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
Pasquale Annicchino	UDG	pannicchino@archimede.ch
Alessio Antonini	OU	alessio.antonini@open.ac.uk
Amera Mojahed	TUD	Amera.Mojahed@uniklinikum-dresden.de
Eva Karaglani	HUA	ekaragl@hua.gr
George E. Dafoulas	TRIK	gdafoulas@e-trikala.gr
Przemyslaw Kardas	MUL	przemyslaw.kardas@umed.lodz.pl
Sergio Guillen	MYS	sguillen@mysphera.com
Eleftheria Polychronidou	ITI	epolyc@iti.gr
Claudio Caimi	HPE	claudio.caimi@hpe.com
Silvio Pagliara	UoW	Silvio.Pagliara@warwick.ac.uk
Jordie de Batlle	CIBER	jordidebatlle@gmail.com
Janire Orcajo Lago	OSA	JANIRE.ORCAJOLAGO@osakidetza.eus
Sergio Copelli	MME	s.copelli@multimediaengineers.com
Francesco Fera	RPU	f.fera@aress.regione.puglia.it
Francesco Giuliani	CSS	f.giuliani@operapadrepio.it
Rosana Angles	ARA	ranglesb@salud.aragon.es

History

Date	Version	Change
25/02/2020	0.1	Creation and circulation of the template for the collection of relevant data from the pilots
23/03/2020	0.2	Creation of the table of content taking into account inputs from the different partners.
6/04/2020	0.3	Work of the partners on the different sections
5/5/2020	0.4	First draft with integrated contributions

15/6/2020	0.5	Integration with contribution from the business cluster
16/7/2020	0.6	Finalization of the first draft
25/8/2020	0.7	Draft sent for review
30/8/2020	1.0	Final version

Key data

Keywords	Data; Data protection; Data management; Datasets; Pilot zones
Lead Editor	Pasquale Annicchino, UDGA
Internal Reviewer(s)	Eugenio Gaeta, UPM; Susan van Hees, UU

Abstract

This document describes the Data Management Plan and serves as a guide for the partners of the GATEKEEPER project.

The deliverable is the first version of the Data Management Plan, in which the framework for data management is defined both at the project and at the pilots' level. It also contains initial identification of the possible datasets that could be collected during the execution of the project. However, due to the early stage in the project and the difficulties brought by the COVID-19 pandemic, updated versions of the deliverable will be provided on a regular base through several iterations with all the involved partners. The consortium will continue to update the document on a regular base and the same will be asked to the pilots for their data management plans. The document is in fact composed by the general Gatekeeper DMP and the DMPs of the different pilots.

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

DMP Data Management Plan

DoA Description of Action

DPIA Data Protection Impact Assessment

DPO Data Protection Officer

EDPS European Data Protection Supervisor

EEA European Economic Area

EHDS European Health Data Space

EU European Union

GDPR General Data Protection Regulation

GK Gatekeeper

GPL General Public Licence

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

PS Pilot Site

RUC Reference Use Case

WP Work Package

Table of contents

List of tables

List of figures

1 Executive summary

This document provides for a data management plan for the Gatekeeper project and for the different pilots. In part one the purpose of the document is explained and its role in the context of the Gatekeeper project. Part two introduces principles and basic notions of data management relevant for the project. Part three defines the allocation of responsibilities within the project. Part four discusses the role of data in the context of Gatekeeper. Part five introduces the Gatekeeper platform, storage and data extraction. Part six discusses IPR rights and licensing in the context of the project. Appendix A includes the data management plans of the different pilots. The figure below illustrates the content of the document. Figure 1 illustrates the mapping of data within the project and offers a visual map of the deliverable.

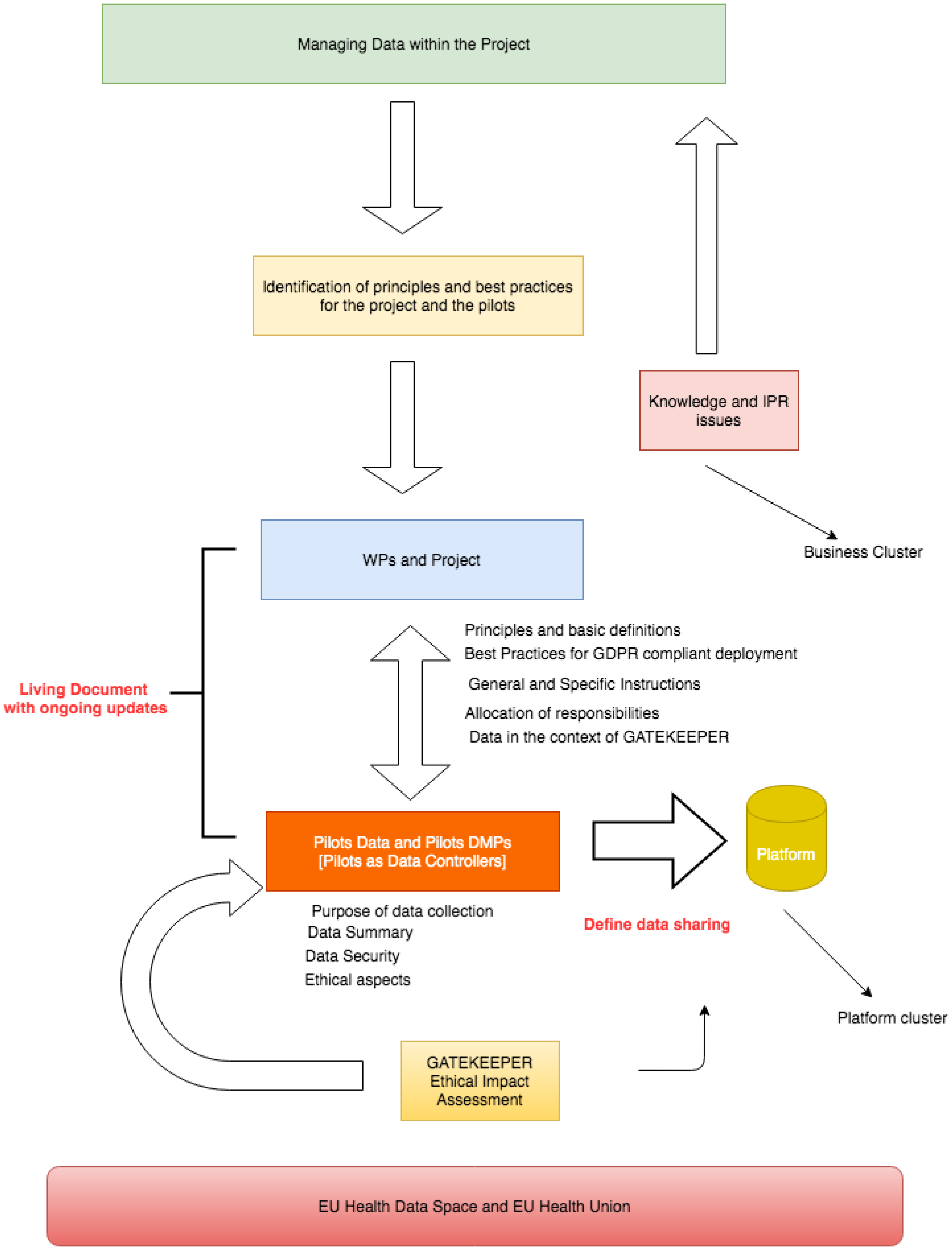

Managing Data within the Project
Identification of principles and best practices for the project and the pilots
Knowledge and IPR issues
Business Cluster
WPs and Project
Principles and basic definitions
Best Practices for GDPR compliant deployment
General and Specific Instructions
Allocation of responsibilities
Data in the context of GATEKEEPER
Living Document with ongoing updates
Pilots Data and Pilots DMPs [Pilots as Data Controllers]
Platform
Define data sharing
Purpose of data collection
Data Summary
Data Security
Ethical aspects
Platform cluster
GATEKEEPER Ethical Impact Assessment
EU Health Data Space and EU Health Union

Figure 1 –Content of D1.4

2 Introduction

2.1 The Gatekeeper project

As we have highlighted in our DoA, the main objective of the project is to "create a GATEKEEPER that connects healthcare providers, businesses, entrepreneurs, elderly citizens and the communities they live in, in order to originate an open, trust-based arena for matching ideas, technologies, user needs and processes, aimed at ensuring healthier independent lives for ageing populations. By the end of the project GATEKEEPER will be embodied in 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; (iv) support value generation through the deployment of advanced business model based on the VBHC paradigm. GATEKEEPER will demonstrate its value by scaling up, through its work plan, towards the deployment of solutions that will involve ca.40,000 elderly citizens, supply and demand side (authorities, institutions, companies, associations, academies) in eight regional communities, from seven EU member states"[1]. We aim also at following very closely all the effort that the Commission is undertaking in the context of the processing of health data by addressing issues such as citizens' access to their personal data or the portability of such data to other operators. Particular attention will be devoted also to the development of the efforts in place towards the creation of a European Health Data Space (EHDS) which aims at fostering the exchange and sharing of different kinds of health data (electronic health records, genomics, registries, etc.) in Europe.

[1] GATEKEEPER DoA.

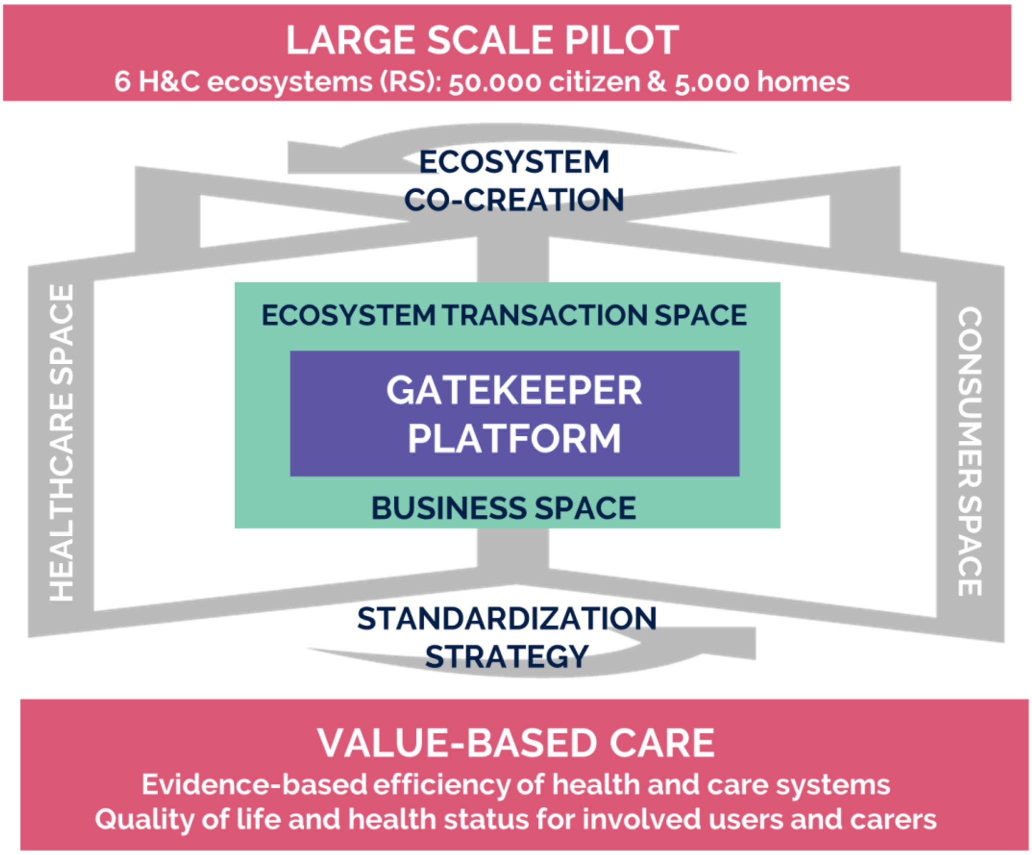

LARGE SCALE PILOT
6 H&C ecosystems (RS): 50.000 citizen & 5.000 homes
ECOSYSTEM
CO-CREATION
HEALTHCARE SPACE
ECOSYSTEM TRANSACTION SPACE
GATEKEEPER
PLATFORM
BUSINESS SPACE
CONSUMER SPACE
STANDARDIZATION
STRATEGY
VALUE-BASED CARE
Evidence-based efficiency of health and care systems
Quality of life and health status for involved users and carers

Figure 2 –The GATEKEEPER Project

2.2 Purpose of the document

2.2.1 Gatekeeper and scientific research

As a research project Gatekeeper will be bound to all the applicable legislation on data protection. The European Data Protection supervisor has recently clarified that the special data protection regime for scientific research applies when the following criteria are met:

1) Personal data are processed;
2) Relevant sectorial standards of methodology and ethics apply, including the notion of informed consent, accountability and oversight;
3) The research is carried out with the aim of growing society's collective knowledge and wellbeing, as opposed to serving primarily one or several private interests[2].

In the field of medical research there are normally two fundamental components of an appropriate approach. According to the European Data Protection Supervisor (EDPS) they are: 1) informed consent and 2) independent ethical oversight. Gatekeeper is built around these two principles. As the EDPS makes clear in his January 2020 opinion: "Human dignity and the right to integrity of the person are recognised in Article 1 and 3 of the Charter for Fundamental Rights of the European Union. Medical research on humans (also known as biomedical research or experimental medicine, including "bench science" and applied research), is strictly subject to ethical standards and controls. Under Article 3(2)(a) of the Charter, the 'free and informed consent of the person concerned' must be respected in the field of biology and medicine"[3]. Gatekeeper will closely follow the developments connected to the newly released European Data Strategy and adapt its approach accordingly. The Covid19 crisis has in fact already revealed how transparency, fair access to data, data sovereignty of the individual, and trust in the use of data are key capabilities to tackle health issues[4]. European institutions, organizations and citizens are today in the process of developing a "truly common European data protection culture"[5].

[2] EDPS, *A Preliminary Opinion on data protection and scientific research*, 6 January 2020, p. 12.

[3] *Ibid*., p. 14.

[4] See also B. Jacobs, J. Popma, *Medical research, Big Data and the need for privacy by design*, Big Data & Society, January-June 2019, pp. 1-5.

[5] European Commission, *Communication from the Commission to the European Parliament and the Council- Data protection as a pillar of citizens' empowerment and the EU's approach to the digital transition – two years of application of the General Data Protection Regulation*, 24/6/2020, p. 5.

2.2.2 Deliverable context

Taking into consideration the development of the project to be undertaken in the different deployment sites and the implementation of the different phases, the current document reflects only the intention of the project partners toward developing the project's datasets. Further developed versions of the document will be prepared and further details will be provided there. The information provided in this document is meant to define common grounds in relation to data management and must be used for the foundations of successful internal and external data management plans in order to motivate participation and collaboration between partners within the Gatekeeper consortium and amongst external partners or participants in the project.

Table 1: Deliverable context

PROJECT ITEM	RELATIONSHIP
Objectives	The deliverables provides the data management plan for the project and for the different pilots
Exploitable results	The deliverable presents a model for data management and will serve as a reference for the consortium
Work plan	The deliverable will be constantly updated according to the DoA
Milestones	Milestones will be set with the pilots in order to plan the update.
Deliverables	The deliverable is to be read in conjunction with D 1.5.
Risks	The constant update of information from the pilots will need to be monitored. Particular attention will need to be devoted to data sharing issues.

2.2.3 The COVID-19 crisis

The COVID-19 pandemic has had and will have important effects on all the activities that concern health, and, of course, e-health. Even within the context of our project we have experienced difficulties because the different hospitals and professionals have been forced to focus on the crisis which has provoked several delays. For the purpose of this document, it is important to highlight that the pandemic will have an impact also on the relationships between data protection and health. The magnitude of the impact can be perceived by pointing at some remarks made by the European Data Protection Supervisor: "(…) our strategy should be adaptable to global game changers. We were confident we had chosen the right, comprehensive methodology. Of course, we maintained some flexibility to be able to cope with entirely unpredictable circumstances. What we had in mind were natural disasters causing unexpected changes in the legislation at both European and national levels. However, we never expected such a tragedy to have occurred so quickly!"[6].This Data Management Plan is understood by the consortium as a living document and therefore will be constantly updated, also taking into consideration the legislative developments brought by the COVID-19 pandemic which will have important effects on the e-health sector.

2.2.4 Findings

This document offers a mapping and a data management plan of the data, especially personal data, that will be managed and processed in the context of the GATEKEEPER project. It creates first a project layer and adds to it the role of pilots of data controllers of the health-related data that will be collected. As the project advances new dimensions of data management will be added and the deliverable updated.

[6]W. Wiewiòrowski, *The moment you realise the world has changed: re-thinking the EDPS Strategy*, 20th March 2020, available at: http://edps.europa.eu/press-publications/press-news/blog/moment-you-realise-world-has-changed-re-thinking-edps-strategy_en.

3 Principles and basic notions of data management

3.1 Definitions

In order to facilitate the understanding of fundamental concepts -also from researchers and people who do not have a legal background-core concepts of the European Data Protection Supervisor are introduced and defined[7]

3.1.1 Accountability

Principle intended to ensure that controllers are more generally in control and in the position to ensure and demonstrate compliance with data protection principles in practice. Accountability requires that controllers put in place internal mechanisms and control systems that ensure compliance and provide evidence-such as audit reports- to demonstrate compliance with external stakeholders, including supervisory authorities.

3.1.2 Automated individual decision

An "automated individual decision" is a decision which significantly affects a person, and which is based solely on automated processing of personal data in order to evaluate this person. Such an evaluation may relate to different personal aspects, such as performance at work, creditworthiness, reliability, conduct, etc.

Article 22 of the GDPR lays down the right for individuals to object to decisions about them and solely based on automated means, unless certain conditions are fulfilled or appropriate safeguards are put in place.

3.1.3 Biometric data

According to article 4(14) of the GDPR "biometric data means personal data resulting from specific technical processing relating to the physical, physiological

[7] The glossary of original definitions is available on the EDPS website at:https://edps.europa.eu/data-protection/data-protection/glossary_en

or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data".

3.1.4 Confidentiality

Confidentiality in general refers to the duty not to share information with persons who are not qualified to receive that information. In a more specific sense, it refers to the confidentiality of communications provided for in Article 5 of the E-Privacy Directive 2009/136/EC.

3.1.5 Consent

In data protection terminology, consent refers to any freely given, specific and informed indication of the wishes of a data subject, by which he/she agrees to personal data relating to him/her being processed (see Article 4 sub 11 of Regulation (EU) 2016/679).

Consent is an important element in data protection legislation, as it is one of the conditions that can legitimise processing of personal data. If it is relied upon, the data subject must unambiguously have given his/her [written or verbal] consent to a specific processing operation, of which he /she shall have been properly informed. The obtained consent can only be used for the specific processing operation for which it was collected, and may in principle be withdrawn without retroactive effect.

3.1.6 Controller

According to article 4(7) of the GDPR: "controller means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided by Union or Member State law".

3.1.7 Cookies

Short text files stored on a user's device by a web site. Cookies are normally used to provide a more personalised experience and to remember user profiles without the need of a specific login. Also, it can be placed by third parties (such as advertising networks) in end users' devices and may be used to track users when surfing across different websites associated to that third party.

3.1.8 Data concerning health

According to article 4(15) of the GDPR "data concerning health means personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status".

3.1.9 Data controller

Under the GDPR, the data controller is the party that, alone or jointly with others, determines the purposes and the means of the processing of personal data. The actual processing may be delegated to another party, called the data processor. The controller is responsible for the lawfulness of the processing, for the protection of the data, and respecting the rights of the data subject. The controller is also the entity that receives requests from data subjects to exercise their rights.

3.1.10 Data minimization

The principle of "data minimization" means that a data controller should limit the collection of personal information to what is directly relevant and necessary to accomplish a specific purpose. They should also retain the data only for as long as is necessary to fulfil that purpose. In other words, data controllers should collect only the personal data they really need, and should keep it only for as long as they need it.

The data minimisation principle is expressed in Article 5(1)(c) of the GDPR.

3.1.11 Data mining

Data mining is the process of analysing data from different perspectives and summarising it into useful information. Data mining software is one of a number of tools for interrogating data. It allows users to analyse data from many different dimensions or angles, categorise it, and summarise the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. It is commonly used in a wide range of profiling practices, such as marketing, surveillance, fraud detection and scientific discovery. Obviously, for data mining to be effective it is necessary to analyse large amounts of previously collected data.

3.1.12 Data protection authority

A Data protection authority (DPA) is an independent body which is in charge of:

-monitoring the processing of personal data within its jurisdiction (country, region or international organization);

-providing advice to the competent bodies with regard to legislative and administrative measures relating to the processing of personal data;

-hearing complaints lodged by citizens with regard to the protection of their data protection rights.

According to Article 51 of the GDPR, each Member State shall establish in its territory at least one data protection authority, which shall be endowed with investigative powers (such as access to data, collection of information, etc.), corrective powers (power to order the erasure of data, to impose a fine or a ban on processing, etc.), and authorisation or advisory powers (issuance of opinions, power to accredit certification bodies, etc..

National data protection authorities have been established in all European countries, as well as in many other countries worldwide.

3.1.13Data Protection Impact Assessment (DPIA)

The data controller shall carry out an assessment of the impact of the envisaged processing operations on the protection of personal data when a type of processing is likely to result in a high risk to the rights and freedoms of natural persons.

This assessment has to be done prior to the processing and, in particular if using new technologies, has to consider the nature, scope, context and purposes of the processing.

3.1.14Data protection officer (DPO)

The DPO is an expert on data protection laws and practices and has to be in the position to operate independently within the organization. The DPO needs to ensure the internal application of the Regulation and that the rights and freedoms of the data subjects are not likely to be adversely affected by the processing operation.

3.1.15Data quality

Data quality refers to a set of principles distinguished in Article 5 of the GDPR namely:

-lawfulness, fairness and transparency;

-purpose limitation;

-data minimisation;

-accuracy;

-storage limitation;

-integrity and confidentiality.

3.1.16 Data retention

Data retention refers to all obligations on the part of controllers to retain personal data for certain purposes. The Data Retention Directive (Directive 2006/24/EC) contains an obligation for providers of electronic communications to retain traffic and location data of communications through telephone, e-mail, etc. The retention takes place for the purpose of the investigation, detection and prosecution of serious crime.

3.1.17Data subject

The data subject is the person whose personal data are collected, held and processed.

3.1.18 Data transfer

Transfers are subject to specific safeguards when the recipient is located in a country outside of the EU/European Economic Area (EEA) according to Chapter V of the GDPR.

3.1.19 Personal data

According to Article 4(1) of the GDPR:"'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person".

The name and the social security number are two examples of personal data which relate directly to a person. But the definition also extends further and also encompasses for instance e-mail addresses and the office phone number of an employee. Other examples of personal data can be found in information on physical disabilities, in medical records and in an employee's evaluation.

Personal data which is processed in relation to the work of the data subject remain personal/individual in the sense that they continue to be protected by the relevant data protection legislation, which strives to protect the privacy and integrity of natural persons. As a consequence, data protection legislation does not address the situation of legal persons (apart from the exceptional case where information on a legal person also is related to a physical person).

3.1.20 Personal data breach

According to article 4(12) of the GDPR "personal data breach means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed".

3.1.21Privacy

Privacy is the ability of an individual to be left alone, out of public view, and in control of information about oneself. One can distinguish the ability to prevent intrusion in one's physical space ("physical privacy", for example with regard to the protection of the private home) and the ability to control the collection and sharing of information about oneself ("informational privacy").

The concept of privacy therefore overlaps, but does not coincide, with the concept of data protection.

The right to privacy is enshrined in the Universal Declaration of Human Rights (Article 12) as well as in the European Convention of Human Rights (Article 8).

3.1.22 Privacy by design

Privacy by design aims at building privacy and data protection up front, into the design specifications and architecture of information and communication systems and technologies, in order to facilitate compliance with privacy and data protection principles.

3.1.23 Processing (of personal data)

According to Article 4(2) of the GDPR, processing of personal data "means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction"

Personal data may be processed in many activities which relates to the professional life of a data subject and of course in the case of health treatments.

3.1.24 Processor

According to article 4 (8) of the GDPR "processor means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.

3.1.25 Processor agreement

Transfer of personal data from a data controller to a data processor must be secured by a data processor agreement. It must meet certain minimum requirements, as set forth by Article 28 of the GDPR.

The contract must stipulate that the data processor shall act only on instructions from the data controller. The data processor must provide sufficient guarantees in respect of the technical security measures and organisational measure governing the processing to be carried out, and must ensure compliance with such measures.

3.1.26 Pseudonymisation

According to article 4(5) of the GDPR: "pseudonymisation means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person".

3.1.27 Retention periods

Data retention refers to all obligations on the part of controllers to retain personal data for certain purposes.

To limit how long you keep personal data is part of data minimisation. The rule of thumb is "as long as necessary, as short as possible", although sometimes legal rules may impose fixed periods. Data that is no longer retained cannot fall into the wrong hands, nor be abused, meaning that defining and enforcing limited conservation periods helps to protect the people whose data are processed.

3.1.28 Right of access

The right of access is the right for any data subject to obtain from the controller of a processing operation the confirmation that data related to him/her are being processed, the purpose(s) for which they are processed, as well as the logic involved in any automated decision process concerning him or her.

The right of access also allows the data subject to receive communication in an intelligible form of the data undergoing processing and information regarding the processing.

3.1.29 Right of information

Everyone has the right to know that their personal data are processed and for which purpose. The right to be informed is essential because it determines the exercise of other rights. The right of information refers to the information which shall be provided to a data subject whether or not the data have been obtained from the data subject. The information which must be provided related to the identity of the controller, the purpose(s) of the processing, the recipients, as well as the existence of the right of access to data and the right to rectify the data. The right of information for the person concerned is limited in some cases, such as for public safety considerations for the prevention, investigation, identification and prosecution of criminal offences, including the fight against money laundering.

3.1.30 Right of rectification

The right of rectification is the right to obtain from the controller the rectification without delay of inaccurate or incomplete personal data. The right of rectification is an essential complement to the right of access and is important to maintain a high level of data quality.

To exercise the right of rectification, the data subject usually has to write to the controller of the processing operation.

3.1.31Right to object

Article 21 of the GDPR gives individuals the right to object to the processing of their personal data at any time. This allows individuals to stop or prevent processing of their personal data. An objection may be in relation to all of the personal data of an individual or only to certain information. It may also only relate to a particular purpose for which data are being processed.

3.1.32 Right to restriction of processing

Article 18 of the GDPR gives individuals the right to restrict the processing of their personal data in certain circumstances. This means that an individual can limit the way in which an organisation uses their data. This is an alternative to requesting the erasure of their data.

3.1.33 Special categories of personal data

Special categories of personal data include data that reveals: “racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade-union membership, genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural’s sex life or sexual orientation (Article 9 of the GDPR).

Processing of such information is in principle prohibited, except in specific circumstances. It is possible to process sensitive data for instance if the processing is necessary for the purpose of medical diagnosis, or with specific safeguards in the field of employment law, or with the explicit consent of the data subject.

3.2 Best practices for GDPR compliant pilot deployments

The research community, in the context of different projects, has identified guidelines which can guide consortium and pilots in the deployment of the different solutions. Gatekeeper aims at being compliant with the standards identified by the research community. Below the core principles on personal data management in the context of LSPs identified in the context of the Large Scale Pilots programmes financed by the European Commission, which encompassed various application domains, are reported. Pilots have been invited to comply with these guidelines[8]

1. Perform a preliminary data protection impact assessment before collecting any data with new technologies. Ensure that you address and mitigate the identified risks;
2. Minimize personal data collection, including by adapting the granularity of the data by processing the data at the edge. Consider data minimization and data protection by design as an opportunity to save costs and to increase the scalability of the system to be deployed. This is a way to leverage the approach to build trust within the organization and towards the different stakeholders;
3. Minimize personal data transfers by prioritizing onsite pre-processing, edge computing and local storage. Decentralized data processing can contribute to enhance both data protection and scalability of the system;
4. Minimize data storage and retention time, which will also save infrastructure costs;
5. Maximize the use of anonymization and retention techniques;

[8] The list of principles is taken by S. Ziegler and others, *Personal Data Protection for Internet of Things deployment: Lessons learned from the European Large-Scale Pilots of Internet of Things*, February 2020, pp. 30-31.

6. Ensure that data processing is lawful and that the amount of personal data collected is proportionate to the legitimate purpose
7. Clarify who are the data controllers and the data processors;
8. Designate a data protection officer;
9. Ensure that the data protection officer can be easily contacted;
10. Formalize your data protection policy;
11. Organize regular communication and training activities on data protection and data management;
12. Write a Data Management Plan;
13. Secure your IoT network;
14. Each IoT mote should be protected by a unique and distinct password;
15. Define and implement a clear access rights policy;
16. Adopt and enforce a strict policy and procedures for updating the firmware;
17. Establish procedures to comply with the data subjects' rights;
18. Exchange and collaborate with other DPOs;
19. Use external certification of compliance with data protection regulation as a mean to reduce liability and to increase trust and transparency with end-users;
20. Identify any cross-border data transfer of personal data and check if they are lawful;
21. Clearly inform and communicate the purpose for data collection, the categories of data processed, who has access and how long the data will be stored online through online applications [i.e. privacyapp.info];
22. Take advantage of online commitment tools to ensure that all partners located in other jurisdictions are committed to respect the same level of data protection [i.e. privacypact.com].

It is also important to take into account recent developments brought by the COVID-19 crisis which has brought new guidelines into play. In this context on April 21st 2020 the European Data Protection board has approved the Guidelines 3/2020 on the processing of data concerning health for the purpose of scientific research in the context of the COVID 19 outbreak.

3.3 Data Ethics in the context of the project

Besides scientific advance, the potential benefit of the project will be on a social, cultural, economic and individual basis. GATEKEEPER can be seen as a fundamental step towards improving independent living and smart living environments for ageing well of older adults and no harm can be foreseen that will arise from the planned WP7 Large scale pilot deployment sites.
The main target user group of GATEKEEPER are older adults (healthy or not). In the pilots (WP7) targeted in GATEKEEPER, the participants will have the competence to understand the informed consent information. In the unlikely case that they are unable to do so, no activity related to the project will be conducted. In order to protect the privacy rights of participants, a number of best practice principles will be followed. These include:

• No data will be collected without the explicit prior informed consent of the individuals under observation. This involves being open with participants about what they are involving themselves in and ensuring that they have agreed fully to the procedures/research being undertaken by giving their explicit consent.

• No data collected will be sold or used for any purposes other than the Gatekeeper project.

• A data minimisation approach is suggested at all levels of the project and will be supervised by each Pilot Demonstration responsible (one per each country involved and managed by the central Ethical Board (T1.4)of the project). This will ensure that no data which is not strictly necessary to the completion of the current project will be collected.

• Any shadow (ancillary) personal data obtained during the course of the pilots will be immediately cancelled.

• Compensation – if and when provided – will correspond to a simple reimbursement for working hours lost as a result of participating in the project special attention will be paid to avoid any form of unfair inducement;

• If employees of partner organizations, are to be recruited, specific measures will be in place in order to protect them from a breach of privacy/confidentiality and any potential discrimination. In particular their names will not be made public and their participation will not be communicated to their managers.

• In addition, any data or information that is disclosed or otherwise made available between GATEKEEPER Parties during the implementation of the projector for any Exploitation activities ("Shared Information"),shall not include personal data as defined by the GDPR. Accordingly, each Party agrees that it will take all necessary steps to ensure that all Personal Data is removed from the Shared Information, made illegible, otherwise made inaccessible (i.e. de-identify) to the other Parties prior to providing the Shared Information to such other Parties.

• Each Party who provides or otherwise makes available to any other Party Shared Information will represent that: (i) it has the authority to disclose the Shared Information, which it provides to the Parties; (ii) where legally required and relevant, it has obtained appropriate informed consents from all the individuals involved, or from any other applicable institution, all in compliance with applicable regulations; and (iii) there is no restriction in place that would prevent any such other Party from using the Shared Information for the purpose of this Action and the exploitation thereof.

3.4 Organization of the project and pilots work

The first task to complete, by each pilot, has been to clarify who is in charge of what, and more specifically, who are the data controller(s) and the data processor(s). This requires clarifying and to define the pilot's organizational scheme. The scheme should clarify: which personal data are/will be collected. This should be done by listing the various data sets that will be collected and assessing their potential qualification as "personal data". As the work on local and general data management plans is understood as an ongoing work, the exercise in data mapping will continue over the course of the project and bring to regular updated of the data management plans. It is suggested to distinguish non-personal data flows from data flows that include personal data. In order to perform this exercise a template has been provided by the Consortium.

The GDPR defines the role and responsibility of a Data Protection Officer (DPO). The DPO is in charge of monitoring the application of the GDPR within an organization and providing strategic advice to it on how to process personal data while respecting individuals' rights. Each Data Controller (pilot) should have a clearly identified DPO. Gatekeeper also has identified a DPO for the project which will be in charge of:

• Establishing common rules and requirements for the consortium data protection policy;
• Coordinating the action and information among the various DPOs and organize, when needed regular calls among the DPOs of the different pilots;
• Serving as an entry point to answer questions and complaints from third parties when addressed to the project as a whole;
• Providing guidance on how to implement the privacy by design and by default principles;
The Local DPOs should report and work in close coordination with the scientific responsible of the different pilots and the project's DPO.

All the researchers to be involved in GATEKEEPER will comply with and follow the principles outlined by the GDPR on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

Medical data will be deposited in anonymised form which is allowed by the informed consent signed by study participants from the user groups. Privacy, confidentiality and security of data are achieved by defining a subject map table. In this table, an automatically generated index number is assigned to a subject's Trial identification number which bears no significance to any confidential characteristic of these subjects. All Trial participants are made aware that the data collected from them will be shared with other research collaborators but that their personal data is kept confidential at all times. The databases that hold confidential data on participating subjects sit on a secure network and do not have an internet (HTTP) connection so as not to compromise the data. Furthermore, technical procedures are in place to monitor what data is entered and exported to ensure there is no breach of this. Measures will be taken to ensure data security at each pilot site.

The project invites the pilots and takes itself into consideration the best practices developed in the context of the other LSPs projects.

3.4.1 General instructions for all processing methods

In the context of GATEKEEPER pilots work, data processing, pursuant to articles 5 and 25 of the GDPR, must comply with the principle of "data minimization", it must take into account the restriction of the quantity of personal data processed, the processing operations, the disclosure time window, the conservation of data, the purposes of processing.

Access will be allowed only to those data that are strictly necessary to complete the task assigned. As far as personal data are concerned, they must be processed in a lawful, appropriate and transparent manner, and be correct and up-to-date.

The person authorized for the processing when performing processing operations is required to:

-ascertain that relevant data protection information are delivered to the party concerned pursuant to article 13 and 14 of the GDPR and verify that each particular processing operation contained therein (for example sharing, disclosing or profiling) is true and complies with the provisions of law and regulations;

-allow the exercise of the rights and powers provided for in chapter III of the GDPR (right to access, right to rectification, right to erasure, right to restriction, right to object, right to human intervention and appeal case of decision based exclusively on automated processing);

-not to transmit to third parties information about personal data processed: communication and disclosure is allowed only if it is functional to the performance of the task assigned or in compliance with regulatory obligations, and with the authorization of the internal data processor;

-ascertain the identity of the data subject before providing information about his/her personal data or the related processing performed (limited to verifying the identification document without having to keep a copy);

-store in locked cabinets any storage devices or documents-even if not final (drafts)- containing personal data at the end of the processing period;

-do not leave prints of documents containing personal data unattended at the photocopiers;

-use the appropriate paper shredders when there is the need to destroy documents containing personal data. When such tools are not available, tear or cut into strips the documents so that they cannot be recreated;

-keep the processed personal data for a period of time not exceeding that necessary for the purposes for which they were collected and processed, in compliance with the terms provided for by the law;

-lower the tone of voice in the conversations and adopt an adequate distance (so-called "courtesy distance") in order to avoid that third parties can, even involuntarily, process personal data and/or professional information;

-for any concerns regarding the processing of personal data, contact the pilot's data protection officer;

-immediately report anomalies, incidents, thefts, accidental losses of data affecting the processing of personal data to the data protection officer in order to initiate the procedure for the communication of the data breach to the data protection authority and the parties concerned;

-fulfill the confidentiality obligation also in the period following the termination, if applicable, of the activities carried out in the context of the project;

- It is recalled hereby that the consultation of the data contained in the databases does not allow any form of data sharing, disclosing and further processing that is not strictly necessary and functional to the fulfilment of the tasks and functions assigned.

-With regard to document flows between the different WPs of the project it is recalled that suitable organizational measures must be adapted to protect the confidentiality of personal data.

3.4.2 General instructions for IT systems-based data processing

With regard to data searches and other processing steps performed by means of IT tools, the authorized processor will have a strictly personal login credential to access data.

The authorized processor will:

-not share his/her credentials with other users, except for the cases expressly allowed;

-not access services that are not permitted;

-not attempt to obtain system administrator privileges;

-verify that the devices used are virus-free;

-not connect devices that allow uncontrollable access to the institution's network devices;

-erase all personal data from any storage devices (disk, USB flash drives, etc.) before reusing them; if this is not possible, the latter must be destroyed.

-All tools must be locked and password protected if left unattended. Whether using online storage, a laptop or some other technology it is important to make your password hard to guess. We recommend using three random words together as a password (eg. "coffeetrainfish" or "walltinshirt"). Make sure you use different passwords for different services too.

-Back up information. Keep a separate copy of any important information to avoid losing access to it. Online storage is an easy way to keep a remote copy of your data should you need it. Or keep a copy on a separate hard drive or USB stick. Just remember to set a strong password to protect your information and lock it away when you are not using it.

3.4.3 General instructions to be followed in case guests or contractors are present

In case of presence of guests or contractor staff, the authorized processor is required to:

-have guests or contractors wait in areas where confidential information or personal data are not present, lowering the volume of voice and/or closing the doors in case of verbal or telephone communication;

-avoid walking away from his/her desk;

-store documents containing personal data and start the screen saver on his/her PC;

-keep his/her login credentials secret and properly stored, in compliance with relevant security measures;

-refrain from revealing or having his/her password entered even by technical assistance staff;

-report any anomaly to the data protection officer;

3.4.4 Specific instructions for the processing of particular categories of data (as per article 9 of the GDPR)

Without prejudice to the foregoing, for the processing of personal data referred to in this paragraph the following additional instructions are prescribed:

-do not provide such data by telephone when not absolutely certain about the identity of the recipient;

-avoid faxing documents containing such data when other identifying information is present: in this case it is preferable to send the documentation without explicit reference to the party concerned (for example, by simply marking the documents with a code);

-replace the name of the data subject with a code and keep the "name-code" association in a separate archive, which access is limited to a small number of authorized processors (so-called "pseudonymization");

-do not leave the documents-including not final ones- or any devices containing such data unattended and keep them in furniture fitted with a lock, whose keys must be properly stored;

-keep documents containing data on health, sex life and sexual orientation in the aforementioned lockable containers, separately from any other documents;

-apply all the necessary organization and security measures to guarantee the security of the datasets collected.

3.5 Findings

The GATEKEEPER project involves different layers of data management. They should all respond to the principle of sounding and compliant data management. We have provided basic principles and instructions that will have to be complemented with the requirements of the different organizations and institutions involved in the project. A parallel attention should be devoted to data ethics issue which have also been addressed in D 1.5.

4 Allocation of responsibilities

4.1 Pilots as data controllers

In the context of GATEKEEPER, data are collected by the participating pilot zones through their own infrastructure and technologies. In a second phase of the project, part of the infrastructure/technology will be focused on the platform and the data sharing from the pilots to the platform. The initial decision to collect, process and share data remains under the direct control of the corresponding pilots' institutions. The participating local responsible organizations are therefore to be understood as data controllers of the data. The sharing of data with the platform will be governed by specific agreements which are still to be defined.

As data controllers, the local pilot organizations are responsible for sharing relevant datasets, also through open access repositories. GATEKEEPER provides a platform where pilot can share their data that can be used by all participating pilots.

Pilots are also responsible for the selection, integrity and compatibility of the data they share with the platform during the project lifetime.

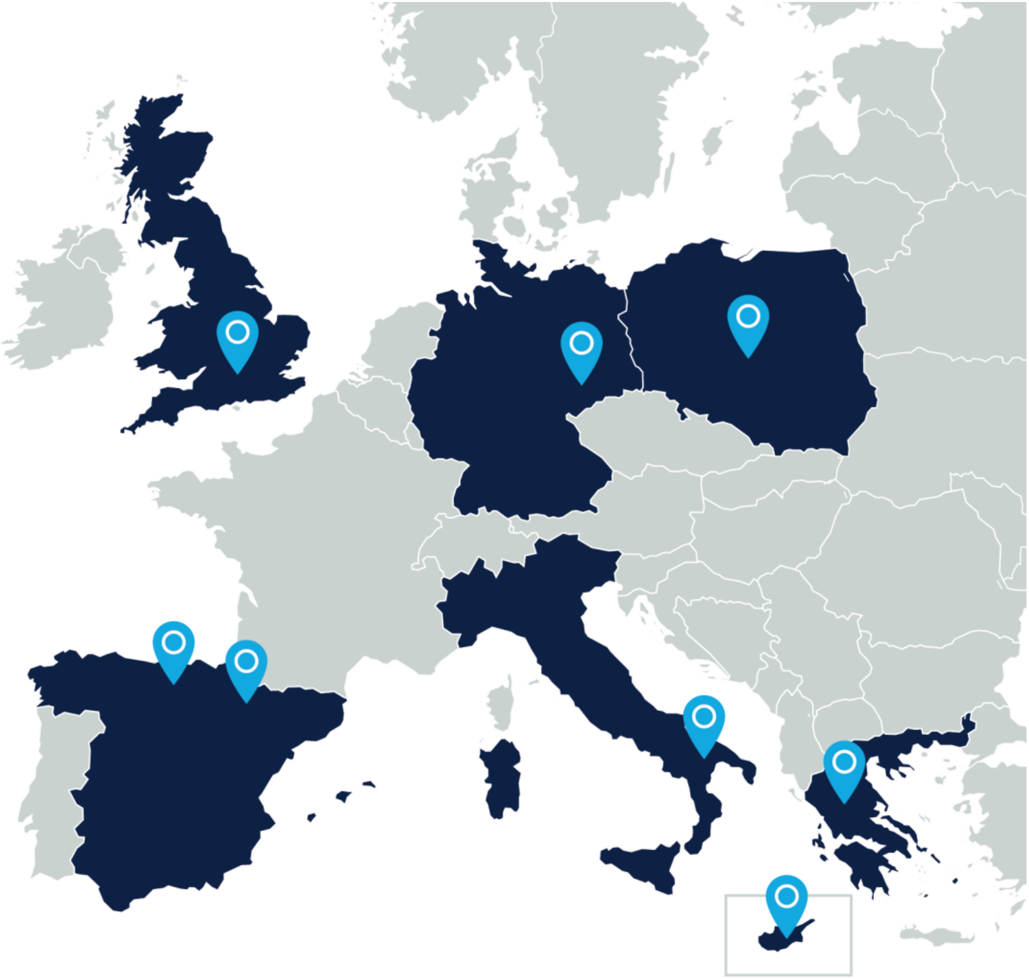


Figure 3 –GATEKEEPER Pilots

In this context each partner is responsible to ensure compliance and respect with applicable regulations:

-Each party is responsible for ensuring its own compliance with all laws and regulations applicable to its activities. Such laws include, but are not limited to, those in respect of data protection, intellectual property rights and healthcare;

-Each party represents that there is patient consent to permit distribution and use of data (including medical data) and any other information provided to other parties;

-Any party which provides any data or information to another party in connection to the project will not include any personal information relating to an identified or identifiable natural person or data subject unless there is a specific agreement which guarantees the application of data protection and security measures;

-To this end, the providing party will anonymize all data delivered to other parties to an extent sufficient to ensure that a person without prior knowledge of the original data and it collection cannot, from the anonymized data and any other available information, deduce the personal identity of the subject.

4.2 Role and authority of the project on data processing

While pilots remain in command for the data processing of their data, the project has specific responsibilities to:

-define the general data management plan;

-support, collectively define and coordinate the data management policy at the project level;

-coordinate with the partners to verify the compliance with applicable legal norms;

-fulfil the function of Data Protection Officer (DPO) at the project level for demands and requests coming from third parties.

The project provides guidance, requirements and a horizontal coordination among the pilots that remain in full control of their respective data collection and processing.

All GATEKEEPER partners are requested to respect the data management policies established at consortium level, including the present DMP.

4.3 Adopting an adequate data management policy

Health providers are taken by two opposite sets of requirements and interests. On one hand there is a direct interest to ease access to data collected by healthcare providers to citizens and third parties who could develop services out of it. Simultaneously, there is a strict need to respect and preserve the privacy and the personal data of the citizens. While the project will support the access to open data by third parties, it will abide to strict personal data protection policy in line with the EU General Data Protection Regulation and any other applicable norms. Personal data protection compliance is part of the project requirements and will guide the architecture design. Personal data protection principles will determine and limit data that can be shared. The project is committed to proactively ensure full compliance with the GDPR through a set of ad hoc policies, mechanisms and tools.

GATEKEEPER will strictly stick to the principle of data minimization by avoiding the collection and processing of any unnecessary personal data.

Personal data can be kept in a form which permits identification of data subjects for no longer than is reasonable, proportionate and necessary for the purposes for which the personal data are processed;

4.4 Findings

In their role as data controllers pilots will play an important role in the context of GATEKEEPER. Other organizations and partners that will manage data for the purpose of the project are also called to ensure the full respect of any applicable law or regulations and to adopt an adequate data management policy. Detailed work will need to be undertaken in the next steps of the project in order to define data sharing issues.

5 Data in the context of Gatekeeper

5.1 Purpose of the data collection and relation to the objectives of the project

In the context of Gatekeeper personal data are processed with the aim of working on solutions with technologies for the better quality of life and care. The purpose of the treatment is to carry out the management of stakeholder participation in the project. Likewise, the data may be processed to develop GATEKEEPER's own dissemination activities or to send information about participation in the project to project users.

Personal data of participants will only be used for the development of the implementation in the region where the GATEKEEPER project is developed, being stored with all the possible guarantees of confidentiality and privacy.

5.2 Data description

Data in the context of GATEKEEPER are collected at the project level and at the pilots' level. At the project level data are mainly collected for the production of deliverables (see table below). Data at the pilot level are described in the annex.

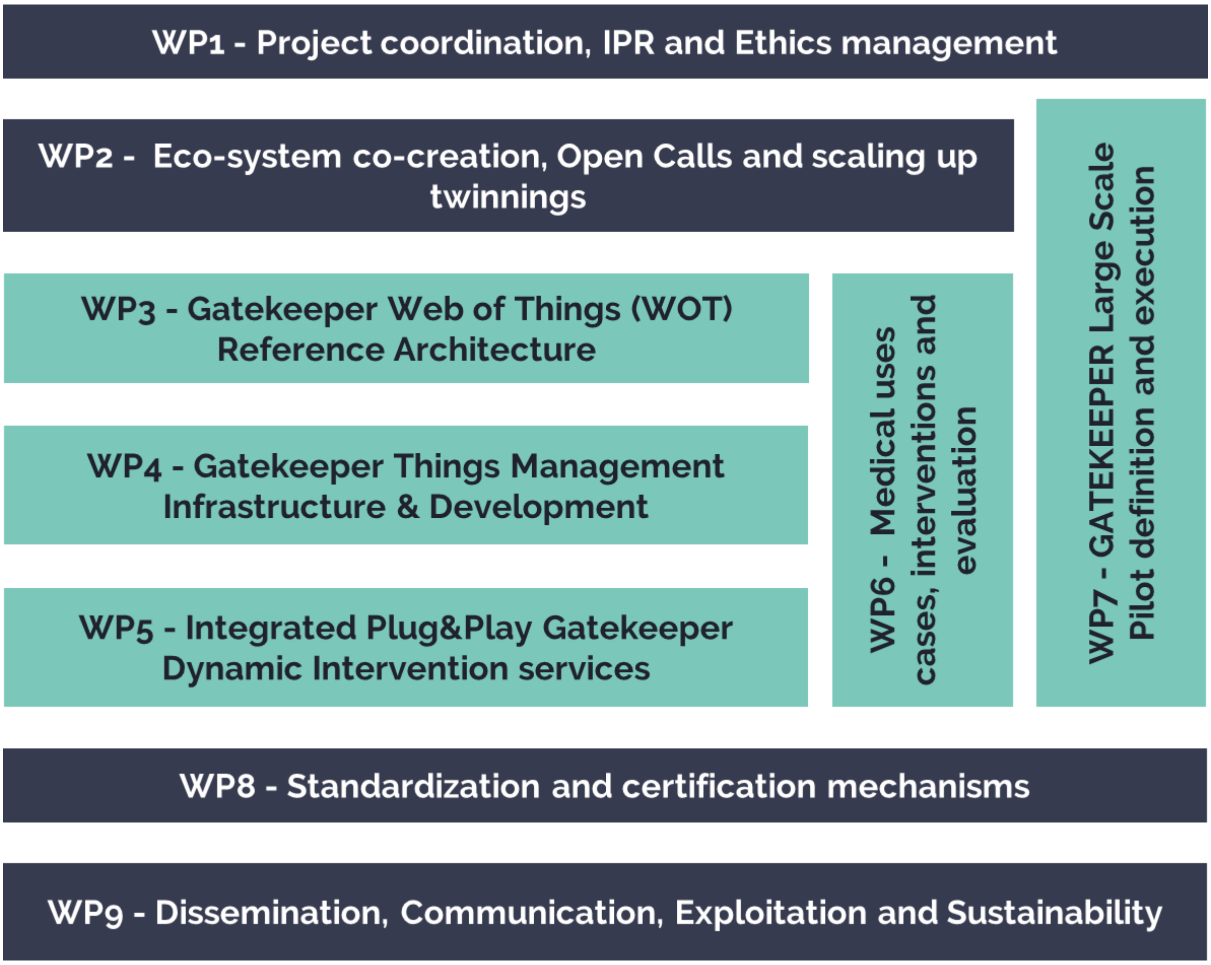

WP1 - Project coordination, IPR and Ethics management
WP2 - Eco-system co-creation, Open Calls and scaling up twinnings
WP3 - Gatekeeper Web of Things (WOT) Reference Architecture
WP4 - Gatekeeper Things Management Infrastructure & Development
WP5 - Integrated Plug&Play Gatekeeper Dynamic Intervention services
WP6 - Medical uses cases, interventions and evaluation
WP7 - GATEKEEPER Large Scale Pilot definition and execution
WP8 - Standardization and certification mechanisms
WP9 - Dissemination, Communication, Exploitation and Sustainability

Figure 4 – Work-packages in the GATEKEEPER project

Table 2: Data in the different Work packages

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[9]	.csv[10]
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

[9] Large scale pilots (LSP) datasets (same as WP 7): This data sets will include a broad range of data from patients (i.e. sensors, health outcomes, use of services, user experience), professionals (i.e. feedback on the system, use of technology) and the environment where LSP will be deployed (i.e. costs associated with implementation, use of healthcare resources, barriers towards implementation). The owners of the datasets will be the partners in each of the pilots sites (specifically those in charge of collecting and analyzing the data), However, and in collaboration with WP 7, there may be initiatives to combine results across sites via meta-analytical approaches (it is not foreseen that this would imply the sharing of data at the individual level)

[10] .csv is a standards database format. The actual format of the databases may be different.

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	.doc; .pdf

	Evolution Report (I to IX)	
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	.doc; .pdf

	assessment reports	
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

5.3 Findings

Data are a central component of the research project. At the project level WPs will mainly managed data related to the production of the different deliverables. Details of the data managed at the pilot level are offered in the Annex with the production of the DMPs of the different pilots.

6 GATEKEEPER platform data, storage and extraction

GATEKEEPER platform is a digital platform that provides AI and data-oriented services for the development of health and care solutions.

GATEKEEPER is based on the concept of digital twins where every platform asset, such as devices, services, data or even other platforms, has a digital replica that is described with a Thing Description in agreement with Web of Thing standard specification.

Within GATEKEEPER the Things are virtual entities that are, decoupled but linked with their physical and/or technological implementation. Based on that, at data level GATEKEEPER allows a high degree of separation between data owner (the physical owner of a database for instance) from the data provider (the service that wraps the data into a digital twin). At conceptual level Gatekeeper could be seen as an interconnected network of things, like a web of things (figure 4).

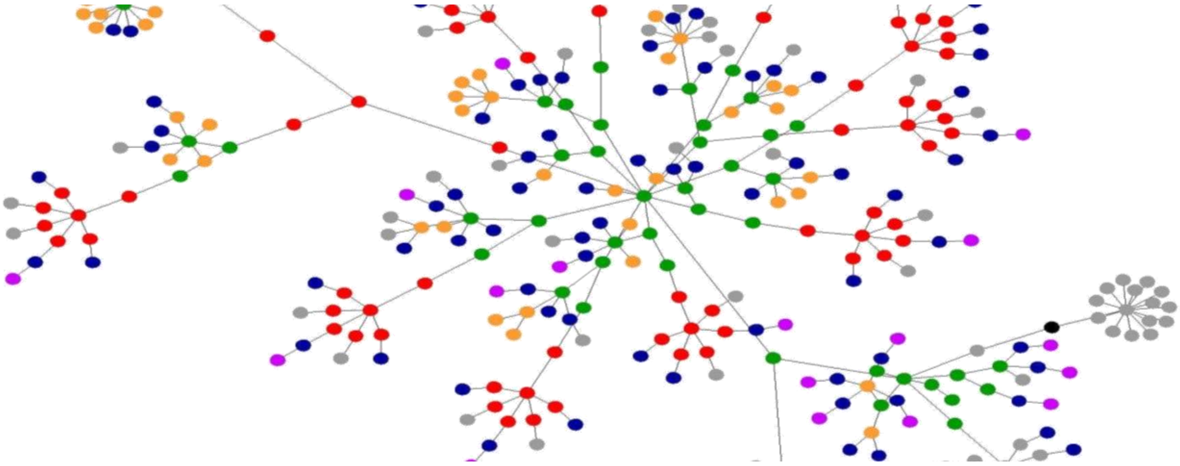


Figure 5 - GATEKEEPER concept platform, interconnected network of think (web of things) similar to World Wide Web structure

The GATEKEEPER core data are the data related to the GATEKEEPER users. GATEKEEPER users are developers and customers, nor patients neither healthcare professionals are expected to be GATEKEEPER user.

Anyway, when a developer builds an application by using GATEKEEPER services, he can associate to several Gatekeeper services sensible data such as personal patients' and/or healthcare professionals' data. This data are private data owned by the developers that GATEKEEPER is hosting for him and for which is granting security and privacy implemented into the platform and the deployment infrastructure. Data stored into the GATEKEEPER platform associated with a user are isolated from any other user.

Data collected and stored by a developer are only accessible and visible by that developer. For the applications developed on the top of the GATEKEEPER services, the developer is responsible to implement the adequate privacy and security mechanisms to avoid data breaches to his applications by using appropriate countermeasures.

Anyway, developers can rely on some core security and privacy services provided by GATEKEEPER that can help them to do this job, such as standard user management services for custom authentication and authorization, secure connections and communications provided by GATEKEEPER infrastructure services, intrusion detection systems (IDS) for network traffic generated by developers' applications, etc.

Furthermore, GATEKEEPER can provide to pilot developers some additional feature in terms of a federation of their physical resources (not only data) into the Gatekeeper platform.

The GATEKEEPER platform by default is deployed into the data centre provided by HPE[11] in Rome. Such data centre provides both physical security (security personnel, access control to the facilities, locks of the physical infrastructure) either IT security such as above-mentioned services IDS, etc. Based on that, by default the storage and data extraction operations are physically done into the HPE data centre.

Anyway, for the GATEKEEPER pilots more flexible deployment models are foreseen in order to join the GATEKEEPER platform:

1. **Standard deployment.** A pilot is part of GATEKEEPER and its data are stored in the same GATEKEEPER cloud owned by the project. In this case, GATEKEEPER is in charge of providing access to the data pilots wants to share. The storage and data extraction operations are physically done into the HPE data centre.
2. **Physical separation of data on private space of** GATEKEEPER **data centre.** A pilot is part of GATEKEEPER but its data are stored in a separate space into the GATEKEEPER data centre. In this case, pilot is in charge of the interface between the data and services around data pilot wants to share. The storage and data extraction operations are physically done into the HPE data centre but in private pilot spaces.

[11] HPE, Hewlett Packard Enterprise, https://www.hpe.com/it/it/home.html

3. **Physical separation of the platform on private space of** GATEKEEPER **data centre.** A pilot runs a copy of GATEKEEPER platform into its own physical space of GATEKEEPER data centre, data and services are not shared with GATEKEEPER platform. Additional developments need to be done by pilot developer in order to share all or some of its data. The storage and data extraction operations are physically done into a private cloud of the pilot.
4. **Physical separation of the platform on pilot premises.** Pilot runs a copy of GATEKEEPER platform into its own premises, data and services are not shared with GATEKEEPER platform. Additional developments need to be done by pilot developer in order to share all or some of its data. Security services such as intrusion detection system, encryption at rest, secure connections and communications, physical security for accessing the infrastructure should be implemented and provided by pilot. The storage and data extraction operations are physically done into the premises of the pilot.

These options allow to pilot to separate at different levels their physical infrastructure from the GATEKEEPER one. As figure 5 shows, the more the pilot separates its infrastructure, the more the management effort related to the maintenance of its deployments is increasing.

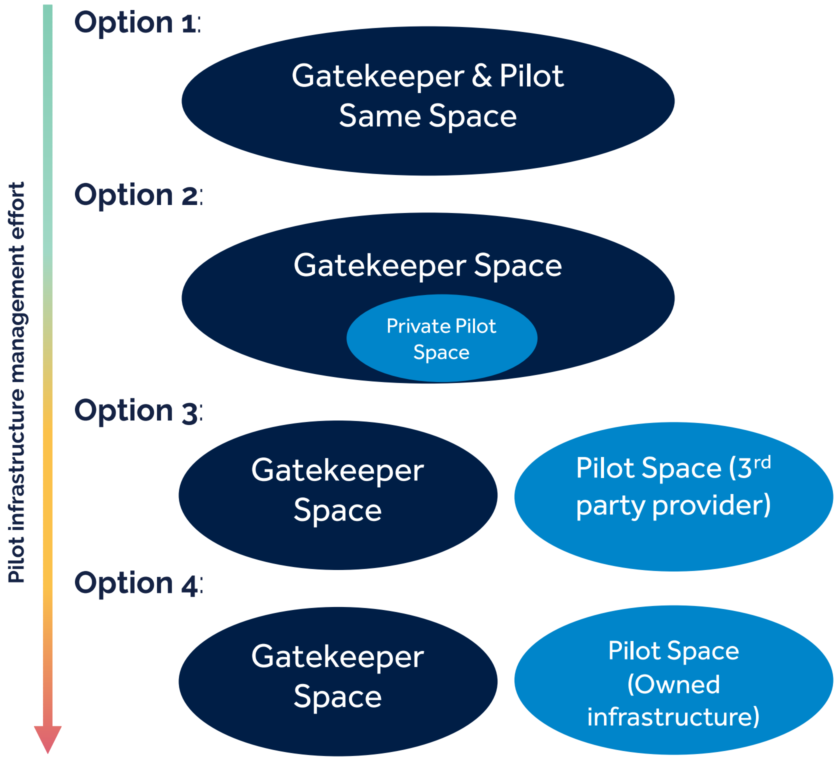

Pilot infrastructure management effort
Option 1:
Gatekeeper & Pilot Same Space
Option 2:
Gatekeeper Space
Private Pilot Space
Option 3:
Gatekeeper Space
Pilot Space (3rd party provider)
Option 4:
Gatekeeper Space
Pilot Space (Owned infrastructure)

Figure 6 - Pilot infrastructure management effort representation based on the separation of Gatekeeper infrastructure and pilot

6.1 Findings

The GATEKEEPER platform will be a key component of the project. As it will manage data also coming from the pilots it will be important to guarantee appropriate legal basis and organizational and security measures for data sharing in order to allow the project to create value from the data processed.

7 IPR Rights and licensing

Standard contracts will regulate the management of IPR (Intellectual Property Rights) in GATEKEEPER, and protect the intellectual property of third parties and beneficiaries involved. The Consortium Agreement, contains provisions regarding access rights.

Results from experiments are owned by the beneficiaries or third parties that generate them. Specifically, the product resulting from an experiment will be owned by third parties. Detailed IPR terms and conditions will be stated in the Consortium Agreement.

7.1 Intellectual Property Rights

IPR is a generic term that encompasses several different issues that are covered by different laws and practices. Generally, all issues related to *copyright*, *patents*, *trademarks*, *trade secrets* and *sui generis database rights* are collectively indicated as IPR.

IPR management is of fundamental importance in GATEKEEPER, because the main software artefacts that the project will release, will be distributed with the intent that Parties, either internal or external to GATEKEEPER, can use it freely. In order to grant Parties these rights, it should be carefully managed about the IPR distribution terms.

In the following paragraphs, we will discuss three basic issues about the knowledge the project are producing:

- Access rights: who will own the basic rights?
- Licences: under which conditions is the project going to exchange it?
- Use and dissemination: how will the project exploit it?

7.1.1 Copyright

Copyright is a corpus of laws that are harmonised in most nations in the world thanks to the Berne copyright convention. Copyright laws establish the rights that the authors have over their work. Copyright applies to most original and non-trivial works, be it writings, painting, music, most works of art and even software, both source and machine-readable code.

Copyright concerns the rights of copying, displaying, performing, printing, publishing, extending, modifying, translating a work. Application of copyright to software involves the rights to copy, modify or distribute the program. It does not involve the right to independently write a program performing the same actions as an original one. Generally speaking, the programmer who writes the program owns the rights. Where there is more than one programmer, the Directive (Directive 2009/24/EC) provides for co-ownership.

Software licences

A software licence is a legal document that accompanies a program. Without a software licence, according to the provisions laid down within the Berne copyright convention, *a program* cannot be distributed or modified *without the explicit permission of the authors*.

There are many kinds of software licences. Broadly speaking, we will divide software licences in two distinct classes: FLOSS licences and non-FLOSS licences. Licences belonging to the latter class are also termed proprietary licences. GATEKEEPER is mostly interested into FLOSS licences because, as stated in the Description of Work, GATEKEEPER execution platforms, tool components and service components shall be released as FLOSS software during the lifetime of the project. However, occasionally some programs or components may be used or written which are not released with a FLOSS licence.

FLOSS licences. FLOSS is an acronym originated in 2001. Its diffusion is mainly due to the Free/Libre and Open Source Software: Survey and Study commissioned by the EC in that same year. The term was coined to encompass all the different terms used to indicate the same class of software copyright licences. Among the FLOSS licences, two broad classes can be identified: *copyleft* and *non-copyleft* licences. In short, a copyleft licence allows the covered program to be redistributed only under the same licence: the licence can be said to be *persistent*. FLOSS licences that do not have this requirement are *non-copyleft* licences.

Copyleft

Copyleft software licences are a class of FLOSS licences that, like all FLOSS licences, permit unrestricted usage, copying and modification of the covered program. Like all FLOSS licences, they also permit modified or unmodified redistribution, but only using the same licence.

In practice, this means that when you obtain a program distributed with a *copyleft* licence, you cannot change the licence while redistributing it, whether or not you have made changes to it. Consequently, any copyleft program you distribute is granted to maintain its FLOSS state, whether it is modified or not. In other words, a copyleft licence *persists* through modifications and redistributions of a covered program.

The first copyleft licence, the GNU General Public License (commonly known as the GPL), is the most famous and by far the most used FLOSS licence, recently revived in its version 3. The Free Software Foundation is responsible for the GPL maintenance.

Non-copyleft licences are those licences that do not require the covered program to be redistributed under the same terms. Take the Apache licence as an example: you can lawfully get a program covered by the Apache licence and redistribute it with any other licence, even a proprietary (non-FLOSS) one, with or without modifications. Often the term *permissive* is used to indicate non-copyleft licences.

7.1.2 Patents

A patent is a set of exclusive rights granted by a national or international body to an inventor for a limited period of time in exchange for a public disclosure of an invention.

The procedure for granting patents, the requirements placed on the patentee, and the extent of the exclusive rights vary widely between countries according to national laws and international agreements. A patent application must include one or more *claims* defining the invention which must be new, non-obvious, and useful or industrially applicable. The exclusive right granted to a patentee in most countries is the right to prevent others from making, using, selling, or distributing the patented invention without permission.

Software patents are an important issue, because they can pose a real danger to FLOSS software. When a software method or algorithm is covered by a patent, the patent office has recognised the inventor's claim that the software method is original (never invented before), and non trivial (a knowledgeable person in the field would not be able to reproduce it from state of the art). The inventors have a 20-years monopoly on the exploitation of the software method, and no one can lawfully use it without their permission.

In Europe the law disallows patents on software per se[12].The legal status of the European software patents is unclear, though. Application of these recommendations depending a big extent on national laws, which are not harmonized.

FLOSS communities are particularly sensible to this risk, and in fact they avoid as much as possible to use software on which patent claims are known or suspected to exist. Modern FLOSS software licences often contain provisions against the most blatant abuses of software patents. The Apache licence, for example, contains some clauses that protect the software against the use of submarine patents: if a contributor's software is covered by a patent, and that contributor makes legal attacks against users of the software, that contributor loses all rights to using the software. The Mozilla, Eclipse and GPL licences all have some sort of protection against software patents.

[12]https://en.wikipedia.org/wiki/Software_patents_under_the_European_Patent_Convention

7.1.3 Trademarks

A trade mark is a distinctive sign, usually a word or a logo. Its usefulness is to give a brand to something and avoid that someone else takes credit for the product using the trade mark or distributes a different version of it with the same name.

7.1.4 Trade secrets

The most generic way of protecting IPR is to just not let slip the knowledge outside of the boundaries of your organization. For its very nature, this practice is utterly incompatible with FLOSS, which is based on openness. Keeping the development secret is a risky choice, because it can easily give the impression to outsiders that the openness of the developers is just a facade, rather than a real overall policy.

In GATEKEEPER, trade secrets would be kept to a minimum, and development should be organised around publicly-available repositories as early as practically feasible.

7.2 IPR Management in GATEKEEPER[13]

The Software IPR Directory will be kept under the responsibility of the Technical Manager. The directory contains entries that are compiled by the Partners who own the copyright on a given software artefact. Each partner delegates an IPR Manager to this purpose. The procedure for entering and validating data into the Software IPR directory will be determined in the next months.

IPR Delegates list

IPR Delegates (one per partner) should be appointed by all partners to the official responsible for the procedures described in this document on behalf of the represented organisation within the consortium. The appointed delegates along with the Technical Manager are the only persons that will have editing permission to the Software IPR Directory to register software components or programs of their authorship.

[13] Parts of the content of this section have been adapted from *Deliverable 9.3-D of universAAL Project*

7.2.1 Software IPR Directory

The Software IPR Directory is the document where we store intellectual property information about software.

From the exploitation and future business perspectives, it is considered of paramount importance that any piece of software used and produced in GATEKEEPER be registered here.

For work the Partners bring inside GATEKEEPER as Background, an entry should be created before it can be used. For work produced inside the consortium as Foreground an entry should be created at the start or as soon as possible from the start date. Foreground and Background are terms more precisely defined in the GATEKEEPER Consortium Agreement.

Entries in the database contain critical information about copyright holders, patents pending, software licence to be used, distribution terms and willingness to contribute it to further possible initiatives that continues the exploitation of GATEKEEPER assets. Consequently, the IPR directory contains important information, including commitments from the Partners. It is therefore critical that the data entered is reliable and non-refutable.

7.2.2 Duties of IPR Delegates

Before starting development of any software artefact, especially in case of *joint development* between Partners, a Partner's IPR Manager should create an entry in the IPR Directory. When filling it, the IPR Manager is taking a commitment on behalf of the Partner.

The critical information in the database is the software copyright licence the Partner plans to use and whether, to the Partner's knowledge, the software is encumbered by any patents.

The reason why the accuracy of the above info is crucial is that the development of the GATEKEEPER platform may in fact depend on it. In fact, Partners working on a specific software artefact should be sure that their plans for software licensing are compatible. Discovering an incompatibility later in the development cycle can produce waste of funded effort. A similar issue applies at a higher level, when the software artefact comes to become part of the GATEKEEPER platform. The Technical Manager and the Exploitation Manager are responsible for assuring that platforms pieces have compatible licensing terms, and that those terms are consistent with the aims of GATEKEEPER before giving clearance to the database entry entered by the IPR Delegates. The Exploitation Manager gives his assent based on the information entered in the database, whose accuracy is therefore critical.

Progressing the status of an entry. The IPR software database contains one entry per software artefact produced or provided by a Partner as Background. IPR Delegates will add entries to the first step, named *Submitted*. The Technical Manager and Exploitation Manager will check it and progress it to the *Verified* status. Normally, entries will stay in this status for the whole duration of the project. Should an entry become obsolete, it is progressed to the *Dismissed* state.

7.2.2.1 Fields of the database

When adding a new entry to the Directory, all the following fields should be filled with appropriate content.

1. *Name*

This is the short name of the software artefact, the one normally used when referring to it. Often short names are acronyms, but there is no general rule.

2. *Long name*

The long name of the software artefact. Often the expansion of an acronym, but not necessarily so. It may be the typical short description of the artefact. On occasion this may be empty.

3. *Version*

Software that has been released even in incomplete or early alpha stages should have a version number or identifier. For software that is yet to be written or in a yet unusable state this field maybe empty or may contain an identifier explaining its state of completion.

4. *© in GATEKEEPER*

This is a semaphore indicator. It is GREEN If all of the copyright holders are Partners of GATEKEEPER. It is YELLOW if some are Partners and some are not. It is RED if none of the copyright holders are Partners of GATEKEEPER.

Software that is Foreground, which means that it is written as part of the project effort and is written from scratch is GREEN. Software based on an already existing code base is GREEN only if the copyright holders of the existing code base are all Partners of the project.

Foreground software that is based on code written by non-Partners is YELLOW. It may switch to GREEN if, in the course of development, all the original code base is removed and rewritten.

5. *© holders*

This is a list of the copyright holders that are member of GATEKEEPER. If some copyright holders are outside the project, they are collectively mentioned as *outside* GK. The complete list of copyright holders will naturally be present in the copyright and licensing information that goes with the software itself.

If a number of Partners are involved in the development, the IPR Delegate of one of them shall enter a record with the name of all partners as copyright holders. Each of the other IPR Delegates are responsible to check that the entry is correctly included.

6. *© year*

This is the year of the latest copyright. It is the most recent year when the software has been modified. Usually, it is also the year when the latest version has been released.

7. Licence

This is the name of the software copyright licence used. As part of a join activity by Platform Cluster and Business Cluster, a **Plan and Guidelines for Software Licencing** for GATEKEEPER will be elaborated. A list of common licences to choose from will be provided and all procedures to determine the best possible licensing type for each software component in the IPR database. Both, the Technical Manager and Exploitation Manager will act as advisors for Partners as far as the choice of software copyright licence is concerned.

8. TM

Generally speaking, software that is part of GATEKEEPER should not come with trademarks, because that would complicate the management of software offering. This means that Foreground should not contain trademarks, and that usage of Background containing trademarks should be first discussed with the Technical Manager.

Put here a GREEN light if the Partner is declaring that there are no trademarks contained in the software, a RED light there are specific trademarks contained in the software. The YELLOW light is for all the less clear-cut cases: suspect trademark status of some given distinctive signs, trademarks whose owners have donated the right to use them and other situations.

9. Maturity

This is an indicator of the quality of the software, and it will normally change during the lifetime of the software, especially for Foreground software.

We use a GREEN light for software that has reached a stable state, where at least some versions exist that can be used with a reasonable confidence that they do not introduce software bugs in the system due to their low quality and instability. We use a YELLOW light for software that is usable and can be made part of a software system, but that is known to be unstable in some cases, or to contain known bugs that affect features used by the system, or that has minor but significant problems such as memory leaks, performance problems and the like. We use the RED light for experimental or alpha stage software that is known to be unstable and that can only be used for limited testing under controlled conditions.

10. Dependencies

This is a free field that should be filled with a list of all the modules, libraries and in general all pieces of software on which the software depends. Each entry on the list should include at least the name and version of the dependency, the date it was released and its software copyright licence. Reference to a web site can be a useful addition. Comments on the licensing terms or any other specific issue should be added of which the Technical Manager should be made aware of.

The accuracy of this list is important for the Exploitation Manager to be able to spot any licensing incompatibilities or generally criticalities in the licensing status of GATEKEEPER software offering. It is also important that it be updated every time the dependencies of the software are changed or updated. Entries should be added or deleted when dependencies change, and updates should be made for the used version number, release date and, most important of all, licensing term changes.

7.3 Findings

As previously highlighted, data are a key component of the project. While the traditional focus is on the management and process of personal and health related data, non personal and commercial data also have and will play an important role especially in the context of the Business Cluster of the project. Taking into account the actions highlighted before, the project will guarantee and appropriate management and valorization of non-personal data.

8 Conclusion and future work

This deliverable includes the best available information on data processes at the project level and at the pilot level in the development of the Gatekeeper project. It is the result of a collective work which has involved representatives of the different WPs and of the different pilot sites. It is to be understood as a living document that will be constantly updated during the next steps of the project and thanks to the contribution of the different participants to the project. Particular attention will be also devoted to the issues arising from data sharing in the context of multisite research[14]. The data management plan of the different pilots will also be constantly monitored and updated. The updated versions of the deliverable will be made available according to what has been agreed in the DoA. Important implications from the work done can be already highlighted:

- Data (personal and non-personal) play an increasingly important role in e-health;
- It is important to guarantee the compliance with legal norms and ethical standards for the treatment of personal data, pilots have therefore produce their own DMPs and also offered preliminary assessments in D1.5;
- The pilots, as data controllers, will have to guarantee the flow of data to the platform. Further work is required here to define and design the data sharing agreements;
- Non-personal data have also to be taken into consideration especially as far as IPR and licensing are concerned. The Business Cluster will be closely involved in all the discussions concerning data management;
- This document is understood by the Consortium ad a living document, it will be constantly updated through the involvement of the pilots and the different clusters.

[14] J.Scheibner, M. Ienca, S. Kechagia, J.R. Troncoso-Pastoriza, J.L. Raisaro, J.P. Hubaux, J. Fellay, E. Vayena, *Data Protection and ethics requirements for multisite research with health data: a comparative examination of legislative governance frameworks and the role of data protection technologies*, Journal of Law and Bioscience, 7, 1, January-June 2020, pp. 1-30.

Appendix A

A.1 Typologies of data in different pilots

A.1.1 Data Management Plan U.K. Pilot

Data summary

State the purpose of the data collection/generation

H2020 GATEKEEPER project is an innovation action and a large-scale pilot (40,000 people in 8 EU countries). The aim of the project is the co-creation and testing of technology for value-based healthcare, combing medical and consumer data in a trust secure ecosystem (the GATEKEEPER). The OU is involved in the role of leader of (1) the UK Pilot site and (2) a task 5.6 concerning Community and Robotic intervention.

Data collection is aimed at recruiting, running and monitoring the GATEKEEPER UK Pilot with the goal of testing and evaluating digital technologies for health and wellbeing of the elderly population. Data collection involves approximately 500 participants, of which 125 participants will be from the elderly population of Milton Keynes and 375 among the close network of the elderly participants (family, friends and community carers).

Explain the relation to the objectives of the project

Data collection is required to evaluate the impact of a set of technological solutions on the ability of elderly participants to live well and independently at home, and to access informal care from their community.

Specify the types and formats of data generated/collected

Data is collected through questionnaires concerning frailty factors related to social relations and their ability to carry out activities concerning their self-sustainment. Data is collected during the enrolment of participants and after six months. Elderly participants are provided with information concerning their close network of informal carers and the type and kind of support they provide. Questionnaires responses are collected as a CSV spreadsheet.

A second type of data collection concerns the test of robot intervention. This data collection is carried out through the sensors of a robotic platform in the participants' household (approximately five). These data are in JSON format.

Specify if existing data is being re-used (if any)

No data is being reused as there are no available datasets concerning the impact assessment of consumer technologies in a community-based care model.

Specify the origin of the data

Pilot participants are the source of data. Data is collected using questionnaires by the project team and an external researcher.

State the expected size of the data (if known)

~50 MB of questionnaires;

~1gb of sensor data from the robot.

Outline the data utility: to whom will it be useful

Data is used by the GATEKEEPER consortium to develop the evaluation of the Large Scale Pilot combining the data from the nine pilot sites. Aggregated data from the Large Scale Pilot is used by the EU Commission to evaluate the cost/benefit of the technology interventions tested in the UK Pilot. The data collection have a direct impact on future policies concerning the adoption of technology-driven healthcare solutions for ageing.

Making data findable, including provisions for metadata [FAIR data]

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.

Making data openly accessible [FAIR data]

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.

Making data interoperable [FAIR data]

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.

Increase data re-use (through clarifying licenses) [FAIR data]

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.

Allocation of resources

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

There are no expected extra costs. Part of the budget is dedicated to research staff with the responsibility to manage data (collection, quality, processing).

The research data that will be made public (aggregated data) will be shared through the OU ORDO repository.

Clearly identify responsibilities for data management in your project

Most of the responsibilities concerning the data management concern the project team.

The project team are responsible for collecting data through questionnaires. This task requires the collection of privacy consent and the consent to process and share data within pilot partners. Furthermore, questionnaire data collection requires monitoring of the quality of collected data and training of the project team supporting pilot participants in filling in the questionnaires.

The project team is responsible for documenting and storing collected data. This task requires creation of datasets, assigning metadata, anonymising records (assigning identifiers and storing personal data in a different repository) and creating and maintaining the project repositories. Lastly, the project team have the responsibility to destroy personal data and individual sensitive data at the end of the project.

An external researcher will have the responsibility to analyse data creating a new aggregated dataset for the evaluation of the pilot.

Describe costs and potential value of long term preservation

There are no expected costs concerning the long-term storage of aggregated data. Individual records concerning sensitive and personal data will be destroyed as soon as possible (after the analysis of data or at the end of the project).

Data security

Address data recovery as well as secure storage and transfer of sensitive data

Data will be kept in encrypted hard drives and stored through a secure network to a secure repository. Data will be uploaded in a secure repository as soon as collected.

Concerning the security guidelines, we refer to the OU Research Data Management: security guidelines http://www.open.ac.uk/library-research-support/sites/www.open.ac.uk.library-research-support/files/files/RDM-Data-security-guidelines.pdf

Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Data collection includes information concerning social activities and the ability and confidence of data subjects in performing basic daily activities. These information partially disclose the cognitive and emotional state of the person. These data will not be made available, but stored in a secure private repository (ORDO), decoupling personal information from sensitive data.

Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

The GATEKEEPER DoA and management includes a centralised unit and centralised management of data protection, privacy, gender and ethics. The rationale behind having project-wise management and standards is to support and monitor the related activities of each pilot site and to apply best practices project-wise.

Furthermore, we comply with the OU guidelines for ethics, data protection and privacy.

A.1.2 Data Management Plan Spain Pilot –Aragòn

This DMP is oriented to describe the specific circumstances that apply to data (mainly from patients) managed in the Aragón pilot of GATEKEEPER.

During the current phase of the project there are still many important decisions that might have a decisive influence on the DMP. The main decision to be made is which KETs are going to participate in the SALUD pilot.

The set of KETs participating in the project is very relevant for data management. There are many issues that have an important role on data management depending on the selected KETs as the technical architecture, the data flows, which data they use / generate or how each of them deal with data transmission and storage. Once the set of KETs selected is known, the description of this DMP will be described in more detail and therefore updated.

Purpose of the document

To provide a description of the data management in the context of the Gatekeeper Project pilot to be held in the region of Aragón (Spain). This document will serve:

- As a guide to SALUD to check specific points on data management;

- As a reference for the Ethical Committee to check the steps followed in data management;
- As a reference for the SALUD Security Committee to check the steps followed in data management.

Data managed during the GATEKEEPER pilot at the Aragón site will be performed mainly by software already running at SALUD premises but also by new technological components – KETs that will capture, manage and store data from SALUD Information Systems and from patients. Strict data protection rules apply to these KETs as even though GATEKEEPER is an Innovation Action and TRLs maybe are not adapted to market, data information belongs to real people so special category data rules apply.

Each company that incorporates a KET to the SALUD pilot and also the companies responsible for the GK platform, will have to provide specific information on data management including

- **Architecture of the solution.** A picture or diagram with information on
    - What would be the inputs / outputs/ transfers of information within the whole module / solution;
    - What are the expected flows of data between the components / subsystems of the software element;
    - Characteristics (distributed / centralized / others) and location of the repositories of data involved in the solution;
    - Annotations in terms of main topics related to data transmission, management and storage of data (e.g. SSL connection, encryption, anonymization procedures, …).
- **Data sets description.** Each dataset is a defined as a collection of information. Each dataset (or group of datasets) that are expected to be used in the solutions should be described by providing the information in the following list
    - **Dataset summary**
        - Dataset designation;
        - Dataset description;
        - Purpose of the dataset collection /generation. Alignment with the pilot and project objectives;
        - Characteristics of the dataset volume, frequency of capture, source / transfer / destination);
        - Data 'utility". To whom might the dataset be useful.
    - **Dataset access and storage**
        - Where is the dataset going to be stored;
        - How is the dataset going to be accessed (e.g. through any standard identification mechanism for location of resources e.g. Digital Object Identifiers?).
    - **Interoperability**
        - Codification. Standard / Non standard. If no standard, mapping / codebook of the format;
        - Metadata on dataset. Description (if any) and standard use.

- **Use of data**
    - Commitment about adequate use of data set during project and no use after the project;
    - Storage length and justification;
    - Resources needed for the preservation of data.
- **Resources**
    - Responsible for data management;
    - Quality assurance process for data management.
- **Security**
    - Provisions in place for data security (data recovery, secure storage, codification / encryption, secure transfer)
- **Ethical aspects. If any to be taken into account**

This collection of this information should also be backed up by a *"Data Privacy and Ethics Statement"* signed by each specific company. SALUD needs this procedure to be followed so as to:

- Ensure the compliance with the existing current legal and ethical context.
- Include information in the informed consent form. The idea is to provide information to the patient on how his/her data is going to be managed so he can make the decision on participation taking into account as much information as possible. Information from companies will be translated into a language easy to understand by patients who may not have a clinical / technological profile;
- Inform the CEICA (Ethical Committee of Research in Aragón) about the data management plan of each company so they can evaluate the appropriateness and compliance with normative;
- Inform the Security Committee of SALUD about the data management procedures of the project so they can check the correct fulfilment of the security and data privacy requirements;
- Inform the technical departments about the data management procedures so they can evaluate if there is any potential breach of the internal normative, and also to monitor and keep record of the software as for other applications used by SALUD.

Data Summary

Purpose of data

The purpose of Data Collection / Generation is the improvement of the (health) care provision to citizens thanks to services that can prevent the appearance of chronic diseases, detect their exacerbations or contribute to their monitorization during the acute phases based on the management of data from people and their provision to healthcare professionals directly or through systems that support them.

Relation to the objectives of the project

GATEKEEPER is built around a network based on European standards, interoperable and secure available to all the developers willing to create combined digital solutions for early personalized detection and intervention.

Types and formats of data generated / collected

- List of information to be collected in low complexity patients:
  - Questionnaires: quality of life e.g. E5QD, healthy habits (food intake, exercise, smoking, alcohol), depression / anxiety e.g. PHQ-9, level of dependency e.g. Barthel);
  - BMI (Body Mass Index);
  - Steps / Physical activity;
  - Sleep.
- List of information to be collected for mild complexity and high complexity patients (the difference between the two levels might strive in the frequency of the measurements and also on the accuracy that should be requested from the devices)
  - Low complexity patients information;
  - Heart Rate;
  - Blood Pressure;
  - Oxygen Saturation;
  - Respiratory Rate;
  - Weight. For instance, the capture of weight should be as automatic as possible and, in the case of high complexity patients a weight scale integrated in the bed of the patient might be very useful to detect the evolution of the liquids retention;
  - Body Temperature;
  - Dyspnoea;
  - ECG;
  - Medication intake.
- Other information that might provide added value to the project
  - Systemic vascular resistance;
  - Sweat level;
  - FEV – Forced Expiratory Volume;
  - Peak Expiratory Flow;
  - Glucose.

Origin of data

The expected contents and sources of information can be seen in the following list:

- **Data from patient** ("real time" / updated). Through sensors / wearables /devices that gather vital signs and other behavioural information (e.g. steps, sleep activity, ..) . All this information might be collected through an app in a tablet or through another device that acts as a hub to be sent to other IS for its management;
- **PROMs** (Patient Reported Outcome Measures). Through questionnaires performed to patients through an app in a mobile device / personal computer through an interface that might be a virtual assistant / digital coach;
- **Clinical data** extracted from the patient EHR ("re-used" data)
- Personal/demographic data (name, age, gender, education, marital status) for the appropriate and personalized management patient care;
- **Social data** extracted from different sources (EHR, SMARTCARE platform, social records, questionnaires) oriented to build the integrated care model that might apply for mild and high complexity patients ("re-used" data).

Expected size of data

The quantity of data will be single records for personal and demographic, social data and a subset of clinical data. During the project lifetime, vital signs and PROMs data will be collected with a frequency that needs still to be defined (it might depend on the characteristics of the KETs and also on the needs that might apply to, for instance, the development of predictive models). Expected and reasonable frequencies and quantities of data could be

- Daily questionnaires with 4/5 items for low complexity (RUC1)
- Weekly vital signs values for mild complexity (RUC2, 5, 7)
- Three times a day vital signs values for high complexity (RUC2, 5, 7)

Data utility

Data will be used

- To monitor end-user vital signs and behaviour so as to feed DSS and predictive models that can generate additional information from patient as probability of a chronic disease to appear, existence/ chances to suffer an exacerbation, clinical (un)stability;
- To feed predictive models that can generate information as the one described in the previous point

These utilities benefit patient, professional and healthcare organisation as well as those companies who can train their predictive models with real data from patients

FAIR Data

The first point to be taken into account is that the main objective of the data management for SALUD is to allow the care provision (the focus is not research).

This is one of the reasons why the FAIR principles are not as important as they could be in a research project.

- **Making data findable**. 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.
- **Making data accessible.** 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.
- **Making data interoperable.** 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.
- **Making data re-usable.** 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.

Allocation of resources

- The innovation unit of Barbastro Healthcare area will be in charge of the data management procedures related to Gatekeeper, they will serve as a link among the main stakeholders that will be implied in this context:
  - The regional ethical committee of research;
  - The regional committee of information security of SALUD;
  - The care staff involved in the project;
  - The patients / citizens involved.

Data security

Data security must be compliant with the legal context at regional, national and European level, including:

- Spanish National Scheme of Security (ENS National Blueprint of Security (ENS) RD 3/2010 8th January / modified RD 951/2015;
- Organic Law 3/2018 of Personal Data Protection and Assurance of Digital Rights. It is the adaptation of the National Data Protection Law due to the publication of the GDPR;
- Royal Decree 994/1999 that establishes the security procedures for the files that might contain personal data;
- Royal Decree 195/2000: is related to the set-up of these procedures defined in 994/1999.

In the context of the Aragón (SALUD) pilot of the GATEKEEPER project, data collected from patients will be stored at the local premises of SALUD. If it is strictly necessary (and justified) for purposes as training artificial intelligence algorithms, some of the data that will be gathered from patients, might be sent to the GATEKEEPER platform, which will be running and store information in servers that belong to the GATEKEEPER consortium. The final goal of the delivery of these data can be *training artificial intelligence algorithms* that may be able to predict worsening or other conditions of patients that have the same characteristics as the participant. In order to meet the requirements of the current legislation in terms of data protection, this data will be *gathered, processed, transmitted and stored* in an anonymized format or, if necessary, pseudo-. If data must be stored in a *pseudo-anonymized* format, SALUD will assign them individual codes. Health data will –using only the before mentioned individual code but not the patient name – sent to the GATEKEEPER platform servers through a secure connection, where pseudo-anonymized data will be stored and processed. Information about which code is assigned to each patient will be stored in the database of SALUD. No third party outside SALUD will have the chance to access this information.

Ethical aspects

Included in the Ethical Impact Assessment.

A.1.3 Data Management Basque Country

Data summary

The healthcare service of the Basque Country (Spain) takes part in GateKeeper project as a Pilot Site made up by two organizations, Osakidetza and Kronikgune, to promote active aging with actions aimed at maintaining the autonomy of people. It proposes comprehensive assessment mechanisms, interdisciplinary management, training in the use of new technologies and promotes volunteering and the maintenance of people in their usual environment. Data collection involves approximately 11,300 elderly participants from Basque Country and their caregivers.

The data generation is the first step on the pathway to information and knowledge. The collection of data allows improvements in personalized medicine and better outcomes for patients.

The data will be created and collected from different digital sources (wearables sensors, questionnaires, medical device sensors or application interaction) to be used within Gatekeeper project, and also from the databases of the Electronic Health Record (EHR) of the Basque Health Service. The variables that might be included in the data set are:

- Demographic characteristics;
- Clinical parameters;
- Diagnoses;
- Laboratory tests;
- Medication;
- Vital signs;
- Life style variables (activity, nutrition);
- Questionnaires (Patient Reported Outcome Measurements).

The variables will be text, numbers, images, codes and others and the format will differ (DOCSX, .TEXT, .PDF, XLSX, CSV format or .XLSX formats and local data models)

However, there are still some open questions:

- How data will be collected from the technical solutions;
- What type of data will they collect;
- In which format will they store the data;
- The estimated size of the data;

Metadata is data that describes other data. Metadata summarizes basic information about data, which can make finding and working with particular instances of data easier. In Gatekeeper, the Basque Country will include the metadata of the datasets within a separate file in a standardized way. Files will be uniquely identifiable and versioned by using a name convention consisting of project name, pilot site name, stakeholder type, participant ID, collection method used, collection date and analysis type.

- **Specify if existing data is being re-used (if any)**

Yes, but it is still undefined.

Origin of the data

The primary data to be used in the use cases will come from:

- Electronic Health Record;
- Personal Health Folder;

- E-prescription module;
- Data coming from the early detection and prevention technologies (wearable sensors, questionnaires, medical devices sensors, applications).

Expected size of the data

The expected size depends on the extent and the nature of the data that are made available. An estimation of the expected size of the data from 133,000 patients will be nearly 2 GIGAS (without images, only with texts and numbers)

Data utility

Aggregated data from the Large Scale Pilot will be used by the Gatekeeper Evaluation Team EU Commission to assess the cost/benefit of the technology interventions tested in the Basque PS. The data collection pretends to have a direct impact on future policies concerning the adoption of technology-driven healthcare solutions for ageing.

In the Basque Country PS, data will only be shared with the research and clinical teams during the project.

Making data findable, including provisions for metadata [FAIR data]

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.

Making data openly accessible [FAIR data]

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.

Making data interoperable [FAIR data]

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 needed in order to provide a mapping to the local vocabulary.

Increase data re-use (through clarifying licenses) [FAIR data]

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).

Allocation of resources

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

Costs for making the project data FAIR are eligible as part of the Horizon 2020 grant. At this current stage, no such costs are foreseeable.

Clearly identify responsibilities for data management in your project

Two Data managers will be designated within the Basque Country PS, one in Biocruces Bizkaia and one in Kronikgune (as each Institute manages the deployment of different use cases). This person will be responsible for the data management of the Basque Country PS. Some of the responsibilities include:

- Design, develop, and modify data management infrastructure to expedite data analysis and reporting.
- Implement policies and guidelines required for the data management.
- Review the data collected from the technological solutions for accuracy and quality.
- Develop standard operating procedures for data handling and archiving.
- Provide guidance in identifying and defining data requirements.
- Design and develop databases with the data collected.
- Maintain internal data asset library.
- Ensure the integrity, confidentiality and security of all datasets.

Describe costs and potential value of long term preservation

There are no expected costs concerning the long-term storage of aggregated data at this stage.

Data security

Address data recovery as well as secure storage and transfer of sensitive data

Osakidetza has established a path for dealing with appropriate incidents and events related to data protection breach. In the confidential document "Security Document For the protection of personal data and the security of OSAKIDETZA's information" for internal use of OSAKIDETZA, the procedures for incident management are detailed and the main roles are described. The "Security Commission" is the highest consultative and decision support body in the field of Information Security, for all OSAKIDETZA centres. The "Security Organisation" is established to identify, notify (to the Basque Data Protection Agency) and to take responsibility for the files affected by the Organic Law on Data Protection. It also includes within its responsibility the definition, implementation and compliance with the security procedures required by OSAKIDETZA. The "Security Officer" is the coordinator of the tasks and activities relating to information security carried out in OSAKIDETZA. In addition, OSAKIDETZA personnel handling Personal Data or Information in the area of Electronic Services to the Citizen, have to observe the security measures to ensure the integrity, availability, authenticity, traceability and / or confidentiality of the information they treat.

Violations of personal data security" are those incidents which, as a consequence, produce any of the following situations:

- Destruction of personal data;
- Loss of data;
- Accidental alteration of data;
- Illicit alteration of data;
- Unauthorized communication or transfer of data;
- Unauthorized access;
- Unavailability of data or lack of access.

For the purpose of the project documentation, the data will be stored according to the Basque Country organizational rules, and regulations and considering the security and privacy processes for the data protection. Storing of personal data will only occur with explicit prior informed consent of subjects, based on the informed consent procedures as laid out in D6.5.

Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Regarding the ethical and legal issues impacting data processing (data collection, data storage, data sharing, data accessibility and data retention), certain provisions have already been implemented in the project ethics work package (WP10). D10.4 provides information on the compliance of the consortium towards the collection of personal data and their handling over the life cycle of the project. In D10.2 information about the vulnerable individuals/groups that will be involved in the project, the measures and the risk of stigmatization is provided.

Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

The GATEKEEPER DoA and management procedures include a centralised unit for management of data protection, privacy, gender and ethics. The rationale behind having project-wise management and standards is to support and monitor the related activities of each pilot site and to apply best practices project-wise.

A.1.4 Data Management Plan-Cyprus pilot

Data Collection

What data will you collect or create?

The proposed pilot testing aims to recruit 400 people with dementia (AMEN) and 1000 patients with cancer (PASYKAF). Demographic information will be collated; however, will not be person-identifiable. Participants with dementia will be recruited from locked wards within the Progressive Neurological Conditions sub-pathway and from Home-based Care from the Archangelos Michael Elderly People Nursing Home / Rehabilitation Centre for patients with Alzheimer (AMEN) and patients with cancer will be recruited from home-based care services offered by the Cyprus Association of Cancer Patients and Friends (PASYKAF).

Once potential participants are identified, the patient's capacity to consent to participate will be established with a capacity assessment undertaken by a registered professional not part of the studies. The proposed pilot testing face a high probability that participants with dementia, or elder patients with cancer, will not be able to consent to participate due to the nature of aging and dementia. Where participants are considered to lack the capacity to consent, a representative will be contacted in the form of a friend, family member, or independent advocate that can consider consent as part of the best interest decision in accordance with the Mental Capacity Act (2005)[1]. Assent will be sought throughout the study for all participants and the procedure will be postponed or terminated if the assent is not evident. Participants who lack capacity will not be able to participate if their representative withdraws from the study. It is justifiable to pursue the proposed research even if individuals are not able to provide consent themselves as the benefits of potentially improving quality of life outweigh the alternative of not taking part. If well-received, similar intervention across healthcare service could be introduced to promote wellbeing in individuals living with dementia and cancer. In addition to the above, people with dementia and cancer will have the right to withdraw their participation or their data at any stage, before, during, and after taking part or providing their behavioural, physiological, and psychological data.

Data will be collected accordingly:

Physiological Data:

Vital signs: pulse;

Patient's mobility progression;

Sleep hygiene;

Psychological Data:

- Subjective report of physical pain
- Subjective reports of psychological symptoms: anxiety, panic, depressive symptoms, post-traumatic stress symptoms.
- Acceptance and Commitment Therapy (ACT) intervention followed by pre- and post- evaluations related to the effect ACT has on the participant.

We use data in the following ways:

All the data will be stored securely in a password-protected computer/online server and only accessible to the Cyprus research team. Once data are encrypted and coded, data analysis will be performed. Once the data analysis will be completed, the stored data will be deleted. Raw data will not be person-identifiable. The raw data will be kept for the project duration (4 years) and destroyed thereafter. Any information that is reported in the research papers and other dissemination practices in the write-up process with not be person-identifiable. Data that must be shared on the plan of Open Research Data will also be no person-identifiable.

The source of the data:

We will use these types of data sources:

- Behavioural data from sensors/ wearable devices/ devices for vital signs capture through a smart phone app;
- History of vital signs data;
- Participants' data will be collected through medical professionals or their familiars.

We will share data with:

- The GK Platform;
- The Cyprus Research team;
- Participants' data will be shared with the pilot partners to monitor the pilot and organize the deployment of technologies;
- Behavioural data will be collected, shared and stored by technical partner.

How will the data be collected or created?

The following quantitative and qualitative data (for which we have participant consent to share in de-identified form) will be collected as part of the project and will be available for sharing in raw or aggregate form. Specifically, any individual-level data will be de-identified before sharing. Demographic data may only be shared at an aggregated level as needed to maintain confidentiality.

The nature of the data is:

- Demographic (name, surname, age, residence address, email, and phone number)
- Physiological
- Behavioural data will be collected continuously from the wearable/sensors/apps and includes sleep analysis, vital signs, walking, etc. These data will be kept by our technology partner beyond the duration of the project in compliance with current GDPR Regulations.

We will be collecting and using the following data:

Physiological Data:

Vital signs: pulse, oxygen, blood pressure;

Patient's mobility progression;

Sleep hygiene.

Psychological Data:

Subjective report of physical pain;

Subjective reports of psychological symptoms: anxiety, panic, depressive symptoms, post-traumatic stress symptoms;

Acceptance and Commitment Therapy (ACT) intervention followed by pre- and post- evaluations related to the effect ACT has on the participant.

The data will be collected as:

1. Patient-level data ("real time" / updated);

- Through questionnaires /sensors/ wearables /devices that gather vital signs and other behavioural information (e.g. steps, sleep activity). All this information will be collected through an app in a tablet / through a web-based platform or through Virtual Reality;
- PROMs (Patient Reported Outcome Measures). Through questionnaires performed to patients through an app in a mobile device / personal computer;
- Clinical data extracted from the patient EHR ("re-used" data);
- Personal/demographic data for the appropriate and personalized management patient care;
- Social data extracted from different sources (HADS, QLQ-C30, EORTC Quality of Life – Core Questionnaire, ACT/ CBT Intervention) oriented to build the integrated care model;
- Health professionals - level data (quality of life questionnaire)
- Informal caregivers – level data (quality of life questionnaire)

Documentation and Metadata

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

No standards are known at the point of the DMP 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.

Ethics and Legal Compliance

How will you manage any ethical issues?

The project will undergo procedural ethical review to ensure its ethical soundness. Any unforeseen ethical issues will be discussed with the WP Coordinator to negotiate resolutions, though identifiable participant data will not be shared in such instances.

Informed consent for data sharing and long term preservation is included during data collection. Sensitive data will be separated as soon as possible and kept secure.

In this respect, both organisations will closely monitor the ethical regulations provided by the European Commission and the national bioethics committee and update our ethics accordingly.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

As indicated inthe ARTICLE 23a — MANAGEMENT OF INTELLECTUAL PROPERTY of the Grant Agreement.

Storage and Backup

How will the data be stored and backed up during the research?

All the data will be stored securely in a password-protected computer/online server and only accessible to the Cyprus research team. Once data are encrypted and coded, data analysis will be performed. Once the data analysis will be completed, the stored data will be deleted. Raw data will not be person-identifiable. The raw data will be kept for the project duration (4 years) and destroyed thereafter. Any information that is reported in the research papers and other dissemination practices in the write-up process with not be person-identifiable. Data that must be shared on the plan of Open Research Data will also be no person-identifiable.

How will you manage access and security?

A specialized GDPR consultancy firm will support us to achieve compliance in the most meaningful manner minimizing risks of GDPR non-compliance and consequently. We will also consult stakeholders with responsibility for information security, IT experts, sociologists, or ethicists.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Which data are you required to keep?

The data that are necessary to validate research findings will be kept.

What is the long-term preservation plan for the dataset?

Data will be kept for at least 10 years. After this time the data may be subject to deletion it has not been reused, accessed, or cited.

Data will be preserved and available for at least 7 years.

Data Sharing

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 (4 years) 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.

Responsibilities and Resources

Who will be responsible for data management?

For the proposed research, Dr. Maria Matsangidou and Mrs. Maria Krini will take the lead and responsibility for coordinating and ensuring data storage and access. However, the IT Department of both organizations will also be involved in managing, storing, and disseminating the results of the project. Both organizations, AMEN and PASYKAF will be responsible for checking that the plan is being followed.

What resources will you require to deliver your plan?

The costs for making the data FAIR are included in the project's budget and will be claimed if compliant with the Grant Agreement's conditions. Associated costs for dataset preparation and data management during the project will be covered by the project itself.

Long term preservation will result in no additional costs other than repository charges for data submission. The dataset will increase in value over the years because of its fundamental impact in the health field, also in the future.

A.1.5 Data Management Plan – Germany Pilot

Data summary

State the purpose of the data collection/generation

GATEKEEPER is a European Multi Centric Large Scale Pilot on Smart Living Environments, with 42 European partners and nine pilots across seven countries. Its main objective is to enable the creation of a platform that connects healthcare providers, businesses, entrepreneurs, and elderly citizens and the communities they live in, ensuring healthier independent lives for the ageing population. The TUD is involved in the role of leader of the Saxony Pilot site. Data collection is aimed at recruiting, running and monitoring the GATEKEEPER Saxony Pilot with the goal of testing and evaluating digital technologies for mental health and wellbeing of the elderly population. Data collection involves up to 10.300participants.

Explain the relation to the objectives of the project

Data collection is required to evaluate the impact of a set of technological solutions on the enforcing of mental health and well-being of elderly participants, promotion of healthy habits and better coping with mental illness, prevention of worsening health implications as well as safe detection and future protection from violent acts, and to enhance activity and social participation.

Specify the types and formats of data generated/collected

The intervention will focus on: Prevention of mental health disorders, screening and early detection of mental health symptoms. Data will be collected through validated questionnaires to detect and track mental health symptoms, improve self-management for prevention or better daily coping, facilitate mental healthcare use, enhance daily activity and social participation. Furthermore, measurements of physiological parameters, differential diagnosis of falls, screening of violence and speech recognition will be conducted to promote healthy habits as to detect and prevent violence and health impairments.

- **List of information to be collected in lowcomplexitypatients:**

  - Frequency of app downloads and frequency of usage;
  - Demographic data;
  - Mental health symptoms

  The frequency of questionnaire usage will be decided by the user

    - EFB (short screening questionnaire with demographics and PHQ4, PC-PCL5, and addiction questions);
    - GPS (International questionnaire on PTSD) **+ special questions on COVID-19 burden;**
    - ITQ (Complex PTSD);
    - LEC-5 (List of traumatic events);
    - PCL-5 (PTSD);
    - PHQ-D (Depression, anxiety and somatisation);
    - RS-13 (Resilience);
    - MoCA (Montreal cognitive assessment);
    - QMCI (Quick Mind Cognitive Impairment - Italy) or MMST (Hogrefe)

- **List of information to be additionally collected in moderate complexity patients**
  All mentioned above as well as the following:

  - Physiological measures (e.g. heart rate);
  - Movement patterns;
  - Sleep patterns;
  - Quality of life;

    - Global Assessment of Functioning (GAF);
    - EQ-5D or SF-36 or SF-8 or SF-12 or SF-6D (to be determined);
  - Level of irritability and mental health symptoms
    - PHQ-15, LEC-5, PCL-5, BDI II, FDS-20, BSI-18, HAQ, RS-13, SWE. The frequency of usage will be decided by the user
  - Social support;
    - The Multidimensional of Perceived Social Support;
  - Days of hospitalization;

- **List of information to be additionally collected in high complexity patients**
  All mentioned above as well as the following:

  - DERS (Difficulties and emotional relation Scale), BriefCOPE, PANAS (Positive and negative affective schedule;

  - Fall detection;
  - Violence screening;
    - Childhood Trauma Screener (CTS), and Conflict Tactics Scale (CTS2S). The frequency of usage will be decided by the user;
  - Health-Voice-on Device: Automatic Speech Recognition (ASR) and Natural Language Processing (NLP) (Audio recordings, typed input);

Specify if existing data is being re-used (if any)

Until now it is not intended to reuse data.

Specify the origin of the data

Pilot participants are the source of data. Data is collected using validated questionnaires and physiological measures by the project team.

State the expected size of the data (if known)

The size of data (per Patient) can only be estimated:

~1 MB of assessment data (questionnaires)

~2-10MB of device-, sensors-related data, speech recognition-related data.

Outline the data utility: to whom will it be useful

Data will be shared within by the GATEKEEPER project frame

Making data findable, including provisions for metadata [FAIR data]

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

Making data openly accessible [FAIR data]

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.

Making data interoperable [FAIR data]

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

Increase data re-use (through clarifying licenses) [FAIR data]

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.

Allocation of resources

All of the responsibilities concerning the data management concern the project team.

The project team are responsible for collecting data through questionnaires. This task requires the collection of privacy consent and the consent to process and share data within pilot partners. Furthermore, questionnaire data collection requires monitoring of the quality of collected data and training of the project team supporting pilot participants in filling in the questionnaires.

The project team is responsible for documenting and storing collected data. This task requires creation of datasets, assigning metadata, pseudonymizing records (assigning identifiers and storing personal data in a different repository) and creating and maintaining the project repositories. Lastly, the project have the responsibility to analyse data creating a new aggregated dataset for the evaluation of the pilot.

Data security

Address data recovery as well as secure storage and transfer of sensitive data

The collections, usage, and deletion of patient data are based on the data protection concept that is designed to ensure compliance with data protection regulations within the given organizational structure of the Faculty of Medicine at the Technische Universität Dresden. The securing and protection of survey data are regulated according to the requirements of §32 GDPR and §78a SGB X as well as its appendix. People involved in the study use 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". The original participants data remain in the corresponding clinic information system and must be archived in accordance with the relevant regulations. The explicitly derived, and pseudonymized data sets and calculation bases are stored on an access-restricted, clinic-internal server area

For further information we refer to the TUD Data Management security guidelineshttps://tu-dresden.de/med/mf/die-fakultaet/caruscampus/datenschutzgrundverordnung

Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

The main target user group of GATEKEEPER are adults over 55 years. Informed consent will be asked from all participants. Consent includes preserving data related to the participant for verification and reuse. In addition the burden of participants is carefully weighed and checked prior to and during the research activities. Ethical procedures are described more thoroughly in the ethical review form.

The participants in this research line will have the competence to understand the informed consent information. In the unlikely case that they are unable to do so, no activity related to the project will be conducted.

Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

This DMP will be embedded in the general GATEKEEPER project DMP. This DMP applies to research activities conducted by members of the TUD research team.

A.1.6 Data Management Plan - Greek Pilot

Data summary

State the purpose of the data collection/generation:

H2020 GATEKEEPER project is an innovation action and a large-scale pilot (40000 people in 8 EU countries). The aim of the project is the co-creation and testing of technology for value-based healthcare, combing medical and consumer data in a trust secure ecosystem (the GATEKEEPER). CERTH is involved in the project as the leader of (1) the Greek pilot sites in Attica and Central Greece, (2) Work Package 3 – "Gatekeeper Web of Things (WOT) Reference Architecture" and (3) tasks 3.1, 4.5, 4.6, 5.7 and 7.5 regarding industry platform requirements, secure data sharing and trusted transactions, marketplace services, technical validation, and technology developments respectively.

Data collection aims at recruiting, running, and monitoring the GATEKEEPER GR Pilot with the goal of testing and evaluating digital technologies for the health and well-being of the elderly population. Data collection involves approximately 1200 participants for the two use cases studied, namely Use Case 1 (UC1) and Use Case 3 (UC3). In the UC1 study, there will be involved 1000 participants, of which 40 healthcare professionals and 960 patients, whereas in the UC3 study 195 patients and 5 healthcare professionals.

Explain the relation to the objectives of the project:

Data collection is required to evaluate the impact of a set of technological solutions on the ability of elderly participants to live well and independently at home, as well as for lifestyle-related early detection and interventions. In UC3 specifically, data collected in a retrospective study will be used for the training and validation of a machine-learning algorithm for the short-term prediction of blood glucose.

Specify the types and formats of data generated/collected:

Data for both Use Cases 1 and 3 will be collected through a digital system, the Clinical platform for the management of chronic conditions. The platform collects sociodemographic data, clinical, data from IoT medical devices and sensors, and patient-reported data. The data are collected manually through the web and mobile interfaces of the platform or automatically through APIs (Application Programming Interfaces). The system stores the data in an HL7/FHIR format and has the ability to exchange data based on XDS and CDA documents. Furthermore, the system analyses the raw data and generates visual analytics with consolidated information. This information can be exported in JSON format. Finally, the system produces reports that can be exported in pdf and HTML formats.

Specify if existing data is being re-used (if any):

No data is being reused as there are no available datasets concerning the impact assessment of consumer technologies in a community-based care model.

Specify the origin of the data:

The source of the data will be either self-reported data by the participants or wirelessly transmitted data as recorded by the ICT technologies of GATEKEEPER. All participants will have previously provided permission (DPP, DPA) for the collection of the relevant data through either one of these sources with their written informed consent form.

State the expected size of the data (if known):

- sociodemographic data ~ MB;
- clinical data ~ MB;
- self-reported data (i.e. questionnaires, nutritional habits) ~ MB;
- data from IoT medical devices and sensors ~ GB;
- visual analytics ~ MB;
- reports ~ MB.

Outline the data utility: to whom will it be useful:

Data collected during the pilot will be used by the pilot users to support the respective objectives provided by GATEKEEPER. Furthermore, collected data will be used by the GATEKEEPER consortium to develop and implement the evaluation framework of the Large Scale Pilot. Aggregated data from the Large Scale Pilot will be used by the EU Commission to evaluate the cost/benefit of the technology interventions tested in the GR Pilot. The data collection has a direct impact on future policies concerning the adoption of technology-driven healthcare solutions for aging.

FAIR data

Making data findable, including provisions for metadata:

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

Making data openly accessible:

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.

Making data interoperable:

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.

Increase data re-use (through clarifying licenses):

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.

Allocation of resources

Explain the allocation of resources, addressing the following issues:

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs:

There are no expected extra costs. Part of the budget is dedicated to research staff with the responsibility to manage data (collection, quality, processing). The research data that will be made public (aggregated data) will be shared through the repository of CERTH in the first phase, and then through the HPE infrastructure.

Clearly identify responsibilities for data management in your project

Most of the responsibilities concerning the data management concern the project team. The pilot managers (HUA, DCCG) are responsible for collecting data through the digital platform. This task requires the collection of privacy consent and consent to process and share data within pilot partners. Furthermore, data collection requires monitoring of the quality of collected data and training of the project team supporting pilot participants in using the system.

The pilot leaders team (CERTH) is responsible for documenting and storing collected data. This task requires the creation of datasets, assigning metadata, anonymizing records (assigning identifiers and storing personal data in a different repository), and creating and maintaining the project repositories. Lastly, the pilot leaders team has the responsibility to destroy personal data and individual sensitive data at the end of the project.

Finally, the pilot leaders team (CERTH) will have the responsibility to analyze data creating a new aggregated dataset for the evaluation of the pilot.

Describe costs and potential value of long term preservation

There are no expected costs concerning the long-term storage of aggregated data. Individual records concerning sensitive and personal data will be destroyed as soon as possible (after the analysis of data or at the end of the project).

Data security

Address data recovery as well as secure storage and transfer of sensitive data

Data will be kept in encrypted hard drives and stored through a secure network to a secure repository. Data will be uploaded in a secure repository as soon as collected.

Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Data collection includes information concerning social activities and the ability and confidence of data subjects in performing basic daily activities. This information partially discloses clinical data related to metabolic syndrome (UC1) and diabetes (UC3) and the emotional state of the person. These data will not be made available but stored in a secure private repository (HPE infrastructure), decoupling personal information from sensitive data.

Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

The GATEKEEPER DoA and management include a centralized unit and centralized management of data protection, privacy, gender, and ethics. The rationale behind having project-wise management and standards is to support and monitor the related activities of each pilot site and to apply best practices project-wise.

A.1.7 Data Management Plan – Puglia Pilot

Data summary

Purpose of the data collection

In the frame of the GATEKEEPER Puglia Pilot, data will be collected to monitor either (i) healthy elderly subjects at risk of frailty and/or mild cognitive impairment as well as (ii) elderly chronic patients at risk of decompensations/exacerbations, depending on the specific Medical Use Cases to be addressed in the Pilot, as specified in deliverable D6.1 Annex A6. The purpose of such data collection - as established in the GATEKEEPER Description of Action - is to early detect possible health decays and enact in correspondence timely early interventions, aimed to mitigate such risks.

In addition to the above, related data on Health-Related Quality of Life (HRQoL) and on healthcare resource usage will also be collected, in order to conduct the cost-effectiveness analysis for the GATEKEEPER technologies involved in such early detection and intervention scheme, in line with the objectives of the project.

Relation to the objectives of the project

As mentioned, the above purpose is fully in line with the objectives of the GATEKEEPER project, that aims to assess the value that early detection and intervention platforms, supporting active and healthy aging, can generate when deployed at large scale across EU Regions.

Types and formats of data generated/collected

At the time of this writing, the following data categories are considered for inclusion in the data collection process:

- Behavioural data (e.g. on mobility, physical activity, socialization, IADLs, sleep quality, nutrition diaries);
- Clinical data related to the chronic conditions targeted by the Chronic Care Model programme enacted in the Puglia Region ("Care Puglia"): COPD, Type 2 Diabetes, Heart Failure, Hypertension (e.g. HR/HRV, blood pressure, oxygen saturation, glycaemia, body weight and composition);
- Geriatric scales (e.g. Lawton scale).

At the "edge", raw data will be collected in specific formats that depend on the specific technologies (e.g. sensing equipment, existing EMR repositories) that will be used in the Pilot. However, all datasets are expected to be stored "at rest" in standard formats compatible with the ones promoted by GATEKEEPER (constituting the GATEKEEPER "data space"), particularly FHIR for clinical data.

Re-usage of existing data

Data related to healthcare resource usage will be extracted from existing administrative databases, related to the Puglia Region's population.

Parts of the experiment that will involve hospitalized patients will also make use of existing hospital EMR data.

Origin of the data

Data will be collected through:

- Consumer devices, such as fitness wristband and apps (for data that will not be used clinically);
- Certified medical devices, such as glucose meters, body composition scales, PPG devices for measurement of various cardiac and other vital parameters, etc. (for data that will be used clinically);
- HRQoL questionnaires;
- Existing hospital EMRs;
- Existing administrative databases, for healthcare resource consumption data.

Expected size of the data

Although the precise size of the datasets to be collected is still to be precisely estimated, in order to provide perspective it can be mentioned that they will regard:

- Up to 10,000 people (healthy subjects), followed up for 12-24 months, for behavioural data;
- At least 500 people (chronic patients), followed up for 12 months, for clinical data, HRQoL data, resource usage data and geriatric scales.

When at rest, data will not be stored in "raw" format but at a higher abstraction level, as previously mentioned (as e.g. FHIR-based clinical data).

Although the above described data can be classified as "big data", since they will be collected longitudinally in a continuous way (periodicity has yet to be defined at the time of this writing), their size is expected to remain in the order of GBs, rather than TBs. Sufficient storage for such volume of data is expected to be available in the frame of the GATEKEEPER data storage infrastructures.

Outline the data utility: to whom will it be useful

During the project, the collected data will be useful to:

- Elderly people and chronic patients, for self-assessment and self-empowerment
- Healthcare professionals that care for those people, for monitoring them and early detect potential signs of health decay and/or

decompensation/exacerbations, in order to enact more effective and timely interventions

After the end of the project, the data will be useful to:

- Healthcare administrators, to assess the value of early detection and intervention technologies for active and healthy aging
- Researchers, to conduct additional investigations, particularly on predictive modelling for early detection of signs of health decay and/or decompensation/exacerbations

FAIR data

Making data findable, including provisions for metadata:

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.

Making data openly accessible:

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".

Making data interoperable:

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.

Increase data re-use (through clarifying licenses):

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.

Allocation of resources

Costs for making data FAIR

During the project, data management and FAIR compliance will be part of the activities covered in the project work-plan, that explicitly include such activities. These costs will be mostly related to data generation and transformation, according to the standards established in the project, as previously described.

Responsibilities for data management

Responsibility with data management will ultimately be with the data controlling entities, that will recruit the participants (e.g. Aziende Sanitarie Locali, i.e. county-level local healthcare authorities), possibly coordinated by "co-controllership" agreements that might be signed among them. The decision process regarding these aspects is on-going at the time of this writing.

Responsibility - as data processor - for data collection, transformation and storage during the project, will be with relevant partners in the GATEKEEPER project team, depending on the technologies and storage infrastructure that will be used, which in turn depend on the architectural design that will be established for the GATEKEEPER platform in WP3, WP4 and WP5.

Costs and potential value of long term preservation

As mentioned before, long term value of the data generated by the Puglia Pilot is expected to arise in relation to future usage for cost-effectiveness assessment of early detection and intervention technologies for active and healthy aging and in further research on predictive modelling for early detection of signs of health decay and/or decompensation/exacerbations.

For this reason - in the context illustrated in subsections before - long term preservation of appropriately anonymized subsets of the generated data will be sought for.

Datasets that are eligible for sharing will be made available through publication on deposition platforms such as Zenodo, with no substantial additional cost to be borne.

Data security

Address data recovery as well as secure storage and transfer of sensitive data

Similarly to all GATEKEEPER Pilot experiments, data management in the frame of the Puglia Pilot experiment will involve the treatment of a significant amount of sensitive, health-related data, as illustrated before.

During the project, such data will be stored and managed through the GATEKEEPER platform components (e.g. the GATEKEEPER data lake), that will include relevant functions for the protection of sensitive of personal data, including through appropriate "trust-building" measures, explicitly covered in the GATEKEEPER work-plan.

On the other side, before deciding on the sharing of such data for research purposes as mentioned in the previous sections, a thorough assessment of data security aspects will be conducted, according to the following principles:

- Personal identifiable information will never be publicly shared, neither will pseudonymized information;

- Before data will be openly shared, anonymization procedures will be defined and executed; such procedures will be selected in order to minimize potential risks for "re-identification";

- A decision to share data will also be based on the data minimization principle of the GDPR, by carefully balancing: (i) the expected advantages that can derive for patients and society by sharing data with the research community; (ii) the actual need of the data under consideration to achieve such advantages;

- Data used by the Puglia Pilot but not generated in the frame of the GATEKEEPER project (e.g. data on healthcare resource usage extracted from administrative databases, hospital EMR data) will not be publicly shared, as the responsibility about such sharing rests with the data originators.

Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Data related to the GATEKEEPER Puglia Pilot will be collected in the frame of an experiment involving human participants and will thus have to be managed accordingly, in particular through the preparation and the submission of the experiment protocols to the relevant Ethics Committees (a process which is on-going at the time of this writing).

Protocols definition will include, in particular, the procedures for obtaining the informed consent of voluntary study participants, according to current regulations.

Regarding the usage of data generated by the GATEKEEPER Puglia Pilot experiment, the informed consent form and the attached project information sheet will make explicit:

- The procedures that will be used to protect the identity of participants (e.g. sharing with the research teams only pseudonymized data, openly sharing only anonymized data);

- The data for which, after the application of the above mentioned protection measures, consent for usage to conduct the GATEKEEPER work will be requested, clarifying the benefits that will derive to the participants themselves and to society as a whole from such work;

- The data for which, after the application of the above mentioned protection measures, consent for sharing with the scientific community to conducting further research work will be requested, including the specification of the type of work that can be conducted and by what kind of institutions;
- All personal data that will not be shared as per previous bullets, will be deleted after the end of the project.

Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Additional specific rules for data management established by the specific local entities that will be involved in participants recruitment and participants' data control (e.g. Aziende Sanitarie Locali, as previously mentioned) will be investigated and duly applied, as necessary. When available, such provisions will be added to the next versions of this document.