This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 732679

ACTivating InnoVative IoT smart living environments for AGEing well

Data Management Plan

Deliverable No.	D1.4	**Due Date**	30-JUN-2017
Type	ORDP	**Dissemination Level**	Public
Version	1.0	**Status**	Release 1
Description	Document reporting the strategy to be adopted within the project regarding the data management, knowledge and IPR issues.		
Work Package	WP1 – Project Coordination, IPR and Ethics Management		

Authors

Name	Partner	e-mail
Sergio Guillén	04 MYSPHERA	sguillen@mysphera.com
Pilar Sala	04 MYSPHERA	psala@mysphera.com
Isabel Perez	04 MYSPHERA	iperez@mysphera.com
Saied Tazari	06 Fh-IGD	Saied.tazari@fraunhofer.de
Kostas Votis	08 CERTH	kvotis@iti.gr
Dimitrios Papageorgiou	08 CERTH	dimpapageorg@iti.gr
Giuseppe Fico	05 UPM	gfico@lst.tfo.upm.es
Martin Serrano	13 INSIGHT	martin.serrano@insight-centre.org
Aqeel H. Kazmi	13 INSIGHT	aqeel.kazmi@insight-centre.org

History

Date	Version	Change
17-May-2017	0.1	Structure of the document and task assignments
28-May-2017	0.2	Structure finalized & section 1 finished
30-Jun-2017	0.3	First draft with integrated contributions
21-July-2017	0.4	Reviewed Version with Contributions and DMP updates
25-July-2017	0.5	Pre-final Version with improvements and Contributions
8-August-2017	0.6	Final draft for internal peer review
23-August-2017	0.7	Final draft for peer review approval
31-August 2017	1.0	Final released version

Key data

Keywords	Data management, data sharing, datasets
Lead Editor	Pilar Sala, 04 MYSPHERA
Internal Reviewer(s)	Mario Diaz-Nava, 02 STM George Dafoulas, 39 DCCG

Abstract

This document defines the methods and conventions alike the recommendations for categorising about the use, manipulation and inclusion of data sets in the ACTIVAGE project. The document refers to regulatory aspects and operational information related to contact, personnel profiles details and about the ownership of the data within the project.

This document describes the data management plan and serves as a guide for the participants of the EU ICT ACTIVAGE Project about the data lifecycle with respect to the creation, identification, caption and description, storage, preservation (including security and privacy), accessibility, discovery and analysis, re-use and transformation of data in the context of the different deployment sites.

This document also includes sections related to intellectual property rights (IPR) with the objective to identify the different important ACTIVAGE's aspects. The data management plan and its contents shall be in accordance with the signed Grant Agreement of the ACTIVAGE project and with respect to EC Horizons 2020 recommendations.

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 management data plans in order to motivate participation and collaboration within the ACTIVAGE consortium partners and amongst external partners or participants in the project.

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.

Table of contents

List of tables

List of figures

1 About This Document

This document describes the initial version of data management plan and serves as a guide for the participants of the EU ICT ACTIVAGE Project about the data lifecycle with respect to the creation, identification, caption and description, storage, preservation, accessibility, discovery and analysis, re-use and transformation of data in the context of the different deployment sites.

This document also includes sections related to intellectual property rights (IPR) and data ownership with the objective to identify the different important ACTIVAGE's aspects. The data management plan and its contents shall be in accordance with the signed Grant Agreement of the ACTIVAGE project and with respect to EC Horizons 2020 recommendations.

This document also defines the strategy and principles as well as recommendations for categorising and defining data sets within ACTIVAGE project. . The document refers to regulatory aspects and operational information related to contact, personnel profiles details and about the data ownership within the project.

Considering the configuration to be done in the different deployment sites and the implementation of the different phases, including the launching of open calls, the current document can only reflect the intentions of the project partners toward developing the overall project's datasets. The second revision (D1.4.2) will be prepared at the end of the project, in which the complete set of datasets generated by the project will be reported.

1.1 Deliverable context

The adequate definition of a data management plan, and the methods and procedures for handling data guarantees the correct handling of information. Furthermore, European commission has requested to establish data policies and methodologies to achieve the goal of data policy across participant projects in the H2020 program. The content of this deliverable shall be aligned with the regulatory statements from the European Commission in terms of openness and distribution of information. 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 management data plans in order to motivate participation and collaboration within the ACTIVAGE consortium partners and amongst external partners or participants in the project. The use of data and its generation within the project is addressing the current demand for information exchange and expose data through electronics and non-electronic means. The following table summarises the overall aspects addressed by this document.

Project item	Relationship
Objectives	This deliverable serves all project objectives as it addressed the management of data generated by all the activities of the project.
Exploitable results	The deliverable contributes to the knowledge assets exploitable results as it will define how research data will be made FAIR, fulfilling EU H2020 guidelines on open access and open data.

Work plan	This deliverable reflects mainly the work done in Task 1.3 (Management of data, knowledge and IPR issues), from WP1 (Project coordination, IPR and Ethics management).
Milestones	MS1 - Build
Deliverables	This deliverable has strong links with: D1.5 Ethical manual, feeding from the guidelines regarding the ethics and legal aspects of personal data protection, that is one type of data that is addressed in this deliverable. D3.3 Security and privacy report, analysis and definition of guidelines and framework related to the security and privacy aspect of the project. D4.2 Data layer support tools, and its updates, serving as an input for the requirements of the tools to be developed and integrated in the AIOTES. D6.2 Tools and processes for the implementation of the evaluation methodology, serving as an input for the requirements of the tools that will be used to collect and process research data and resulting KPIs. D9.1 Detailed experiment and KPI definition, providing guidelines and templates for the definition at the level of the DS of the data sets and data management procedures.
Risks	Rk3. This deliverable contributes to identify and define data management processes that enable the correct application of the relevant regulations for data protection and takes, in the future versions, the risks analysis on security and privacy to be performed in the Task 3.2.. Rk11. This deliverable contributes to the definition of the procedures for data collection and for data quality assessment.

1.2 The rationale behind the structure

This document is organised in such way that can be used as a guide for understanding the data nature in the different phases of ACTIVAGE project. At the same time, it refers to the following three different levels of operation of the ACTIVAGE project:

1. **Operational Level - Deployment Site(s)**
   The operational level concerns the deployment sites where data is directly manipulated according to local methodologies and requirements,
2. **Technical Level - ACTIVAGE**
   The technical level concerns the ACTIVAGE level where selected data will be processed to enable global methodologies and overall requirements and the third one,
3. **Public Level – Users and Consumers**
   The public level is where the data is manipulated and can potentially be exploited and distributed for innovation/business purposes.

1.3 Version-specific notes

This document is a confidential document and for the sole use of the ACTIVAGE partners' consortium. This deliverable addresses the following audiences:

– Researchers and developers within the ACTIVAGE consortium, who will take into account the various procedures and data quality mechanisms in order to conduct research, designs and implementations within the ACTIVAGE project at deployment sites level and contribute to the global vision of the AIOTES implementation.
– Members of the European Commission, who can find in this document a readily available management and administration procedures and the quality control methodologies defined in the context of an international collaborative project.
– Other individuals that for nomination and approval from the ACTIVAGE coordination and/or representing the European Commission, who are authorized, invited or assigned to get access to this document.

This deliverable is the first version of the Data Management Plan, in which the framework for data management is being defined, as well as an initial identification of the possible datasets that could be collected during the execution of the experiments in each DS. However, due to the early stage in the project, and the existence of Open Calls incorporating new applications and services late in the project, an updated version of deliverable is foreseen at the end of the project (M42) where the final list of datasets will be provided with all needed details.

2 Principles of Data Management

2.1 ACTIVAGE Data management

ACTIVAGE project is aware of both the difficulty of exchanging variety of data models and at the same time profiting the data diversity of application requirements at the level of different deployment sites. Existing IoT data models vary widely in the form in how data is manipulated and how it is accessed using query data models, thus the problem is more particular on defining globally how to offer APIs, shared functionalities, and cross-optimization capabilities.

Semantic Web addresses many of the technical challenges of enabling interoperability among data from different sources. Likewise, Linked Data enables information exchange among distributed entities, i.e., distributed data providers, data-processing engines, data consumer mechanism with computer power to process meaning (semantics) of IoT data. In ACTIVAGE project the requirements and principles that will form the basis upon which the ACTIVAGE data management plan (DMP) will be defined based on those local and overall needs.

A DMP describes the data management life cycle for the data to be collected, processed and/or generated by the ACTIVAGE Horizon 2020 project. Data Management Plans (DMPs) are a key element of good data manipulation and exchange.

More specifically, the DMP shall be generated based on the EU Commission guidelines regarding the management of data requirements coming from projects funded by the H2020[1]. According to these guidelines, the data that is going to be shared for scientific, experimental and commercial purposes should be easily discoverable, accessible, assessable and intelligible. Usability goes beyond the original purpose of the objective of data collection and interoperability is related to ensuring following appropriate quality standards.

Furthermore, and due to the oriented nature of the ACTIVAGE project, additional but equally important attributes have to be considered, such as the **data security** and the **preservation of the participants' privacy**. In this direction, all the collected sensitive data of users will be protected from any unauthorized access, but also, they must be carefully anonymized to be shared through the proposed open data management platform of the project. In this sense, the guidance, support and results that Task 3.2 will provide, will be used through AIOTES, for the pilot execution.

Publication of data should always conform to the ethical guidelines of the ACTIVAGE project, defined in the Ethics Privacy Protection Manual (D1.5) and in the Ethical and legal reports (D1.6, D1.10, D1.11 and D1.12).

2.2 EU Commission Guidelines for data Management

The EU Commission has published some guidelines for appropriate data management plans (DMP) in Horizon 2020 projects. This guide for defining DMP is structured as a series of questions that should be ideally clarified for all datasets produced.

[1] ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

The following Table presents the main components of the DMP, the related issues to be addressed, and the actions taken in order to comply with them.

Table 1: EC guidelines to develop Data Management Plan

DMP Component	Issues to be addressed	Action taken in ACTIVAGE DMP
Data Summary	– State the purpose of the data collection/ generation – Explain the relation to the objectives of the project – Specify the types and formats of data generated/collected – Specify if existing data is being re-used (if any) – Specify the origin of the data – State the expected size of the data (if known) – Outline the data utility: to whom will it be useful	This component is addressed in Chapter 2 of this deliverable. Details of data set descriptions are given by each DS in Deliverable 9.1
FAIR Data. Making data findable, including provisions for metadata.	– Outline the discoverability of data (metadata provision) – 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? – Outline naming conventions used – Outline the approach towards search keyword – Outline the approach for clear versioning – Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how	This is partially addressed in Chapter 2 and 3. Further details will be given by each DS in Deliverable 9.1. Final description is done in Deliverable D1.13
Making data openly accessible	– Specify which data will be made openly available? If some data is kept closed provide rationale for doing so – Specify how the data will be made available – 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)? – Specify where the data and associated metadata, documentation and code are deposited – Specify how access will be provided in case there are any restrictions	This is addressed in Chapter 4. Further details will be given in D6.2 It will be further refined in Deliverable D1.13
Making data interoperable	– Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.. – 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?	This is addressed in Chapter 2. Further details will be provided in several deliverables of WP3 and WP5, dealing with semantic interoperability It will be further refined in Deliverable D1.13
Increase data re-use	– Specify how the data will be licenced to permit the widest reuse possible	This will be addressed in Deliverable D1.13

(through clarifying licences)	– Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed – 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 – Describe data quality assurance processes – Specify the length of time for which the data will remain re-usable	
Allocation of resources	– Estimate the costs for making your data FAIR. Describe how you intend to cover these costs – Clearly identify responsibilities for data management in your project – Describe costs and potential value of long term preservation	This will be addressed in Deliverable D1.13
Data security	– Address data recovery as well as secure storage and transfer of sensitive data	This is partially addressed in Chapter 2 and 3. Further details will be given by each DS in Deliverable 9.1. Final description is done in Deliverable D1.13
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	This is addressed in Deliverables D1.5 and D1.10
Other	– Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)	Not applicable

2.3 Approaches for Personal Data Management

The EU guidelines referenced above do stand that focus is not only given on personal data but that privacy concerns are also taken into account. However, the only mention about personal data is limited to a question from an ethical perspective and as follows: "*Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?*" In addition, under the rubric "Data Security", the above table requires that the transfer of sensitive data shall be addressed in data management plans.

In ACTIVAGE, the LSP on Smart Living Environments, it is mandatory to have a very clear strategy with regard to the personal data and privacy protection, mainly because it is expected that such environments have the following characteristics:

– Smart Living Environment are assistive in areas of private daily life, which means that these environments have to collect related sensitive data in order to be able to recognize situations in which assistance is needed;
– Smart Assisted Environments are most likely to take over critical tasks, such as turning on and off a cooker/whole electricity and opening the entrance door. Without the right management, these actions could wrongly allow the physical access to the environment to external people.

ACTIVAGE project has investigated and studied the data regulatory aspects in current European frameworks and follows closely the new European General Data Protection Regulation (GDPR) process and its outcomes that will enter in force on 25 May 2018. The data management plan introduced in this document is designed and described considering the coming modification in the European GDPR.

"The EU General Data Protection Regulation (GDPR) replaces the Data Protection Directive 95/46/EC and was designed to harmonize data privacy laws across Europe, to protect and empower all EU citizens' data privacy and to reshape the way organizations across the region approach data privacy"[2].

In ACTIVAGE, four roles are so far identified that are used in the context of the Data lifecycle and that will be reflected in the ACTIVAGE architecture design and implementation alike the IOTES services and tools suite. These roles are the following:

1. **Individuals:**

   Any person that participates directly and/or indirectly in the ACTIVAGE project at the local environments (deployment sites) and/or actively executing the ACTIVAGE overall consortium activities and that could be the generators or consumers of data, ACTIVAGE services and tools operators.

2. **Data Operators:**

   Any individual with granted permissions and capacities to manipulate and perform ACTIVAGE-related activities to support, deploy and maintain data services.

3. **Data sources:**

   Those generators of data (individuals and/or physical/virtual elements) that by being part of the ACTIVAGE ecosystem are able to follow the data formats and regulations to produce data defined in ACTIVAGE and

4. **Users of Data Services:**

   The persons that will consume data, services and make use of the tools from ACTIVAGE.

The flow of consents or permissions to use the data is separated from the actual lifecycle that defines the flow of data following the principle of security by design and aligned with the current European regulations in terms of data protection.

Based on the existing guidelines and approaches for data ACTIVAGE decided to use "MyData – A Nordic Model for human-centered personal data management and processing"[3] as an important related reference for the project.

In ACTIVAGE the rights and the level of control that individuals have over their personal information can vary. The minimum requirement for MyData is that individuals have the right to access and use their personal data.

The Helsinki Institute for Information Technology (HIIT) is maintaining an open project for the detailed specification of the related concepts under https://github.com/HIIT/mydata-stack. This provides ACTIVAGE with the opportunity to become a contributor that may be able to influence the further refinement of the specification in order to make sure that ACTIVAGE requirements are fully met. Section 4.1.2 on "Data as managed by AIOTES" will get back to this point later.

[2] http://www.eugdpr.org

[3] A white paper funded by the Finnish Ministry of Transport and Communications, available under http://urn.fi/URN:ISBN:978-952-243-455-5.

Figure 1 describes MyDate Principles (1, 2, 3 and 4) and depict the data flow and defined the roles that ACTIVAGE will follow in order to implement the data lifecycle and data management plan described in this document.

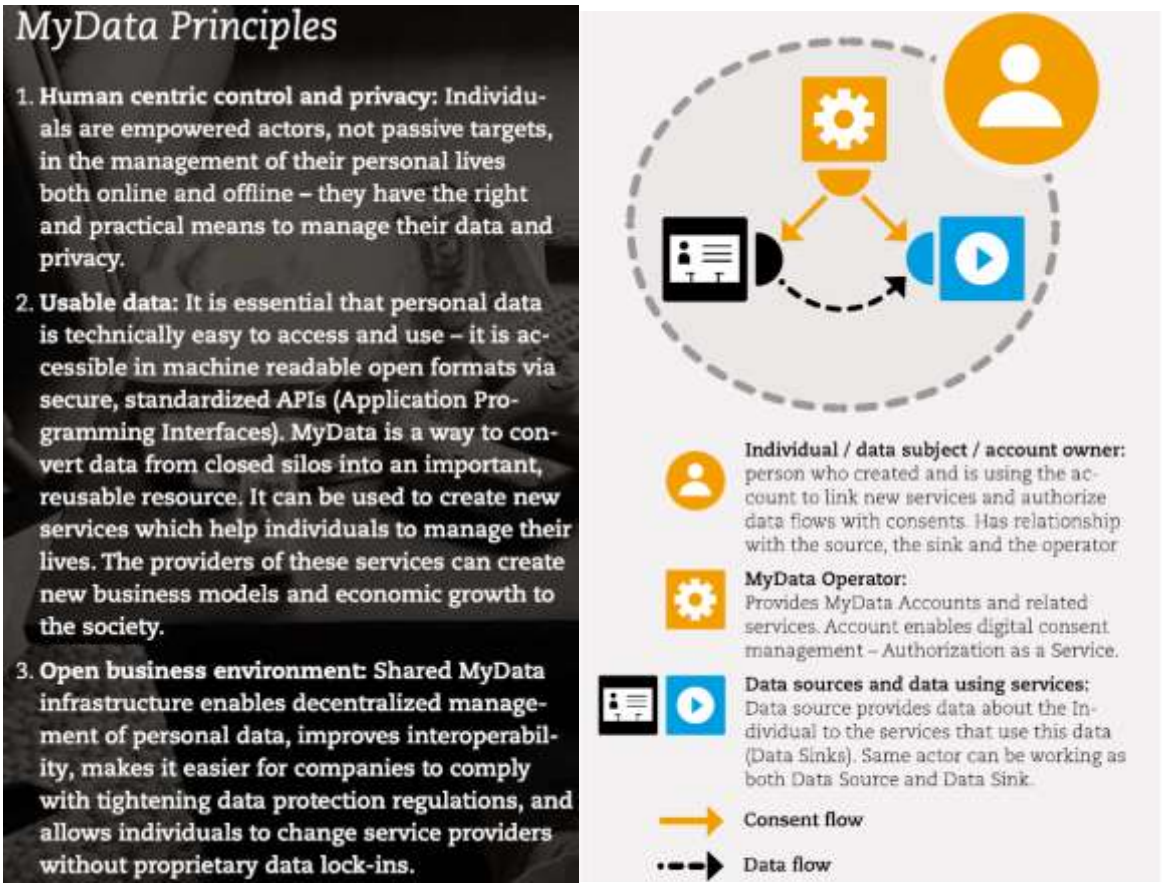

Figure 1: MyData Principles and Data flow adopted by ACTIVAGE

Figure 2 shows the data rights, an extract from the MyData White Paper, on what the related concepts are about data protection and rights in terms of data ownership.

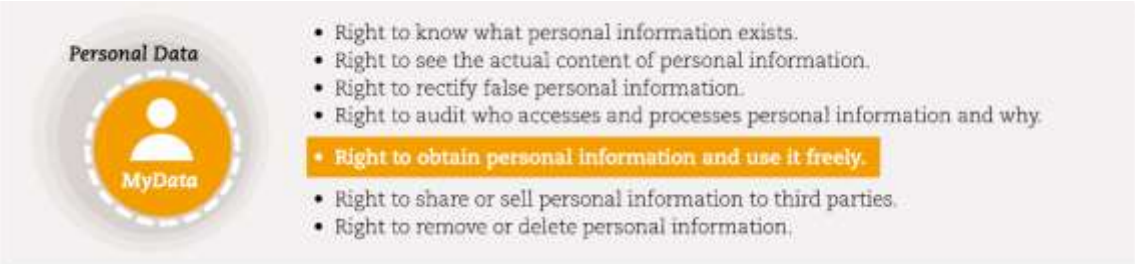

Figure 2: Personal Data Rights according to MyData and adopted by ACTIVAGE

3 ACTIVAGE Data

The Data Management Life Cycle in ACTIVAGE follows a user-data centric approach that describes all the process for data creation, data storage, data processing, data anonymization, across the data value chain from Deployment Sites operations to making data openly accessible.

Figure 3 shows the project data management life cycle. Next sub-sections of this chapter describe more in detail each one of the elements of this cycle.

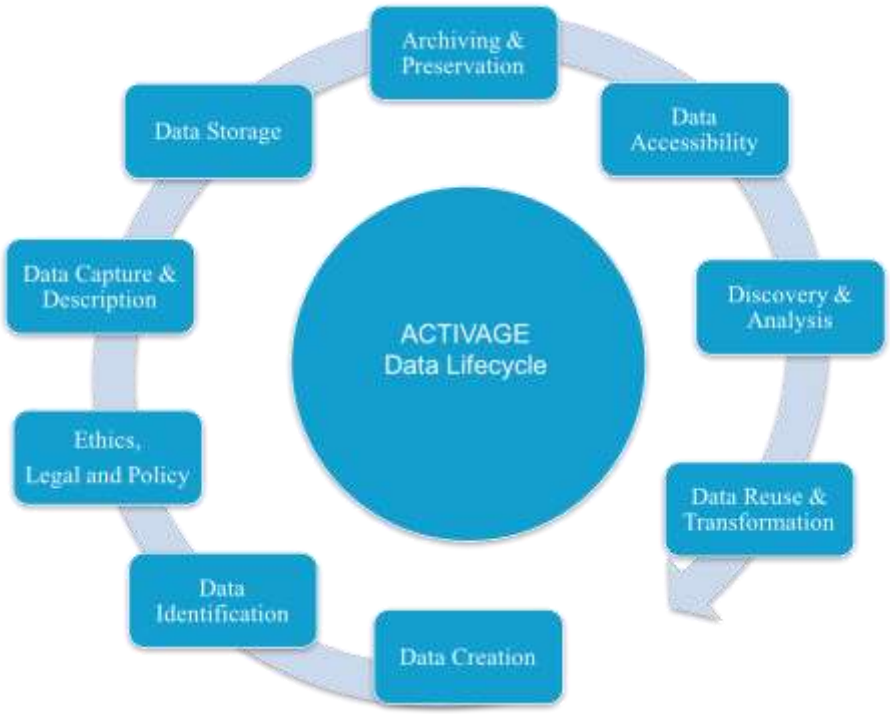

Figure 3: ACTIVAGE Data Life Cycle

Based on data use, data characteristics could be different, or at least such ones necessaries in the data lifecycle that has been identified. In ACTIVAGE five inherent characteristics of the data are the following:

- Data is the result of a design process
- Data follows the nature of the content by means of its purpose
- Data is part of a data management plan (formats, storage etc.)
- Data follows the principal objective of the plan / consent for sharing
- Data captures and creates metadata.

3.1 ACTIVAGE Data Lifecycle Management Model

In ACTIVAGE, once the data lifecycle has been defined, the next priority is to design and implement a management model following the European recommendations as much as possible and addressing the identified requirements in the different ACTIVAGE deployment sites (described in D2.1).

The ACTIVAGE Data Life Cycle Management Model is composed of three main layers, as illustrated in Figure 4, and described as follows:

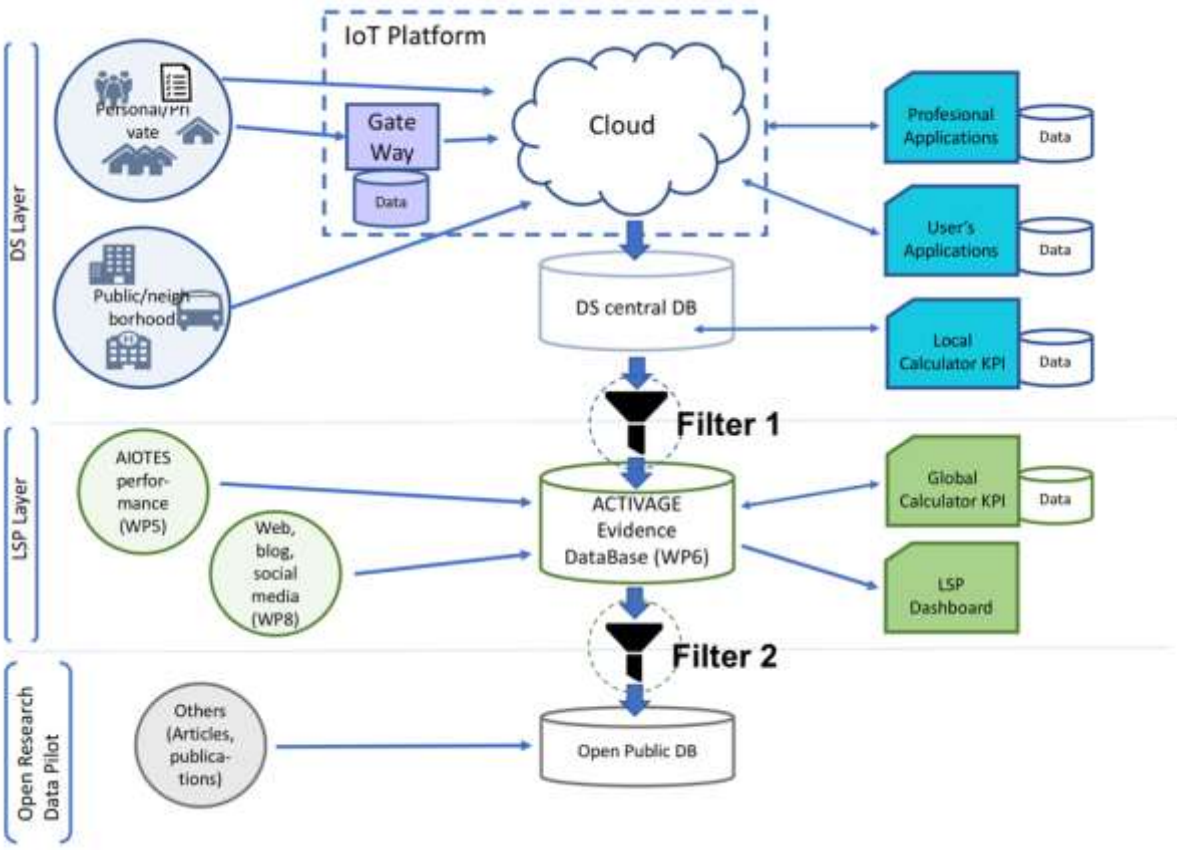

Figure 4: Life cycle Management Model

1. **Deployment Site layer** is the major producer of data. Data is produced through both, a creation process of raw and computed data (see section 2.1.1); and the KPI calculator process that computes Local Key Performance Indicators. Data will be computed at different levels of the creation process: i) at sensor level (raw sensed data), ii) at the edge level (computed data), by built-in IoT platform services, e.g. fall detection, in-doors location and so on; iii) at cloud level by built-in IoT platform services, e.g. out-door location, geo fencing and more; iv) at application level; and v) at AIOTES level. All these computed data will be stored in the ACTIVAGE DS # DATA BASE, which will be built, managed, secured and protected according with the local regulation under the responsibility of the Deployment Site members and Local Data Managers (see section 2.1.3 for more info regarding legal and ethical aspects).

   One important data processing stage in the DS level is the FILTER 1 Process. Filter 1 performs the conditioning of data for transference to the lower layer, such as data anonymization, annotation, description, standardization of formats and identification. The complete procedure and details about how data will be conditioned and transferred to WP6 shall be negotiated between DS and WP6 representatives: the first version will be described in D6.2 and D9.1 in Month 9.

2. **Large Scale Pilots Layer (WP6 layer)** is where GLOCAL KPIs are computed and interpreted (see D6.1 for more details about GLOCAL KPIs). The sources of data are multiple and varied: i) the data coming from the upper layer; ii) data collected from project activity in social networks and web media, provided by WP8.3; iii) data provided

from AIOTES, provided by WP5. All these data are further processed to derive new KPIs to assess the evidence and the impact of the project. Similarly to the upper layer, a FILTER 2 process will deal with the conditioning of data to transfer to the Public Portal layer.

Open public access platform layer consists of a data repository inherited from H2020 project "MYAIRCOACH", where filtered and conditioned data coming from WP6 will be made accessible under the defined conditions for re-use by third parties on research and innovation projects. In addition to this, other projects results will be made available through the portal, like deliverables, published papers, presentations, blog posts, opinion articles and so on. Furthermore, the H2020 Open data management portal will be also taken into account.

3.2 Data Creation

Data creation is the initial step in any lifecycle. In ACTIVAGE, data is created and transformed in new data, or information or both, in different parts of the life cycle process. Data is created by many different sources within this process and at this initial point data emerges in the form of an icon, a symbol, a formal/informal representation or verbal.

In ACTIVAGE, the primarily sources are sensors that measure physical or biological data. These are usually called RAW data, however in our domain we will consider raw all input data used to produce computed data or more higher abstraction level information. Computed data is the results of aggregation and computing of raw data (we do not consider internal computing at a sensor level, i.e., a GPS.

3.2.1 At Deployment Site layer

Deployment sites are the first layer at which data is going to be created in ACTIVAGE project, initiating the process of data collection.

Data to be created at DS layer is highly heterogeneous, and it is mainly the result of two processes:

– Operation of an IoT ecosystem, where sensors and other data sources will generate data.
– Execution of the experiment, where research data will be created from questionnaires, interviews and other instruments.

The initial and NON-EXHAUSTIVE list of types and categories of data that will be produced in ACTIVAGE across the three layers of the data management lifecycle model is:

– Personal data: It is created by the interaction of a person within ACTIVAGE and it could come from different sources, such as:
  - Obtrusive data: it requires the direct interaction of the person, either with the IoT infrastructure or with the researchers, the individuals are then aware of the fact that they are being studied, which can influence their responses or behaviour. Some examples of this are data coming from:
    - Questionnaires – Interviews
    - Wearable sensors
    - Medical device sensors
    - Application interaction

    - Unobtrusive data: it is created indirectly by the person, it doesn't require an explicit interaction, subjects are then not aware of the fact that they are being studied and therefore its research does not affect their response or behaviour.
        - Presence sensors
        - Location sensors
        - Etc.
- Environmental data: It is created without intervention of a person, such as:
    - Home environment sensors (gas, CO2, water, room temperature, etc.)
- Public Data: It is data already available in public repositories, such as:
    - Smart City data (traffic, transport, public events agenda, etc.)
    - Smart Environment (real-time pollution, forest fire detection, weather forecast, etc.)

All DS will define in their D9.1 report the data that will be created in their deployment site with all required details.

3.2.2 At LSP Layer

Data can be managed and created within the LSP layer. The main categories of data generated are:

- Local KPIs: information that is collected at DS level, but is used in the context of the Evaluation Framework, to verify the achievement of goals and outcomes that are relevant at local level and that will be grouped into "Boxes" to answer to "Big Research Questions", as described in D6.1
- Global KPIs: information that is collected at Local level that is common to all DS, this means that it has a global and a local relevance at the same time
- GLocal KPIs: information that enable to see at a glance how "an item is measured or a big question is managed throughout the different DSs and, consequently, in those different European ecosystems" (from D6.1).
- IoT KPIs: information that is generated by AIOTES and that provides values and indicators that can be used to demonstrate the "impact" of an IoT approach to the AHA domain
- Social network and media: information that is collected during the Dissemination and Communication activities that can be used to assess the project impact, in line with the activities defined in D8.1.

3.2.3 At Public Layer

The ACTIVAGE project participates in the Pilot on Open Research Data launched by the European Commission along with the Horizon 2020 programme. The consortium strongly believes in the concepts of open science, and in the benefits that the European innovation ecosystem and economy can draw from allowing reusing data at a larger scale. Therefore, all data produced by the project can potentially be published with open access after ensuring anonymization for private data. According to the "Guidelines on Open Access to Scientific Publication and Research Data in Horizon 2020": Research data refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, and observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form."

Thus, some of the following project datasets will be available through the project:

- ACTIVAGE Project results - deliverables.
- ACTIVAGE Scientific publications
- ACTIVAGE Project results - Software tools and AHA services
- ACTIVAGE Project results – Data sets
- ACTIVAGE Dissemination material for AHA and Interoperability platforms

3.3 Data Identification

The data identification points to where data will be used, as well as providing any needed generalization and/or categorization of the data. Considering that data is a representation of a fact (virtual or physical), its identification and intended usage can modify the way it is communicated.

Identification is a process where the data is not only identified but categorised and classified with the purpose of outlining:

- the discoverability of data (metadata provision)
- the use of identifiers of data and refer to standard identification mechanism. E.g. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- naming conventions used
- the approach towards search keyword
- the approach for clear versioning

3.3.1 At Deployment Site layer

Data identification at DS will be done jointly with the process of specification of the datasets, initially described in D9.1 and updated and maintained in the following versions of this document. Main categories of data assets to be identified by DS comprise both data created during the operation of the IoT system and research data for evaluation purposes. In all cases, some aspects that will be addressed during the specification of the datasets to facilitate identifiability and discoverability of the data are:

- Source of the data
- Category of data, as listed in previous section
- Provision of metadata, including keywords
- Reference to standard identification mechanisms, making use as much as possible of persistent and unique identifiers (DOI)

3.3.2 At LSP Layer

The definition of a *KPI* is by default constituting an attempt to aggregate different sources and pieces of information (*Performance Indicators*) in a meaningful way (*Key*). From a data identification perspective, this means that a metadata description will be used to associate and link the different elements described in the previous section. WP6 partners are currently analysing whether an ontology has to be created to provide a semantic description of the data generated in this layer.

In terms of data identifiability and standards identification, those indicators that are coming from subjective information, such as questionnaires or scales, are based on existing standards instruments and methods, therefore also their representation and codification will be standardized. As regards data coming from “objective information” such as sensors, device logs, user interaction with systems, standardized indicators and instruments will be

used, such as the ISO 9241 covering ergonomics and Human Computer Interaction for usability analysis.

As regards "search-ability", not only keywords will allow user to access and search the different elements through web-like navigation, but different views will allow different type of users to access and see the information according to their profile (e.g. Demand Side partners will see KPIs and outputs to verify whether their service is being successful in terms of cost-effectiveness and user acceptance; the *Supply Side* partners will see whether the learning curve of their device is lowering appropriately over time).

As regards versioning, names and coding will allow to handle progressive and incremental versions of the information generated. Critical, in this sense, will be the granularity of information that has to be described in the study protocol, in particular about the end-points and the phases of the study (baseline, demonstrate, expand, growth and sustainability).

3.3.3 At Public Layer

At public layer we can identify the following broad categories of data to be identified:

- **ACTIVAGE Project results - deliverables** as they will be produced throughout the project and summarize the important results, main findings, lessons learned and strategies selected.
- **ACTIVAGE Scientific publications** as they will translate the results of the project to scientific knowledge to be exploited by researchers and AHA technology specialists
- **ACTIVAGE Project results - Software tools and AHA services**, refers to all software tools produced by the project and its potential refinement to create commercial products.
- **ACTIVAGE Project results – Data sets,** refers to the research data that will pass through the selection filters across the previous layers and that supports the scientific research and associated publications.
- **ACTIVAGE Dissemination material for AHA and Interoperability platforms** as they will be used by the project for the dissemination of the objectives and results of the project as well as the increase of the ACTIVAGE user base

3.4 Ethics, Legal and Policy

The definition of ethics is clear and straightforward, in terms of users' data but in its relationship with IoT up to now the ownership is not so obvious. In the context of Europe regulations currently there is an intense discussion among policy and regulatory bodies on how to define adequate best practices or regulation for data management in large IoT ecosystems especially when the data centres are cross-border. Recently, 9 leading UK universities formed a research consortium (IoTUK[4]), to study the ethics of IOT adoption within businesses and the public sector. Also, the findings from ETICA[5] project from Jeroen van den Hoven (Chair of the WG Ethics in the IoT EG) as well as the BUTLER EU project[6] indicated the possible controversial ethical topics within the IoT ecosystem and compared the legal regulations with the customer's expectations. Recently the 5th Annual Workshop (2015) of the IGF Dynamic Coalition on IoT (through the funded IERC project SMART-ACTION) indicated the "IoT GOING ETHICAL" motto in order to get a common

[4] https://iotuk.org.uk/

[5] *www.etica-project.eu/*

[6] http://www.ehcobutler.eu

understanding of what it takes to go forward with IoT in a sustainable way-what "ethical" in the sense of IoT means.

In ACTIVAGE, all stages of the data management lifecycle must follow a number of ethics, legal and policy requirements steaming from the European and National legislation and guidelines, being the more relevant:

– The Universal Declaration of Human Rights and the Convention 108 for the Protection of Individuals with Regard to Automatic Processing of Personal Data and
– Directive 95/46/EC & Directive 2002/58/EC of the European parliament regarding issues with privacy and protection of personal data and the free movement of such data. This one will be replaced next year by the EU General Data Protection Regulation, thus all activities involving data management will take this one into consideration.
– The new EU General Data Protection Regulation (GDPR) that will replace the Data Protection Directive 95/46/EC to harmonize and re-shape all data privacy laws across Europe, to protect and empower all EU citizens.

3.4.1 At Deployment Site Layer

ACTIVAGE and its IoT relevant dimension might be the cause of ethical & legal problems. According to Managing the digital world report[7], an ethical behaviour requires a) enforce the property rights on information; b) ensure the access to information; c) ensure the integrity of the information; and d) enforce the right to privacy. In 2018 the European Commission will issue a recommendation amongst the participants' states in the union for Data protection, a report that ACTIVAGE project will follow closely. In ACTIVAGE at the deployment site every individual shall needs to be ensured that they will be protected by effective IoT technical solutions. In this direction, trust is a fundamental issue since the IoT environment is characterized by different devices which have to process and handle the data in compliance with user needs and rights. More details on the ethical aspects of each Deployment Site can be found in Section 3.3 of ACTIVAGE Ethics Manual (D1.5).

This will be done in full compliance with any European and national legislation and directives relevant to the country where the data collections are taking place, according to the ACTIVAGE Ethics Manual (D1.5)

3.4.2 At LSP Layer

The current measures have been already taken, resulting from an interaction between UPM personnel involved in ACTIVAGE and the independent Ethical Committee at UPM, when reviewing the conditions for the LSP layer creation:

All digital information will be stored using cryptographic methods, applied both at edge and server levels. Only users will have access to edge information and only a subset of this information will be sent to UPM servers. The information that is planned to be permanently stored will be further encrypted through DES/ECB/PKCS5Padding, as a first approach. Additionally, the project Task 3.2 will perform a risk security and privacy analysis, to verify if this method is solid or new measures should be taken, according to the new GDPR and current state of the art in terms of security and privacy. Additionally, the project will adopt recommendation and standards provided by ENISA (https://www.enisa.europa.eu/).

[7] Valacich, J., Schneider, C., Information Systems Today. Managing in the Digital World, 4th Edition, Pearson Publishing House, Boston, 2010, p. 484.

3.4.3 At Public Layer

In accordance with the ethical and legal requirements regarding data obtained from human participants, the ACTIVAGE public datasets will be initially available to the Consortium Members and only after its careful anonymization. Personal information regarding the participants will be kept strictly private.

Furthermore, as the project progresses and the collected data are used for the research and development processes of the project they will become available for the project's open data platform after the approval by the local ethics committees of the ACTIVAGE project. The inclusion of a individual's data in the public part of this dataset will be done on the basis of appropriate informed consent to data publication.

A recent Ethical design for the Internet of Things[8], was designed providing a new approach for users' interaction with the IoT, based on "Ethical Design" implemented through a policy-based framework (the SecKit policy- based framework), which can give more control to the user. More details about the SecKit policy- based framework and some ideas about its implementation in ACTVAGE can be found in Chapter 7 of Ethics Manual (D1.5). Last but not least the EU General Data Protection Regulation (GDPR) shall also be taken into consideration.

3.5 Data Capture and Description

In ACTIVAGE annotation procedures shall be used to coordinate at the deployment sites on how to collect the data locally and Globally (GLOCAL). In some cases, it is a questionnaire to the be presented to the user before and after the pilot; in other cases, is the result of tracking the usage of the system. It is a process for gathering the information according to timesheets from care professionals to assess the available data and make it machine understandable; etc. Therefore, it is expected that at local level each deployment site will have its own data format and collection method that will be connected that will be connected to ACTIVAGE Global Evidence Open Data Base collection tool. All sites will share the same tool for the same data formats, as well as the same procedure to collect the data. The following are the envisage steps/phases involved in data capture and description.

- Data Generation
  - Data and information that is generated from ACTIVAGE Project that will be consumed and/or handled within the ACTIVAGE project by means of electronic and non-electronic forms, e.g. data sets that will be produced as result of management and operations with other sources of information.
- Data Registration
  - Data and external information to be sued by the project that will be derived to the ACTIVAGE project by means of electronic and non-electronic forms, e.g. data sets that will be provided to the ACTIVAGE project to store, handle and or processing.
- Data Exposition
  - Data and information that has been registered and or produced in the ACTIVAGE project and could be distributed by means of electronic and non- electronic forms, e.g. data sets that will be accessible via interfaces to get the information used and/or

[8] Gianmarco Baldini, Maarten Botterman, Ricardo Neisse, Mariachiara Tallacchini, Ethical Design in the Internet of Things, Science and Engineering Ethics 2016,pp 1-21.

generated by ACTIVAGE allowing that those data sets are used for informative process.

– Data Management
  - Data and information that can be registered and produced within ACTIVAGE mediums, by means of, electronic and non-electronics forms that are used within the consortium for processes, operations and or simply usage for the normal activity within the project.

3.5.1 At Deployment Site layer

At this layer, there are two main categories of data assets that will be captured:

– Data captured by the IoT infrastructure: In this case, each DS is defining the mechanisms for data capture and management, with the particularity that in the case of the data generated by the IoT system, it is following a semantic description approach based on standard formats.
– Data captured by the evaluation protocol: this data will be captured both in paper and by electronic means and in most cases in local languages, so the DS are defining the processes to digitize, transcribe and translate as needed, to prepare the data to be aggregated and transformed in local KPIs.

3.5.2 At LSP Layer

In the definition of the Evaluation Tool (D6.2), the mechanisms for data transfer, quality check and semantic annotation of the information between the main data sources and the LSP data layer will be provided in detail, about the main categories (DS, AIOTES, and Social media & channels). Based on the study protocol, on the IoT KPIs, and on the impact attainment strategy, orchestration mechanisms and rules for data aggregation and quality control will be performed to derive Local, Glocal and Global KPIs. Accordingly, notification and alerting triggers will be defined to inform users about correct and coherent generation of information to be then captured and described in the LSP layer.

3.5.3 At Public Layer

During the specifications of the open data platform, mechanisms for automatic capture of data from the upper layer will be defined, including the generation of the metadata needed to appropriately identify and describe the information made available through the platform. Additionally, manual capture of data is enabled by the use of web-based templates and forms in the user interface, to allow uploading the main data assets generated at this layer (deliverables, publications, dissemination material, etc.)

3.6 Data Storage

The Data is an asset that itself has a value, storing this data provides the capability to preserve value. Data storage is the process when data needs to be stored for a later managing or processing purpose apart of the initial one to describe and communicate. When data has to be stored there are implications that until this point in the data lifecycle didn't exist, the most relevant are the format and how the data will be stored, some characteristics in this process can be summarize as follow:

– Format to be used for storing data
– Duration of the storage

- Plan for reusing the data
- Intended community that uses this data
- Data protection for security and privacy reasons

3.6.1 At Deployment Site layer

In ACTIVAGE according to differences in the background and reference ecosystem among the ACTVAGE deployment sites, specific Pilot targets shall be detailed at local level according to local background and expected impact. Evaluation Data concerning every GLOCAL KPI per sites will be collected in a global data-repository called "ACTIVAGE Evidence Open Data Base" that will be specifically implemented and exploited by ACTIVAGE project. This repository is fed mainly from all the KPIs being collected at each site apart from the evidence that supports those KPIs: questionnaires, usage statistics, usability analysis, costs analyses, personal opinions from the end users about the services and solutions. Everything is collected in this repository that becomes essential for the success of ACTIVAGE mission.

3.6.2 At LSP Layer

The objective of the project is to generate evidence and produce impact, especially during the last phase of the project, which is sustainability. In this perspective, the LSP layer is planned to go beyond the project duration, and serve as a basis for future enrichment in terms of data and functionalities. In order to achieve this, the information that is going to be stored should be kept in these records and re-used. The conditions for duration and re-usage have to be negotiated with the different stakeholders, which in this case are:

- Local Ecosystem within each DS
- International Ecosystem, represented by the key partners of the Consortium, i.e. those partners making part of the Steering Board

Therefore, it is likely that WP9 and WP8 will interact with WP6 to achieve such agreements in terms of IPR, Exploitation Plans at local and global levels, as well as in terms of AHA and IoT ecosystems.

3.6.3 At Public Layer

ACTIVAGE shall provide the Public Evidence Website, a user interface to be accessed openly by any person or professional interested in accessing the underlying ACTIVAGE Evidence Open Data Base, by accessing pre-defined reports or making queries that allow creating composition of underlying data, and project data analytics. This tool is foreseen to remain active and available after project's end in order to offer a rich interface to the whole ACTIVAGE evidence.

3.7 Data Preservation

Preserving the data sometimes is equally treated as the process of storing data, however in ACTIVAGE and following data preservation best practices, this is a process to guarantee a durable threat for the stored data. In a wider definition preservation of data defines the process of formulating and defining the ways to that guarantee the data will remain during the time and that will be useful for the same purpose which it was created and stored, some general characteristics about the data in this process can be summarize as follow:

- Migrate the data to best format

- Migrate the data to suitable medium
- Back-up stored data to preserve the information
- Create metadata and documentation for the stored data
- Archive data in a defined physical or virtual medium
- Secure mechanism against threats to delate, change, stole, … data

3.7.1 At Deployment Site layer

The DS partners will identify the data to be preserved based on the level of interest for the research community. On one side, subsets containing intermediate research data, coming both from the IoT solutions or the evaluation protocol, will be preserved and shared for validation purposes, improving the availability of current public datasets and ground-truth in the field of Active and Healthy Ageing.

Preserving these datasets and making them publicly available will require in the first place, the selection of the data to be shared and the preparation of the research data associated with it, this will be done when transferring this data to the LSP layer. In a second place, it will also require the maintenance of the data and the management of their access that will be done at the level of the Public layer by means of the use of the Open Data Platform.

As for the rest of the data managed by the DS, preservation is not necessary, meaning it is already preserved by other partner or that they do not provide an added value for the research community (internal data generated during the project development). This does not avoid that partners preserve it for themselves in their archives if they consider it useful for their research and innovation activities.

In order to preserve data, the security and privacy guidelines to be defined in the D3.3 will be followed in each DS as well as in the implementation of the ACTIVAGE architecture.

3.7.2 At LSP Layer

Being great part of the data based on the information generated through the execution of the pilot, the conditions that have to be complied at each DS level to run the pilot activities are paving the way for the strategy that the LSP layer will adopt to preserve the information. In other words, and taking into account the new GDPR directives, Ethical Committees are already imposing that proper backup and preservation strategies are in place during the pilot execution.

UPM as one of the main partners in charge of the LSP layer will likely be hosting the tool in their premises. Further details will be provided in D6.2.

The ethical committee of UPM has required that concrete details about the preservation of the data are provided. The following preservation mechanisms have been defined so far, and approved by the UPM's Ethical Committee (translated from the report):

"Servers will be physically installed in the facilities of the UPM ETSI Telecommunication Faculty, in the computing centre of the faculty, with limited access. These servers are configured as Virtual Machines of a DMZ, behind the firewalls of the LST research group, telecommunication faculty and university. Only those ports that are needed for the connection with the project's client will be open and enabled".

3.7.3 At Public Layer

This layer is the main recipient of the data to be preserved, and will rely on the use of the Open Data Platform described in section 5.

Regarding the particular data assets that are generated in this level, public reports, including public deliverables and open-access papers on journals or conferences will be also preserved and shared whenever possible (when they do not limit future exploitation plans), along with the research data necessary for validating the published results.

All dissemination material produced during the project, including demonstration videos and project public presentations, will be preserved and made it public as soon as possible to let the research community know about ACTIVAGE solutions and results in a more graphical way.

For the public reports and dissemination material, no much extra effort is considered for its preservation beyond the act of publishing them both in the website of the project and in the Open Data Platform.

3.8 Data Accessibility

Providing access or define the adequate ways to access the data is one of the main objectives in ACTIVAGE. Data Accessibility is a complex process because based on the demand for the data is the way to facilitate data access. If the demand is coming from multiple parts the access process is multiplied, first because there are activities associated that are specific for every part demanding access to the data and second because if the data is not prepared to be used by multiple parts it has to be pre-processed to repeat parts of the lifecycle in generating new data. Data access also defines the ways to:

- Distribute Re-distribute the data
- Share and format of the data
- Security and privacy mechanisms to control the accessibility
- Establish copyright for the data
- Promote data

3.8.1 At Deployment Site layer

At Deployment Site level, accessibility will be defined according to the following roles:

- **Producer:** They are the stakeholders from each DS value chain that produce or generate the data that is managed by the IoT system or by the research activities. Unless otherwise agreed, those are the data owners and thus they will have full accessibility to the data they have produced.
- **Controller:** They are the stakeholders that receive the data produced and maintain it according to the regulations, taking into account that it could exist several controllers depending on the data produced (i.e. a technical service provider could be the controller for the data collected from the IoT system, and the service providers for the evaluation data – questionnaires, interviews, etc.)
- **Processor:** They are the stakeholders that process the data, generating new data, knowledge or KPIs, taking into account that it could exist several processors depending on the processes (i.e. an operator of the IoT system is in charge of generating behaviour patterns and generate alerts, while a public authority will generate the Local KPIs).
- **Consumer:** They are the stakeholders that consume data, taking into account that it could exist several consumers depending on the different data flows and IoT applications (i.e. it could be the relatives, that consume alerts and daily activity of the elderly through the application, but also the service providers that can visualize patterns, or the technical provider that can visualize technical system performance). Access will be granted to these consumers based on the explicit consent of the data owners.

– **Supervisor:** It is the stakeholder that is in charge of ensuring full compliance with ethical and legal regulations regarding data protection including security and privacy.

3.8.2 At LSP Layer

Three main categories of users have been defined so far:

– Users of the ACTIVAGE Public Evidence Website: lay users and professionals that can benefit from being informed about the project's results, evidence and impact.
– Users of the LSP Dashboard: this is the view that provides access to project partners in charge of evaluation activities, at DS level (i.e. WP9 leader and DS leaders) and at AIOTES level (WP3, WP4 and WP5 leaders)
– Users of the AHA Advisor: third parties and end-users, informing about services, solutions available that provide benefits and values to them

Accordingly, access to data and meta-data will be granted to these users.

3.8.3 At Public Layer

In accordance with Grant Agreement Article 25, data must be made available upon request, or in the context of checks, reviews, audits or investigations. If there are on-going checks etc. the records must be retained until the end of these procedures.

Access Rights:

– Project partners: Current ACTIVAGE partners must give each other access -on a royalty -free basis -to data needed to implement their own tasks under the action, where is legally and practically possible
– Future ACTIVAGE partners, must give each other access –under fair and reasonable conditions (Article 25.3)–for exploiting their own results to data, where is legally and practically possible
– Affiliated entities (from open calls, etc.): Unless otherwise agreed, access must be given, under fair and reasonable conditions, and where is legally and practically possible.

3.9 Data Discovery and Analysis

ACTIVAGE focuses on providing tool for data discovery and this Discovery is one of the main objectives while Analysis has more implications based on a specific demand of the data discovery following the original representation and basic information provided when the data was created. Discovery can also look at the other processes in the lifecycle in order to enrich the capacity to look at those characteristics. Discover an Analysis reside together in the ACTIVAGE lifecycle simply because the demands for the information is to generate a meaningful outcome that helps to inform, observe, visualize features of the data.

3.9.1 At Deployment Site Layer

At ACTIVAGE deployment sites the Data Store will integrate a Data Discovery Engine to capture real time data from devices and sensors and perform advanced analytics to discover and extract data value from the different deployment sites (WP9) Pilots data. Advanced large-scale data Analytics mechanisms to extract maximum value from both historical and real-time data stored in the Data Store will be analysed and integrated too.

3.9.2 At LSP Layer

In ACTIVAGE define and implement a 'data lake' infrastructure for Data Discovery operations processes; including cataloguing, indexing and searching of all ACTIVAGE IoT data (services, applications, devices) is crucial for the operation and running of the ecosystem. The Data Discovery and Analysis is by design in AIOTES Architecture cloud-based data repository that provides both open and private Pilot data to authorised users by a Web interface and API's. In particular, on predictive data analysis mechanism – finding patterns and associations in the data, which enable us to predict issues, spot anomalies, and understand relationships between different important factors. The adoption of the visual analytics multi-functional platform (developed by CERTH-TRL7), offering multi-criteria analysis and detection of spatiotemporal patterns that are important for IoT management optimization operations will be also analysed and integrated.

3.9.3 At Public Layer

ACTIVAGE developers and technology integrators will be provided with a set of tools and functionalities that support the development and the deployment of new services and applications within AIOTES (WP5) to improve the reusability, interoperability and sharing of ACTIVAGE cross pilot IoT services/applications.

ACTIVAGE aims at providing innovative features and Advanced Analytics tools that will allow the in depth understanding of the data collected from ACTIVAGE deployment sites (WP9), and also help developers optimize their services/applications. Alike it provides all the necessary tools for the creation of services/applications from users with minimum technical training.

3.10 Data re-Use and Transformation.

The re-usability of the data and the ways to transform it into meaningful results is the last process in ACTIVAGE Data Management Lifecycle. By providing a mechanism for transformation, the correct collection of data can be performed, by enabling data re-usability the data can be repurposed. At Deployment Site layer

ACTIVAGE will make the best use of this process by means of offering re-usability and transformation tools to annotate, describe and share data as follow:

- Follow-up data along other lifecycles
- New Data lifecycles
- Undertake information about the produced data
- Find points for improving data
- Learn from the shared data
- Identify better mechanisms for sharing data
- Define post processing beyond re-using and transforming data

3.10.1 At LSP Layer

Depending on the type of use and on the strategy defined in the project, different policies and licences will be granted at LSP layer level.

As a first consideration, the following type of usages is foreseen:

- Some data will be destroyed after the project is finalized (e.g. data collected during the pilot in devices, sensors, etc.)
- Data will be shared for public purposes (meta-indicators related to the evaluation activities at a Global level)
- Data will be shared for awareness purposes (e.g. AHA Advisor and Public Evidence)
- Data can be used to create alliances within partners (e.g. LSP Dashboard data)
- Data can be shared with third parties (e.g. AIOTES parameters and information to be used during Open Calls and then for creating the AIOTES community and outreach strategy)
- Data can be "sold" to third parties (e.g. a DS partner open/sell their data and use the ACTIVAGE data model to enable new services after the project)

3.10.2 At Public Layer

In accordance with Grant Agreement Article 25, data must be made available upon request, or in the context of checks, reviews, audits or investigations. If there are ongoing checks, etc. the records must be retained until the end of these procedures. Regarding affiliated entities (from open calls, etc.), unless otherwise agreed, access must be given, under fair and reasonable conditions and where is legally and practically possible.

4 ACTIVAGE Data Management Plan

This chapter provides guidelines for the data management at the three levels of the data lifecycle, according with the general indications and definitions in Chapter 2.

4.1 Data sets

4.1.1 Data on Pilot Sites

All DS will define as part of D9.1 their planned datasets according to the following template that includes the main aspects identified by the EU guidelines to ensure FAIR data:

Table 2: DS Dataset description template

Dataset name	**DSx.number DESCRIPTIVE_NAME**
Data identification	
Data set description	*Where are the sensor(s) installed? What are they monitoring/registering? What is the dataset comprised of? Will it contain future sub-datasets?*
Source (i.e. which device?)	*How will the dataset be collected? What kind of sensor is being used?*
Partners responsibilities	
Owner of the device	*What is the name of the owner of the device?*
Partner in charge of data collection (if different)	*What is the name of the partner in charge of the device? Are there several partners that are cooperating? What are their names?*
Partner in charge of data analysis (if different)	*Partner name*
Partner in charge of data storage (if different)	*Partner name*
Standards and metadata	
Info about metadata (Production and storage dates, places) and documentation?	*What is the status with the metadata so far? Has it been defined? What is the content of the metadata (e.g. datatypes like images portraying an action, textual messages, sequences, timestamps etc.) How will it be created?*
Standards, Format, Estimated volume of data	*Has the data format been already decided? What will it look like?*
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	
Data access policy /	*Example text:*

Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The full dataset will be confidential and only the members of the DS and/or Consortium will have access on it. Furthermore, if the dataset or specific portions of it (e.g. metadata, statistics, etc.) are decided to become of widely open access, it will be uploaded to the ACTIVAGE open data platform. Of course, these data will be anonymized, in order to avoid any potential ethical issues with their publication and dissemination.*
Data sharing, re-use and distribution (How?)	*Has the data sharing policies already been decided? What requirements exists for sharing data? How will the data be shared? Who will decide what to be shared?*
Embargo periods (if any)	
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Who will own the information that has been collected? How will it adhere to partner policies? What kind of limitation are put on the archive?*

In addition to that, ACTIVAGE project will design and implement the "ACTIVAGE Public Evidence website" as a user interface to be accessed openly by any person or professional interested in accessing the underlying ACTIVAGE Evidence Open Data Base. The way this shall be work is by accessing pre-defined reports or making queries that allow creating composition of underlying data, and project data analytics. This tool is foreseen to remain active and available after project's end in order to offer a rich interface to the whole ACTIVAGE evidence.

4.1.2 Data as Managed by AIOTES

AIOTES, as the ACTIVAGE solution for platform interoperability across the deployment sites, will have to take into account the different deployment architectures at the different DSes. The underlying platforms used by the DSes are following different data storage, protection, access and processing principles, from pure Cloud-based to pure Edge-based or combinations of them. These platforms provide access to critical data and functionality that can cause safety and security threats to the users if appropriate mechanisms are not supported by them.

However, ACTIVAGE does not aim at intervening in the local deployment architecture of the platforms used by the DSes. Instead, AIOTES will support the implementation of such use cases that involve at least two distinct platforms. In its most general form, this will involve the sharing of both data and functionality across platforms, which in turn will force AIOTES to deal with all challenges of data management in the whole data lifecycle.

As a first principle, the DMP presented in this document requires that any data repository designed by AIOTES for the purpose of sharing data and functionality across different IoT platforms has to be agnostic with regard to the deployment strategy, in order to accommodate the freedom of the DSes in the selection of their deployment architectures.

Furthermore, from the relevant references addressed in Section 2, this DMP identifies the MyData model as the one most relevant for AIOTES. Therefore, we recommend the involved project tasks[9] to

[9] Project tasks in charge of taking care of these challenges include the whole WP3, esp. T3.1 and T3.2, and T4.2.

– clarify the technical position of ACTIVAGE towards the MyData model
– examine possibilities for becoming a contributor to the related open specification project addressed in Section 2.3
– align AIOTES mechanisms with the resulted specification for the storage, protection, access and processing of personal data, if it is eventually asserted that it meets related ACTIVAGE requirements[10].

From a policy perspective, this DMP requires that the MyData model is complemented in a way that it allows to automate the "MyData Operator" role as far as possible and host the related software system both on Edge and in the Cloud in order to make sure that conflicts with the local deployment architectures at the ACTIVAGE DSes can be avoided.

Last but not least, another requirement on AIOTES from a DMP perspective is about the support for data anonymization and pseudo-anonymization, in order to avoid privacy issues, especially in all cases where data is extracted from its actual application context and transferred to other repositories with different data ownerships as the original one.

4.1.3 Data as used for evaluation in ACTIVAGE

Figure 5 gives an overview of the main information generated at LSP-layer level as described in D6.1.

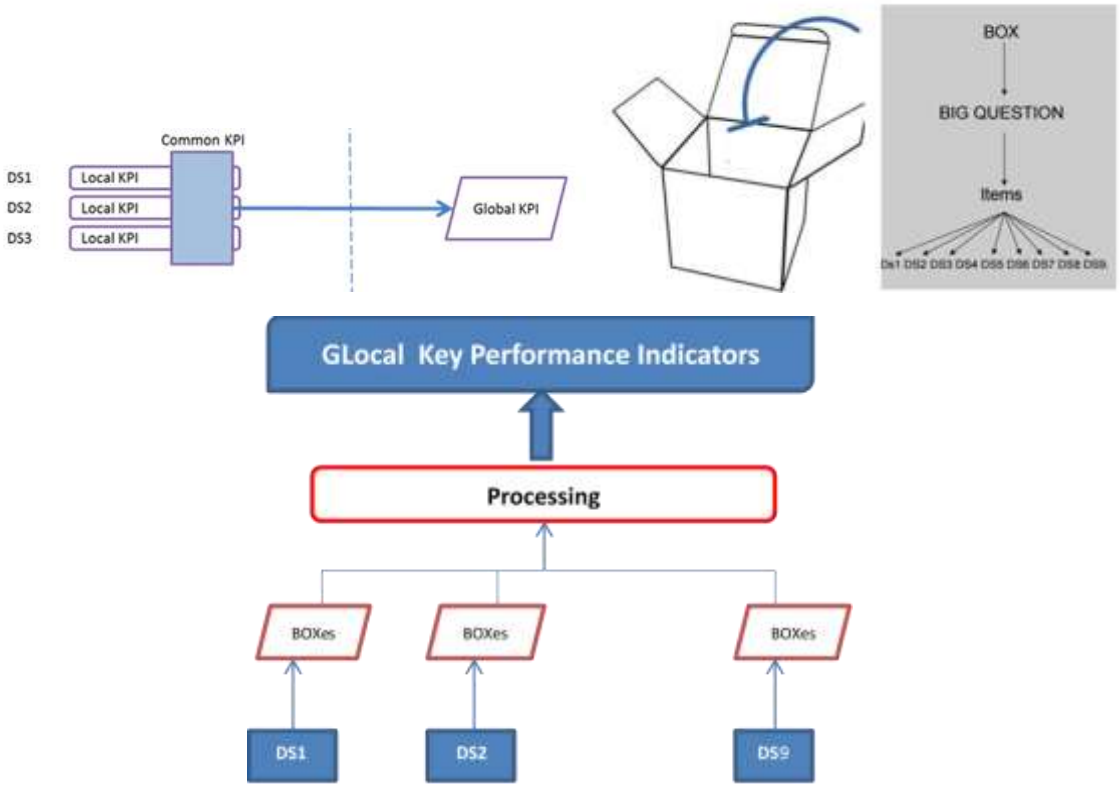

Figure 5: Overview of the main information generated at LSP-layer level (taken from D6.1)

[10] The recommended contribution should actually ensure this.

In the coming months, negotiation and interaction between WP6 and DS partners will take place, to decide whether the "boxes" will be physically created and stored at DS level or whether they can be virtual elements belonging to the open database. A first agreement will be done and reflected in D9.1 and D6.2.

4.1.4 Data as contributed to open data

An ACTIVAGE specific goal is to raise the "ACTIVAGE Evidence Open Data Base" as Reference Framework for evaluating IoT solutions in the AHA domain, acting as dashboard to collect evaluation data and as well as advisor tool for stakeholders and company that want to access ranking of IoT solutions in the AHA domain. ACTIVAGE will collaborate with the AHA key players to make ACTIVAGE Evidence Open Data Base a sustainable service.

The ACTIVAGE Evidence Open Data Base aims to provide orientation and advisory services to a variety of users, including of course end-users of Pilots, and motivate fieldwork players and manufacturing and product distributor companies to participate in enriching the collective knowledge that it represents. The outcome of such ACTIVAGE Public Evidence website is to be able to provide an evaluation data collection tool and at the same time to make available quantitative and qualitative KPI values reported by pilots and different initiatives.

4.1.5 Self-Audit Process

This section in the data management plan describes the self-audit process and its approach to be implemented in ACTIVAGE. The process that every DS in ACTIVAGE will follow to assess the compliance of the DMP is shown in Figure 6. The initial steps concern to the concepts around the data sets that will be generated, registered and published in ACTIVAGE. In this self-audit process the definition and any practical pre-validation, post validation, certification and self-audit will be defined as part of the Deployment Site Strategy but will be aligned with the ACTIVAGE project overall strategy. Self-audit activities will be described and the result of their implementation as part of Deployment site quality control measurements. Figure 6 is used to represent and describe the logic flow of the process for self-audits in ACTIVAGE.

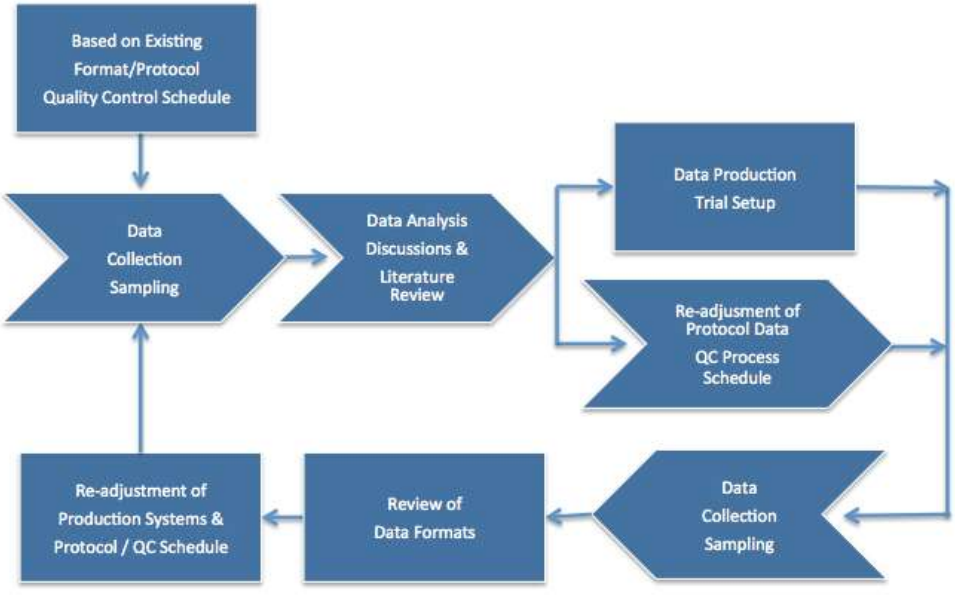

Figure 6. ACTIVAGE Self-Audit Flow Diagram.

The ACTIVAGE project has a selected group of experts acting as steward of the project, amongst them the Data Analytics and IPR Manager will also support the activity and execution of self-audits to the data that will be collected, managed and provided at the different deployment sites. ACTIVAGE self-audit process will aim at effectively managing data and identifying its potential, conditions and value of its data assets. Conducting a data self-audit will provide valuable information, raising awareness of collection strengths and data issues to improve overall strategy. A data self-audit will highlight duplication of effort and areas that require additional investment, allowing putting its resources to best use. It is most importantly for highlighting inadequacies in data creation and curation practices, suggesting policy change to lessen the risks. An organisation that is knowledgeable about its data puts itself in a position to maximise the value of its collections through continued use. In ACTIVAGE, the implementation of a self-audit methodology envisages to bring three main benefits a) Prioritisation of data resources which leads to efficiency savings; b) Ability to manage risks associated with data loss and irretrievability and c) Realizing the value of data through improved access and reuse.

The self-audit process in ACTIVAGE follows an effective way to guarantee consistency in the data, a controlled process and the best methodologies at the different ACTIVAGE levels, i.e. Device, Platform and Application, The following is the list of additional activities that are overseen in order to execute the self-audit process:

- Planning
  - Plan and Set-up the Self-Audit
  - Collect Relevant Documents
- Identification, Classification and Assessment of Datasets
  - Analyse Documents
  - Identify Data Sets
  - Classify Data Sets
  - Assess Data Sets
- Report of Results and Recommendations
  - Collate and analyse information from the audit
  - Report on the compliance with the Data Management Plan
  - Identify weaknesses and decide on corrective actions

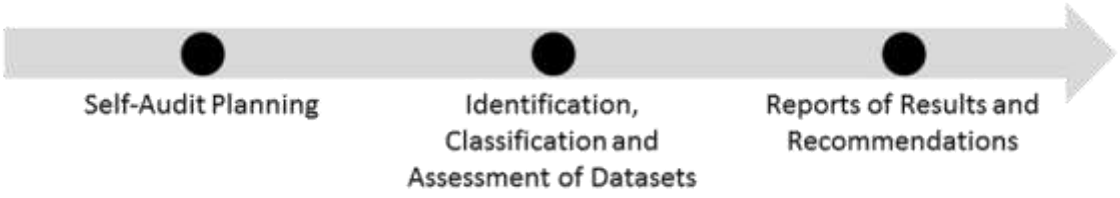

Figure 7. ACTIVAGE Self-Audit Process.

4.2 Other Considerations

4.2.1 Linked Data

Until recently the landscape of data on the Web was comprised of a plethora of self-contained data repositories. Basically, each Web application or platform maintained its own repository, even if there was a significant overlap between these datasets and data that was

publicly accessible. From a knowledge and information retrieval perspective, however, the integration of different kinds of data sources yields significant added values. The use of different formats and different technologies made such an integration challenging in the past.

These challenges spurred the development and success of the concept of Linked Data. This concept describes a method of publishing not only documents but also all kinds of structured data so that it can be interlinked and become more useful. It builds upon standard Web technologies such as HTTP and URIs (Uniform Resource Identifier), but rather than using them to serve web pages for human readers, it extends them to share information in a way suitable for reading them automatically by computers. This enables data from different sources to be connected and queried. By studying linked data lifecycle, we can have a comparative between FIESTA-IoT project[11] data lifecycle and the above-described ACTIVAGE data lifecycle with enrichment, interlinking/merging, repair, etc.

The term Linked Data refers to a set of best practices for publishing and interlinking structured data on the Web. These best practices were introduced by Tim Berners-Lee in his Web architecture note Linked Data and have become known as the Linked Data principles. These principles are the following:

4.2.2 URIs and Names Identification

This principle advocates using URI references to all things, i.e. extending the scope of the Web from online resources to encompass any object or concept in the world. Thus, things are not just Web documents and digital content, but also real world objects and abstract concepts. These may include tangible things such as people, places and cars, or those that are more abstract, such as the relationship type of knowing somebody, the set of all green cars in the world, or the colour green itself.

To publish data on the Web, the things need to be uniquely identified. As Linked Data builds directly on the Web architecture (Jacobs and Walsh, 2004), the Web architecture term resource is used to refer to these things of interest, which are, in turn, identified by HTTP URIs. Linked Data uses only HTTP URIs, avoiding other URI schemes such as Uniform Resource Names (URN[12]) and Digital Object Identifier (DOI[13]). The benefits of HTTP URIs are: (a) they provide a simple way to create globally unique names in a decentralised fashion, and (b) they serve not just as a name but also as a means of accessing information describing the identified entity.

4.2.3 HTTP and URIs

The HTTP protocol is the Web's universal access mechanism. In the classic Web, HTTP URIs are used to combine globally unique identification with a simple, well-understood retrieval mechanism. Thus, this Linked Data principle advocates the use of HTTP URIs to identify objects and abstract concepts, enabling these URIs to be dereferenced (i.e., looked up) over the HTTP protocol to obtain a description of the identified object or concept. As a result, any HTTP client can look up the URI using the HTTP protocol and retrieve a description of the resource that is identified by the URI. This applies to URIs that are used to identify classic HTML documents, as well as URIs that are used in the Linked Data context to identify real-world objects and abstract concepts.

[11] http://www.fiesta-iot.eu

[12] http://tools.ietf.org/html/rfc2141

[13] http://www.doi.org/handbook_2000/DOIHandbook-v4-4.1.pdf

In case of URIs identifying real-world objects, it is essential to distinguish these objects themselves from the Web documents that describe them. It is, therefore, common practice to use different URIs to identify the real-world object and the document that describes it, in order to be unambiguous. This practice allows separate statements to be made about an object and about a document that describes that object. For example, the creation date of a person may be rather different to the creation date of a document that describes this person. Being able to distinguish the two through use of different URIs is critical to the coherence of the Web of Data.

4.2.4 URIs for providing useful information, using the standards

In order to enable a wide range of different applications to process Web content, it is important to agree on standardised content formats. The agreement on HTML as a dominant document format was an important factor that made the Web scale. The third Linked Data principle therefore advocates use of a single data model for publishing structured data on the Web – the Resource Description Framework (RDF).

RDF provides a graph-based data model that is extremely simple on the one hand but strictly tailored towards Web architecture on the other hand. RDF itself is just describing the data model, it does not address the format in which the data is eventually stored and transferred. To be published on the Web, RDF data can be serialised in different formats. The two RDF serialisation formats most commonly used to publish Linked Data on the Web are RDF/XML and RDFa.

4.2.5 Include links to other URIs

This Linked Data principle advocates the use of hyperlinks to connect not only Web documents, but also any other type of thing. For example, a hyperlink may be set between a person and a place, or between a place and a company. Hyperlinks that connect things in a Linked Data context have types, which describe the relationship between the things. For example, a hyperlink of the type “friend-of” may be set between two people, or a hyperlink of the type “based-near” may be set between a person and a place. Hyperlinks in the Linked Data context are called RDF links in order to distinguish them from untyped hyperlinks between classic Web documents.

The forth Linked Data principle is to set RDF links pointing into other data sources on the Web. Such external RDF links are fundamental for the Web of Data as they are the glue that connects different data repositories into a global, interconnected data space. This, in turn enables applications to discover additional data sources by following RDF links between different data sources, for example to navigate through the Web of Data using Linked Data browser or to crawl data as input for Linked Data search engines.

4.2.6 Querying Linked Data

SPARQL[14] (SPARQL Protocol and RDF Query Language) is the most popular query language to retrieve and manipulate data stored in RDF, and it became an official W3C Recommendation in 2008. Depending on the purpose, SPARQL distinguishes the following for query variations:

– SELECT query: extraction of (raw) information from the data

[14] http://www.w3.org/TR/rdf-sparql-query/

- CONSTRUCT query: extraction of information and transformation into RDF
- ASK query: extraction of information resulting in a True/False answer
- DESCRIBE query: extraction of RDF graph that describes the resources found

Given that RDF forms a directed, labelled graph for representing information, the most basic construct of a SPARQL query is a so-called *basic graph pattern*. Such a pattern is very similar to a RDF triple with the exception that the subject, predicate or object may be a variable. A basic graph pattern matches a sub graph of the RDF data when RDF terms from that sub graph may be substituted for the variables and the result is RDF graph equivalent to the sub graph. Using the same identifier for variables also allow combining multiple graph patterns. To give an example, the SPARQL query returns the name of all pairs of people where `? person1` knows `? person2` (note that `foaf:knows` is not defined as symmetric relation)

5 ACTIVAGE Open Access Platform

ACTIVAGE will integrate a data management portal as part of its website and based on CERTH's existing data management infrastructure implemented within the myAirCoach H2020 project as it offers:

– A Members Area for registered users
– Open access area for the access of project outcomes and anonymized datasets
– Introduction to the purpose of the platform and the rights/obligations of its users
– Links to the open access and open data guidelines of H2020

This portal will provide to the public, for each dataset that will become publicly available, a description of the dataset along with a link to a download section. The portal will be updated each time a new dataset has been collected and is ready for public distribution. However, the portal will not contain any datasets that should not become publicly available. This data management portal will enable project partners to manage and distribute their public datasets through a common infrastructure.

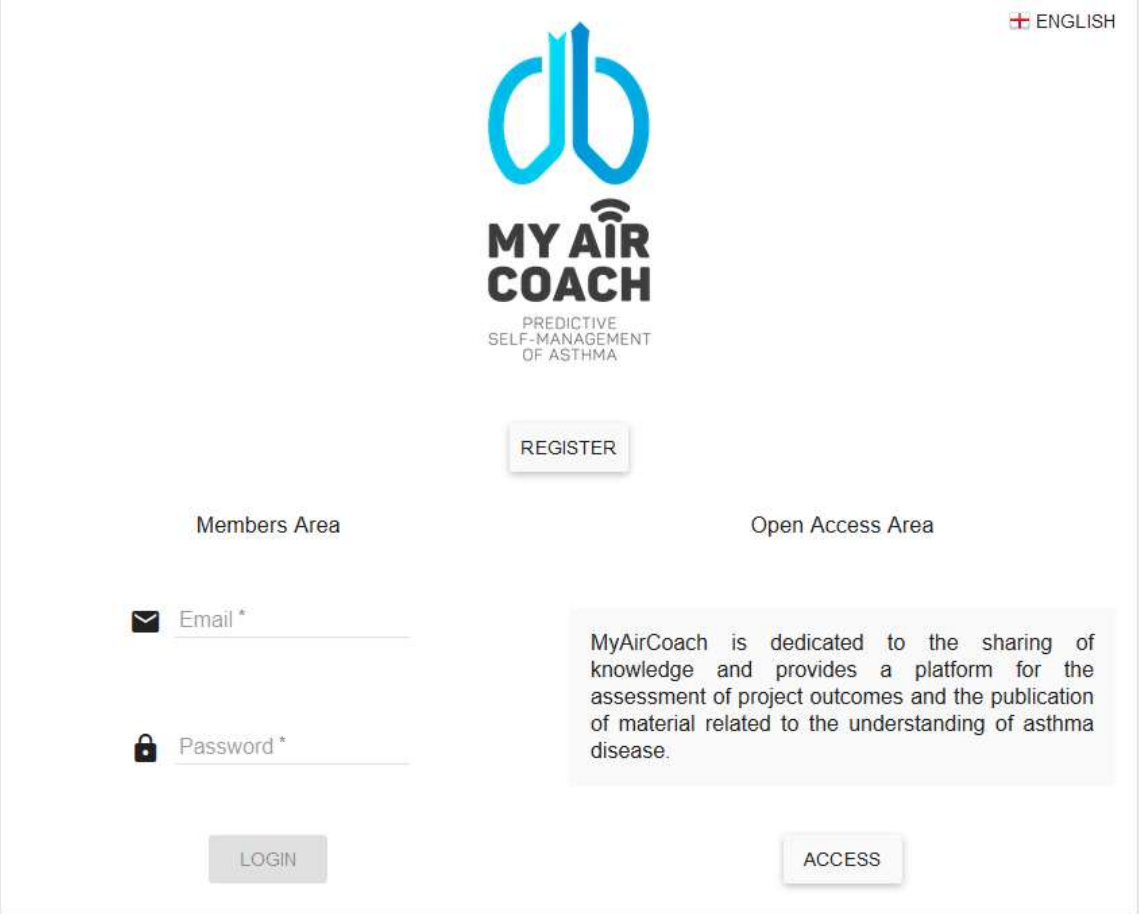

Figure 8: Login page of the MyAirCoach Platform[15]

[15] accessed by http://160.40.50.91:40000/myaircoach/app/#/opendata

The open access platform will be integrated as a component of the final online platform of the ACTIVAGE web site and as such has offers two fundamental views. The first one is addressed to registered members of the system who in addition to the data of their users will be able to access anonymized datasets and the knowledge generated within the ACTIVAGE project. Furthermore, these users will be able to upload data to the open access framework and share them with the entire research community. The second view of the system is intended for unregistered users who need to get access to the datasets and publications of ACTIVAGE without registering as a user. In this case only anonymized data will be made available to them and they will not be able to upload any type of data to the system.

After login, the users will be presented with the functionalities of the system that will be different for based on whether the user is registered or not the system.

Figure 8 illustrated the login page of the MyAirCoach platform showing the two different ways of accessing the data of the project.

The documents repository of the platform will be used in order to access the outcomes of the project and more specifically it will include as described above, project deliverables and documents (Scientific publications, etc.), Dissemination material, etc.

In order to support the usability, usefulness and accessibility of the data a metadata template will be used for the description of every uploaded document. It should be underlined that only the creators of the document and the system administrator have the right to edit and change the provided information or delete the document from the repository.

6 Conclusion / Future Work

The Data Management Life Cycle in ACTIVAGE has been designed, introduced and described in this document following the user-data centric approach. The data management lifecycle describes all the process for data creation, data storage, data processing, data security, across the data value chain from Deployment Sites operations to making data openly accessible.

The purpose of the ACTIVAGE Data Management Plan described in this document is to support the data management life cycle for all data that will be generated, collected, and processed by the ACTIVAGE project in the different deployment sites and also in the ACTIVAGE system.

This deliverable is the first edition of the Data Management Plan and it is a working document, which shall evolve during the lifespan of the ACTIVAGE project. This document is expected to mature during the phase of architecture/building ACTIVAGE AIOTES and when the requirements collection phase will be finalized. The next version of the Data Management Plan shall include the update(s) concerning the outcomes of above mentioned activities.

.