DELIVERABLE 1.3
Data Management Plan

RISA
October, 2019

www.eco-bot.eu

D1.3: Data Management Plan

Summary

This document presents the Data Management Plan for the Eco-Bot project. The aim of this deliverable is to describe the plan for managing the data generated and collected during the project. Specifically, this deliverable describes the datasets that will be collected or generated and how they will be managed during the project and beyond its completion. Moreover, the document presents how the FAIR principles, data security, and ethical aspects are addressed in the project. The Data Management Plan that has been devised is in compliance with the General Data Protection Regulation that came into effect in May 2018.

DELIVERABLE NUMBER	WORK PACKAGE
D1.3	WP1

LEAD BENEFICIARY	DELIVERABLE AUTHORS
RISA	Aggeliki Giakoumaki (RISA) Stephanos Camarinopoulos (RISA)

QUALITY ASSURANCE	
Reviewer 1: Sarantis Kotsilitis	PLEGMA
Reviewer 2: Dora Karali	ERRA

PLANNED DELIVERY DATE	ACTUAL DELIVERY DATE
31/3/2018	22/10/2019

DISSEMINATION LEVEL	□ PU = Public □ PP = Restricted to other programme participants x CO = Confidential, only for members of the consortium

Table of contents

List of Tables

List of Acronyms and Abbreviations

API: Application Programming Interface

CLI: Command-Line Interface

CO: Confidential

DDoS: Distributed Denial-of-Service

DOI: Digital Object Identifier

DMP: Data Management Plan

DPO: Data Protection Officer

EC: European Commission

GDPR: General Data Protection Regulation

IPR: Intellectual Property Rights

MITM: Man-In-The-Middle

NILM: Non-Intrusive Load Monitoring

OA: Open Access

PU: Public

REST: Representational State Transfer

SDK: Software Development Kit

SSH: Secure Shell

WP: Work Package

Executive summary

This deliverable is the Data Management Plan for the Eco-Bot project. It has been prepared taking into account the guidelines available at the template "Horizon 2020 FAIR Data Management Plan (DMP)". The document describes the datasets that will be collected or generated during the project, their origin, format, and accessibility level, as well as how they will be managed during the project and beyond its completion. Moreover, the document presents how the FAIR principles, resources allocation, data security, and ethical aspects are addressed in the project. The Data Management Plan that has been devised is in compliance with the General Data Protection Regulation that came into effect in May 2018. It is a living document that will be updated throughout the project's lifetime.

1. Introduction

This deliverable is the Data Management Plan of the Eco-Bot project. The document has been prepared taking into account the guidelines available at the template "Horizon 2020 FAIR Data Management Plan (DMP)".

The deliverable is structured as follows:

Section 2 describes the datasets that are collected or generated in the project and provides information about their origin, format, sharing level, archiving and preservation.

Section 3 presents how the FAIR principles are applied to the project.

Section 4 indicates foreseen allocation resources, as well as the persons responsible for defining the data management strategy and supervising its implementation and compliance with the GDPR.

Section 5 outlines the security policy followed in Eco-Bot.

Section 6 presents how compliance with the GDPR and ethical principles is achieved.

The Data Management Plan is a living document and will evolve during the lifetime of the project, according to the status of the project reflections on data management and security.

2. Data Summary

This section presents the datasets that are collected or generated during the project. Table 1 outlines the datasets and indicates their origin.

Table 1: Eco-Bot datasets and origin

#	Dataset	Origin
1	User-specific data	Pilots
2	Smart meter data	Pilots
3	NILM data	USTRAT
4	Behavioural survey data	Primary data from users
5	Appliance survey data	Primary data from users
6	Time diary data	Primary data from users
7	Expert system rules	Eco-Bot expert system
8	Recommendations lists	UEKAT
9	Users Recommendation Profile	System generated
10	User-system interaction data	User-system interaction

#	Dataset	Origin
11	User-system dialogues	User-system interaction
12	Weather data	Third party

The following tables describe each dataset and provide information about their origin, format, data sharing, as well as regarding their archiving and preservation.

Table 2: User-specific data

Dataset Name
User-specific data
Description and Purpose
User-specific data are collected from the pilots' backends and are used in order to enable proper system configuration and access to anonymised data of the specific user from the pilot's backend.
Origin
The user-specific data are collected from the pilots.
Format
.json
Data Sharing
User-specific data include the information that is necessary in order to customise the frontend and to access the required data from the pilot's backend. In compliance with the GDPR and the 'data minimisation principle', personal data are only limited to what is necessary and no redundant data are collected. To this end, no other personal data of the participants, except from their locale, are collected. All data collected and processed by the system regarding a specific user, are anonymised and linked to him/her through a user-specific UID, and only the pilot can link the UID to the user's real identity.
Apart from the user's locale, this dataset includes also information about the pilot the user belongs to, the tokens and API keys required to access the necessary user's data from the pilot's backend, as well as the following, pilot-specific data: for SEC, if the user has smart meter or not; for EYPESA, the frequency of the user's smart meter; and for DEXMA, the number, type and size of buildings the user handles. The above anonymised information is available to both PLEGMA and ERRA to enable proper system configuration.

Archiving and Preservation

This dataset is maintained and stored in the Eco-Bot backend which is hosted in a virtual server on Hetzner and in ERRA's dedicated server at the company's premises, accessible under credentials only by the entitled researchers. The information will be kept up to two years, however it will be used only for the Eco-Bot research purposes.

Table 3: Smart meter dataset

Dataset Name

Smart meter data

Description and Purpose

The Eco-Bot backend collects smart meter data from the pilots to forward them to USTRAT's NILM module, in order to perform energy disaggregation and provide energy consumption information on an appliance level. Smart meter data are also used by the system in order to send notifications and alerts to the users, as well as to respond to users' inquiries on energy consumption.

Origin

Energy consumption data of the users/buildings are collected from the pilots.

Format

.json

Data Sharing

The Eco-Bot backend collects energy consumption data from the pilots in predefined time intervals, in order for them to be shared with USTRAT's NILM module that will provide the disaggregated results. The energy consumption data are shared with USTRAT through JSON-based RESTful API.

Energy consumption data that the user might request and which are not anymore stored in the Eco-Bot backend, are retrieved from the pilots upon system request and forwarded to the Eco-Bot frontend, using JSON-based RESTful APIs.

Energy consumption data are anonymised and are linked to the user through his/her UID, therefore, neither of the partners having limited-time access to them, i.e. PLEGMA, USTRAT and ERRA, can trace the user's real identity.

Archiving and Preservation

Energy consumption data are received by the Eco-Bot backend in predefined time intervals from the different pilots. These data are processed in the Eco-bot backend in order to check data validity (i.e. detect negative values, gaps, NaN values, etc.) and are then shared with USTRAT's NILM module, which is responsible for providing the disaggregated results. Energy consumption data are stored in the Eco-Bot backend for a limited time period and then they are deleted from the system. Specifically, non-high frequency energy consumption data (i.e. 1-hour and 1-min data from EYPESA and 15-min resolution data from DEXMA) are stored for one month, while high frequency data (10-sec measurements from SEC) are stored for one week. This anonymised dataset is also used by USTRAT, which stores temporarily the data in a dedicated server at the university's premises, only during the course of processing them to provide the NILM results, and then deletes them. Energy consumption data that the user might request and which are not stored anymore in the Eco-Bot backend, are retrieved from the pilots upon system request and forwarded to the frontend to be shown to the user. JSON-based RESTful APIs developed by Eco-Bot partners are used for data transfer.

Table 4: NILM dataset

Dataset Name
NILM data
Description and Purpose
The Eco-Bot backend receives disaggregated energy consumption data from USTRAT's NILM module to forward them to the pilots for long-term storage. NILM data are also used by the system in order to send reports and notifications to the users, respond to users' inquiries on their appliance-level energy consumption, and provide users with appliance-related recommendations.
Origin
Energy consumption data of the users/buildings are collected from the pilots.
Format
.json
Data Sharing
USTRAT's NILM module sends to the Eco-Bot backend disaggregated energy consumption data of users/buildings in predefined time intervals. The backend forwards

these data to the pilots for long-term storage, and also stores them for a limited time period, after which they are deleted from the system.

NILM data that the user might request and which are not stored anymore in the Eco-Bot backend, are retrieved from the pilots upon system request and forwarded to the Eco-Bot frontend, using JSON-based RESTful APIs.

Disaggregated energy consumption data are anonymised and are linked to the user through his/her UID, therefore, neither of the partners having limited-time access to them, i.e. PLEGMA, USTRAT and ERRA, can trace the user's real identity.

Archiving and Preservation

The Eco-Bot backend receives disaggregated energy consumption data from USTRAT's NILM module and forwards them to the pilots for long-term storage. USTRAT also stores these anonymised data in a dedicated server at their premises for further processing and NILM enhancement, where it will be kept up to two years after the project's completion, however it will be used only for the Eco-Bot research purposes.

These data are also stored in the Eco-Bot backend for a limited time period and then they are deleted from the system. Specifically, NILM outputs of non-high sampling frequency consumption data (i.e. 1-hour and 1-min data from EYPESA and 15-min resolution data from DEXMA) are stored for one month, while NILM outputs of high sampling frequency consumption data (10-sec measurements from SEC) are stored for one week.

Disaggregated energy consumption data that the user might request and which are not stored anymore in the Eco-Bot backend, are retrieved from the pilots upon system request and forwarded to the frontend to be shown to the user. JSON-based RESTful APIs developed by Eco-Bot partners are used for data exchange.

Table 5: Behavioural survey dataset

Dataset Name

Behavioural survey data

Description and Purpose

This dataset contains the responses of the participants of SEC and EYPESA pilots to the survey that has been created by UEKAT and is filled in by them upon their registration to the system. The responses of each participant are used by the behavioural module that is integrated into the system, in order to assign the participant to a specific segment, and based on this, to create his/her initial recommendation profile.

This dataset will be updated during the course of the project, as participants will be asked to respond to an extended behavioural survey, thus enabling both the identification of a

potential shift of the user to a different segment and the evaluation of the behavioural model.

Origin

This dataset is created from primary data given from the participants of SEC and EYPESA pilots and includes the participants' responses to the behavioural surveys created by UEKAT.

Format

.json

Data Sharing

This dataset is created in the Eco-Bot backend and includes the users' responses to an online questionnaire that has been developed by PLEGMA based on the survey questions given by UEKAT.

During his/her registration at the pilot's frontend, the user is given a link (specific to the user's UID) of the online questionnaire, and is asked to respond to it in order to get access to Eco-Bot. The responses of the user are retrieved and stored by the Eco-Bot backend. The behavioural module that is integrated into the system, takes these responses as input and assigns the user to a specific segment, thus enabling the creation of the initial recommendation profile of the user.

This dataset is shared with UEKAT, in order to enable further training, enhancement and evaluation of the behavioural model. The behavioural survey data are shared with UEKAT through a JSON-based RESTful API.

All dataset records are anonymised to protect users' privacy and confidentiality. All user's responses to the behavioural survey are linked to the user through his/her UID, therefore, neither PLEGMA, nor UEKAT, the only partners that have access to this dataset, can link them with the user's real identity. SEC and EYPESA, which can trace the user's real identity based on his/her UID, do not have access to this dataset, and their only involvement in the behavioural survey is that they provide the user with the link of the online questionnaire during his/her registration.

Archiving and Preservation

The behavioural survey dataset is stored in the Eco-Bot backend which is hosted in a virtual server on Hetzner, as well as in UEKAT's dedicated server at the university's premises, and accessible under credentials only by the entitled researchers. The behavioural module that is integrated into the system uses these data to assign each user to a segment. This anonymised dataset will be kept up to two years after the project's completion, however it will be used only for the Eco-Bot research purposes.

Table 6: Appliance survey dataset

Dataset Name
Appliance survey data

Description and Purpose
This dataset contains information about the household appliances of the participants of SEC and EYPESA pilots. Users are asked during their registration to fill in an appliance survey, and their household appliances information will be used to help establish the ground truth for NILM. This dataset will be updated during the course of the project, as users will be enabled to update their appliances list by indicating their intention to the chatbot.

Origin
This dataset is created from primary data given from the participants of SEC and EYPESA pilots and includes information about their household appliances.

Format
.json

Data Sharing
This dataset is created in the Eco-Bot backend and includes information about the users' household appliances, collected through an online appliance survey that has been developed by PLEGMA based on USTRAT's guidelines. During his/her registration at the pilot's frontend, the user is given a link (specific to the user's UID) of the online appliance survey, and is asked to respond to it in order to get access to Eco-Bot. The information given by the user regarding his/her appliances is retrieved and stored by the Eco-Bot backend. Upon user's request, the appliance list is retrieved and shown to him/her, enabling the user to update it. This dataset is shared with USTRAT, in order to help establish the ground truth for NILM. All dataset records are anonymised to protect users' privacy and confidentiality. All user's appliances information is linked to the user through his/her UID, therefore, none of the partners having access to this dataset, i.e. PLEGMA, ERRA and USTRAT, can link them with the user's real identity. SEC and EYPESA, which can trace the user's real identity based on his/her UID, do not have access to this dataset, and their only involvement in the appliance survey is that they provide the user with the link of the online survey during his/her registration.

Archiving and Preservation

The appliance survey dataset is stored in the Eco-Bot backend which is hosted in a virtual server on Hetzner, as well as in USTRAT's dedicated server at the university's premises, and accessible under credentials only by the entitled researchers. This anonymised dataset will be kept up to two years after the project's completion, however it will be used only for the Eco-Bot research purposes.

Table 7: Time diary dataset

Dataset Name
Time diary data

Description and Purpose
This dataset contains information about the hourly intervals during which each household appliance was used in the participant's home for a specific period and will be used for establishing the ground truth and for further NILM training as well as evaluation purposes. Participants of SEC and EYPESA pilots will be asked upon their first login into Eco-Bot to fill in the time diary for one week, to help establish the ground truth. This dataset will be updated during the course of the project, as users will be asked from time to time to fill in the time diary for one day in order to receive enhanced results as well as for evaluation purposes. During the course of the project, if a user updates his/her appliances list with a new appliance, he/she is asked by the system to fill in the time diary again for one week, but only for the new appliance. Users can voluntarily fill in the time diary for more days in order to receive more accurate NILM results, by indicating their intention to the chatbot.

Origin
This dataset is created from primary data given from the participants of SEC and EYPESA pilots and includes information about the hourly intervals during which each household appliance was used for a specific period.

Format
.json

Data Sharing
This dataset is created in the Eco-Bot backend and includes information about the time slots during which the user's household appliances were used for a specific number of days. This information is collected through an online time diary that has been developed by PLEGMA based on USTRAT's guidelines.

Upon their first login into Eco-Bot, users are given a link (specific to their own UID) of the time diary, and they are asked, on a voluntary basis, to fill it in for one week in order to help establish the ground truth and receive enhanced NILM results. Upon either USTRAT's or user's request, the user will be given the link to the online time diary to fill it in also during the course of the project. The information given by the users regarding the time when their appliances were used is retrieved and stored by the Eco-Bot backend.

This dataset is shared with USTRAT, in order to help establish the ground truth for NILM, as well as for training and evaluation purposes.

All dataset records are anonymised to protect users' privacy and confidentiality. All user's information about the time when his/her household appliances were used, is linked to the user through his/her UID, therefore, none of the partners having access to this dataset, i.e. PLEGMA and USTRAT, can link them with the user's real identity. SEC and EYPESA, which can trace the user's real identity based on his/her UID, do not have access to this dataset.

Archiving and Preservation

The time diary dataset is stored in the in the Eco-Bot backend which is hosted in a virtual server on Hetzner, as well as in USTRAT's dedicated server at the university's premises, and accessible under credentials only by the entitled researchers. This anonymised dataset will be kept up to two years after the project's completion, however it will be used only for the Eco-Bot research purposes.

Table 8: Expert system rules dataset

Dataset Name
Expert system rules
Description and Purpose
This dataset contains the knowledge base of the system, which conveys the knowledge obtained from the experts through the form of rules. The dataset includes the recommendation rules that define the personalised recommendations that will be sent to the users, as well as other rules that the expert system uses in order to provide users with additional features including notifications and alerts. This dataset is expandable, as the system user interface enables the addition or modification of rules so as to enhance personalisation and user engagement.
Origin

This dataset has been created by PLEGMA that developed the Eco-Bot expert system, based to a large extent on the knowledge provided by UEKAT to form the knowledge base. It will be updated during the project based on the addition or modification of rules by system users through the system user interface.

Format
JSON

Data Sharing
This dataset is used by the Eco-Bot expert system in order to provide personalised features to users. It will be available only to PLEGMA, that developed the rule based expert system. This dataset will be updated with new or modified rules, enabling the system to expand and enhance the offered services. System users, according to their privileges, are enabled to add new rules or modify their own existing ones, through the system user interface described in D4.1. System users from the pilot partners can only access and edit the rules regarding their own pilot.

Archiving and Preservation
This data set is in the Eco-Bot backend which is hosted in a virtual server on Hetzner and will be kept for long-term use, as it is expected to be used for Eco-Bot exploitation purposes.

Table 9: Recommendations lists dataset

Dataset Name
Recommendations lists

Description and Purpose
This dataset contains the lists of recommendations (both behavioural and investment) per pilot and per segment. It is used by the system as a reference base to define the personalised recommendations that will be sent to the users according to their recommendation profiles. Recommendations are sent to users by rule firings. This dataset also contains the relevant statistics, i.e. how many times each recommendation was sent, how many times it was found useful based on the users' feedback, how many users followed it, etc.

Origin

This dataset was created internally in the system, based on the recommendation lists that were prepared and provided by UEKAT in Excel files. The system incorporates the behavioural module developed by UEKAT that enables the user's assignment to a segment and thus to a segment-tailored recommendation list.

This dataset is updated with relevant statistics based on user-system interaction and users' feedback on the recommendations.

Format

.json

Data Sharing

This anonymised dataset is used by the system as a reference base to define the personalised recommendations that will be sent to the users according to their recommendation profiles. This dataset will also be made available to UEKAT, through a JSON-based RESTful API, to enable evaluation of the behavioural model and of the recommendations.

Some aggregated, anonymised information of this dataset will also be shared with the pilots (SEC, EYPESA, DEXMA) in order to help them identify the responsiveness and engagement of their customers to recommendations, on a segment group level, and will also be shared publicly through the public deliverable D5.5 (Validation Results including Lessons Learned and societal impact) and through Open Access publications that will describe the project outcomes and the results of the system evaluation.

Archiving and Preservation

The data repository is maintained and stored in the Eco-Bot backend which is hosted in a virtual server on Hetzner and accessible under credentials only by the entitled researchers. PLEGMA software components (rule engine of the expert system) access this anonymised dataset to create the personalised recommendations lists based on the users' segments, and update the dataset with evaluation metrics data based on users' feedback. The information will be kept up to two years but it will only be used for the Eco-Bot research and exploitation purposes.

Table 10: User recommendation profiles dataset

Dataset Name

Users Recommendation Profiles

Description and Purpose

This dataset contains all the information related to the recommendation profile of each user, which determines the personalized recommendations that will be sent to them. The user's initial profile is created when he/she responds to the behavioural survey and is accordingly allocated to a specific segment, based on which the user is assigned to a segment-tailored recommendation list. This profile is updated as it contains also data records such as recommendations the user indicates that he/she followed, user's feedback on the usefulness of each recommendation, energy saving events that the user indicates that he/she performed, etc.

Origin

This dataset is created internally in the system, which incorporates the behavioural module developed by UEKAT that enables the user's assignment to a segment and thus to a segment-tailored recommendation list, and is updated based on user's feedback.

Format

.json

Data Sharing

This anonymised dataset is used by the system for the delivery of personalised recommendations to the users. Recommendations are sent to users by rule firings.

This dataset will also be made available to UEKAT, through a JSON-based RESTful API, to enable evaluation of the behavioural model and of the recommendations.

All dataset records are anonymised to protect users' privacy and confidentiality. All user's data records are linked to the user through his/her UID, therefore, PLEGMA, the developer of the Eco-Bot backend platform, does not have access to the user's real identity. The same applies to ERRA, the developer of the Eco-Bot front-end, which receives user's feedback on the recommendations and energy saving events and forwards it to PLEGMA. Likewise, UEKAT has access to these anonymised data records for evaluation purposes, without being able to link them to the user's real identity. The pilots, which can trace the user's real identity based on his/her UID, do not have access to this dataset.

Archiving and Preservation

The data repository is maintained and stored in the Eco-Bot backend which is hosted in a virtual server on Hetzner, and accessible under credentials only by the entitled researchers. PLEGMA software components (rule engine of the expert system) process this anonymised dataset to offer the personalised recommendations to the users. The information will be kept up to two years but it will only be used for the Eco-Bot research and exploitation purposes.

Table 11: User-system interaction data

Dataset Name
User-system interaction data

Description and Purpose
This dataset contains different data that are collected automatically by the system during the user-system interaction. These data include logs recording the user's activity, such as number of user's logins, date/time and duration of user's sessions, number of messages sent and received per session, number and type of user's inquiries, as well as aggregated information including pilot statistics, system response metrics, number of users online, etc. This dataset is used for system usage, performance, and user engagement monitoring.

Origin
User-system interaction data are collected and updated automatically by the system.

Format
.json

Data Sharing
This dataset is used by the system and pilot administrators to monitor system usage and performance, as well as from the pilot users to monitor their users' activity and engagement in using the system.
These data are collected, stored and updated automatically by the Eco-Bot backend and frontend. This dataset, accessed by PLEGMA and ERRA, is anonymised, as all recorded data are linked to the users only through their UIDs. SEC, EYPESA and DEXMA pilot users and pilot administrators can monitor user and system activity, according to their privileges, through the system user interface described in D4.1.
Some aggregated, anonymised information of this dataset will also be shared publicly through the public deliverable D5.5 (Validation Results including Lessons Learned and societal impact) and through Open Access publications that will describe the project outcomes and the results of the system and user engagement evaluation.

Archiving and Preservation
This dataset is maintained and stored in the Eco-Bot backend which is hosted in a virtual server on Hetzner, as well as in ERRA's dedicated server at the company's premises, and accessible under credentials only by the entitled researchers. The information will be kept

up to two years, however it will be used only for the Eco-Bot research, dissemination and exploitation purposes.

Table 12: User-system dialogues

Dataset Name
User-system dialogues
Description and Purpose
This dataset contains the dialogues between the users and the chatbot and is used for chatbot performance analysis and evaluation, training, and response optimisation.
Origin
This dataset is created and updated internally in the system based on the user-system interaction and includes the dialogues taking place between the users and the chatbot.
Format
.json
Data Sharing
This dataset is created internally in the system based on the user-system interaction and records the dialogues taking place between the users and the chatbot. This dataset is available only to ERRA, that developed the chatbot, and will use the dialogues for chatbot performance analysis and evaluation, training, and response optimisation. All dataset records are anonymised to protect users' privacy and confidentiality. All user's dialogues with the chatbot are linked to the user through his/her UID, therefore, ERRA, the only partner that has access to this dataset, cannot link the dialogues to the user's real identity. SEC and EYPESA, which can trace their customers' real identity based on their UID, do not have access to the dialogues.
Archiving and Preservation
The user-system dialogues are stored in ERRA's dedicated server at the company's premises, and are accessible under credentials only by the entitled researchers. This anonymised dataset will be kept up to two years after the project's completion, however it will be used only for the Eco-Bot research purposes.

Table 13: Weather data

Dataset Name
Weather data
Description and Purpose
This dataset contains weather information for the areas where the participants are located. The weather data are collected from a third party and are used by the system in order to provide users with weather-related recommendations on energy efficiency. Weather-related recommendations are sent to users by rule firings.
Origin
This dataset is retrieved from a third party, namely Yr.no, using its API.
Format
Origin XLM stored in JSON
Data Sharing
This dataset is used by the Eco-Bot backend in order to send weather-related recommendations on energy efficiency to users, based on their location. Weather-related recommendations are sent to users by rule firings. Weather information is publicly available by the Norwegian Meteorological Institute and the Norwegian Broadcasting Corporation on Yr.no website.
Archiving and Preservation
This dataset is stored in the Eco-Bot backend which is hosted in a virtual server on Hetzner and will be kept up to two years after the project's completion to be potentially used for Eco-Bot research purposes.

All the data produced or used in the project should follow the basic principles of FAIR data, as described in the EC's "Guidelines on FAIR Data Management in Horizon 2020". According to these principles, data should be:

- Findable
- Accessible.
- Interoperable.
- Re-usable

The following sections present how these principles are addressed in the project.

3.1.   Findable data

The data produced or used in the project have different levels of access, and each dataset, according to its nature and purpose, is available only to the project partners that are entitled to access it, solely for the Eco-Bot research purposes, as described in Section 2. Project partners exchange the information needed for the different tasks using JSON-based RESTful APIs they have developed.

The consortium partners have decided that none of the datasets will be uploaded to an open repository and the data will not be available for re-use, hence it is not foreseen to create metadata and keywords for the datasets. Nevertheless, certain aggregated, anonymised data will be extracted from the previously described datasets and these will be openly available at Zenodo, which registers and preserves Digital Object Identifiers (DOIs), thus making the data uniquely identifiable and persistently locatable. Moreover, metadata that will be created for these subsets using Zenodo's metadata format, as well as keywords, will make them easily discoverable.

3.2.   Accessible data

According to FAIR principles, data should be as open as possible and as closed as necessary; open so as to enable re-usability and to accelerate research, but at the same time closed as needed to protect privacy and confidentiality of the involved individuals and parties.

To this end, the consortium has decided to treat the Eco-Bot datasets, as far as open availability is concerned, as presented in Table 14. The table shows which datasets will be partially openly available, as well as the reasons for excluding the rest of the datasets from public disclosure.

Table 14: Eco-Bot datasets availability

#	Dataset	Openly available	Justification
1	User-specific data	No	Privacy and confidentiality reasons
2	Smart meter data	No	Proprietary rights of SEC, EYPESA and DEXMA
3	NILM data	No	Privacy and confidentiality reasons
4	Behavioural survey data	No	Privacy and confidentiality reasons
5	Appliance survey data	No	Privacy and confidentiality reasons
6	Time diary data	No	Privacy and confidentiality reasons
7	Expert system rules	No	Exploitation reasons: the knowledge base of the expert system is the backbone of Eco-Bot.
8	Recommendation lists	No (only certain aggregated, anonymised data will be openly available)	Intellectual property of UEKAT
9	Users Recommendation Profile	No	Privacy and confidentiality reasons
10	User-system interaction data	No (only certain aggregated, anonymised data will be openly available)	Privacy and confidentiality reasons
11	User-system dialogues	No	Privacy and confidentiality reasons
12	Weather data	Yes (already openly available by a third party)	N/A

As far as the availability of the Eco-Bot datasets to the project partners is concerned, sharing information per dataset is given in Section 2 (Data Summary). As shown in the description of the datasets in that section, accessibility of Eco-Bot datasets follows the above-mentioned principle, and each dataset is available only to those partners that are entitled to access the data in order to perform the required tasks.

All user-related data are anonymised and they will be stored and preserved by the responsible consortium partners for as long as it is needed for Eco-Bot research and exploitation purposes, but no longer than two years after the project's completion.

No specialised method, software tool and/or documentation is needed to access the data; the consortium partners have developed JSON-based RESTful APIs to be used for data transfer, and the API documentation is available on the project's internal website research data repository (https://redmine.risa.eu/projects/eco-bot).

In order to safeguard privacy and confidentiality of the involved participants, only anonymised and aggregated data will be openly available. The consortium plans to make these data available through the public deliverable D5.5 (Validation Results including Lessons Learned and societal impact), which will be available on the project's website, as well as through Open Access (OA) publications that will describe the project outcomes and the results of the system and user engagement evaluation. For Open Access publications, the Zenodo repository will be used, which is automatically linked to OpenAIRE.

3.3. Interoperable data

According to FAIR principles, data produced in the project should be interoperable, in order to allow data exchange and re-use. In accordance with this principle, the interoperability of all datasets that will be created during the course of the project will be ensured by using an open-standard file format, namely JSON.

Given that no re-use of the created datasets is foreseen, each dataset will be available only to the consortium partners entitled to access it. The sharing will be done directly between the authorised partners and the datasets will not be uploaded to any open repository. JSON-based RESTful APIs have been developed by the partners to be used for data exchange.

As the produced datasets will not be uploaded to any open repository, it is not foreseen to create metadata for them. Especially as regards the anonymised and aggregated data that will be openly available at Zenodo, their interoperability will be achieved using Zenodo's metadata format, i.e. JSON-format according to a defined JSON schema, which allows metadata to be exported in several standard formats such as MARCXML, Dublin Core, and DataCite Metadata Schema (https://about.zenodo.org/policies/).

3.4. Re-usable data

The consortium partners have decided that no dataset will be made openly available, due to privacy and confidentiality compliance, IPR issues, and exploitation purposes. The reasons for restricting access to each one of the datasets, are presented in Table 14 in Section 3.2.

Moreover, the entitled project partners will store and preserve the datasets only for as long as it is needed, solely for Eco-Bot research and exploitation purposes, but no longer than two

years after the project's completion. Hence, no re-use of the created datasets is foreseen to take place.

4. Allocation of Resources

Long-term preservation is foreseen only for the expert system rules dataset, which is the backbone of the Eco-Bot expert system and knowledge base and is expected to be used for exploitation purposes. Given that this dataset will not be openly available and will be stored only in the designated repository, while all other datasets will be kept for no more than two years after the project's completion and will then be deleted, no resources are foreseen to be needed for long term preservation.

The foreseen costs as regards making data FAIR, involve the costs of Open Access publications, and will be claimed as part of the Horizon 2020 Grant.

The Data Management Plan is authorised by the Steering Committee and all subsequent changes / revisions should also be approved and authorised by the Steering Committee. RISA's Data Protection Officer has been appointed by the consortium partners as responsible for supervising the data management strategy and implementation, and for ensuring compliance with the GDPR and with the ethical principles of the Grant Agreement.

5. Data Security

The Eco-Bot backend follows a strict privacy policy; security is ensured via the Azure Web app used to offer the API interface and the user-interface for system users. Azure is the cloud infrastructure of Microsoft that ensures the following:

- Communication of secrets (such as connection strings) between an app and other Azure resources (such as SQL Database) stays within Azure and does not cross any network boundaries. Secrets are always encrypted when stored.
- All communication over the App Service connectivity features, such as hybrid connection, is encrypted.
- Connections with remote management tools like Azure PowerShell, Azure CLI, Azure SDKs, REST APIs, are all encrypted.
- 24-hour threat management protects the infrastructure and platform against malware, distributed denial-of-service (DDoS), man-in-the-middle (MITM), and other threats.

Service availability is ensured through cloud components, cached data and failsafe mechanisms provided by Windows Azure Web App Service ensuring 99.9% availability.

Windows Azure App Service allows securing apps with HTTPS. When an app is created, its default domain name (<app_name>.azurewebsites.net) is already accessible using HTTPS. Hence secure communication between systems is ensured by HTTPS protocol using SSL certificates and SSH-tunnelling for database security. The database configuration is in a remote server provided by Hetzner and allows connections only from the App Service of the

Azure infrastructure with strict rules in the firewall and the configuration of the MongoDB. The communication is encrypted over SSH and the communication with the database is based on SSH-tunnelling and database accounts that have been created. Enhanced security features of Microsoft Azure web app such as identity management, auto software update and network isolation are used.

6. Ethical Aspects

The consortium partners have revised the Data Management Plan so as to ensure compliance with the General Data Protection Regulation (GDPR) that came into effect in May 2018.

Article 5 of the GDPR defines the main principles relating to processing of personal data as follows:

- lawfulness, fairness and transparency
- purpose limitation
- data minimisation
- accuracy
- storage limitation
- integrity and confidentiality
- accountability

In compliance with the above GDPR principles, personal data are processed in a lawful, fair and transparent manner in relation to the involved subjects. Data subjects will receive detailed information in an easy to understand way regarding their personal data that will be processed, the purposes of this processing, and the envisaged period for which they will be stored.

Moreover, personal data will be collected only for specified, explicit and legitimate purposes, i.e. research and statistical, and will not be processed further in a way incompatible with those purposes.

In accordance with the 'data minimisation' principle, collected personal data will be adequate, relevant and limited only to what is necessary in relation to the purposes for which they are processed, as described in Section 2.

As mentioned in the relevant datasets description in Section 2, personal data that are collected by the system are only limited to the locale, which is used for system configuration, and the energy consumption data. The system performs data validity check of the energy consumption data received from the pilots so as to ensure accuracy of the data, and in case any data are found to be inaccurate, they are deleted or rectified without delay.

Personal data collected by the system will be stored and preserved in the designated repositories only for as long as they are needed in order to be used solely for the Eco-Bot

research purposes, and in any case they will not be kept for more than two years after the project's completion.

Personal data that the system collects from the pilots have already been anonymised at the pilots before being made available to the system, and the only information associating the personal data to the user is his/her UID, which can be linked to the user's real identity only in the pilots' backends. Moreover, personal data that the system receives directly from the participants through their responses in the behavioural and appliance surveys and in the time diary, are also linked to them only through their UID, and in order to protect privacy and confidentiality, these data are not shared with the pilots. Additionally, in compliance with the principle of integrity and confidentiality, the system follows a strict privacy and security policy, as described in Section 5, and proper security measures are used to ensure protection against unauthorised or unlawful processing and against accidental loss or damage.

Moreover, the consortium partners have appointed RISA's Data Protection Officer (DPO), Mrs. Katerina Papasileka, to oversee the data management strategy and implementation, and her responsibility is to ensure compliance with the GDPR and with the ethical principles of the Grant Agreement.

In accordance with Article 6 of the GDPR regarding the lawfulness of processing, data subjects will be asked to give consent to the processing of their personal data for the purposes that will be clearly explained to them.

Article 7 of the GDPR defines the conditions for consent and states that the data subject has the right to withdraw his/her consent at any time, and that this should be as easy as to give consent. Participants will be informed before giving their consent about their right to opt out at any time and information about how they can revoke their consent will also be written in the consent form.

In compliance with Article 12 of the GDPR, participants will be provided with all necessary information relating to their personal data that will be processed in a concise, transparent, intelligible and easily accessible form, using clear and plain language. The information will be provided in writing, or by other means, including, where appropriate, by electronic means. When requested by the participant, the information may be provided orally, provided that the identity of the data subject is proven by other means.

Article 15 of the GDPR defines the right of access by the data subject. Participants will be informed about their right to access their personal data that are processed.

In compliance with Article 17 of the GDPR regarding the 'right to be forgotten', participants will be informed about their right to obtain the erasure of their personal data without undue delay, and upon any participant's request, the respective content will be erased from the system within up to five working days.

Article 32 of the GDPR defines the requirements regarding security of processing of personal data, including among others pseudoanonymisation and encryption of personal data, confidentiality, integrity, availability and resilience of the system. Section 5 outlines the security policy that is followed in Eco-Bot. Moreover, as mentioned above, the personal data that the system collects from the pilots have already been anonymised at the pilots before

being made available to the system, and the only information associating the personal data to the user is his/her UID, which can be linked to the user's real identity only in the pilots' backends, where all personal data are encrypted.

All participants will sign an informed consent during the recruitment phase that will describe in detail the data that will be collected from them, the way these data will be handled, who will have access to them, for how long, and for which purposes. The consent form will also inform them about their right to withdraw their consent and their right to be forgotten. By signing the informed consent form, participants will agree to authorise access to all of their collected data (raw, aggregated, anonymised) and to the usage of their anonymised and aggregated data for research and exploitation purposes. The template of the consent form to be used in the recruitment phase for the pilot demonstrations will be presented in D5.1.

It should be noted that the pilot partners have decided not to connect Eco-Bot with social media and provide access to it only though their own websites. Therefore, sensitive data such as social and everyday habits information regarding the users, will not be available to the system. All data regarding the users will be collected in a transparent way, in compliance with the GDPR and with the ethical principles of the Grant Agreement.

7. Conclusion

This deliverable describes the Data Management Plan of Eco-Bot and specifies the datasets that are collected or generated and how they are managed during the project and after its completion. The document also describes how the FAIR principles are applied to the project, and presents how data security requirements and ethical principles are met, referring also to how the compliance with the GDPR is achieved.

The Data Management Plan is a living document and will be updated during the project's lifetime.