



## **Deliverable D7.3**

### Data Management Plan (v.1)

V1.0



**This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815301**

## D7.3 –Data Management Plan (v.1)

### 1 Document Information

<b>Programme</b>	Horizon 2020 – RE-COGNITION
<b>Project acronym</b>	RE-COGNITION
<b>Grant agreement number</b>	815301
<b>Number of the Deliverable</b>	<b>D7.3</b>
<b>WP/Task related</b>	[WP7 / T7.3]
<b>Type (distribution level)</b>	CO Confidential
<b>Date of delivery</b>	[25-09-2019]
<b>Date due according to GA</b>	[30-09-2019]
<b>Status and Version</b>	Version 0.1
<b>Number of pages</b>	32 pages
<b>Document Responsible</b>	Dimitrios Kourempapas – IDE
<b>Author(s)</b>	Ioannis Kompougias – IDE
<b>Reviewers</b>	POLITO CERTH E@W

## 2 Revision History

Version	Date	Author/Reviewer	Notes
0.1	20/06/2019	Ioannis Kompougias & Dimitrios Kourempapas / Authors	First Draft of the Deliverable
0.2	26/09/2019	Ioannis Kompougias & Dimitrios Kourempapas / Authors	Revision of the Draft
1.0	30/09/2019	Dimitrios Kourempapas	Final Revision and Consolidation

### 3 Executive summary

The RE-COGNITION solution stands as an ICT integration framework, on-top of established and newly developed, innovative RE-based technologies, representing a highly inter-disciplinary concept for empowering the transition towards a renewable building landscape. The framework will enable the building integration of both variable renewable energy (VRE) sources (e.g. solar, wind) and dispatch-able energy sources (e.g. bio-gas, geothermal) along with energy storage and advanced energy efficient technologies to ensure that the highest possible share of the buildings energy demand is satisfied. Actually, the proposed solution goes a step further by setting the ambition towards zero-energy or even energy positive buildings, with emphasis on the small and medium-sized building stock in Europe. The synergy among the RES-based technologies is achieved by means of the platform that consists in the following main parts:

- 1) An Automated Cognitive Energy Management Engine (ACEME), that allows harmonizing production and demand side of various energy vectors (power, heating and cooling), by proposing optimal set of control strategies, depending of the demand and the external conditions. ACEME would be able to integrate data coming from different smart meters, operating within multiple time-scales to optimally cover short-term and middle-term needs.
- 2) iGateway, the intelligence that allows the communication of the ACEME with the RES-based devices.
- 3) Building Energy Plant Planning Tool BE-PLATO, a system that assists during the planning phase; this provide decision support to the various stakeholders for selection of RE-based systems, with the aim of minimizing the operational costs by selecting RES with high utilization for the specific installation, by keeping the planning complexities as low as possible. Financial analysis has been performed considering information about equipment available on the market.

The final goal of the platform developed is to converting conventional buildings into zero energy or energy-positive buildings by means of advanced RES-based technologies. In order to support the implementation of these activities, a Data Management Plan is proposed.

The Data Management Plan will be updated, based on the effective progress of the activities, and depending on the evolution of data to be distributed.

## 4 Introduction

### 4.1 Scope of the document

This document, D7.3 – Data Management Plan (DMP) is a deliverable of the RE-COGNITION Project, which is funded by the European Union’s Horizon 2020 Program under Grant Agreement #815301.

This document provides the general plan for managing the data generated and collected during the project. It covers:

1. The handling of research data during and after the project;
2. What data will be collected, processed or generated;
3. What methodology and standards will be applied;
4. Whether data will be shared/made open and how;
5. How data will be curated and preserved.

The DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur.

### 4.2 Notations, abbreviations and acronyms

Table 1 reports the list of abbreviations adopted in the entire document.

**Table 1 - Acronyms list**

CO	Confidential
DMP	Data Management Plan
BE-PLATO	Building Energy Plant Planning Tool
ACEME	Automated Cognitive Energy Management Engine
BEMS	Building Energy Management System
RES	Renewable Energy Source
ESCO	Energy Service Company
DBMS	Database Management System

DPO	Data Protection Officer
R&D	Research and Development

## 5 RE-COGNITION Data Summary

Data play a key role for an optimum operation of a RE-COGNITION platform, interconnecting the major network components in an efficient way. As major components can be considered:

- **Field Devices:** This is the group of sensors, metering units, control units (BEMS, alarms etc.), Storage and Conversion devices, RES, RE-COGNITION platform needs to succeed the goal of zero energy or energy-positive buildings. These units can be either new or already installed on the field and their number and type may vary from platform to platform. The interconnection with the rest system components shall be achieved via local controllers.
- **iGateway:** It is the device that redirects data between ACEME and Field Devices
- **ACEME, RE-COGNITION Repository and Supervisory Dashboard:** ACEME, the engine described before, shall be framed with a dedicated Repository for storing the input data and generated results of ACEME and a specialized dashboard for generating supervision diagrams for users on demand. Depending on the installation site requirements there will be two options for the aforementioned components configuration:
  - **Option 1:** ACEME, RE-COGNITION Repository and Supervisory Dashboard will be installed locally on secure, certified servers, while an extra cloud-based Repository will store safety backup of the RE-COGNITION Repository, public data and critical data for BE-PLATO tool
  - **Option 2:** ACEME, RE-COGNITION Repository and Supervisory Dashboard will be cloud-based applications installed on secure, certified servers
- **BE-PLATO and BE-PLATO Repository:** BE-PLATO, the tool described before, will be a cloud-based application installed on secure, certified server. BE-PLATO Repository shall store the user profiles and the generated planning scenarios.

Data can be distinguished according to their path and functionality at the following categories:

1. Data between Field Devices and iGateway
2. Data between Cloud Services and iGateway
3. Data between iGateway and Local Server
4. ACEME, RE-COGNITION Repository and Supervisory Dashboard Data
5. BE-PLATO and BE-PLATO Repository data

Data categories and their flow in the proposed RE-COGNITION platform are depicted in Figure 1.

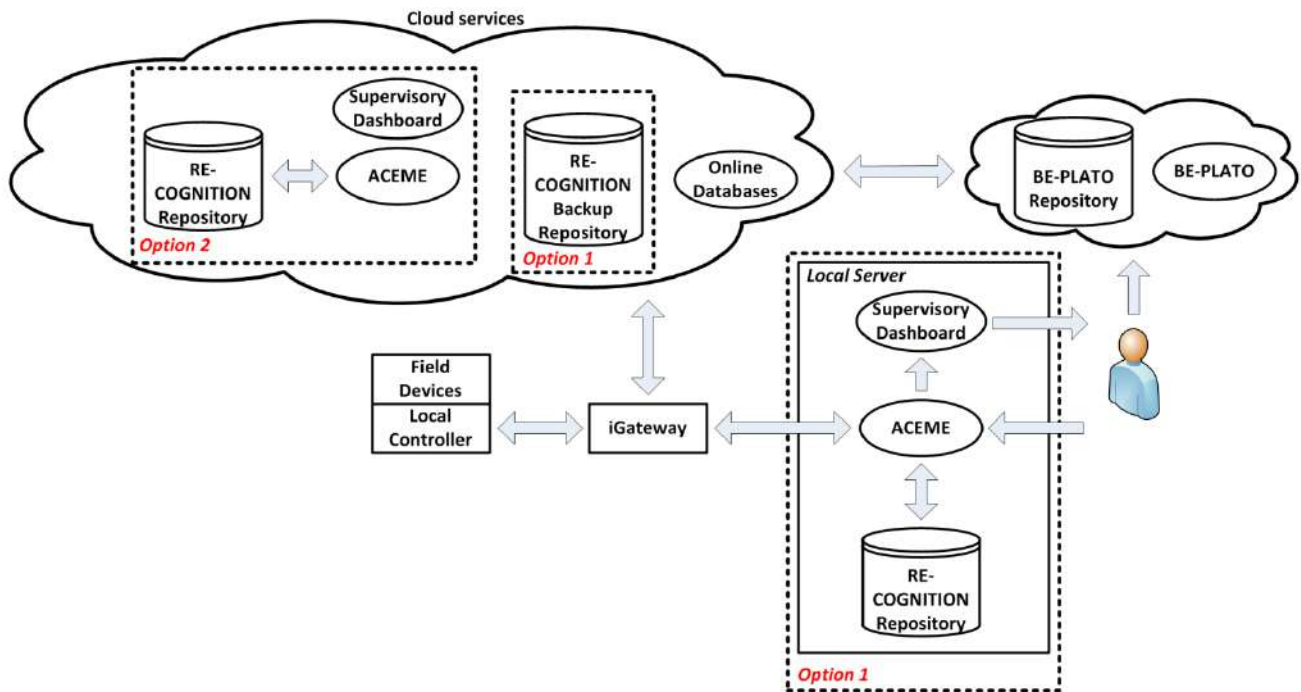


Figure 1– RE-COGNITION Data flows and Interactions

A detailed description of each data category is provided on the following paragraphs. The following parameters - based on the existing knowledge on data characteristics- are examined:

- Purpose of the data collection/generation in relation to the project objectives;
- Types and formats of data generated collected;
- Data re-usability;
- Data Origin;
- Data Size;
- Data Utilization.



## 5.1 Data between Field Devices and iGateway

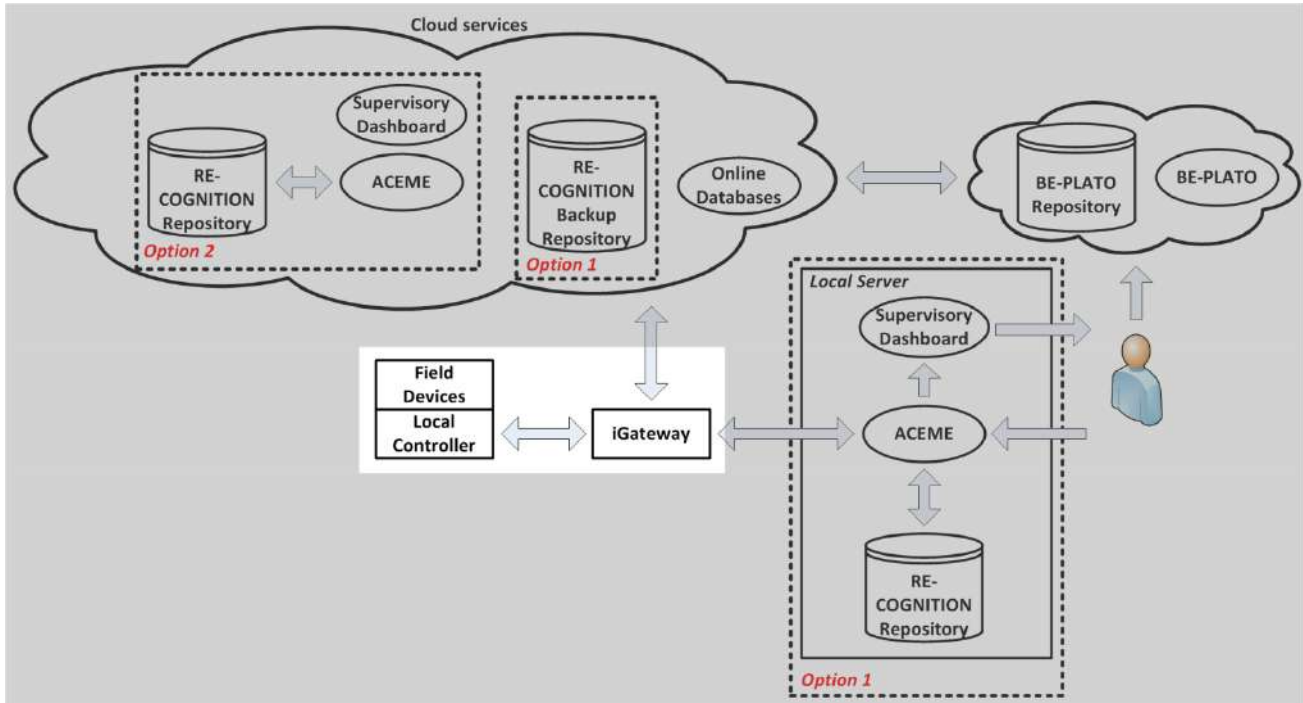


Figure 2 – RE-COGNITION Data between Field Devices and iGateway

### 5.1.1 Purpose of the data collection/generation in relation to the project objectives

These bidirectional data shall be useful for achieving a fully functional/efficient RE-COGNITION Platform operation according to ACEME results. The data flow shall include:

- Control signals between Field Devices (RES, BEMS, alarms etc.) and ACEME as a final destination
- Measurements from installed sensors (ambient conditions, building occupancy etc.), metering points, Storage and Conversion devices and RES
- Status of all Field Devices

### 5.1.2 Types and formats of data will the project generate/collect

The exact data types shall vary for each Field Device. The main type categories shall be:

- Device Status;
- Current Measurements;
- Warnings / Errors;
- Control data;
- Measurement records.

The applied filename format shall provide critical information such as timestamp, corresponding

Field Device and platform. Due to the dependence on ACEME and RE-COGNITION Repository architecture the filename format is going to be determined on following project tasks.

Regarding the communication protocols the data format shall be compatible with Modbus RTU and Wi-Fi (IEEE 802.11x).

### 5.1.3 Re-Use of Data

Data between Field Devices and iGateway will be real-time data which won't be stored in their current form and re-used.

### 5.1.4 Data Origin

Measurement and Status data shall be generated by Field Devices, gathered by the local Controller and sent to iGateway.

Control data shall be generated by both Field Devices and ACEME, passing through iGateway.

### 5.1.5 What is the expected size of the data?

The corresponding data category refers to real time data which are not stored and reused in their current form. Thus the total data size shall not be considered as a critical parameter.

On the other hand the Data Baud Rate shall depend on:

- Number and type of the installed Field Devices per platform: More information about this shall be available in following work packages, when the Pilot site equipment shall be finalized.
- The Baud Rate ACEME requires for a steady efficient platform operation: This request shall be determined on following work packages about ACEME development
- The Baud Rate limits of the applied protocol, which may vary in each pilot site

### 5.1.6 Data Utility

Field Devices shall use the control data sent by ACEME through iGateway

Measurement, control and status data from Field Devices will be received by iGateway and redirected to ACEME or RE-COGNITION Repository.

## 5.2 Data between Cloud Services and iGateway

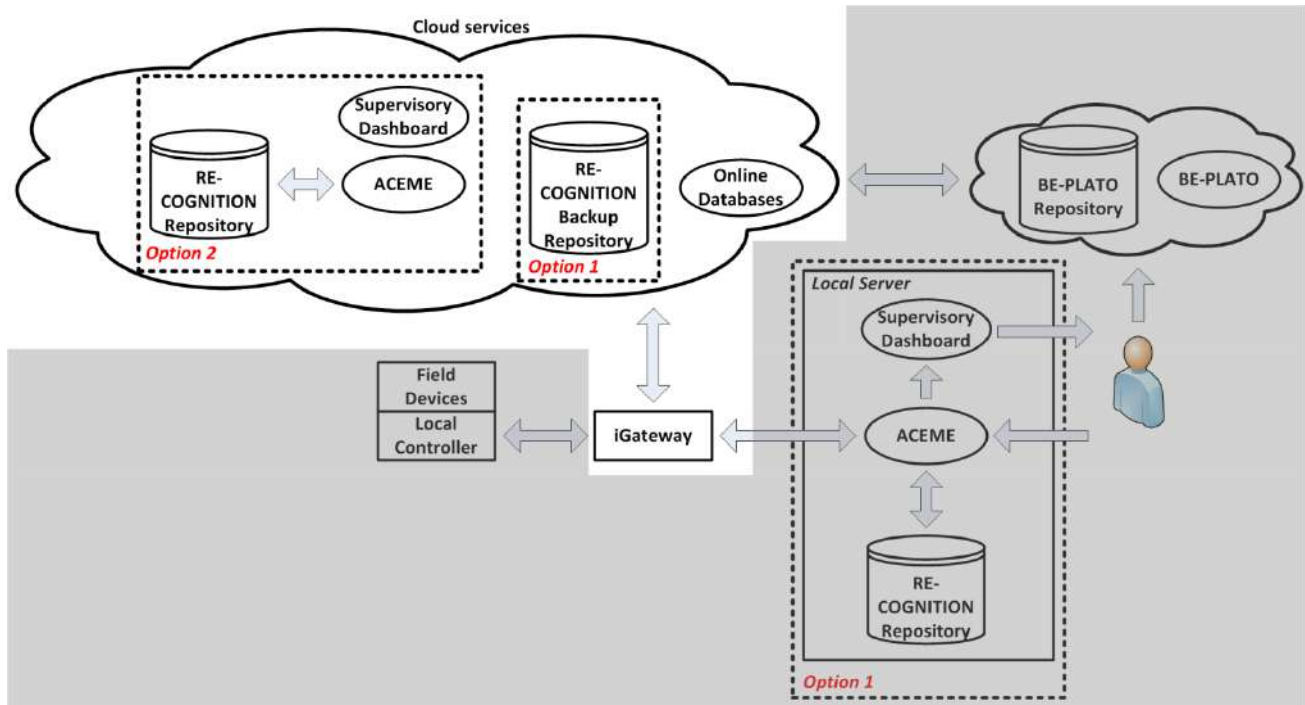


Figure 3 – RE-COGNITION Data between Field Devices and iGateway

### 5.2.1 Purpose of the data collection/generation in relation to the project objectives

The purpose of the data between iGateway and the cloud services is different for each of the two platform configurations.

According to Option 1 (local-based ACEME, RE-COGNITION Repository and Supervisory Dashboard) the data flow between iGateway and Cloud shall include:

- Data from online databases, required by ACEME to evaluate the short-term optimal plan
- Data for the cloud-based RE-COGNITION Backup Repository that are distinguished in:
  - Data for secure storage and
  - Shared data generated from the local-based ACEME

In case of Option 2 (cloud-based ACEME, RE-COGNITION Repository and Supervisory Dashboard) the transmitted data shall be the ones in 5.1.1, redirected by iGateway.

### 5.2.2 Types and formats of data will the project generate/collect

#### 5.2.2.1 Option 1 Data

Data from online databases shall include dynamic information such as:

- Energy prices
- Meteorological records and prediction
- Field Device estimated efficiencies and productivity

The selection of the used online databases and the data format (CSV, XML, JSON etc.) shall be determined in future work packages related to ACEME development.

The applied filename format shall provide critical information such as timestamp, corresponding online database, magnitude or Field Device. Due to the dependence on ACEME and RE-COGNITION Repository architecture the filename format is going to be determined on following project tasks.

Regarding the communication protocols the data format shall be compatible with TCP/IP and Wi-Fi (IEEE 802.11x).

Data for the cloud-based RE-COGNITION Backup Repository shall be an image of the local-based RE-COGNITION Repository. A detailed presentation of RE-COGNITION Repository data type, filename and data format shall be presented in 5.4.1 and 5.4.2.

Regarding the communication protocols, the data format shall be compatible with MQTT over TCP/IP and Wi-Fi (IEEE 802.11x).

#### **5.2.2.2 Option 2 Data**

The exact data types and filename formats shall be the same as presented in 5.1.2. Regarding the communication protocols the data format shall be compatible with MQTT over TCP/IP and Wi-Fi (IEEE 802.11x).

#### **5.2.3 Re-Use of Data**

Data in this category are either real-time data which won't be stored in their current form or data transferred between databases. Thus data of this path shall not be re-used.

#### **5.2.4 Data Origin**

##### **5.2.4.1 Option 1 Data**

Online Databases data shall belong to third-party open-access, cloud-based libraries which will be determined in further project tasks.

Data for the RE-COGNITION Backup Repository shall be generated by ACEME.

##### **5.2.4.2 Option 2 Data**

The transmitted data shall be the same as presented in 5.1.4, redirected by iGateway.

### 5.2.5 What is the expected size of the data?

This information shall be provided in future work packages, when more information are available regarding the pilot site equipment, ACEME and RE-COGNITION Repository architecture, stakeholder agreement and used online databases.

### 5.2.6 Data Utility

#### 5.2.6.1 Option 1 Data

Online Databases data shall be used by ACEME, passing through iGateway.

Data for the cloud-based RE-COGNITION Backup Repository shall be used by:

- Platform Users regarding the RE-COGNITION Repository secured backup
- Third party users regarding the Open Data, publicly shared generated by the local-based ACEME, under the agreement of the platform user
- Consortium regarding the confidential data generated by the local-based ACEME, under the agreement of the platform user

#### 5.2.6.2 Option 2 Data

Field Devices shall use the control data sent by ACEME through the iGateway.

Redirected from iGateway, measurement, control and status data from Field Devices shall be used from the cloud-based ACEME or RE-COGNITION Repository.

### 5.3 Data between iGateway and Local Server

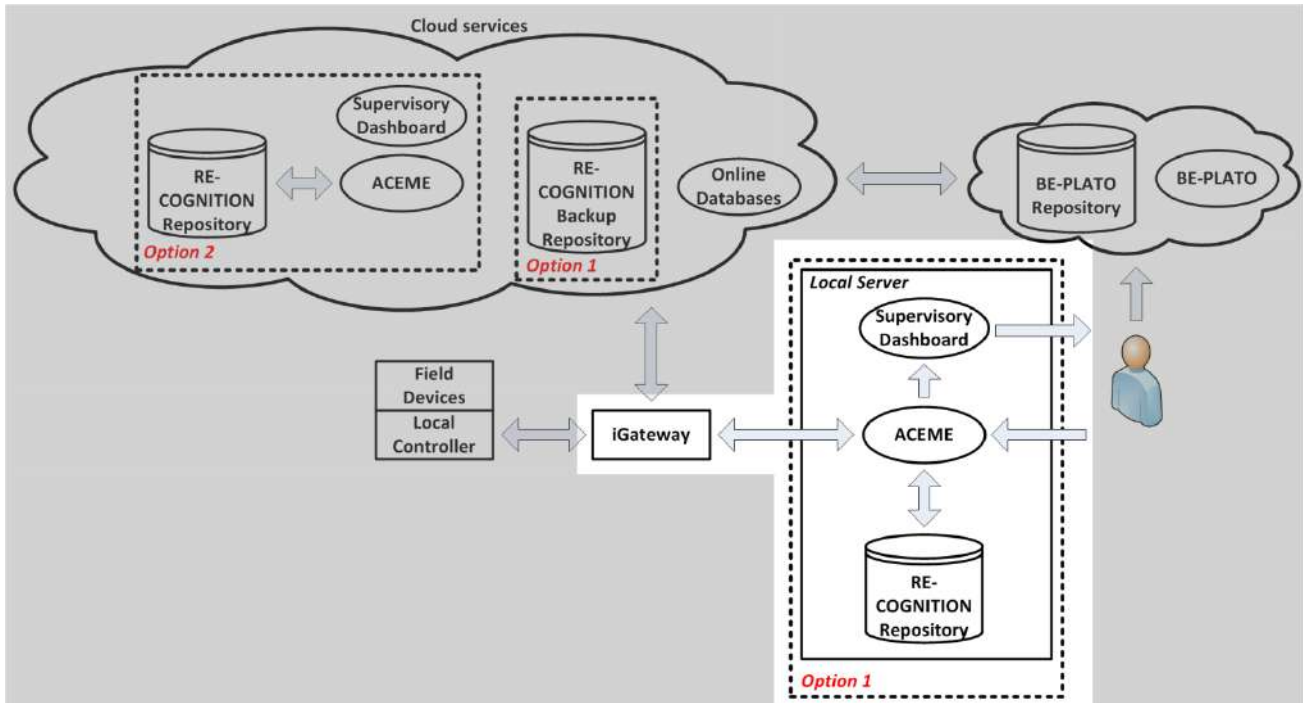


Figure 4 – RE-COGNITION Data between Field Devices and iGateway

#### 5.3.1 Purpose of the data collection/generation in relation to the project objectives

The current data flow shall be considered valid only under the Option 1 platform configuration, namely a platform with ACEME, RE-COGNITION Repository and Supervisory Dashboard installed on a Local Server.

The data flow shall consist of:

- Control signals generated by ACEME, collected by iGateway and redirected to the corresponding Field Devices
- Measurement, control and status data from Field Devices, received by iGateway and redirected to the Local Server, supporting ACEME and RE-COGNITION Database operation
- Online Database data, received by iGateway and redirected to ACEME in order to evaluate the short-term optimal plan
- Data generated by ACEME, received by iGateway and redirected to the cloud-based RE-COGNITION Backup Repository for secure storage and data-sharing (Open data and Consortium confidential data)

### 5.3.2 Types and formats of data will the project generate/collect

Information about the type, filename and data format of this data flow has been already presented in details in previous subchapters. In more details:

- Information about Control signals from ACEME and Measurement, control and status data from Field Devices are presented in 5.1.2
- Information about the Online Database data and cloud-based RE-COGNITION Backup Repository data are presented in 5.2.2.1

Regarding the communication protocols the data format shall be compatible with TCP/IP and Wi-Fi (IEEE 802.11x).

### 5.3.3 Re-Use of Data

Data in this category shall be either real-time data which won't be stored in their current form or data transferred between databases. Thus data of this path shall not be reused.

### 5.3.4 Data Origin

Information about the data origin has been already presented in details in previous subchapters. In more details:

- Information about Control signals from ACEME and Measurement, control and status data from Field Devices are presented in 5.1.4
- Information about the Online Database data and cloud-based RE-COGNITION Backup Repository data are presented in 5.2.4.1

### 5.3.5 What is the expected size of the data?

This information shall be provided in future work packages, when more information are available regarding the pilot site equipment, ACEME and RE-COGNITION Repository architecture, stakeholder agreement and used online databases.

### 5.3.6 Data Utility

Information about the data utility has been already presented in details in previous subchapters. In more details:

- Information about Control signals from ACEME and Measurement, control and status data from Field Devices are presented in 5.1.6
- Information about the Online Database data and cloud-based RE-COGNITION Backup Repository data are presented in 5.2.6.1

## 5.4 ACEME, RE-COGNITION Repository and Supervisory Dashboard Data

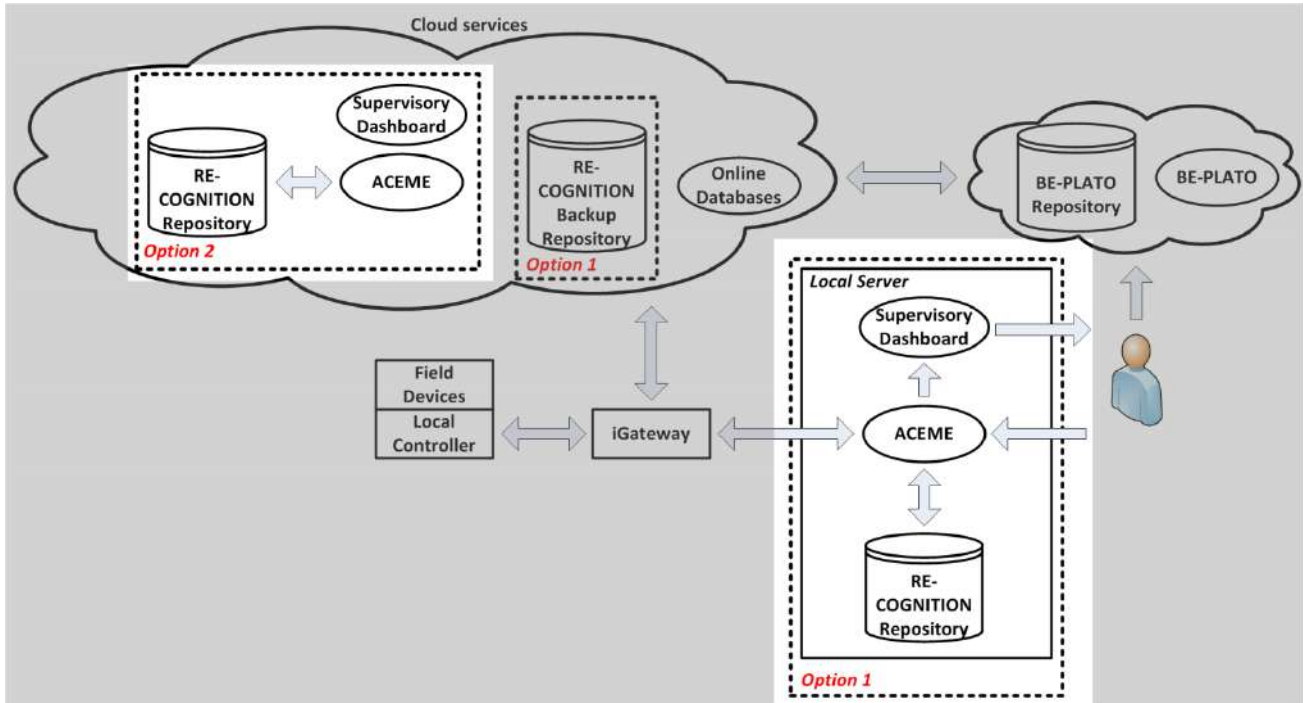


Figure 5 - ACEME, RE-COGNITION Repository and Supervisory Dashboard Data

The information about the data of ACEME, RE-COGNITION Repository and Supervisory Dashboard, presented in 5.4, remains the same regardless the platform configuration (Option 1 or 2).

### 5.4.1 Purpose of the data collection/generation in relation to the project objectives

ACEME shall use the data provided from the authorized users and iGateway in order to design its optimization algorithms for the optimum power flow dispatch. First of all, a baseline energy consumption will be calculated based on the collected data. At the same time, the forecast of load and RES generation will be determined through the exploitation of prediction algorithms which will also use available site and environmental collected data along with historical data. The visualization of the above mentioned will be presented in the Supervisory Dashboard (e.g. graphs, diagrams) and stored in the RE-COGNITION Repository files. In addition ACEME, under the agreement of the platform user, shall generate and store in the Repository Open Data publicly shared and Consortium confidential data.

In more details, authorized platform users, through special user interfaces (wizards), shall be able to request graphs from Supervisory Dashboard, providing information such as key performance indicators (KPI), field assets' characteristics (constraints), RES or magnitudes of interest.



Depending on the platform configuration Users shall have access to user interfaces either directly on the Local Server (Option 1) or using an internet browser (Option 2).

ACEME shall use the input data to:

- request the appropriate data from the Online Databases
- organize data from Field Devices and online Databases, create meta-data and store all of them in a central database, in RE-COGNITION Repository, applying appropriate form for ACEME algorithm process
- generate control data for Field Devices
- generate special graphs and plots for the Supervisory Dashboard according to Users request. Graphs, plots and updated user profiles shall be stored in the RE-COGNITION Repository.
- generate data for the RE-COGNITION Repository, categorized in:
  - Pilot sites generated datasets shared between the Consortium partners
  - Pilot sites generated datasets that are used for individual partner purposes (Private)
  - Pilot sites generated datasets shared to the public (Open Data)
  - Research findings and outcomes, which should be publicly disseminated (Open Data)
  - The aforementioned data availability can be divided in three categories:
    - Open Data: Data that shall be publicly shared for re-use and exploitation
    - Consortium: Confidential data that shall be available only to the members of the consortium and the EU Commission services and subject to the project Non-Disclosure Agreement (NDA)
    - Private: Data that shall retained by individual partners for their own processes and tests

#### **5.4.2 Types and formats of data will the project generate/collect**

Information about the type, filename and data format of the data on the path to/from iGateway has been already presented in details in previous subchapters (5.1.2, 5.2.2.1).

Types, filename and data formats of the Supervisory Dashboard graphs, RE-COGNITION Repository and user interface files depend on ACEME architecture, that will be finalized in future work packages. Such types would mainly include time series data (energy measurements / power quality measurements -if needed- from the pilots' metering infrastructure, control set-points), but

also static information such as key performance indicators and field assets' characteristics (constraints).

Regarding the file formats, they may include CSV, XML, JSON formats .etc. The RE-COGNITION Repository shall be a Relational Based SQL Repository, using database management system (DBMS) such as Postgres, MySQL and InfluxDB. However the Repository architecture shall be explicitly determined by the end of RE-COGNITION Common Information Model T3.1.

#### **5.4.3 Re-Use of Data**

All data saved on the RE-COGNITION Repository shall be available for re-use, complying with the platform user agreement and stakeholders' engagement.

#### **5.4.4 Data Origin**

Information about the origin of the data on the path to/from iGateway has been already presented in details in previous subchapters (5.1.4, 5.2.4.1).

User interface data shall be provided by authorized platform users.

All the aforementioned data, stored in the RE-COGNITION Repository or presented in the Supervisory Dashboard shall be generated by ACEME.

#### **5.4.5 What is the expected size of the data?**

The expected data size shall be a critical parameter for the RE-COGNITION Repository and will depend in several parameters such as:

- The number and type of Field Devices per pilot site
- The measurement sampling rate ACEME shall require for an optimum control
- Size of the measurement datasets and RE-COGNITION Repository storage duration to cover the system overview and system debugging needs, platform devices warranty services
- The maximum allowed Supervisory Dashboard and user profile size
- data exchange intervals
- included meta-data

The aforementioned parameters shall be finalized and available on future project tasks.

## 5.5 BE-PLATO and BE-PLATO Repository data

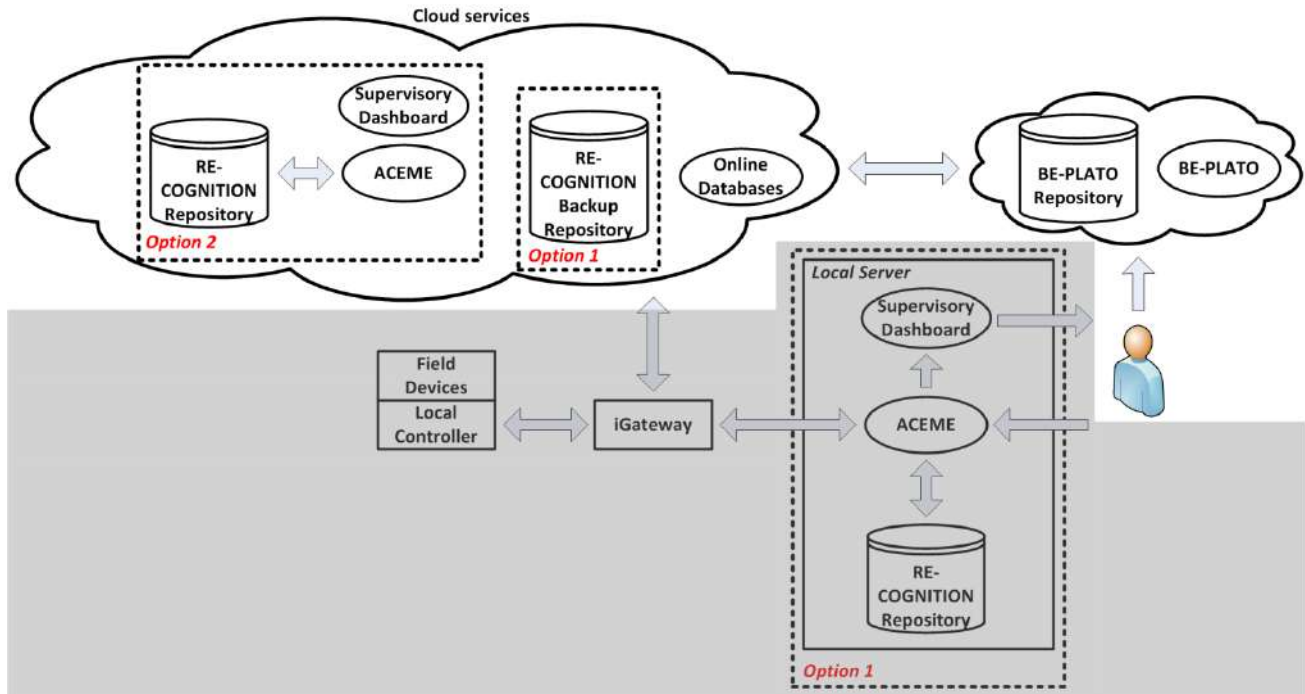


Figure 6 - BE-PLATO and BE-PLATO Repository data

### 5.5.1 Purpose of the data collection/generation in relation to the project objectives

The main objective of BE-PLATO is to provide information about economic viability and effects (e.g. reduction CO<sub>2</sub> emissions) of either introduction or modification of RES in buildings, optimizing the design of the RES installation.

Two versions of cloud-based BE-PLATO portals shall be available regarding the user access rights on the RE-COGNITION backup Repository data. The first portal version shall be dedicated to stakeholders, platform owners and consortium members who shall have access on Consortium and Private data presented in 5.4.1. Third party users shall use a second PE-PLATO portal with limited access only to Open Data (5.4.1).

All authorized Users shall be able to login to RE-COGNITION marketplace, provide information to the dedicated BE-PLATO portal about the building and then BE- PLATO, using publically available RES potential datasets (Online Databases) for solar radiation and Global Wind Atlas for wind potential, will provide an optimum design of RES installation. User profiles and generated data shall be stored on a dedicated cloud-based repository (BE-PLATO Repository).

### 5.5.2 Types and formats of data will the project generate/collect

Information about the type, filename and data format of the Online Databases data has been already presented in details in 5.2.2.1.

RE-COGNITION Backup Repository shall use the same types, filename and data formats with RE-COGNITION Repository, as presented in 5.4.2.

BE-PLATO shall contemplate several forecasting models for scenario analysis of electric and thermal energy generation from different renewable energy systems such as: (1) BIPV and photovoltaic in general, (2) variable geometry wind turbine, (3) energy battery, (4) biogas micro-CHP system, (5) solar cooling unit and (6) thermal storage to provide suggestions for optimal configuration of RES improving user awareness of the benefits of installing renewable systems. To this end, also specific models for weather forecasts and for the consideration of the economic constraints will be used. The filename and data format shall be determined in future work packages, relevant to BE-PLATO and BE-PLATO Repository development.

BE-PLATO Repository shall be a Relational Based SQL Repository using database management system (DBMS) such as Postgres, MySQL and InfluxDB. The supported file formats may include CSV, XML, JSON formats etc. As in case of RE-COGNITION Repository, the architecture shall be explicitly determined in future project task.

The User Interface data filenames and data format shall be determined in future work packages, relevant to BE-PLATO portals development.

### 5.5.3 Re-Use of Data

All data saved on the RE-COGNITION Backup Repository and BE-PLATO Repository shall be available for re-use, complying with the platform user agreement and stakeholders' engagement.

### 5.5.4 Data Origin

Information about the origin of the data on the Online Databases has been already presented in details in 5.2.4.1.

User interface data shall be provided by authorized platform users.

RE-COGNITION Backup Repository data shall be generated by ACEME.

All data stored in the BE-PLATO Repository shall be generated by ACEME.

### 5.5.5 What is the expected size of the data?

RE-COGNITION Backup Repository, as an image of RE-COGNITION Repository, shall have the same expected data size as presented in 5.4.5.

The expected data size for the BE-PLATO Repository will depend in several parameters such as:

- The maximum allowed user profile size
- data exchange intervals
- included meta-data

The aforementioned parameters shall be finalized and available on future project tasks.

### 5.5.6 Data Utility

BE-PLATO data shall be used:

- Building Occupants
- Facility managers & owners
- ESCOs
- Commercial and Residential Customers
- Stakeholders at the Pilot Sites
- General Public

## 6 FAIR Data (Findability, Accessibility, Interoperability, and Reusability)

### 6.1 Making data findable, including provisions for metadata

The platforms repositories (RE-COGNITION Backup, RE-COGNITION and BE-PLATO Repositories) shall be designed in a way to not only to help the efficient and fast data process but also optimum findability, accessibility, interoperability and reusability.

For this reason well-known, certified database technologies shall be applied such as Relational Based SQL databases and database management system (DBMS) (Postgres, MySQL and InfluxDB etc.).

Applied filename formats shall be relevant to critical information such as Field Devices, Pilot sites, User profile, timestamps and version numbers.

In addition indicative metadata shall be generated which shall include:

- Description of the pilot setup (e.g. location, date, etc.) and procedure that led to the generation of the dataset
- Specific event and state of this event, etc.

Filename formats and meta-data, supported with special keywords and filters, shall help the extraction of the desired information and its transfer among subsystems and platforms more efficiently.

However detailed description about the aforementioned features shall be feasible in future work packages. By then datasets will be accompanied by a detailed documentation of their contents.

### 6.2 Making data openly accessible

In general, a data access committee shall be introduced, separated in two groups concerning their role about the data handling activities; the data controllers and the data processors. The data controllers shall be the ones imposing the purposes and rules of data processing, while the processors shall be the ones processing the personal data on behalf of the controller. The data controllers shall be the Pilot Responsibles and the Data Protection Officers (DPOs), whereas the data processors shall be all the technical partners who will need to analyze data from the pilot use cases.

The data access committee shall define how data will be shared and more specifically the access procedures, the necessary software and other tools for enabling re-use, for all data-sets that will be collected, generated or processed in the project.

Within the RE-COGNITION platform, ACEME and BE-PLATO are the units that shall organize input data and generate new ones that will be stored in the platform repositories so as to be re-used. As stated before the generated datasets can be categorized in:

- Pilot sites generated datasets shared between the Consortium partners
- Pilot sites generated datasets that are used for individual partner purposes (Private)

- Pilot sites generated datasets shared to the public (Open Data)
- Research findings and outcomes, which should be publicly disseminated (Open Data)
- User profile datasets

The aforementioned data availability can be divided in three categories:

- Open Data: Data that shall be publicly shared for re-use and exploitation. The open access to the public may be granted on a fully anonymized dataset, under specified conditions by the consortium members
- Consortium: Shared datasets with confidential data shall be allowed between the consortium members for the purpose of fulfilling the project's objectives, For this reason, and under specified conditions, NDAs would be introduced and signed among involved data processors and controllers before distributing these data within the consortium members.
- Private: Data that shall retained by individual partners for their own processes and tests

Data access rights shall be determined in future project tasks taking into account:

- the determination of the required datasets of Consortium partners, stakeholders, platform users, third party users
- stakeholders engagement and participation, based on the pillars of Responsible Research and Innovation (RRI)
- ethical, rules of personal data, intellectual property, commercial / privacy-related, security-related reasons.

With the exception of ACEME users at Option 1 platform configuration, who shall use directly the user interface on the Local Servers, authorized users shall be able to access datasets, meta-data and information mainly stored in cloud-based repositories.

In addition RE-COGNITION will explore the possibility to utilize online platforms such as ResearchGate and OpenAIRE's Zenodo repository to disseminate public deliverables, technical reports, working papers and conference papers, which are not scientifically peer reviewed, openly accessible via the project website.

Authorized users shall be able to use internet browsers to login to RE-COGNITION marketplace and use the corresponding portals, wizards and other user interfaces. The authentication shall be verified with secure, well-known tools such as SSH and OATH 2.0.

Instructions and documentation about the user interfaces and dataset content use and access rights / limitations shall be provided.

### 6.3 Making data interoperable

The RE-COGNITION information exchange will be based on well-known standards such as ETSI SAREF and SAREF4ENER and common industry protocols such as MQTT and Modbus over TCP/IP

As for the data models/ontologies which would be developed in RE-COGNITION, these would be maintained and found in secure project servers which will be determined in further steps.

### 6.4 Increase data re-use (through clarifying licenses)

When the corresponding project work packages for:

- determining the required datasets of Consortium partners, stakeholders, platform users and third party users
- defining the stakeholders engagement and participation based on the pillars of Responsible Research and Innovation (RRI)
- gathering ethical, rules of personal data, intellectual property, commercial / privacy-related, security-related reasons

are finalized, the data access committee shall be able to determine the initial data classification and access rights in a way to promote an increased re-use of Open and Consortium Data.

However both classification and data access shall be dynamic and be adapted on changing needs (pending patents, publication review periods, platform hardware changes, changes of ownership etc.). The duration of these changes shall vary in each case.

The committee shall be responsible for the data management as long as the data are available to authorized users for re-use. This period is not determined yet but may overcome the RE-COGNITION project end for up to five years.

Quality Plan D7.1 describes in Chapter 9 the data quality assurance process that shall be applied.



## 7 Resource Allocation

The costs making data FAIR in RE-GOGNITION project can be distinguished in three groups:

- Cost of developing the platform repository infrastructure
- Costs for the data management
- Costs for data re-usability

Costs for developing the platform repository infrastructure include the Research and Development (R&D) costs for the RE-COGNITION repositories, ACEME and BE-PLATO, which shall be covered by the corresponding budget of each partner, as declared in Grant Agreement.

Data management costs refer to Manhours the data access committee shall spend, costs that are already included to partners approved budget.

Costs for data re-usability refer to operational costs keeping data available to authorized users on Repositories and online platforms. These costs would depend on the amount of data and storage duration, parameters that shall be determined in further stages of RE-COGNITION project.

## 8 Data security

All RE-COGNITION Repositories shall be hosted on certified Servers. The storage means shall be in RAID configuration (greater than 1) and encrypted with certified third party software, which is not determined yet.

The data shall be stored for a long term preservation and curation, which duration shall be determined in future steps

## 9 Ethics, Security and Legislation

### 9.1 Ethics and Security

Ethics and security issues will be thoroughly investigated during the project lifetime in other specific but interrelated tasks. A preliminary list of main ethical risk factors is listed in the following table. An elaborated version of these will be delivered within WP9 in the context of defining the RE-COGNITION Ethics Policy.

Table 2 - RE-COGNITION Ethics Policy

#	Ethical Risk	Description of Risk	Foreseen Risk Management Measures
1	Loss of Privacy Control	Storage and process of human data towards detection of abnormal patterns and flows within Pilots sites	For human presence privacy-preserving sensors will be utilized. Any original records or data will be destroyed after that, if this is not forbidden by law of the country in which the information was collected, stored and analyzed.
2	Data Security	Difficulty in ensuring the security of household/personal data in the pilot sites	Special attention will be given to ensure confidentiality and for incorporating PET technologies to ensure protection from data breaches. RE-COGNITION partners have the capacity and the experience to cope with the delivery of security mechanisms, if needed.
3	Storage and process of personal data, Confidentiality	Measurements from various sensors and applications will be transmitted wirelessly Difficulty in ensuring the security of human data collected during the execution of the Pilots	RE-COGNITION consortium has the expertise and the know-how from similar past and ongoing research projects, towards providing the necessary ethical guidelines that should be adopted during the execution of the Pilots. Local ethical committee (and the National committee, if needed) will be informed towards getting an official permission for the execution of the selected Pilots.
4	Lack of Transparency	Work of professionals (Workers/Employees in selected Pilots, visitors, etc.)	An ethics manual will be delivered for the Pilots towards all activities performed to be in compliance with National and European legislation. Prior the execution of the pilots the local ethical committees will be informed for the data collection as part of the RE-COGNITION experimentation and the necessary documents will be created by the consortium in order to get an ethical approval.
5	Delegation of control Privacy Incidental Findings	Need to Notify proper Pilot authorities	Within <b>T1.1</b> , a sub-activity has been included to address local and European legislation. In that context, all the Pilots will be performed according to them and relevant data protection authorities will be informed on time.
6	Use of IT Equipment	Installation and use of equipment on the pilot sites	The consortium partners have the expertise to make the appropriate installation for the purposes of the pilots. Most of the partners have participated in several National and European projects related to integration of sensors for research purposes and their use in ethical compliance with National and European

			legislations.
--	--	--	---------------

Any data collection involving humans will be strictly held confidential at any time of the research. This means in detail that:

- All the test subjects, and their legal guardians where applicable, will be informed and given the opportunity to provide their consent to any monitoring and data acquisition process that all the subjects will be strictly volunteers and all test volunteers receive detailed oral information.
- No data will be collected without the explicit informed consent of the individuals under observation and their legal guardian where applicable. This involves being open with participants about what they are involving themselves in and ensuring that they have agreed fully to the procedures/research being undertaken by giving their explicit consent.
- No personal or sensitive data will be centrally stored. In addition, data will be scrambled where possible and abstracted in a way that will not affect the final project outcome.
- No data collected will be sold or used for any purposes other than the current project.
- A data minimisation policy will be adopted at all levels of the project and will be supervised by the Ethics Panel. This will ensure that no data which is not strictly necessary to the completion of the current study will be collected.
- Any shadow (ancillary) personal data obtained during the course of the research will be immediately cancelled. However, the plan is to minimize this kind ancillary data as much as possible. Special attention will also be paid to complying with the Council of Europe’s Recommendation R(87)15 on the processing of personal data for police purposes, Art.2 : “The collection of data on individuals solely on the basis that they have a particular racial origin, particular religious convictions, sexual behaviour or political opinions or belong to particular movements or organisations which are not proscribed by law should be prohibited. The collection of data concerning these factors may only be carried out if absolutely necessary for the purposes of a particular inquiry”.
- Specific measures will be in place in order to protect the pupils from a breach of privacy/confidentiality and any potential discrimination; In particular their names will not be made public and their participation will not be communicated to. Any incidental findings will be kept strictly confidential and erased from files under request from the enrolled subject.

The research to be conducted will be in full compliance with the principles and guidelines of “Ethics for Researchers” to Facilitate Research Excellence, prepared by the EC Governance and Ethics Unit in 2007

## 9.2 Legislative Framework

The RE-COGNITION consortium is aware that a number of ethical, privacy and data protection issues will be raised by the activities to be performed in the scope of the project. The project involves the carrying out of specific data collection in a set of **five (5) pilots in United Kingdom, Italy, Greece and Romania**, and may also perform similar activities in Greece in the context of pre-validation activities (one (1) pre-pilot in Greece) in order to assess the overall impact, performance, cost-effectiveness and cost-efficiency of the proposed solution. For this reason, human participants will be involved in certain aspects of the project and data will be collected. This will be done in full compliance of the main legislation and directives and more specifically:

- The Universal Declaration of Human Rights and the Convention 108 for the Protection of Individuals with Regard to Automatic Processing of Personal Data;
- 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;

In addition, RE-COGNITION will fully comply to specific (and in certain cases more strict) national legislation of the pilot countries involved in the project. As previously mentioned, specifically focused pilot activities (pilot demonstrators) are foreseen in United Kingdom, Italy, Greece and Romania under the leadership of the coordinating partners (EC, POLITO, Lizzanello (see Letter of Intent), HELPE and TUCN correspondingly, as well as CERTH for pre-pilot in Greece).

An indicative list of major national legislation acts for these countries is provided hereafter:

- **Italy:** The protection of personal data laid down in particular in the following legal standards: **Italian Law 196 – 30th June 2003** “Code concerning the protection of personal data” and the further General Authorisations Issued for the Processing of Sensitive Data (as currently in force) and reported in the following
  - Authorisation No. 1/2014 Concerning Processing of Sensitive Data in the Employment Context
  - Authorisation No. 2/2014 Concerning Processing of Data Suitable for Disclosing Health or Sex Life
  - Authorisation No. 3/2014 Concerning - Processing of Sensitive Data by Associations

and Foundations

- Authorisation No. 4/2014 Concerning - Processing of Sensitive Data by Self-Employed Professionals
  - Authorisation No. 5/2014 Concerning - Processing of Sensitive Data by Various Categories of Data Controller
  - Authorisation No. 6/2014 Concerning Processing of Sensitive Data by Private Detectives
  - Authorisation No. 7/2014 Concerning Processing of Judicial Data by Private Entities, Profit-Seeking Public Bodies and Public Entities
  - Authorisation No. 8/2014 for the Processing of Genetic Data
  - Authorisation no. 9/2014 - General Authorisation to Process Personal Data for Scientific Research Purposes
- **Greece:** The General Data Protection Regulation (EU) 2016/679: “For the protection of individuals with regard to the processing of personal data and the free movement of such data and repealing Directive 95/46 / EC” prevails in Greece, as the dominant legislation for data protection.

Relevant national legislation includes:

- Law 3471/2006: Protection of personal data and privacy in the electronic communications sector and amendment of Law 2472/97

Whereas relevant authorities include:

- Regulatory authorities and ethical committees: Hellenic Data Protection Authority (<http://www.dpa.gr/>)
- **United Kingdom:** The relevant legislation includes:
- “**Data Protection Act 1998**”, <http://www.legislation.gov.uk/ukpga/1998/29>
  - “**Framework for Research Ethics, Economic and Social Research Council**”, <http://www.esrc.ac.uk/about-esrc/information/framework-for-research-ethics/> (Latest Version 2015)

Relevant national legislation includes:

- United Kingdom Ethics Committee Authority (UKECA), <http://www.eurecnet.org/information/uk.html>

- The Information Commissioner’s Office, <https://ico.org.uk/global/contact-us/>
- Association for Research Ethics, <http://www.arec.org.uk/>

➤ **Romania:**

Relevant national legislation includes:

**“The Romanian Law no. 121/2014, on Energy Efficiency is the transposing of the EU DIRECTIVE 27/2012 on the national level”**

- *Annex 8 Energy efficiency criteria for power network regulation and for electricity network tariffs*
- *Annex 9 Energy efficiency requirements TSO’s and DSO’s*
- **“Law No. 677/2001 on the protection of individuals with regard to the processing of personal data and the free movement of such data, English Translation of”**, <http://www.dataprotection.ro/servlet/ViewDocument?id=174>

Relevant authorities include:

- The National Supervisory Authority for Personal Data Processing, <http://www.dataprotection.ro/>

Summarizing, core ethical issues within RE-COGNITION will be addressed by fully complying with EU and national legislation. The **RE-COGNITION Data Management Plan**, will evolve around these main principles:

- Ensure transparency on all data collection and management practices performed by the project and notify all people and stakeholders involved
- Confirm the (explicit and written) Informed Consent of occupants involved in the project pilot evaluation phase, while option to withdraw will be available at any time
- Safeguard data protection, security and privacy issues through an integrated security and ethics management policy throughout technologies as well as data management practices in the project’s field of research (household energy consumption data & respective profiles)

## 10 Conclusion

This Data Management Plan provides an overview of the data that RE-COGNITION will produce together with related challenges and constraints that need to be taken into consideration.

The analysis contained in this report allows anticipating the procedures and infrastructures to be implemented by RE-COGNITION to efficiently manage the data it will produce.

Nearly all project partners will be owners or/and producers of data, which implies specific responsibilities, described in this report.

The RE-COGNITION Data Management Plan will put a strong emphasis on the appropriate collection – and publication should the data be published – of metadata, storing all the information necessary for the optimal use and reuse of those datasets.

Specific attention will be given to ensuring that the data made public breaks neither partner IPR rules, nor regulations and good practices related to personal data protection. For this latter point, systematic anonymization of personal data will have to be provisioned.

As stated in the beginning of this document, The Data Management Plan is a living document and will be updated, based on the effective progress of the activities, and depending on the evolution of data to be distributed between the partners