SIT4EnerGY

Project Acronym:	**SIT4Energy**
Project Full Title:	**Smart IT for Energy Efficiency and Integrated Demand Management**
Project Duration:	**36 months (01/03/2018 – 28/02/2021)**

DELIVERABLE 3.6
Report and data personalization and privacy issues for SIT4Energy

Work Package	**WP3 – Design and Analysis of Technology Pillar**
Task	**T3.6 – Personalization and privacy of data**
Document Status:	**Version 0.1**
Due Date:	**31.08.2019**
Submission Date:	**31.08.2019**
Lead Beneficiary:	**CERTH**

Dissemination Level	
Public	**X**

European Union
European Regional
Development Fund

Co-financed by Greece and the European Union

Authors List

Leading Author			
First Name	**Last Name**	**Beneficiary**	**Contact e-mail**
Adamantia	Chouliara	CERTH	achouliara@iti.gr
Co-Author(s)			
#	**First Name**	**Last Name**	**Beneficiary**	**Contact e-mail**
1	Christopher	Schneider	SHF	Christopher.schneider@stwhas.de
2	Kostas	Peppas	CERTH	kpeppas@iti.gr

Legal Disclaimer

The SIT4Energy has received funding from the German Federal Ministry of Education and Research (BMBF) and the Greek General Secretariat for Research and Technology (GSRT) in the context of the Greek-German Call for Proposals on Bilateral Research and Innovation Cooperation, 2016. The sole responsibility for the content of this publication lies with the authors. It does not necessarily reflect the opinion of the funding agencies or the European Commission (EC). Funding Agencies or the EC are not responsible for any use that may be made of the information contained therein.

Copyright

Executive Summary

This deliverable presents the measures taken to preserve data confidentiality and secure the participants of the SIT4Energy pilots from personal data misuse and privacy violation. It introduces the guidelines followed to tackle their privacy issues and the methods of data collection, storage and manipulation during all the phases of the project for preserving privacy user rights. The considerations required for exchanging data between the project participants are described in detail.

More specifically, the Deliverable D3.5's main objectives are:

- Explanation of guidelines concerning the data collection and management set out in the Cooperation Agreement.
- Clarification of the pilot enrolment procedure and the burdens on behalf of the participants
- Description of data collection, storage and process protective measures and limitations
- Specification of the pilot duration
- Identification of storage limitation and the user right to access, amend, correct and delete their personal data
- Commitment of accountability by the project's consortium regarding detailed information about pilot procedures in compliance with the GDPR offered to the participants

Table of Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

Term	Description
DSO	**Distribution System Operator**
SEMD	**Smart Energy Management Dashboard**

1.   Introduction

1.1   *Scope and objectives of the deliverable*

The purpose of this report is to produce a document that will guide the project consortium to comply with EU data protection and privacy principles along with regulating issues concerning privacy and security. Moreover, it ensures the neutrality of the recommendation systems and their appropriateness for various groups of users tackling all the related privacy issues. It also highlights the methods of data collection, storage and manipulation during all the phases of the project for preserving privacy user rights.

This document will be revised if new information arises during the project development.

1.2   *Structure of the deliverable*

The deliverable is built as follows:

An introduction to the deliverable is given in *Section 1*.

Section 2: presents the guidelines set by Cooperation Agreement ensuring that all data collection and management procedures will preserve the privacy of the user

Section 3: presents the policy of enrolment in pilot trials and the precise information offered to the participants through signed content forms

Section 4: examines the methods of anonymization, encryption and security applied strictly during data collection, storage and process

Section 5: declares the pilot trials planned duration and foreseen lifetime

Section 6: outlines the storage limitation regarding the nature and expected lifetime of data used

Section 7: indicates the user awareness that they have the right to access, amend, correct and delete their personal data

Section 8: refers to the project consortium's accountability providing the participants detailed information about pilot procedures in compliance with the GDPR

Section 9: presents the conclusions and summarizes the document

1.3   *Relation to Other Tasks and Deliverables*

This deliverable introduces the data protection, security and privacy principles and regulations.

The main dependencies with other deliverables are concentrated in WP2 where the legal issues are addressed as well as WP3 and WP4 where end-users and pilot participants are involved as briefly outlined below:

The basis on which this deliverable forms its guidelines creating a further extension is D2.4 *Legal/Regulatory barriers and Energy End User Rights for SIT4Energy*. This deliverable analyses legal regulations regarding Energy End users' privacy and data protection rights and it constitutes the legal baselines D3.1 *SIT4Energy activity tracking algorithms and micro-moments creation* will involve pilot trials ensuring confidentiality of user provided data. In D3.2 *SIT4Energy mobile recommendation system DNA*, the recommendation systems designed will employ pilot participants following strict personal data guidelines D3.3 *SIT4Energy Adaptive incentivization service* engages consumer and prosumer end-users ensuring strict privacy policies. D3.4 *SIT4Energy Context-aware attention triggering service* will ensure consumers' and prosumers' consent and will take into account all guidelines about the protection of their personal rights in data handling. D3.6 *SIT4Energy Smart Energy Management Dashboard* will design and develop the Smart Energy Management considering privacy issues and following strict data protection policies D4.1 *Global SIT4Energy framework* will provide the unified project's framework in full accordance to participating users' data privacy and rights.

Moreover, the deliverables D4.2 *Pilots setup*, D4.3 *Greek smart home pilot outcomes*, D4.4 *Greek academic campus pilot outcomes*, D4.5 *German household pilot outcomes* deal with pilot employment issues and regulations and are fully coordinated with the privacy of user information and user personal data rights.

2.   Guidelines

The SIT4Energy project has followed the appropriate guidelines set out in the Cooperation Agreement. These guidelines aim at preserving the privacy of the user, protecting his/her private data and limiting the risk of improper use to the minimum. The following requirements have been fulfilled to ensure data integrity throughout the project activity:

- The pilot supervisor (for all foreseen pilot premises) will inform the participants with clarity about the procedure of the pilot tests, the system operation and the objectives, the data retrieval and storage and the exact dates the tests will be running.
- No sensitive personal data shall be collected.
- A data minimization policy will be adopted at all levels of the project and will be supervised by the Ethic Helpdesk of the project in line with European legislation (GDPR) and current national legislations of Germany and Greece, as explained and documented in D2.4 [1]
- This data minimization policy will ensure that data not strictly necessary for the completion of the current research pilot will not be collected.
- No data will be collected without the explicit written consent of the individuals taking part in the pilot test. Consent forms have already been drafted and distributed for the pre-pilot tests.
- No collected data will be sold or used for any purposes other than the SIT4Energy project.
- The research will be conducted in full compliance with the principles and guidelines of EU ethics for research projects, namely:
  - Horizon 2020 Rules for Participation: Ethics Reviews (Article 14)
  - Horizon 2020 – Regulation of Establishment: Ethical principles (Article 19)

The SIT4Energy consortium will be transparent in the collection of personal data. This means when collecting the data, a Research Information Letter and a consent form will be provided to the users (See D2.4) [1]. This document describes the kind of information, the manner in which it will be collected and processed, and how, and for which purpose it will be disseminated. Furthermore, the subjects will have the possibility to request what kind of information has been stored about them and they can request up to a reasonable limit to be removed from the results.

3.   Participation in the trial

On a voluntary basis, end-users of the pilot facilities can enrol on the trial only after they have read, accepted and signed the consent form.

Personnel working or residing in the selected pilot sites have already been informed about the project's goals and the objective of the trials and the extraction procedures that will take place during the pilot tests. Those who have given their preliminary consent in participating in the pilot procedures will be contacted again before the actual pilot preparation phase.

Participants will enrol in the pilot test as part of their daily activities, as no further work will be requested for their participation in the STI4Energy project.

The burden upon participants will not be superior to that of just installing a mobile application or accessing a web-based framework and follow tips to improve energy efficiency. Participants will be adequately informed and assured that they have the possibility to opt out of the pilot trial at any point, if they wish, without any consequences. They will be able to exercise their right to access their data as well as their extracted profiling parameters and delete their data (up to a reasonable limit) at any point throughout the duration of the trial test. If employees of partner organisations, or university students serving in any partner university, are to be recruited, specific measures will be in place in order to protect them from a breach of privacy/confidentiality or any potential discrimination; Additionally, in

the case of enrolment of employees of any partner organisation, the recruitment method and informed consent procedures will ensure no coercion (not even soft or indirect) is exerted.

4.   Data Collection, Storage and Process

For the need of the pilot realization and validation, a set of different approaches will be followed at the different pilot premises, based on the use cases foreseen (residential, tertiary, etc.). For the SIT4Energy goals, only monitoring of aggregated energy-related information will take place at the various buildings, utilising existing metering infrastructure. In addition, the mobile app will have access to the mobile sensors, but analysing data on the fly and without storing them outside the device. There will be no raw data stored from these sensors and only certain features will be exposed to the SIT4Energy components as it will be explained below.

All meters used during the deployment of the GreenSoul framework will not acquire any sensitive personal data nor violate end-users' privacy, following Regulation (EU) 2016/679 of the European Parliament and of the Council, and current relevant national legislation in force where the pilot building is located not overlapping the GDPR´s scope of application. Accordingly, all SIT4Energy software components will not store or expose any data of that sort that may compromise personal data or violate end-users' privacy.

To this regard, and in order to preserve privacy of users, protect their private data and limit the risk of interception to the minimum, all pilot end-users that will participate in the pilots will be appropriately informed with very specific information disclosed about their data collection and processing.

Data on activity tracking, micro-moments, habits and preferences of building users that affect energy consumption (e.g. time of arrival, working hours, HVAC preferred set-points etc.) will be collected and stored, including: a) data concerning the number of people that reside into the building spaces, b) data concerning individual and aggregated energy consumption habits (i.e. lights, HVAC, appliances consumption and time of use); in the latter case, specific devices (e.g. PC monitors) could be identified to each user (anonymously), so as to extract personalized preferences if applicable based on existing infrastructure.

Any personal data will need to be stored safely and in a secure environment, secure communication over a computer network. Finally, authorization and authentication features are foreseen and will be supported during the project lifecycle among the SIT4Energy components, whereas part of the databases that has information regarding the users (i.e. user profiles) will be also encrypted.

Most (if not all) information exchange between the mobile and web frameworks and other SIT4Energy components will be based on HTTPS. Each pilot site will ensure secure communication through HTTPS. HTTPs is used for authentication of the accessed App(s) and protection of the privacy and integrity of the data that are being exchanged while in transit. With this standard protocol we protect the communications against man-in-the-middle attacks. Through a bidirectional encryption of communications between client and server, HTTPS protect against tampering and eavesdropping of the communication.

In particular, security measures are used that guarantee: pseudonymisation or encryption of the data; the confidentiality, integrity, and availability of the data as well as the resilience of the systems and services that process them; the ability to restore data in the event of a data breach.

In order to be able to offers to the research subjects the ability to exercise their right to withdraw their data within 48 hours of the experiment, pseudoanonymization of data will be done. The key between the pseudonymised file and the list of participants will be stored in the pilot sites by data controllers on a separate physical location from the original files. Only data controllers and data processors will have access to this personal information. Do keep in mind that the research subjects should be able to withdraw their data within 48 hours of the experiment.

Access to this secure environment can be granted or revoked by either the researchers responsible for the data with the approval of the project manager, or the project management on a case to case basis and will not be given out by default to all researchers. All users that are granted access to the database with personal information will sign a confidentiality agreement about the data.

In order to be able to provide a more detailed presentation of the expected (since all components are still at the prototype level) data collection, storage, and processing, the basic SIT4Energy components and their data management follows.

4.1   User Activity Tracking Engine

All user data collected during the relevant pilot trials will be anonymised and encrypted before database storage. The key information concerning the user's physical activity is to verify whether he is in the location of interest and if so, to check for being still or in moving condition and will not be collected without user permission. The necessary data for inferring the physical activity of the user are the ones from the accelerometer sensor, which will be used but not stored. Finally, the micro-moments detection tool will receive an acknowledgement whether the user is in the target location and is performing a stationary activity, but not the data. The consortium will fully inform the participants about the procedures and their rights throughout the experiment providing them with detailed consent forms before their enrolment.

4.2   Occupancy Inference Tool

In the pilots involved, energy related information will be gathered. In particular, the data collected will be restricted in the total energy consumption, measured in regular time intervals using smart meters installed in the pilot sites. This data will be analysed for occupancy inference with the use of machine learning algorithms. This type of data will refer to energy consumption and generation at building and room level, occupancy patterns etc and will be associated only with a device ID, that will be the Meter ID without implying any direct correlation with consumer ID. The data produced from the three pilot sites activities will be anonymized using a combination of relevant techniques in order to eliminate any possibility of reidentification. After this process, they will be transferred to SIT4Energy platform. All the Consortium members are committed to treat the data anonymously, while taking into account the reidentification risks. All participants will be fully informed about the procedures and their rights through detailed consent forms given to them before the experiment.

4.3   Micro-Moments Detection Tool

In the respective pilots, experimental data will be collected by the participants regarding their mobile activity which will be the moments in which we detected a micro-moment. The detection of the micro-moment depends on the location, physical and mobile activity of the user. While the information whether the location and physical activity are the target ones is received from the the user activity tracking engine, the inference of the mobile activity needs the usage statistics of the mobile device for which the user needs to manually give permission for, but again this data is not stored. When a micro-moment is detected, the time stamp will be collected and stored. All data will be anonymised and encrypted before stored in database and will be used for selecting the right time to send notifications to the users without being interruptive. All users will be asked for data access permission through the mobile application and no data will be collected without user agreement. Moreover, all participants will be fully informed about the procedures and their rights through detailed consent forms given to them before the experiment.

4.4   Recommendation Engine

In the current phase of the Greek pilot, the accumulated data from HUA premises are transmitted to a local installation of a Home Assistant1 server and consist of:
   a)   Smart meter consumption data.
   b)   Smart sensors data (Luminosity/Temperature Sensors, Motion Detection Sensors for room occupancy etc.)
   c)   Smart Switches data (Bearing only information about their on/off state without consumption data)

This data is properly classified to depict the type of energy consumption data (HVAC, lights) as well as their topology (in order to associate energy meters with their relevant smart sensors). Pending the completion of the development of an api (application program interface) that will be used to query the HUA Home Assistant database, a partial database dump has been incorporated in the mobile app backend to support the initial testing phase.

Based on this data, the initial recommendation engine implementation is built on a rule-based system with predefined conditions that trigger a recommendation to be sent to the mobile app when these prerequisite conditions are met. In addition, the user must have given his/her consent for receiving recommendations. If these rules are only partially satisfied, the engine selects among eligible general recommendations to be sent to the app, that are still relative to the collected energy consumption type. (i.e. "If there is adequate sunlight and light energy consumption then suggest to the user to close the lights, else provide a recommendation related to energy saving of type Lights: "replace light bulbs with led to save energy")

To support the above the present recommendation engine database holds the following basic attributes

Field Name	Field Type	Description
RecomID	Integer	The unique id of a recommendation
EnergyType	Integer	The type of energy consumption the recommendation relates to
Content	Text	The recommendation content
Induction	Boolean	The type of recommendation (General/Specific)
ClusterType	Integer	User type: egoistic, altruistic, biospheric
Stage	Integer	pre-contemplation, contemplation, and action

(The ClusterType and Stages fields have been added to align with the Adaptive incentivation classification deriving from D3.3.1 for the next version of the app)

Following the value system already implemented in the first prototype of the mobile app for the Tips functionality, in the future the user's feedback regarding the accuracy and feasibility of a suggested recommendation will be collected. This feedback will be combined with other characteristics and metrics (i.e. number of office occupants, level/type of consumption) so as to train a Machine Learning model in order to accomplish higher precision in the delivered recommendations.

The precision and validity of the recommendations will be further strengthened by combining the user's feedback regarding his/her intention to follow a recommendation with the comparison of related changes in his/her energy consumption compared with historical consumption data of the user.

4.5   Context-aware attention triggering service

In first prototype, the context-aware attention triggering service does not yet collect, store or process personal information such as individual meter data or user profile information apart from the names and the E-Mail addresses of the users participating in the pilot, to which the service sends awareness-

---

[1] https://www.home-assistant.io/

raising triggers. The service in its first implementation only uses context-information coming from publicly available webservices (stock market prices, weather forecast information) and will run physically on SHF's servers protected by state-of-the-art security measures in place and fully respecting the company's data privacy and protection policies which complies with the respective legislation (see section 4.5 for further details). All users participating in the SIT4Energy pilot will be asked to give their explicit informed consent for the system to receive notifications from the context-aware triggering service and only on this consent being given will they receive notifications.

4.6   Smart Energy Management Dashboard

In the German pilot a data collection of consumption, production and dashboard usage will be made. For end-users (Prosumers) this will happen on building level. The processed data for the utility dashboard will contain of the end-user data and additionally grid operational data (e.g. LV-measurement of substations, measurement of high scale production plants or industrial/commercial customers). For both applications external providers will be used to gather environment data (e.g. weather data).

No data, which is not needed for the project, will be collected. Only data related to the triggering service and detection of the actual use case end-users, namely employees, residents, and/or visitors, will be collected. Since the collected data in the latter case may involve sensitive personal data, all provisions for data management will be made in compliance with the national and EU legislation. In particular, the data collection process will always be made by means that guarantees the user's anonymity. Further measures will be taken to ensure that all meter data, which involve personal information, will not be possible to be intercepted. Finally, special measures will be taken with user generated data, such as consumption data, in order to avoid any capacity of their use in any form of ethnic, racial, and/or gender discrimination profiling.

The collected data will be stored in secure server systems and will be anonymized. In this context, all pilot participants will be granted with an identification number (ID) based on each participant's role in each of the pilot use case, namely role ID, allowing mapping of participants' actions during the use case execution and pilot realization phase. The key to link the participant's name to the code, which identifies the data file will be provided to no one; hence, the privacy of the personal data will be fully protected. Nevertheless, data will be kept only for the period of time that is necessary in order to accomplish the goals and objectives of the SIT4Energy project.

Data is protected according to ISMS (DIN ISO 27001) and GDPR/DSGVO guidelines in SHF's premises. Access to personal data of customers is arranged and controlled strictly. The operations division of SHF can only access anonymised data of the several grid connections points. The sales division, only when asked by the customer itself or an authorisation is given.

Access to the Utility SEMD will only be given to advised employees of the control centre, that are part of the ISMS team. The data itself is stored and processed on servers in SHF's sister company "Rechenzentrum Haßfurt GmbH" in the same building, where unauthorized access to data is being prevented with state-of-the-art security technology.

4.6.1 Data Integration (data flows)

This section provides a complete but schematic overview of the main steps involved in reading, handling, organizing and forwarding the consumption data from **SHF's** smart meters and the forecast/recommendation conditions from external services (e.g. weather data) in the German pilot, as envisioned.
Figure 2 describes schematically the envisioned data flow for the Prosumer SEMD of the German pilot. A more detailed description of each step can be found down below.
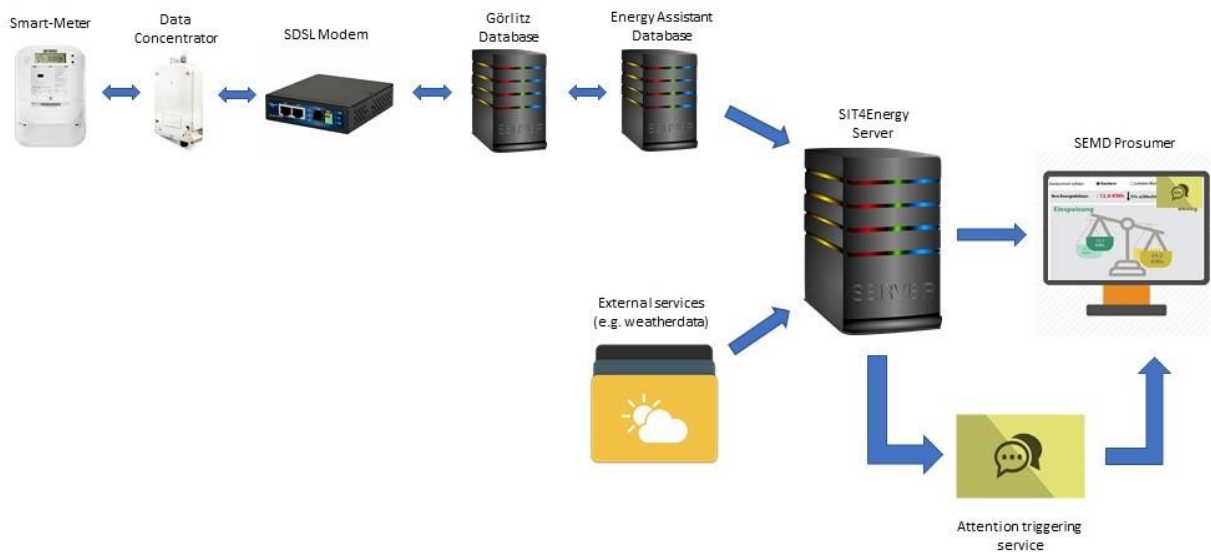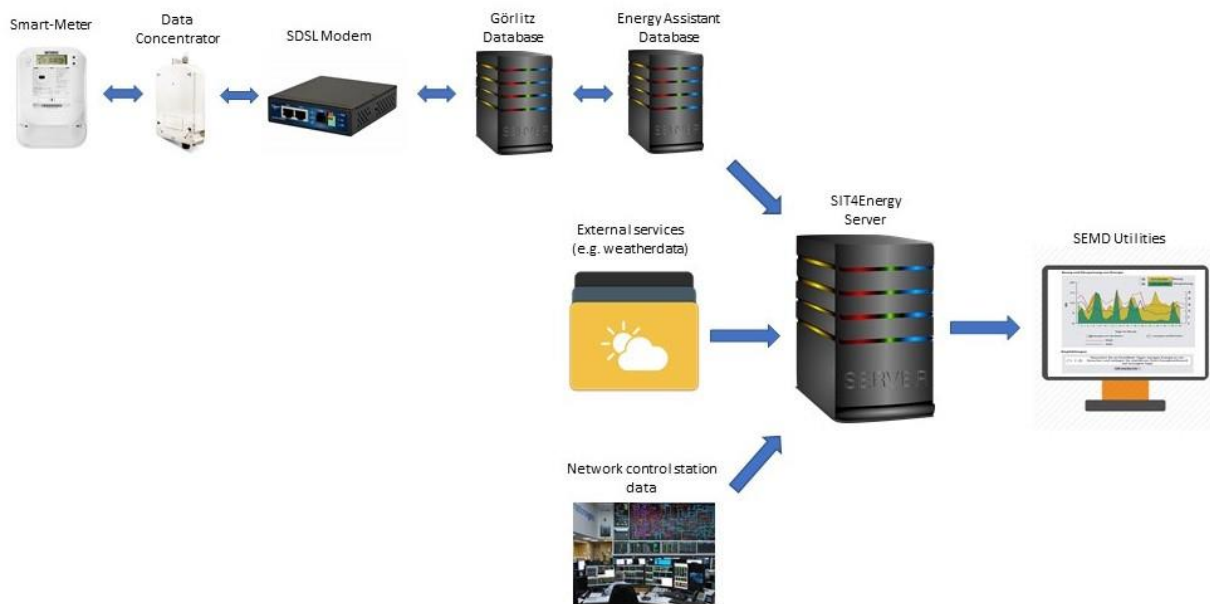
Figure 1: Data integration Prosumer SEMD - German pilot

The data reading procedure of the **end-users in Germany** follows these steps, going left to right in the flow for Prosumer SEMD in Figure 1**: Data integration Prosumer SEMD - German pilot**.

- Smart meters installed at the points of delivery read the power consumption (or production) of the client in 60-minute intervals.
- Data concentrators scattered around the power network retrieve the data from the smart meters connected to them, using power-line communication (installed since 2008).
- The smart meter management system (Görlitz Database) is a software running on SHF's servers. Periodically the concentrators push the gathered information and the data is stored in the Görlitz Database. This data exchange runs either on the national GSM network or an isolated broadband connection network in SHF's property via SDSL modems.
- The Server of the Görlitz database is located in SHF's headquarters. From there the data of the individual pilot participants is pushed to the Energy Assistant Database, that is a management system for SHF's customer portal (signed up customers can view their consumption/production data, deposited with the individual electricity tariff).
- On request of a project participant, the smart-meter-data as well as the current weather data and forecasts are imported and visualized to the customer in his smart energy management dashboard (SEMD).
- In addition, the participant receives advice on how to optimally regulate his energy consumption and to use the electricity, which is produced by the own PV-system with the attention triggering service.

Figure 2 describes schematically the envisioned data flow for the Utility SEMD of the German pilot. A more detailed description of each step can be found down below.

Figure 2: Data integration utilities SEMD - German pilot

The data reading procedure of the **grid operation data** follows these steps, going left to right in the flow for Utility SEMD in Figure 2:

- The steps for the integration of the end-user data as well as the data of the external services are identical to the data integration for the Prosumer Dashboard.
- In Addition, resource data and data form industrial and commercial customers as well as from generation plants will be integrated from the Network Control Station of SHF.

5.   Pilot Duration

The project's pilot trials will take place throughout the duration of the project at selected intervals depending on the status of the components that need to be tested and the implementation of the different scenarios foreseen in the project. The pre-pilots have already commenced at CERTH/ITI premises and in selected cases in Hassfurt and HUA pilots. The actual pilots are foreseen to commence in the first quarter of 2020.

6.   Storage Limitation

The gathered personal data will be kept during the lifetime to the project lifetime and for the time period expected by the funding programme.

Data related to the energy consumption will be completely anonymized so that it can be used for research and statistical purpose. All personal identification data will be removed. The data will be kept for statistical purposes, first removing personal identification data.

7.   Right access, amend, correct and delete personal data

The person concerned has the right to access their personal data, to request its rectification in case personal data is not accurate or, in its case, to request their removal when, among other reasons, data is no longer necessary for the purposes it was collected for. The user can submit a request to the Pilot controller. Information about this contact appears in the consent form. The user can request to delete personal information up to 48 hours after the experiment has taken place. Furthermore, the user can request that no additional data collection of the user will take place starting immediately from the time of request. More information is available in D2.4.

8.   Accountability

In every pilot site, the **Pilot supervisor** will inform the participants with clarity about the procedure of the pilot tests, the system operation and objectives, data retrieval and storage, and exact dates the test will be running.

The **data controller** shall be responsible for, and be able to demonstrate compliance with the GDPR. At project level, the project manager is responsible for the correct data management within the project. Whether the partners follow the guidelines for data protection will be regularly checked by the Ethic Helpdesk. For each data set, a responsible person has been appointed at partner level, who will be held accountable for this specific data set. The **data processors** will use the collected information to fulfil scientific and technical objectives. Each data processor will need to make a mention of a dataset with personal information to their Data Protection Officer, in line with the GDPR regulations, to ensure that the technical and organizational security measures has been implemented.

 A list of pilot supervisors, data controllers and data processors for the pilot is featured below:

Pilots	Pilot supervisors	Data controllers	Data processors
CERTH	**Dr. Stelios Krinidis** (**krinidis@iti.gr**)	**Mr. Kostas Peppas** (**kpeppas@iti.gr**)	**Dr. Stelios Krinidis** (**krinidis@iti.gr**) **CERTH**
HUA	**Mr. George Bravos** (**gebravos@itml.gr**) **ITML**	**Mr. George Tsakirakis** (**gtsa@itml.gr**) **ITML**	**Prof. Dr. Jasminko Novak** (**jasminko.novak@hochschule-stralsund.de**) **HOST**
Hassfurt	**Mr. Christopher Schneider** (**sit4energy@stwhas.de**)	**Ms Ana Grimaldo** (**ana.grimaldo@uabc.edu.mx**) **HOST**	**Mr. George Tsakirakis** (**gtsa@itml.gr**) **ITML**

9.   Conclusions

The purpose of this report is to present the measures taken for preserving personal data confidentiality and sensitive information privacy on behalf of pilot participants. It presents the guidelines considered by the project consortium to comply with EU and national data protection principles along with regulating privacy issues concerning all phases of data collection, storage and manipulation. Moreover, it ensures the neutrality of the recommendation systems and their appropriateness for various groups of users tackling all the related data security issues. The key features of cooperation framework are

outlined by the pilot data management methodology, which guarantees that all SIT4Energy partners will have access to the necessary data for the completion of the research, while at the same time no sensitive personal data will be distributed outside the SIT4Energy project.

References

[1]. SIT4Energy D2.4: Legal/Regulatory barriers and Energy End User Rights for SIT4Energy, 2019