



D 0.3 – Data Management Plan WPO – Project Management

16 November 2017

VERSION 1.0

Revision: Final Draft
Authors: Pradeepan Sarma
Submitted to: VUB - IES
Keywords:



EXECUTIVE SUMMARY

A data management plan (DMP) is a short, general outline of a project's policy for data management. It describes the data management life cycle for the data to be collected, processed and/or generated by a project. The present deliverable constitutes the DMP for the PARENT project.

Six different datasets have been identified at this stage of the project. In Sections 2-3 we provide a detailed presentation of the data that is expected to be collected, processed, or generated by the PARENT project along with the respective data management policies.

The first dataset (PARENT PP PD Dataset) will consist of all contact details collected from the participants of the pilots and living labs in each of the respective cities.

The second dataset (PARENT PLL Dataset) will consist of all recordings and transcripts that will be created during the respective pilot meetings and living labs.

The third dataset (PARENT US ID Dataset) consists of a listing of the Smappee device id number given to each of the participants.

The fourth dataset (PARENT PC SD Dataset) will consist of all the filled-in questionnaires that will be collected within the context of T4.2 "User Clustering".

The fifth dataset (PARENT ELSA IA SD dataset) will consist of all filled-in questionnaires that will be collected within the context of the PARENT Impact Assessment Report.

Finally, the sixth dataset (PARENT SHC Dataset) will consist of consumption data downloaded by the PARENT Consortium generated by the household electricity consumption of all the participants of each respective pilot through the means of the Smappee Energy Monitor.

The six datasets mentioned above are specified in this document and details are given regarding their contents, purpose, sharing, storage, and archiving and preservation.

Version Log			
Issue Date	Rev No.	Author	Change
24/09/2017	V0.5	VUB - LSTS	First draft
02/10/2017	V0.9	VUB - LSTS	Inputs and internal review feedback to v0.5
31/10/2017	V1.0	VUB - LSTS	Updated DMP with relevant changes
15/11/2017	V1.0	VUB - LSTS	Input and internal review feedback

TABLE OF CONTENTS

1	Introduction	7
1.1	Document structure	7
2	Data Management Strategy	8
2.1	Data to be collected/generated	8
2.2	Data sources.....	8
2.3	Types of data	9
2.4	Standards / Metadata	9
2.5	Exploitation, availability of data and re-use	9
2.6	Storage and backup	9
2.7	Archiving and preservation.....	10
2.8	Ethics and legal compliance	10
3	Project Datasets	10
3.1	Pilot participant personal data	11
3.1.1	Dataset reference and name.....	11
3.1.2	Dataset description	11
3.1.3	Dataset sharing.....	12
3.2	Pilot and living labs transcripts and recording data	13
3.2.1	Dataset reference and name.....	13
3.2.2	Dataset description	13
3.2.3	Data sharing	14
3.3	Unique Smappee identifier data	14
3.3.1	Data sharing	14
3.3.2	Dataset description	14
3.3.3	Dataset sharing.....	15
3.4	Participant clustering survey data	16
3.4.1	Dataset reference and name.....	16
3.4.2	Dataset description	16
3.4.3	Dataset sharing.....	17
3.5	ELSA impact assessment survey data	17
3.5.1	Dataset reference and name.....	17
3.5.2	Dataset description	18
3.5.3	Dataset description	19
3.6	Smappee household consumption data	20
3.6.1	Dataset reference and name.....	20
3.6.2	Dataset description	20
3.6.3	Dataset sharing.....	22
4	References	28

LIST OF TABLES

TABLE 1	OVERVIEW OF THE RESEARCH DATA OF THE PARENT PROJECT	7
TABLE 2	PILOT PARTICIPANT PERSONAL DATA	11
TABLE 3	PARENT PP PD DATASET DESCRIPTION	12
TABLE 4	PILOT AND LIVING LABS TRANSCRIPTS AND RECORDING DATA	13
TABLE 5	PARENT PLL DATASET DESCRIPTION	13
TABLE 6	UNIQUE SMAPPEE IDENTIFIER DATA	14
TABLE 7	PARENT US ID DATASET DESCRIPTION.....	15
TABLE 8	PARTICIPANT CLUSTERING SURVEY DATA.....	16
TABLE 9	PARENT PC SD DATASET DESCRIPTION.....	17
TABLE 10	ELSA IMPACT ASSESSMENT SURVEY DATA	17

TABLE 11 PARENT ELSA IA SD DATASET DESCRIPTION.....	19
TABLE 12 SMAPPEE HOUSEHOLD CONSUMPTION DATA	20
TABLE 13 PARENT SHC DATASET DESCRIPTION	21

List of Abbreviations

Abbreviation	Description
API	Application Programming Interface
DMP	Data Management Plan
ELSA	Ethical and Social Acceptability Elements
PARENT	PARticipatory platform for sustainable ENergy management
PARENT PP PD Dataset	Pilot Participant Personal Data
PARENT PLL Dataset	Pilot and Living Labs Transcripts and Recording Data
PARENT US ID Dataset	Unique Smappee Identifier Data
PARENT PC SD Dataset	Participant Clustering Survey Data
PARENT ELSA IA SD	ELSA Impact Assessment Survey Data
PARENT SHC Dataset	Smappee Household Consumption Data
WP	Work Package
Partner Abb.	Description
VUB-IES	Vrije Universiteit Brussel – Institute for European Studies
BLP	Blue Planet AC
UU	Universiteit Utrecht
RES	Resourcefully
SVT	University of Bergen, Centre for the Study of the Sciences and the Humanities
ENR	Enerbyte

1 Introduction

The present deliverable constitutes the Data Management Plan (DMP) of the PARENT project, i.e. a short, general outline of the project policy for data management.

DMPs are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include information on:

- the handling of research data during and after the end of the project
- what data will be collected, processed and/or generated
- which methodology and standards will be applied
- whether data will be shared/made open access and
- how data will be curated and preserved (including after the end of the project).¹

The data management plan is a living document, and consequently this present deliverable is expected to mature during the project lifetime and will be updated accordingly. This is because, among other reasons, as the project develops there will likely be new datasets identified. At this stage of the project, six different datasets have been identified.

1.1 Document structure

This document provides a description of the DMP within the PARENT project through the following sections:

Section 2 defines the data management strategy to be followed during the PARENT project according to the Guidelines on Data Management in Horizon 2020² published by the European Commission.

Section 3 details the six datasets have been identified, the main exploitation perspectives for each of those datasets, and the major management principles the project will implement to handle those datasets.

Table 1 gives an overview of all the data that are expected to be collected, processed or generated by the PARENT project.

Table 1 Overview of the Research Data of the PARENT Project

Dataset ID	Dataset Name
1. PARENT PP PD Dataset	Pilot Participant Personal Data
2. PARENT PLL Dataset	Pilot and Living Labs Transcripts and Recording Data
3. PARENT US ID Dataset	Unique Smappee Identifier Data
4. PARENT PC SD Dataset	Participant Clustering Survey Data
5. PARENT ELSA IA SD	ELSA Impact Assessment Survey Data

¹ Interpolated from

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

² See above.

2 Data Management Strategy

While the PARENT project is not a Horizon 2020 project, its general strategy for data management plan will be based on the Guidelines on Data Management in Horizon 2020. Consequently, the PARENT project's data management plan will be based on the identification and classification of data generated and collected, standards and metadata to be used, exploitation and availability of data as well as how the data will be shared and archived, the preservation of the information, as well as its ethical, social acceptability, and legal compliance. The PARENT DMP will cover all the data life cycle in accordance with the H2020 guidelines regarding Open Research Data.

2.1 Data to be collected/generated

The data to be collected/extracted/generated in PARENT can be grouped in the following categories:

- **Identificatory data:** This qualitative data is comprised of basic identification-based personal data (names, email addresses, city location, written consent, etc.);
- **Raw audio/visual/textual information:** Audio and or visual-related data collected during the pilot and/or Living Labs meetings, including transcripts of said meetings;
- **Survey Data:** Qualitative data about participant attitudes; qualitative data about Consortium partners;
- **Processed data:** Primarily quantitative structured information related to household electricity consumption and its related metadata.

2.2 Data sources

PARENT involves primary data gathering from different sources: 1) Sign-up and consent forms; 2) Pilot meetings and Living Labs meetings, including focus group discussions; 3) Smappee Energy monitor; 4) Survey questionnaires.

1. *Sign-up and consent forms:* Forms asking for basic information from participants necessary for their participant in the project, including their names, emails, city location, written signatures, participant Smappee ID information.
2. *Pilot meetings and Living Labs meetings:* These are meetings that take place in the context of PARENT T6.2 "Brussels Pilot"; T6.4 "Bergen Pilot"; T6.5 "Amsterdam Pilot"; T7.2 "Living Labs". These involve participants giving information related to their experiences participating in the pilot and involving their collaboration in real-life testing and deliberation (Living Labs). These meetings and workshops will be used to collect information from participants to simultaneously assess the impact of the PARENT platform while helping to make further refinements to the platform itself and the learning experiences supported by it.
3. *Smappee Energy Monitor:* These are devices that participants agree to install in their homes and which will gather information about their household energy consumption and that of particular appliances within their households. This information will be aggregated by Smappee and transmitted to members of the PARENT consortium.

4. *Survey Questionnaires*: These are questionnaires given to participants within the context of the respective pilots (T6.2 “Brussels Pilot”; T6.4 “Bergen Pilot”; T6.5 “Amsterdam Pilot”), WP4.2 “User Clustering” and the related deliverables (particularly D4.2 “User Questionnaire” and D4.3, “User Analysis Report”).

2.3 Types of data

The type of data to be collected and processed can be either:

Non-personal data: these datasets are not subject to national or European data protection legislation.

Personal data: According to European data protection law, personal data is “any information relating to an identified or identifiable natural person”.³

The use of personal data in the PARENT project follows all relevant national and European legislation. All personal data in the PARENT project is collected and/or processed only after gaining the express permission of the participants concerned and will only be stored for the duration of the project. All personal data will be only used in compliance with the guidelines derived from T2.1 and T2.2 and manifested in D2.3 “The Impact Assessment Report”. For further details, please refer to the Impact Assessment Report.

2.4 Standards / Metadata

Metadata is only generated in the context of the PARENT SHC Dataset, and this is generated through the Smappee API used in the creation of the participatory platform. This will consist of the following:

GetServiceLocationInfo: This call provides the latitude and the longitude of the Smappee Device, revealing its location, but the generation of this metadata is only at the behest of the participant.

All datasets are stored using currently accepted and interoperable standards.

2.5 Exploitation, availability of data and re-use

PARENT project intends to make its non-personal data available for open access and use by other researchers. Any personal data which will be made available (particularly PARENT PC SD Dataset) will be anonymized before being made available for reuse. In furtherance of this goal, PARENT will deposit its research data into a digital repository ([Zenodo](#) or a repository with similar security and data protection measures).

2.6 Storage and backup

All research data will be stored and backed up regularly through existing back-up mechanisms in place at the relevant partner institutions. Qualitative data will be backed up and secured by the coordinator and data controller, VUB IES, on a regular basis and metadata will include clear labelling of versions and dates.

With respect to the datasets themselves:

³ Regulation (EU) 2016/679

- PARENT PP PD Dataset is stored on the local harddrives of individual researchers of each of the respective projects. These researchers make locally-stored backups on a regular basis using Time Machine and CrashPlan Pro.
- PARENT SHC Dataset is stored in a database using the PostgreSQL database system which is hosted on a virtual private server (VPS) rented by VUB from Hetzner Online GmbH, a data center operator with its datacenters located in Germany. As part of the agreement with the Service Provider, a local backup is provided which backs up configuration files. A process monitoring system is in place which stops processes with excessive running times and/or RAM exploitation. System failure is handled by Hetzner Online GmbH. A backup system for the PARENT SHC Dataset is still at the proposal stage. It is proposed to use Amazon S3 for the back-up of this dataset, the servers of which are located in Germany and Ireland.
- With respect to the other datasets, all data is stored on the institutional servers of the respective PARENT Partner. VUB stores its datasets and subsets on institutional Microsoft OneDrive servers, which are based in Germany. UiB stores its datasets on institutional servers which are based in Norway. These folders are backed up daily.

2.7 Archiving and preservation

It has yet to be decided by the PARENT Consortium which project datasets will be preserved, the duration of this preservation period, and the costs that such preservation will entail.

Regarding the selection of a research repository, it has been proposed to utilize VUB's institutional repository for EU project data management. However, this repository is still in development. More details on this proposal will be available as the project develops and will be duly reported in subsequent versions of the present deliverable.

The PARENT consortium will take its final decisions regarding the sharing, access, and licensing strategies for all PARENT Datasets in due time, when the datasets are available. All relevant details will be duly reported in subsequent versions of the present deliverable.

2.8 Ethics and legal compliance

All details about the ethical and legal compliance of the PARENT project, particularly with respect to current domestic and EU legislation, will be included in D2.3 "The Impact Assessment Report".

Additionally, the PARENT Consortium Agreement is to be referred to for further details on the ownership, management of intellectual property and access.

3 Project Datasets

Each dataset is, where needed, divided into subsets, based on an identifiable characteristic, which collectively make up the dataset. For each dataset, the following is provided:

- Dataset Reference and Name
- Dataset Description

Dataset Reference and Name

In order to be able to distinguish and easily identify datasets, each dataset is assigned a unique short-form ID. A reference to each dataset is given where possible. For each dataset, a table listing the name of dataset and all data subsets, the dataset ID of the dataset, and a reference to the dataset, is provided.

Dataset Description

Each dataset that will be collected, processed or generated within the project will be accompanied by a brief description, along with a table summarizing the key attributes, characteristics and other information of the dataset.

3.1 Pilot participant personal data

3.1.1 Dataset reference and name

Dataset and Subset IDs, names and references are shown in Table 2.

Table 2 Pilot Participant Personal Data

Dataset ID	Dataset Name	Dataset Reference
PARENT PP PD Dataset	Pilot Participant Personal Data	Given the confidential nature of this dataset, a reference link cannot be provided.
1st Level Subset ID	1st Level Subset Name	1st Level Subset Reference
PARENT BXL PP PD Subset	Brussels Pilot Participant Personal Data	Same as above.
PARENT BRG PP PD Subset	Bergen Pilot Participant Personal Data	Same as above.
PARENT AMS PP PD Subset	Amsterdam Pilot Participant Personal Data	Same as above.

3.1.2 Dataset description

PARENT PP PD Dataset will consist of all contact details collected from the participants of the pilots and living labs in each of the respective cities (Brussels, Amsterdam, Bergen) by the respective partner. The contact details consist of: name, email address, commune, address, and consent form signature. Given that the content of this dataset is of a personal nature, it also contains written signatures of the participants authorizing PARENT to collect the aforementioned data.

The PARENT PP PD Dataset will be processed in order to:

1. Facilitate the conducting of the respective pilots (T6.2 “Brussels Pilot”; T6.4 “Bergen Pilot”; T6.5 “Amsterdam Pilot”) and the Living Labs (T7.2). The dataset is the basis on which participation of local citizens can be conducted.
2. The dataset will further be drawn upon for the completion of the reports that result from the above tasks, namely, D6.2 “Report from Brussels pilot”; D6.4 “Report from Bergen pilot”; D6.5 “Report from Amsterdam pilot”; D7.2 “Report (1) from Living Labs”; D7.3, “Report (2) from Living Labs”; D7.4, “Report (3) from Living Labs”.
3. The PARENT PP PD Dataset will be used to facilitate (electronic – email) communications between the PARENT partner and the Pilot and Living Lab participants, whether in the form of generalized

newsletters or the scheduling of local meetings or individualized responses to inquiries by participants.

4. This dataset will be processed, in conjunction with PARENT US ID Dataset and PARENT SHC Dataset, for the use of individualized emails to participants through the use of Mailjet. In this case, the email information contained within the dataset will be used in conjunction with the unique Smappee device ID contained in PARENT US ID Dataset to generate individualized emails to participants about consumption habits through the use of the Mailjet email service provider.

Key attributes, characteristics and other information regarding PARENT PP PD Dataset are presented in Table 3.

Table 3 PARENT PP PD Dataset description

Description of the data that will be generated or collected	Qualitative personal data about participants based on sign-up and consent forms.
Data origin (in case it is collected)	Content acquired in the context of the pilots, living labs, and participatory platform of the PARENT project.
Data nature	Text-based filled-in spreadsheet documents in Excel *.xlsx and *.csv format.
Data scale	Filled-in spreadsheet documents containing less than 100 entries each.
To whom could the data be useful	Solely to the researchers facilitating the respective pilots and living labs, due to both the contextual nature of the data specific to the PARENT project and the confidential nature of the data being of a personal nature.
Do the data underpin a scientific publication	To be determined in due time. ⁴
Information on the existence (or not) of similar data	Collecting personal data on participants in social science research studies is routine, and as such, similar data exists elsewhere.
Information on the possibilities for integration and reuse of the data	This data cannot be reused for any other purposes. The PARENT pilot leaders are contractually bound not to share this data or to use it for any other purposes other than within the context of PARENT.
Where the data is stored and how is it backed-up	Data is stored on local hard drives of individual researchers from respective PARENT partner. Individual project members make locally-stored backups on regular basis using 'Time Machine' and 'Crash Plan Pro'.

3.1.3 Dataset sharing

While the personal data aspects of each of the datasets will be more thoroughly examined in D2.3 "The Impact Assessment Report", where all ethical and legal aspects will be thoroughly reported on, it is already evident that the contents of this dataset cannot be shared. The PARENT consortium provided guarantees to the participants of the respective pilots that it would not share personal contact details that comprise this dataset to any parties outside of the context of the PARENT project. Consequently, given the personal data aspects of this dataset it is unlikely that any sharing of this dataset would be possible for both ethical and legal reasons.

⁴ Any developments will be duly reported in subsequent versions of the present deliverable.

3.2 Pilot and living labs transcripts and recording data

3.2.1 Dataset reference and name

Dataset and Subset IDs, names and references are shown in Table 4.

Table 4 Pilot and living labs transcripts and recording data

Dataset ID	Dataset Name	Dataset Reference
PARENT PLL Dataset	Pilot and Living Labs Transcripts and Recording Data	<i>Link will be provided at a later stage of the PARENT project.</i>
1 st Level Subset ID	1 st Level Subset Name	1 st Level Subset Reference
PARENT BXL PLL Subset	Brussels Pilot and Living Labs Transcripts and Recording Data	<i>Same as above.</i>
PARENT BRG PLL Subset	Bergen Pilot and Living Labs Transcripts and Recording Data	<i>Same as above.</i>
PARENT AMS PLL Subset	Amsterdam Pilot and Living Labs Transcripts and Recording Data	<i>Same as above.</i>

3.2.2 Dataset description

PARENT PLL Dataset will consist of all audio-visual recordings and transcripts that will be created during all pilot meetings and living labs for each of the respective pilots (Bergen, Brussels, and Amsterdam), in the context of T6.2 “Brussels Pilot”, T6.4 “Bergen Pilot”, T6.5 “Amsterdam Pilot”, as well as T7.2 “Organising and Conducting Living Labs” and T7.3 “Pilot and Living Lab Validation”.

At the present time only the PARENT BRG PLL Subset can be described in any detail. With respect to the PARENT BRG PLL Subset, the dataset includes interview transcripts and recordings with pilot participants, transcripts and recordings of focus groups within the pilots, and recordings and notes of the Living Lab meetings.

Key attributes, characteristics and other information regarding the PARENT PLL Dataset are presented in Table 5.

Table 5 PARENT PLL dataset description

Description of the data that will be generated or collected	Transcripts and audio recordings taken place during the respective pilots and Living Labs meetings.
Data origin (in case it is collected)	Content acquired in the context of the Pilots and Living Labs. <u>PARENT BRG PLL Subset</u> : Interviews with participants of the Pilot and Living Lab meetings.
Data nature	<u>PARENT BRG PLL Subset</u> : (1) Audio recordings in *.wav/.ogg and captured using a digital audio recorder; (2) Transcript files in *.doc format.
Data scale	<u>PARENT BRG PLL Subset</u> : (1) Audio files less than 200 mb each; (2) Text files less than 1 mb each.
To whom could the data be useful	(1) To social science researchers conducting research related to energy conservation; (2) market research firms.
Do the data underpin a scientific publication	To be determined in due time. ⁵

⁵ Any developments will be duly reported in subsequent versions of the present deliverable.

Information on the existence (or not) of similar data	As far as we know similar data does not exist.
Information on the possibilities for integration and reuse of the data	Provided the data is properly anonymized, it is possible that this data could be reused. Potentially it can be integrated with similar qualitative data from social science research.
Where the data is stored and how is it backed-up	PARENT BRG PLL Subset: Datasets are stored on the individual hard drives of the researchers involved. The data is backed up on institutional storage spaces allocated for individual researchers by the University of Bergen. Access to data is only granted to researchers involved at local level.

3.2.3 Data sharing

The decision on publicly sharing parts or the entirety of the PARENT PLL Dataset will be taken in due time – when the dataset is available – and will be duly reported in subsequent versions of present deliverable. In case part or all of the dataset is decided to be shared, careful anonymization will take place in order to remove any personal data.

3.3 Unique Smappee identifier data

3.3.1 Data sharing

Dataset and Subset IDs, names and references are shown in Table 6.

Table 6 Unique Smappee identifier data

Dataset ID	Dataset Name	Dataset Reference
PARENT US ID Dataset	Unique Smappee Identifier Data	<i>Given the confidential nature of this dataset, a reference link cannot be provided.</i>
1st Level Subset ID	1st Level Subset Name	1st Level Subset Reference
PARENT BXL US ID Subset	Brussels Unique Smappee Identifier Data	<i>Same as above.</i>
PARENT BRG US ID Subset	Bergen Unique Smappee Identifier Data	<i>Same as above.</i>
PARENT AMS US ID Subset	Amsterdam Unique Smappee Identifier Data	<i>Same as above.</i>

3.3.2 Dataset description

PARENT US ID Dataset consists of a listing of the 5-character unique Smappee device id number generated by the PARENT partner and the corresponding participant e-mail address to which it is linked. This 5-character unique identifier corresponds to a particular device and is the means by which the PARENT partner can link household consumption data from the PARENT SHC Dataset that the PARENT partner receives from Smappee to an individual participant. To ensure accuracy the code has a checksum letter which ensures its accuracy.

This dataset will be used as part of the participatory platform, particularly for WP5 and its associated deliverables. Here, the dataset will be used to provide individualized emails to participants on their individual energy usage, comparison to aggregate data, along with recommendations on the reduction of energy consumption which forms the core of the PARENT project.

Given the crucial role this dataset plays much of the Parent project’s activities, the dataset will likely be drawn upon in other WPs as the project develops. Specifically, the PARENT US ID Dataset will

be processed, in conjunction with the PARENT SHC Dataset and the email information of the individual participant from the PARENT PP PD Dataset, for the use of individualized emails to participants through the use of the Mailjet email service. This France-based service is a platform that provides SMTP servers to send out automated, individualized emails to participants.

Accordingly, this dataset serves as the link to ‘de-pseudonymize’ the PARENT SHC Dataset that is otherwise non-identifiable, associating data containing otherwise ‘pseudonymized’ energy consumption habits of participants to the personal data (email addresses largely) contained in PARENT PP PD Dataset. As such, it is explicitly personal data in the context of PARENT. This issue will be delved into more thoroughly in D2.3 “The Impact Assessment Report”.

Key attributes, characteristics and other information regarding the PARENT US ID Dataset are presented in Table 7.

Table 7 PARENT US ID dataset description

Description of the data that will be generated or collected	Quantitative personal data about participants consisting of the unique 5-character identifier linking a Smapppee device to an individual participant in the respective PARENT pilot (Brussels, Amsterdam, and Bergen).
Data origin (in case it is collected)	Content generated by Blue Planet from an algorithm with a checksum letter for verification and associated with data collected from PARENT PP PD Dataset in the context of the pilots and participatory platform of the PARENT project.
Data nature	Text-based filled-in spreadsheet documents in Excel *.xlsx and *.csv format.
Data scale	Filled-in spreadsheet documents containing less than 100 entries each.
Data storage and backup	Data is stored on institutional ‘One Drive’ servers run by Microsoft (hosted in Germany). Data is backed-up on local hard drive of individual researcher.
To whom could the data be useful	To the researchers facilitating the respective pilots and living labs, due to both the contextual nature of the data specific to the PARENT project and the confidential nature of the data being of a personal nature.
Do the data underpin a scientific publication	To be determined in due time. ⁶
Information on the existence (or not) of similar data	To the best of our knowledge, such data is not used by other researchers in past projects.
Information on the possibilities for integration and reuse of the data	This dataset is used in conjunction with other datasets (‘PARENT PP PD Dataset’ and ‘PARENT SHC Dataset’) within the PARENT project to carry out the aims of the PARENT project. This data cannot be reused for any other purposes. The PARENT pilot leaders are contractually bound not to share this data or to use it for any other purposes other than within the context of PARENT.

3.3.3 Dataset sharing

While the personal data aspects of each of the datasets will be more thoroughly dealt with in D2.3 “The Impact Assessment Report”, where all ethical and legal aspects will be thoroughly reported on, the Unique Smapppee Identifier Data has significant personal data aspects which would make the sharing of this dataset not possible. Further, given that its use is limited to the context of the PARENT project, it is unlikely to serve any use.

⁶ Any developments will be duly reported in subsequent versions of the present deliverable.

3.4 Participant clustering survey data

3.4.1 Dataset reference and name

Dataset and Subset IDs, names and references are shown in Table 8.

Table 8 Participant clustering survey data

Dataset ID	Dataset Name	Dataset Reference
PARENT PC SD Dataset	Participant Clustering Survey Data	<i>Link will be provided at a later stage of the PARENT project.</i>
1st Level Subset ID	1st Level Subset Name	1st Level Subset Reference
PARENT PC HC SD Subset	Housing Characteristics Survey Data	Same as above.
PARENT PC SC SD Subset	Socioeconomic Characteristics of Primary Households Survey Data	Same as above.
PARENT PC PA SD Subset	Personal Attitudes Towards Energy & Environment Survey Data	Same as above.

3.4.2 Dataset description

The PARENT PC SD Dataset will consist of all the filled-in questionnaires that will be collected within the context of the respective pilots, within the context of WP4.2 “User Clustering”, and the related deliverables (particularly D4.2 “User Questionnaire” and D4.3, “User Analysis Report”).

The purpose of this data is to “serve as a basis to understand the different segments of the participant groups and to conduct an analysis of key characteristics at the end of the pilot period-based on their responses and their engagement in the pilots and living labs.”⁷

Analytically, the collected data is as follows:

1. “Housing Characteristics”: This data subset consists of general questions about the housing situation of each participant – housing situation, energy class certification, type of heating system, and estimation of energy consumption.
2. “Sociodemographic characteristics of primary households”: This data subset consists of sociodemographic information relating to the participant and his/her household, including age group, gender, level of education, etc.
3. Personal Attitudes on Energy and Environment: This data subset consists of questions relating to the participants’ attitudes towards various issues concerning the environment and energy savings.

The dataset is expected to be processed in the context of the pilots in each of the cities (WP6), the Living Labs (WP7), and for research and analysis in WP8 “Results and Analysis” and WP9 “Recommendations for Increasing Energy Efficiency”. Only the “Housing Characteristics” data subset is further expected to be used for user clustering (WP4.2 “User Clustering”) during the pilots and for possible gamification strategies (WP4.4) later on during the course of the project, while the two other data subsets (“Sociodemographic characteristics of primary households” and “Personal Attitudes on Energy and Environment”) will not be used for these purposes.

⁷ D4.2 “User Clustering”, page 1.

Key attributes, characteristics and other information regarding PARENT PC SD Dataset are presented in Table 9.

Table 9 PARENT PC SD Dataset description

Description of the data that will be generated or collected	Qualitative data about participants, some of a personal nature, based on a filled-in questionnaire.
Data origin (in case it is collected)	Content acquired in the context of the respective pilots (Bergen, Brussels, and Amsterdam) of the PARENT project.
Data nature	Text-based filled-in spreadsheet documents in Excel *.xlsx and *.csv format.
Data scale	Filled-in spreadsheet documents containing less than 100 entries each.
Data storage and backup	Data is stored on individual researcher's hard drive. Individual researcher makes locally-stored backups on regular basis using 'Time Machine' and 'Crash Plan Pro'.
To whom could the data be useful	To market research firms, such as Ipsos, or Datamonitor, and to researchers for qualitative research and analysis on consumer habits and attitudes toward the environment and energy consumption.
Do the data underpin a scientific publication	To be determined in due time. ⁸
Information on the existence (or not) of similar data	It is likely that similar data exists elsewhere.
Information on the possibilities for integration and reuse of the data	This data consists of files in standard text formats, enabling easy re-use. This dataset can be integrated with other data collected on attitudes toward the environment and energy conservation, such as that produced in the DOMINO project, ⁹ and used to produce research on a range of topics, from 'choice architecture' to energy use.

3.4.3 Dataset sharing

After all ethical and legal aspects of the dataset are thoroughly examined in T2.2 "Impact assessment against ELSA requirements" and reported on in D2.3 "The Impact Assessment Report", the PARENT consortium will carefully consider if any part of the Participant Clustering Survey Data can be shared with specific groups of individuals or with the public and for what reason. In case of a positive decision, careful anonymization will take place in order to remove any personal data.

3.5 ELSA impact assessment survey data

3.5.1 Dataset reference and name

Dataset and Subset IDs, names and references are shown in Table 10.

Table 10 ELSA impact assessment survey data

Dataset ID	Dataset Name	Dataset Reference
PARENT IA SD Dataset	ELSA Impact Assessment Survey Data	<i>Link will be provided at a later stage of the PARENT project.</i>
1 st Level Subset ID	1 st Level Subset Name	1 st Level Subset Reference

⁸ Any developments will be duly reported in subsequent versions of the present deliverable.

⁹ See <https://www.dominoenergy.eu/en/project> for more details.

PARENT PQ ELSA IA Subset	Preliminary questions to ELSA Impact Assessment	<i>Same as above.</i>
PARENT DPP IA SD Subset	Legal – Data Protection and Privacy Impact Assessment Survey Data	<i>Same as above.</i>
PARENT ES IA SD Subset	Ethical and Societal Impact Assessment Survey Data	<i>Same as above.</i>

3.5.2 Dataset description

PARENT ELSA IA SD dataset consists of all filled-in questionnaires that will be collected within the context of the PARENT Impact assessment report, specifically within T2.2 "Impact Assessment against ELSA requirements".

The content of the dataset reflects determinations from deliverable D2.2 "Framework for Impact Assessment Against ELSA Requirements" that in the context of smart metering systems, a "strong emphasis on the fundamental rights of privacy and personal data protection" is crucial. As such, the collected data will comprise written answers concerning the effects of PARENT activities and its platform on the privacy and personal data protection rights of the participants involved, with a particular focus on the web platform given the potential for commercialization.

Drawing upon the methodologies identified and operationalized in D2.2 "Framework for Impact Assessment Against ELSA Requirements", the PARENT ELSA IA SD dataset will facilitate the assessment of PARENT project through a specially developed IA framework.

The PARENT IA framework consequently draws upon:

1. European Commission's Recommendation on a Data Protection Impact Assessment template (DPIA) for smart grids and metering environments
2. Legal obligation inscribed in article 35 of the GDPR that introduces an obligation of performing data protection impact assessments
3. Guidelines on DPIAs, most notably Article 29 Data Protection Working Party's "Guidelines on Data Protection Impact Assessment (DPIA)" (WP 248)¹⁰ and CNIL's PIA Manual¹¹.
4. An evaluation methodology integrating the ELSA and RRI principles (as identified in D2.1 and 2.2) and taking into consideration the specific context of the PARENT project.

The assessment drawn from this dataset will be used to formulate recommendations on how to improve the observance of data protection and privacy in the PARENT project as well as on improving the observance of applicable ethical and societal principles in the project.

The PARENT ELSA IA SD dataset will also be used within T2.3 "Continuous monitoring of the impacts of PARENT upon ELSA requirements" in order to monitor the impact of PARENT upon the requirements identified in D2.1 "Definition of the ELSA benchmark", specifically the legal requirements identified therein, with its focus on adherence to both national and EU-privacy and personal data protection legislation. This will culminate in D2.3 "The Impact Assessment Report".

¹⁰ WP29, Guidelines on data protection impact assessment (DPIA) and determining whether processing is likely

to result in "high risk" for the purposes of Regulation 2016/679, WP 248, 04 April 2017, http://ec.europa.eu/newsroom/document.cfm?doc_id=44137.

¹¹ This can be found here: <https://www.cnil.fr/fr/PIA-privacy-impact-assessment>.

Finally, another purpose of the collection of this dataset is to alert all partners to potential issues concerning ELSA matters that may not be immediately obvious, especially partners who do not have expertise in the area concerned.

Key attributes, characteristics and other information regarding PARENT ELSA IA SD Dataset are presented in Table 11.

Table 11 PARENT ELSA IA SD dataset description

Description of the data that will be generated or collected	Quantitative and qualitative survey data from filled-in ELSA Impact Assessment questionnaires.
Data origin (in case it is collected)	Content acquired in the context of the ELSA Impact Assessment of the PARENT project.
Data nature	Filled-in text-based questionnaires in Word *.doc format.
Data scale	Filled-in spreadsheet documents containing less than 100 entries each.
Data storage and backup	Stored in individual researcher's hard drive.
To whom could the data be useful	Researchers involved in projects similar to PARENT, i.e. participatory platforms concerning smart metering technologies or other technologies with a strong personal data component.
Do the data underpin a scientific publication	To be determined in due time. ¹²
Information on the existence (or not) of similar data	The dataset involves information particular to the context of the PARENT project. However, data protection and privacy Impact Assessment questionnaires are used often for a variety of reasons (social sciences research, internal or external evaluation of security systems etc.), and particularly for compliance with the EU General Data Protection Regulation ¹³
Information on the possibilities for integration and reuse of the data	The data could be reused for comparison with projects of a similar nature or for use in studies analyzing data protection/privacy and the smart grid and/or participatory platforms.

3.5.3 Dataset description

The decision on publicly sharing parts or the entirety of the PARENT ELSA SD Dataset will be taken in due time — when the dataset is available — and will be duly reported in subsequent versions of present deliverable. The data contained in this dataset will give away full profile of the parent participatory platform in terms of privacy protection, data protection, and ethics. Therefore, the PARENT consortium will carefully examine to what extent the PARENT ELSA SD Dataset can be publicly shared.

In case it is decided to publicly share parts of the PARENT ELSA SD dataset, this will be realized through their inclusion as an annex to public version of D2.3 “The Impact Assessment Report”.

¹² Any developments will be duly reported in subsequent versions of the present deliverable.

¹³ Regulation (EU) 2016/679

3.6 Smappee household consumption data

3.6.1 Dataset reference and name

Dataset and Subset IDs, names and references are shown in Table 12.

Table 12 Smappee household consumption data

Dataset ID	Dataset Name	Dataset Reference
PARENT SHC Dataset	Smappee Household Consumption Data	n/a
1 st Level Subset ID	1 st Level Subset Name	1 st Level Subset Reference
PARENT SHC BRG Subset	Smappee Household Consumption Data - Bergen	Same as above.
PARENT SHC BXL	Smappee Household Consumption Data - Brussels	Same as above.
PARENT SHC AMS	Smappee Household Consumption and Solar Energy Generation Data - Amsterdam	Same as above.

3.6.2 Dataset description

The PARENT SHC Dataset will consist of consumption data downloaded by the PARENT Consortium generated by the household electricity consumption of all the participants of each respective pilot (Brussels, Bergen, and Amsterdam) through the means of the Smappee Energy Monitor each participant has been individually provided.

The Smappee energy monitor, through its connection to a current clamp attached to the phase wire in the participant's fuse box and the participant's wifi network¹⁴, collects data on the electricity consumption of the various appliances located in the household, at 5 minute intervals, and transmits this data to Smappee, which stores the data, in aggregate form, in *.json format, on a remote server hosted by Amazon Web Services (AWS) in Ireland. In the case of the Amsterdam pilot, the energy monitor additionally measures the yield of any solar panels and its contribution to household energy consumption. Individual participants contractually agree to provide access to this data to the respective PARENT Consortium partner running the respective pilot. The respective PARENT partner, through the use of the Smappee API Smappee has provided,¹⁵ can then access, and download, the respective participant's household consumption data to a database hosted by the PostgreSQL database system.

As this data forms the core of the PARENT project, it will be processed in conjunction with several WPs:

1. WP4 "Strategies for Stakeholders Engagement" (especially T4.2 "User Clustering", T4.3 "Behaviour Change Model", T4.4 "Gamification Strategies", and T4.5 "Key Performance Indicators", given that this data will be basis on which the strategies developed in this work package will be designed;
2. WP5 "Development of the Virtual Energy Advisor", as the software developed with the context of this WP will allow participants to engage with the data generated from the PARENT SHC Dataset.

¹⁴ See https://www.smappee.com/be_en/faq/ for more details.

¹⁵ See <https://smappee.atlassian.net/wiki/spaces/DEVAPI/overview> for more details.

3. WP6 “Pilots”, as the PARENT SHC Dataset will be processed specifically within the context of the respective pilots in each of the respective cities.

Key attributes, characteristics and other information regarding PARENT SHC Dataset are presented in Table 13.

Table 13 PARENT SHC dataset description

Description of the data that will be generated or collected	Quantitative data on the electricity consumption of the participant in whose household the Energy Monitor is installed. The Amsterdam subset additionally provides data on solar power generation.
Data origin (in case it is collected)	While the data is generated by the Smappee Energy Monitor, it is derived from the household energy consumption patterns of the individual participant in whose household the Smappee Energy Monitor is installed. This data is accessed, and downloaded, with explicit permission of the participant, through the use of an API that Smappee has provided. Access to this data is provided in the context of the PARENT project.
Data nature	Data is in *.json format (JavaScript Object Notation) ¹⁶ , an open standard file format using human-readable text and can be parsed using a variety of programming languages.
Data scale	Data is measured in megabytes. The estimated size for the data generated for a household over the course of 12 months is 3.6 megabytes.
Data storage and Backup	The data is hosted in a database using the PostgreSQL database system which is hosted on a virtual private server (VPS) rented by VUB from Hetzner Online GmbH, a data center operator with its datacenters located in Germany.
To whom could the data be useful	The data can be useful to a variety of academic studies monitoring energy consumption, to public utilities providers, and to the private sector.
Do the data underpin a scientific publication	To be determined in due time. ¹⁷
Information on the existence (or not) of similar data	The data which this dataset concerns is merely a small segment of the consumption data stored by Smappee. Other companies in the same sector generate (and store) similar data. In the context of existing scientific research, other research projects (i.e. DOMINO ¹⁸) generate similar data.
Information on the possibilities for integration and reuse of the data	This dataset is used in conjunction with other datasets (particularly ‘PARENT PP PD Dataset’ and ‘PARENT US ID Dataset’) within the PARENT project to carry out the aims of the PARENT project. As this data is largely non-identifiable (although legally pseudonymous rather than anonymous) on its own it does not trigger the same personal data protection concerns that the other datasets trigger on their own. However, this is contingent on the other datasets which will be processed in conjunction with this one remaining strictly confidential. The data could be reused for scientific research and can be used in conjunction with existing consumption data from other studies (i.e. DOMINO) generating similar data.

¹⁶ See <http://json.org/> for more details on this format.

¹⁷ Any developments will be duly reported in subsequent versions of the present deliverable.

¹⁸ See <https://www.dominoenergy.eu/en/project> for more details.

3.6.3 Dataset sharing

The decision on publicly sharing parts or the entirety of the PARENT SHC Dataset will be taken in due time — when the dataset is available — and will be duly reported in subsequent versions of the present deliverable. It is particularly important for the results of D2.3 “The Impact Assessment Report” to be released before any decision on the sharing of the PARENT SHC Dataset can be made.

This is because, as stated in D1.2 “Mapping Stakeholders’ Needs”, smart metering technology makes it possible to process data on household consumption which can give unique insight into the activities of households. As such, the household consumption dataset carries significant personal data and privacy dimensions that make it subject not only to significant ethical considerations but also significant legal considerations.

Therefore, the PARENT consortium will carefully examine to what extent the PARENT SHC Dataset can be publicly shared, and this decision will be made no earlier than the results of D2.3 “The Impact Assessment Report” are released, disseminated, and properly deliberated upon.

