bEhaVioral Insights anD Effective eNergy policy acTions

Project No. 957117

Project acronym: EVIDENT

Project title:

Behavioral Insights and Effective Energy Policy Actions

Deliverable 3.3

Data collection and management

Programme: H2020-LC-SC3-EE-2020-1

Start date of project: 01 December 2020

Duration: 36 months

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957117

Document Control Page

Deliverable Name	Data collection and management
Deliverable Number	D3.3
Work Package	WP3 “Intervention preparation and execution”
Associated Task	T3.2 “Data collection and management”
Covered Period	M13-M35
Due Date	M35 – October 31, 2023
Completion Date	October 31, 2023
Submission Date	October 31, 2023
Deliverable Lead Partner	DUTH
Deliverable Author(s)	Ioannis Pragidis (DUTH), Karypidis Paris (DUTH), Tilemahos Efthimiadis (JRC), Paul Liston (TCD), Christos Ntoumanopoulos(CERTH), Vasileios Melissianos (PPC), Alexios Lekidis (PPC)
Version	**V1.0**

Dissemination Level		
PU	Public	**X**
CO	Confidential to a group specified by the consortium (including the Commission Services)	

Document History

Version	Date	Change History	Author(s)	Organisation
0.1	February 28, 2023	Table of Contents	Paris Karypidis	DUTH
0.2	May 23, 2023	Initial version of Section 1 and Section 3	Paris Karypidis	DUTH
0.3	May 25, 2023	Initial version of Section 4.2	Paris Karypidis	DUTH
0.4	July 27, 2023	Initial version of Section 4.1	Tilemahos Efthimiadis	JRC
0.5	August 26, 2023	Initial version of Section 2.1.1	Paris Karypidis	DUTH
0.6	September 01, 2023	Initial version of Section 2.1.3	Ioannis Pragidis, Paris Karypidis	DUTH
0.7	September 10, 2023	Initial version of Section 2.3 and Section 2.4	Paul Liston, Paris Karypidis	TCD, DUTH
0.8	October 15, 2023	Final version of Section 3 and Section 4 and initial version of Section 2.2 and Section 2.1.2	Paris Karypidis, Christos	DUTH, CERTH, PPC

			Ntoumanopoulos, Vasileios Melissianos	
0.9	October 19, 2023	Final version of all sections, ready for internal review	Paris Karypidis	DUTH
1.0	October 31, 2023	Final Version of the Deliverable after Review	Paris Karypidis	DUTH

Internal Review History

Name	**Institution**	**Date**
Alexios Lekidis	PPC	October 27, 2023
Tilemahos Efthimiadis	JRC	October 20, 2023

Quality Manager Revision

Name	**Institution**	**Date**
Dimosthenis Ioannidis	CERTH	October 23, 2023

Legal Notice

Table of Contents

List of Figures

List of Tables

Acronyms

Acronym	Explanation
APIs	Application Programming Interfaces
BCTs	Behaviour Change Techniques
CW	CheckWatt
DCE	Discrete Choice Experiment
DUTH	Democritus University of Thrace
DX.X	Deliverable X.X
EED	Energy Efficiency Directive
eHER	Electronic Home Energy Report
EPCs	Energy Performance Certificates
EU	European Union
EU	European Union
EVIDENT	Behavioural insights and effective energy policy actions
FP	Framework Programmes
HER	Home Energy Report
IEA	International Energy Agency
kWh	Kilowatt Hour
NGO	Nongovernmental Organization
PPC or DEH	Public Power Corporation
RCT	Randomized Controlled Trials
RIA	Research and Innovation Action
RIA	Research Innovation Action
SHT	Social House Tariff
TX.X	Task X.X
VPN	Virtual Private Network
WPx	Work Package X
WTA	Willingness-to-accept
WTP	Willingness-to-pay

Executive Summary

The EVIDENT project is a comprehensive research initiative aimed at studying the various factors affecting energy conservation among consumers. Through the use of field trials, surveys, and serious games, the project seeks to identify the behavioural biases and heuristics that may be hindering the adoption of energy-efficient practices by households and businesses alike. By understanding these factors, the project aims to propose evidence-based policy measures that can help reduce energy consumption and accelerate the uptake of energy-efficient technologies.

Through its field trials, the EVIDENT project gathers data on the actual behaviour of consumers in real-world settings, providing valuable insights into the factors that may be driving or hindering energy conservation efforts. Survey data, on the other hand, allows the project to gather information on consumer attitudes, perceptions, and beliefs regarding energy use and conservation. Serious games, which are interactive simulations designed to educate and engage players on various topics, are used to test consumer responsiveness to different types of energy efficiency interventions and to assess the potential impact of these interventions on energy consumption.

D3.3 "Data collection and management" is the last deliverable of WP3 "Intervention preparation and execution". WP3's main objective is the implementation of the experiments designed in WP2 "Policy interventions and pilots design" including tasks such as the creation of the related informational materials (use cases 1 & 2), the preparation of the three quasi-experiments (use cases 4 & 5) and their implementation through the EVIDENT and the Qualtrics platforms. WP3 splits into two discrete tasks, T3.1 "Preparatory actions for RCTs and surveys", which encloses the preparatory actions for the experiments that take place in use cases 1, 2, 4 and 5, and T3.2 "Data collection and management", which is devoted in the data collection and management as a processes step for analytical WP4 "Econometric analysis and policy evaluation".

This deliverable is organized into three main sections. The first section, "EVIDENT Project Datasets and Data Structures", provides information about the data used in all EVIDENT project use cases. It aims to present the corresponding data schema for either the data provided by the EVIDENT partners or collected through the EVIDENT platform during the project's lifespan. Moreover, this section includes information about the data generation mechanism and the pre-processing steps followed to ensure that the analysis can take place in the tasks of WP4.

The second section, "EVIDENT Data Services", presents a flexible architecture for scale-in/scale-out infrastructure developed to facilitate data collection, storage and processing within the EVIDENT project. This proposed architecture is documented within this deliverable, serving as a valuable set of lessons learned and best practices, and showcasing effective collaboration among partners with diverse expertise and analytical capabilities.

The final section, "Support Multi-Domain Exchange of Information", aims to support a multi-domain exchange of information and to link the project's empirical results and tools developed with applications to other policy domains. The ultimate objective is to bring a new perspective and novel insights into the research area, while promoting an open-code and open-data policy to support future research and initiatives.

1. Introduction

1.1 Purpose of the Deliverable

T3.2 outputs D3.3, which includes four main objectives. The first objective is to describe the data used to implement all five EVIDENT use cases, either provided by the EVIDENT partners (mainly the two energy companies, PPC and CW) or collected by the EVIDENT consortium during the project's lifetime. The second objective is to describe the corresponding data services developed for the analytical WP4. More specifically, this includes the creation of a flexible architecture for big data storage and data retrieval adopted during the project's lifecycle. The last objective is the support of a multi-domain exchange of information to link the project's empirical results and tools developed with applications to other policy domains. This also includes collecting data from previous relevant surveys that can assist the EVIDENT consortium in enhancing the empirical results of the project.

1.2 Relation with other Deliverables and Tasks

Deliverable 3.3 is informed by *D3.1 'Specifications of preparatory actions for RCT, surveys and serious game'* and *D3.2 'Implementation of preparatory actions for RCT, surveys and serious game'*. The data collected through the actions of D3.3 have already been used in the context of *D4.3 'Updated econometric methodologies and robustness tests'*, *D4.4 'Analytical usage handbooks for tools and datasets'* and *D5.2 'Updated impact evaluation plan and policy measures'*.

1.3 Structure of the Document

This deliverable is structured as follows:

- Section 1 – Introduction: This section introduces this deliverable.
- Section 2 – EVIDENT Project Datasets and Data Structures: Presents the data schema of the final datasets of EVIDENT's five large-scale scenarios and presents their generation, collection and transformation processes.
- Section 3 – EVIDENT Data Services: Describes the big data analytics architecture designed and followed by the EVIDENT partner to safely store the available data and be able to retrieve them to perform the corresponding analyses.
- Section 4 – Support Multi-Domain Exchange of Information: Presents data collected from prior relevant surveys for enhancing the empirical result of the EVIDENT project and describes the steps followed by the EVIDENT consortium to support information exchange and knowledge sharing into other domains.
- Section 5 – Conclusion: Concludes the deliverable.

2. EVIDENT Project Datasets and Data Structures

This section presents data used in the context of all five EVIDENT use cases. Section 2 is organised into four sub-sections, one for each use case The data for use cases 1 & 2 are both presented in section 2.1 since the same data were used. In addition, the feedback mechanisms leveraged (consumption feedback and peer comparison) were utilized in a unified home energy report (eHER). Each sub-section consists of a small introduction about the data and how these data were leveraged in each use case, the data schema, the data generation mechanism, the pre-processing steps followed, and the descriptive analytics.

2.1 EVIDENT Use Cases 1 & 2 Datasets

2.1.1 CW's dataset

2.1.1.1 Data Description, Usage and Schema

The data presented in this section refers to the CW's natural field experiment designed and implemented by the EVIDENT consortium. The experiment focuses on the effects of electronic home energy reports on a cohort of approximately 850[1] *prosumers,* that is, individuals who both consume and produce electricity through photovoltaic panels. The customers were randomly assigned between a control and a treatment group (300 and 500 customers respectively).

Every two weeks the treatment group receives an email with an electronic home energy report. The eHER, presented in Figure 1 (English translation), consists of several main sections designed to provide valuable information to consumers regarding their energy consumption. Two diagrams showcase energy feedback and peer comparisons of energy consumption and energy bought. Additionally, the eHER includes a section that categorizes household efficiency into three levels: 'Great,' 'Good,' and 'Moderate.' Furthermore, it provides a quantified format of household energy consumption. Within the report, consumers can find their rank in comparison to their neighbours. An 'injunctive' frame box is also featured, presenting messages about social norms and energy usage. This box encourages consumers to adopt more energy-efficient practices. Lastly, the eHER offers practical energy-saving tips that consumers can follow to reduce their energy consumption and improve efficiency. The neighbour classes, created to provide peer comparison information, describes customers with similar characteristics and are formed based on Sweden's climate zones where the house is located, the house heating area and heating appliance.

[1] The exact number may vary since at the moment of writing, the experiment is still ongoing.

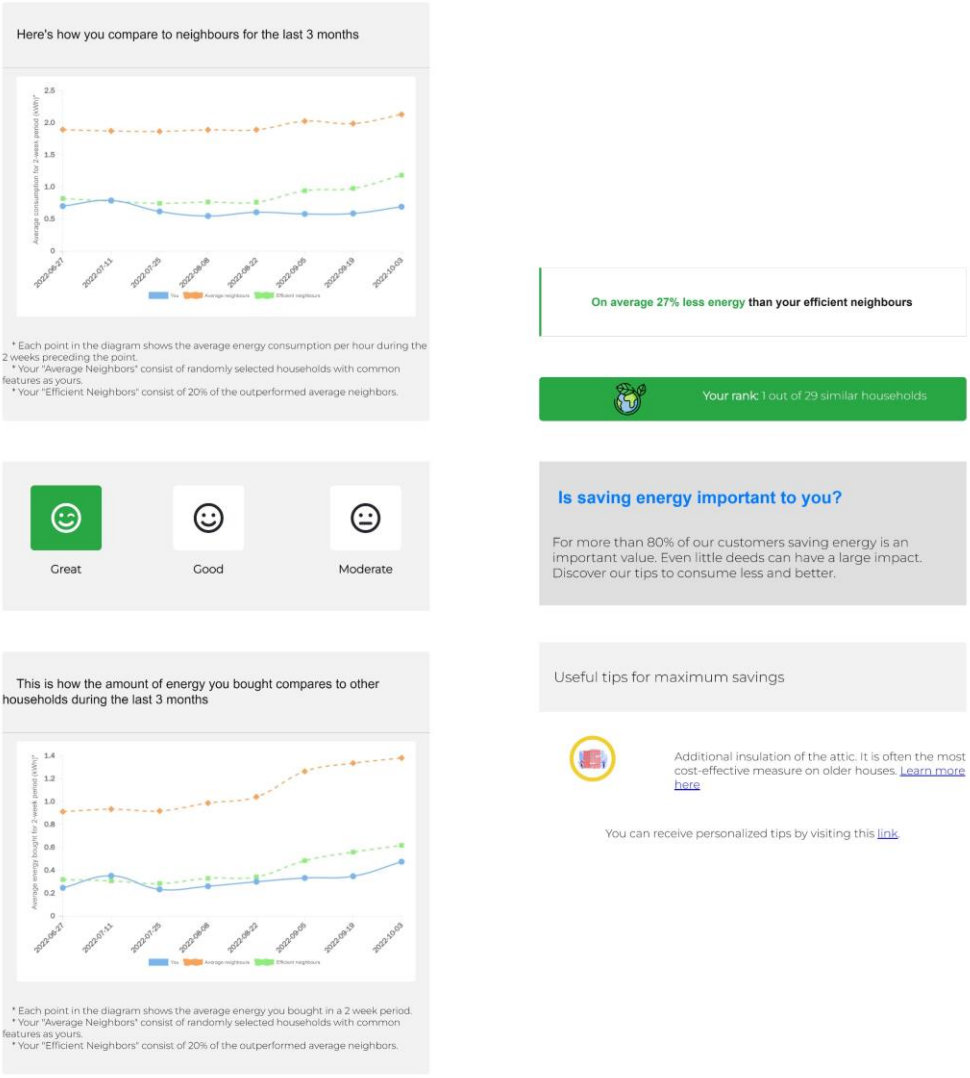


Figure 1: eHER for CW's natural field experiment

CW's experiment commenced in early December 2021 and is expected to conclude in November 2023. Following the completion of the experiment, CW intends to transition the research-oriented framework, initially developed to cater to the requirements of the EVIDENT project, into a production system. This transition will ensure the continued dissemination of eHERs to all customers, extending the project's impact beyond the experimental phase[2].

[2] This paragraph is intended provide a concise overview of the main aspects of CW's natural field experiment that has been described in the deliverables of WP2 and the previous deliverables of WP3 in details. The list of WP2 and WP3 deliverables can been found in EVIDENT project website here.

Table 1, an updated version of Table 1 initially included in *D3.2 'Implementation of preparatory actions for RCT, surveys and serious game'*, presents the number of eHERs received by each customer in the treatment group for each month.

Table 1: Number of eHERs received by CW customers in each month (updated)

Month	Number of eHERs received by each customer	Month	Number of eHERs received by each customer
December 2021	2	**December 2022**	2
January 2022	2	**January 2023**	2
February 2022	2	**February 2023**	2
March 2022	2	**March 2023**	2
April 2022	2	**April 2023**	2
May 2022	2	**May 2023**	2
June 2022	2	**June 2023**	3
July 2022	1	**July 2023**	0
August 2022	1	**August 2023**	1
September 2022	2	**September 2023**	3
October 2022	1	**October 2023**	2
November 2022	2		
TOTAL	**42 eHERs over 3 years**		

Based on the experiment time plan presented on *D3.2 'Implementation of preparatory actions for RCT, surveys and serious game – Section 2.1.2.2 Experiment time plan'* the experiment has passed phase 4 ('Include additional 150 customers in the treatment group in November 2022') and it's heading towards the estimation of the intervention effect on the additional 150 targeted customers[3].

Provided data

To design and implement CW's natural field experiment, a plethora of data is needed. Thus, CW provided a rich dataset of electricity measurements and demographics for the customers participating in the experiment. The data are used in two ways, (a) to create personalized eHERs to the treated customers and (b) to analyse the effect of the experiment upon customers' energy conservation. More information regarding the data transformation is described in section 2.1.1.2.

Table 2 depicts the data schema related to the energy measurements provided by CW. The descriptive statistics presented in Table 3 are based on the provided data for the period for the period December

[3] The selection of the additional 150 customers have been made based on the findings of use case 3 presented in *D4.2, "Econometric analysis and robustness tests – Section Big data analytics for optimizing RCT"*. D4.2 can be found here.

2020 to October 2023 after we removed the outliers. We identify the outliers as the measurements with total consumption greater than the 90th percentile or less than 50 watt-hours (Wh).

Table 2: CW's energy measurements dataset schema

Field Name	Data Type	Description	Constraints
id	String	Unique record id (primary key)	• No duplicates allowed
clientid	String	Unique client id	• Unique for each client
hour	Datetime	Date and time of the observation (1-hour period)	• Format: YYYY-MM-DD HH:MM:SS
bought	Double	Total energy bought from the grid (watt hours)	• Can't be negative • Can be zero
produced	Double	Total energy production (watt hours)	• Can't be negative • Can be zero
sold	Double	Total energy sold to the grid, electricity produced but not used (watt hours)	• Can't be greater than the produced • Can't be negative
totalconsumed	Double	Total consumption calculated as $bought + produced - sold$	• The value should be verified based on the data

Table 3: Descriptive statistics for CW's energy measurements

	Total consumption (kWh)	Produced (kWh)	Bought (kWh)	Sold (kWh)
Mean	40.369	30.579	30.539	18.186
Standard deviation	34.761	28.979	21.264	14.956
Min value	0.05	0.05	0.05	0.05
25%	11.662	4.571	12.570	3.940
50%	32.439	22.504	25.750	15.545
75%	60.486	49.827	45.707	30.536
Max value	145.561	109.935	82.010	49.150

Except the energy related measurements, CW provided a rich set of customers' metadata. Table 4 presents the dataset schema for the provided customers' metadata.

Table 4: CW's customers metadata dataset schema

Field Name	Data Type	Description	Constraints
id	String	Unique record id (primary key)	• No duplicates allowed
clientid	String	Unique client id	• Unique for each client
age	Integer	Customer's age	-
gender	Categorical	Customer's gender	-

education_level	Categorical	Customer's education level	-
employment_status	Categorical	Employment status of the consumer	-
yearly_income	Double	Customer's yearly income in thousand euros	-
remote_work	Boolean	Whether or not the customer works remote	-
city_zipcode	String	Customer's city	-
nuts_2_code	Categorical	Nomenclature of Territorial Units for Statistics	-
house_year_built	Datetime	The year the house was built	• Format: YYYY
heating_area	Double	The heating area in square meters	-
household_members	Integer	Number of members in the household	-
building_type	Categorical	The type of the building	-
water_consumption	Double	Household water consumption in cubic meters	-
extra_insultation	Boolean	Whether extra insulation is installed to the house	-
extra_insultation_year	Datetime	The year extra insulation is installed to the house (if yes)	• Format: YYYY
electric_car_charger	Boolean	Whether there is an electric car charger installed	-
electric_car charger_year	Datetime	The year the electric car charger were installed (if any)	• Format: YYYY
sauna	Boolean	Whether a sauna is installed in the house	-
pool	Boolean	Whether a pool is installed in the house	-
heating_type_primary	Categorical	Primary heating type	-
heating_type_secondary	Categorical	Secondary heating type	-
renewed_heating	Boolean	Whether there is a renewed heating	-
renewed_heating_year	Datetime	The year the heating system were renewed (if any)	• Format: YYYY
cooling_system	Boolean	Whether there is a cooling system installed	-
cooling_system_year	Datetime	The year the cooling system were installed (if any)	• Format: YYYY

2.1.1.2 Data Generation, Collection and Transformation

Data Generation and Collection

CW offers a comprehensive service package to its customers, including the provision and installation of photovoltaic panels. At the same time, the company equips their customers with smart meters and provides a smartphone application to optimize energy production and consumption management. The data generated in the field (each smart meter) are then collected into company's central database.

Every customer gains access to a personalized profile within CW's application interface, granting them control over their services. To utilize the smartphone application, customers are prompted to complete a brief demographic questionnaire.

Data Transformation

In order to transform the raw data described in the previous section into valuable insights, such as home energy reports or data suitable for analysis, a series of pre-processing steps are essential. Figure 2 illustrates the sequential process used to generate the home energy reports received by the treatment group. Meanwhile, Figure 3 outlines the pre-processing procedures applied to the data, enabling the developed analytical tools to effectively analyse the information.

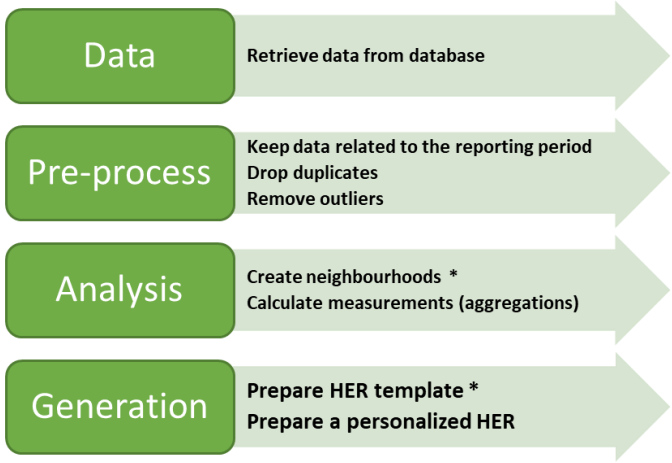


Figure 2: Use case 1 & 2, steps followed to generate the home energy reports

The primary step of the process involves retrieving data from the database. This data, provided by CW, are stored within a private database accessible solely through a virtual private network (VPN). This method is a standard procedure, adopted to ensure data integrity and security. Further insights into this phase are elaborated upon in section 3, where EVIDENT's flexible architecture for scale-in/scale-out infrastructure is presented.

Following the data retrieval is the pre-processing steps. This step includes actions such as data removal based on the reported period, removal of duplicate measurements and outliers' removal. The thresholds applied to outliers' measurements were selected to exclude periods of low and extreme energy usage and concentrate on the active periods of the day. Following data retrieval, the subsequent phase encompasses pre-processing steps. These steps include various actions, such as data isolation based on the reported period, removal of duplicate measurements, and the identification and removal of outliers. For the latter, the thresholds applied for outliers' measurements identification, were selected to exclude periods of low and extreme energy usage and concentrate on the active periods of the day.

The next step involves a comprehensive analysis aimed at two objectives:

(a) Formulating neighbourhoods and

(b) Computing measurements for individual customers.

For the formulation of neighbourhoods, this process initiates at the experiment's outset and whenever a new customer joins the treatment group[4]. Randomly selected customers in the treatment group are allocated to neighbourhoods based on three defining characteristics: house location, size, and primary heating appliance. Additional details on neighbourhood creation can be found in *D3.2, "Implementation of Preparatory Actions for RCT, Surveys, and Serious Game"*[5].

Continuing the process, we calculate the energy consumption for each customer. This entails aggregating measurements from hourly to biweekly intervals, computing consumption feedback, peer comparisons, customer rankings, performance metrics, and energy consumption tips. Ultimately, this personalized information is visualized in the home energy reports that each customer receives. The outcome of this analysis is stored in a file that involves all information destined for inclusion in the eHERs.

The final step in this process is the eHER generation. To achieve this, a standardized eHER template has been designed, including several placeholders ready to be replaced with unique customer-specific data.

The interested reader can find further details about EVIDENT's field experiment preparation and execution in deliverables across work packages 2 and 3. In addition, technical resources can be found in deliverable *D4.4, "Analytical Usage Handbooks for Tools and Datasets"*, as well as the blog post "Creating an Electronic Home Energy Report using Python" authored by DUTH[6].

This remainder of this section describes the steps followed to transform the raw information into personalized home energy reports.

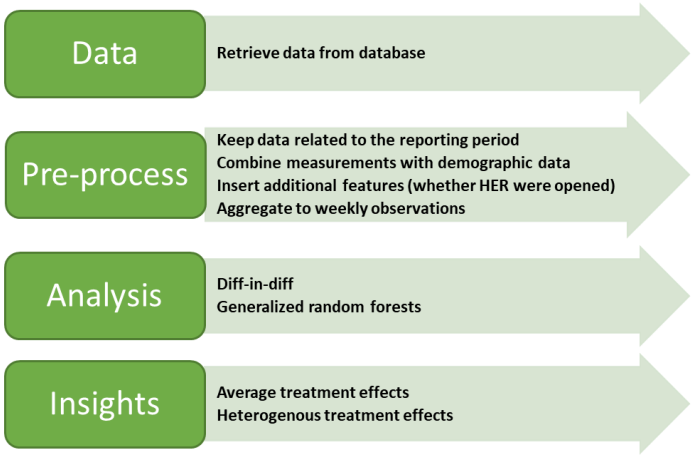


Figure 3: Use case 1 & 2, steps followed to prepare the data for analysis

Figure 3, illustrates the step-by-step process undertaken to pre-process the provided data, rendering it suitable for analysis. This process encompasses four distinct stages, though this section will place particular emphasis on the first two steps. Comprehensive insights and analyses can be found in *D4.3, "Updated Econometric Methodologies and Robustness Tests"*.

Initiating once more from the data source, the first step in this process involves extracting data from CW's database. The preliminary pre-processing step is to

[4] However, and based on the objectives of the application, this might differ. For example, in case of a user application that provides similar services (consumption feedback, peer comparison, etc.) this mechanism might be more dynamic and might executed based on period intervals and automations.

[5] EVIDENT *D3.2 "Implementation of preparatory actions for RCT, surveys and serious game"* can be found here.

[6] All EVIDENT's publicly available deliverables can be found here. "Creating an electronic Home Energy Report using Python" blog post can be found on EVIDENT website here.

precisely isolate data spanning one year before the initiation of experiment (December 2020). This data is then fused with energy measurements, represented as time series data, along with demographic information.

Given the experiment's multi-round nature, the insertion of additional information (i.e., new eHER every two weeks) is very important. Information about customer interactions with the eHERs, specifically whether they open and read their reports, is crucial for the ensuing analysis. To capture this data, the EVIDENT consortium has developed a custom service for tracking email opens. As the final act of pre-processing, data is aggregated into weekly observations.

2.1.2 PPC's dataset

2.1.2.1 Data Description, Usage and Schema

This section presents a detailed description of the PPC's dataset. The scope of this analysis focuses on providing energy efficiency reviews and energy efficiency tips based on the information the customers share on the "MyEnergyCoach" online platform. This data is partially filled in by the consumer directly, which leads to various degrees of completeness within the dataset and poses challenges in providing accurate energy efficiency insights. To tackle this, data filtering and pre-processing techniques are employed to create a more compact and informative dataset. The "MyEnergyCoach" service provides a dataset through an online demographic template that is updated every four months and documents total consumption, household characteristics, household devices, device usage, and information about the household members, as presented in Figure 4. The generated dataset offers a wide range of demographic and quantitative information with impactful and non-impactful features.

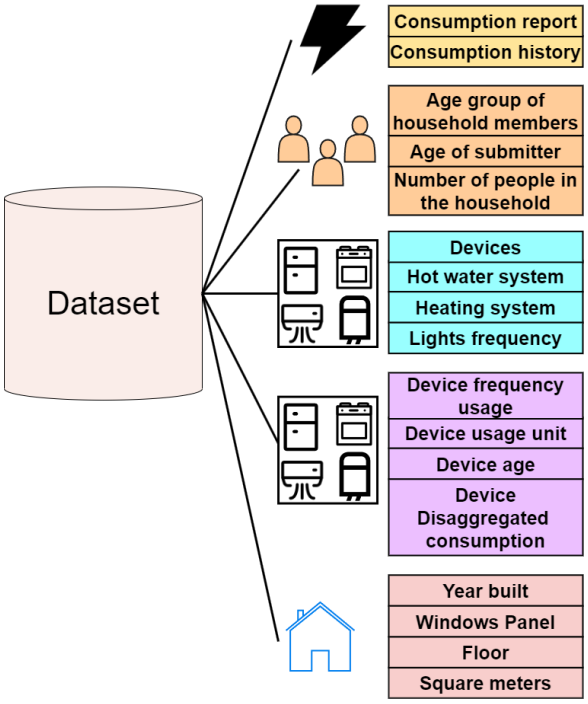


Figure 4: Description of PPC's dataset. The dataset includes various types of information. The Demographic data refer to the orange, blue and red fields, while the Quantitative data refer to the yellow and purple fields.

Since the dataset offers a variety of information, all of its raw data are imported into a pre-processing phase. In this phase, the selected demographic information is chosen as a characteristic set of features that heavily affect the usage of the electrical devices of a household. It also provides a solid insight that the core of the consumption routine is based on specific characteristics (e.g. the number of people living in a household). On the other hand, the quantitative data used for the analysis acted as input to a set of mathematical operations to review and assess the consumption behaviour of every consumer on the devices they possess (e.g. stove, refrigerator), provide frequency device usage suggestions and device replacement options to reform the consumption routine of the consumer in a more energy efficient and eco-friendly way. PPC's dataset schema is presented in Table 5, while its descriptive statistics are shown in Table 6.

Field Name	Data Type	Description	Constraints
meter_id	String	User unique ID	Unique for each client
Question_id	String	Question parameter that each id represents a specific question	-
Questionnaire_id	String	Question parameter that represents to which meter_id the question_id was applied.	-
Answer_id	String	Answer parameter that each id represents a specific answer	-
Reply_id	String	Answer parameter that represents to which meter_id the answer_id was applied.	-

Reply_datetime	Date	Chronological datetime of the reply.	Format: YYYY-MM-DD HH:MM:SS
Question_description_el	Categorical	Question parameter that represents the question in Greek	-
Question_description_eng	Categorical	Question parameter that represents the question in English	-
Answer_description_el	Categorical	Answer parameter that represents the answer in Greek	-
Answer_description_eng	Categorical	Answer parameter that represents the answer in English	-
User_m2	String	Square meter value of each meter_id household	-
User_m2_id	String	Square meter parameter that represents to which meter_id the square_meter value was applied	No negative value
con_created_datetime	Date	Chronological datetime of the consumption.	Format: YYYY-MM- DD HH:MM:SS
Con_total	String	Consumption value of the total consumption of a specific period	No negative value
Con_day	String	Consumption value that represents the consumption during the day	No negative value
Con_night	String	Consumption value that represents the consumption during the night	No negative value
Con_neighbor	String	Consumption value of the neighbour	No negative value
con_neighbour_top	String	Highest consumption value of the neighbour	No negative value
Con_id	String	Consumption parameter that represents to which meter_id the consumption value was applied	-
Bill_id	String	bill parameter that represents to which meter_id the bill value was applied	-
Start_date	Date	Chronological datetime of the starting period of consumption.	Format: YYYY-MM- DD
End_date	Date	Chronological datetime of the ending period of consumption.	Format: YYYY-MM- DD
Appliance_id	String	Electric Appliance id that each id represents a specific device	-
appliance_created_date	Date	Chronological datetime of the archived appliances.	Format: YYYY-MM- DD
appliance_description_el	Categorical	Each appliance_id represents a specific electrical device in Greek	-
appliance_description_eng	Categorical	Each appliance_id represents a specific electrical device in English	-

Appliance_to_meter_id	String	Appliance parameter that represents to which meter_id the electric applianace refers to	-
appliance_to_meter_year_of	String	Year of the device	-
appliance_to_meter_total_use	Integer	Number of total uses per unit	No negative values
appliance_to_meter_total_use_unit	Categorical	Appliance usage on its unit. (eg. Washing machine 2 uses per day)	-
Disaggregation_id	String	Disaggregation parameter that represents to which appliance_id the disaggregation_consumption refers to	-
Dis_device	Categorical	Refers to the name of the electric appliance	-
Dis_consumption	Float	Disaggregated consumption value of the specific electric appliance	No negative value

Table 5: Dataset information overview

	Total consumption (kWh)	Consumption period (days)	Number of Electrical Devices	Consumption in the Day (kWh)	Consumption in the Night (kWh)
Mean	5331.12	162.42	11.23	4413.42	1618.15
Standard deviation	4538.54	83.57	1.72	3479.4	2032.97
Min value	1018.31	1.0	6	575.86	2.5
25%	3219.6	102.5	10.0	2770.51	718.91
50%	4518.36	161.0	11.0	3843.69	1181.53
75%	6122.91	216.0	12.0	5215.27	1847.68
Max value	78982.35	669.0	16	65973.89	28923.15

Table 6: Descriptive statistics of the dataset

2.1.2.2 Data Generation, Collection and Transformation

Data Generation and Collection

PPC's dataset was provided by "MyEnergyCoach" through an online platform hosted by myDEH[7]. The submitted forms are then archived in another data centre, "DEHMyEmporia", which converts the online form into an Excel format. The nature of the data provided comes in a set of independent forms where each form represents a specific field (Consumption, Bills, Appliances, etc.) with the submitter ids not sorted.

[7] PPC's full name is Public Power Corporation S.A., which in Greek is «Δημόσια Επιχείρηση Ενέργειας - ΔΕΗ». The acronym DEH comes from the Greek acronym name ΔΕΗ.

Data Transformation

The scope of the analysis seemed demanding over the proposed dataset since it grants a plethora of information. Although the dataset comprises 24,980 consumers, the demographic form is complete for only a small portion of the consumers. The majority of consumers provided partially filled or unrelated data. Thus, this analysis focused on a small-sized compact dataset with 675 filled forms out of 24,980 submissions.

A series of pre-processing procedures were implemented to carry out the dataset trimming. The dataset's information was a total of many forms, each containing the '**meter_id**' parameter representing a consumer and, therefore, a consumer's information about the specific form, as shown in Figure 5.

Figure 5: PPC's dataset, main pre-processing schema

The first step was to find the common meter_ids of all forms and then discard the meter_ids with at least one absence from any of these forms (see Figure 5, Phase 2 the red boxes). Afterwards, the next procedure was to eliminate the rest meter_ids with any missing values (NULL) or inappropriate values that are marked with grey colour (see Figure 5, Phase 3; the grey boxes disappeared along with the respective meter_ids), resulting in a more trimmed dataset but with full information in any field needed. Finally, the data were converted to a list array for archiving the valuable information, as shown in Figure 6.

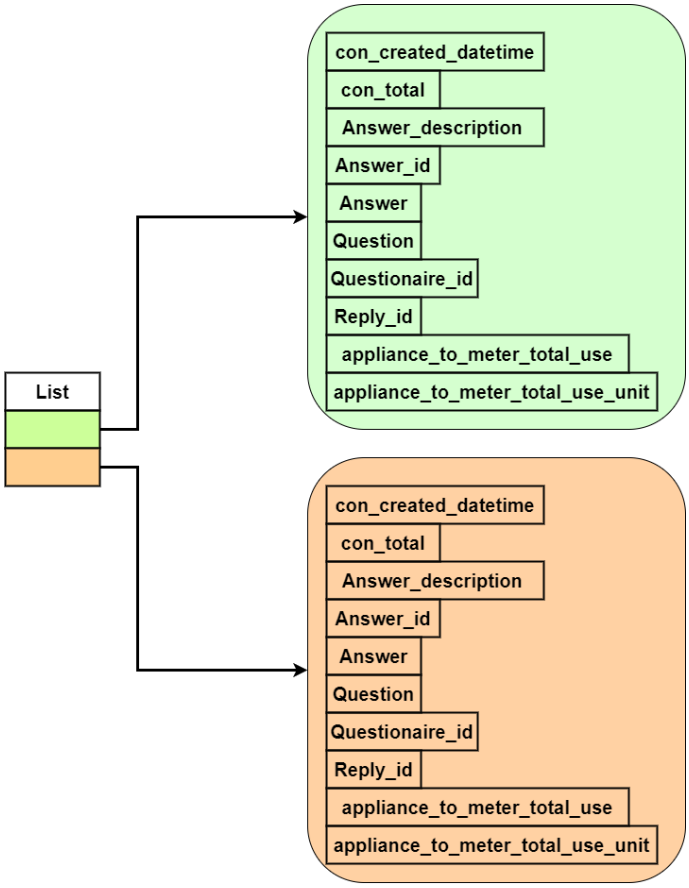


Figure 6: Trimmed version of PPC's dataset. Each meter_id has its specific values and the resulted dataset hold only fully submitted information.

The third and final step was to apply a feature selection on the demographic data to determine which features are essential for understanding consumer behaviour. To achieve that, a statistical analysis called cross-correlation algorithm (Feature Relation Impact) was applied where the parameters are matched by considering the consumption feature.

2.1.3 Protergia's dataset

2.1.3.1 Data Description, Usage and Schema

This section provides an overview upon the data used for examining customers' degree of attention when choosing an electricity contract. This analysis is based on a dataset provided by Protergia, a large energy utility in Greece. A suboptimal contract choice is costly for consumers while it drives electricity consumption away from an optimal level, while at the same time, from the utility's perspective, this means a higher consumers dropout rate. From the other hand, the fixed consumption and the social house tariff (SHT) contracts induce customers to consume within their contract limits since there is an overcharge for overconsumption. In this analysis we are estimating whether customers systematically underestimate their future electricity consumption across different contract groups.

The energy utility provides three different consumption tiers to its customers. The first is for an annual consumption up to 2,500kWh, the second is up to 4,000kWh and the third is up to 6,500kWh. If a customer

stays as close as possible to her consumption limits in the selected tier, then a discount is applied relative to the regular tariff contract. A detailed analysis will be presented in D4.3 since in this document we are presenting the data schema of the provided data.

Provided data

The unbalanced panel dataset is taken from Protergia, a large energy utility in Greece covering the period from January 2016 to December 2021. A customer may switch to different contracts and thus our analysis focuses on contract spells. Similarly, to (Della Vigna and Ulrike 2006), a contract spell starts when a customer signs for a contract and ends whenever the customer switches to a different contract or changes provider.

The sample includes information for 15,047 customers (both active customers and customers that have switched energy provider) and 25,694 spells. Overall, it contains information for three different contract types, a fixed consumption contract (1,690 spells), a regular tariff contract (10,416 spells), and a Social Housing Tariff (SHT) contract (13,588 spells). The Tier-1 contract has almost three times more customers than the Tier-3 contract. All Tier-type contracts have shorter duration than the regular tariff and the SHT contract by approximately 7 and 4 months respectively. It should be noted that this difference in contract duration is probably higher since most regular tariff and SHT customers were still active when the sample period ended while most fixed contract customers have switched provider at some point within the sample period. Protergia's dataset schema is depicted in Table 7 while descriptive statistics are presented in Table 8.

Table 7: Protergia's dataset schema

Field Name	Data Type	Description	Constraints
ID	String	User unique ID	• Unique for each client
meter	String	Meter unique ID	• Unique for each client
contract	Categorical	Selected contract for the billing period	-
billtype	Categorical	Type of bill (In advance/Clearance)	-
startDate	Datetime	Billing period, start date	• Format: YYYY-MM-DD HH:MM:SS
endDate	Datetime	Billing period, end date	• Format: YYYY-MM-DD HH:MM:SS
dueDate	Datetime	Bill due date	• Format: YYYY-MM-DD HH:MM:SS
PaymentStatus	Categorical	Ontime payment or delayed payment	-
BillDate	Datetime	The date the bill was paid	• Format: YYYY-MM-DD HH:MM:SS
consumptionDay	Double	Consumption in watthours during daily tariff	-
ConsumptionNight	Double	Consumption in watthours during nightly (off peak) tariff	-
priceRegular	Double	Regular price, no discounts	-

priceReduced	Double	Discounted price (off-peak consumption)	-
costRegular	Double	Regular price paid times electricity consumption	-
costReduced	Double	Discounted price paid times electricity consumption	-
distributionTaxes	Double	Energy provider distribution tax	-
estateTax	Double	Municipality estate tax	-
discount	Double	Discount on invoiced amount	-
lastPaymentDate	Datetime	Last payment date so the bill won't be overdue	• Format: YYYY-MM-DD HH:MM:SS
overDueDays	Integer	Overdue days for bills paid after the lastPaymentDate	-
municipality	String	Municipality the household is located	-
postalcode	String	Postal code the meter is located	-
expirationDate	Datetime	Contract expiration date	• Format: YYYY-MM-DD HH:MM:SS
size	Double	Size of the house	-
provider	String	Previous provider	-
myprotergia	Boolean	Boolean variable for "myProtergia" application: 1-Active, 0-Not active)	-
meterstatus	Categorical	Meter status: Active/Stopped	-
contactBill	Categorical	How the bill is contacted to the customer	-
Invoicedamount	Double	Invoived amount (amount on bill)	-
Previousunpaidamount	Double	Unpaid amount from previous bills	-
Nightcontract	Boolean	Boolean variable indicating if the customer has a contract with nightly tarrif (off-peak) available	-
fixedcontract	Boolean	Boolean variable indicating if the customer has a fixed contract	-
socialtarriff	Boolean	Boolean variable indicating if the customer has a social tarrfi contract in which case he receives a 70% discount	-
N_obs	Integer	Total consumer's observations (bills received)	-
consumptiondays	Integer	Consumption period in days	-
averagedailycons	Double	Average daily consumption during daily tariff	-
averagenightcons	Double	Average daily consumption during nightly (off-peak) tariff	-
Contract_cat	Categorical	Contract category	-
regions	Categorical	Region of Greece the meter is located	-
consumptionlimit	Double	Consumption limit based on the selected contract type. This only applies to fixed contracts. Under a	-

		fixed contract a customer buys a fixed amount of KWh.	

Table 8: Descriptive statistics for Protergia's dataset

	Sample: Genius type			Sample: Regular tariff	Sample: SHT
	Tier-1	Tier-2	Tier-3	Regular tariff	Social Housing Tariff
Number of spells	914	538	238	10,416	13,588
Aver. daily consumption (kWh)	7.97 (4.15)	11.36 (4.69)	16.79 (6.56)	10.61 (5.33)	11.16 (5.40)
Aver. consumption limit (kWh)	6.84	10.95	17.88	-	-
Aver. daily cost (€)	1.97 (0.65)	2.60 (0.74)	3.84 (1.22)	2.24 (1.10)	2.16 (1.24)
Cost (€/kWh)	0.247	0.228	0.215	0.211	0.193
Contract duration (days)	532.50 (158.32)	519.17 (157.57)	516.61 (151.60)	733.55 (439.07)	620.04 (373.56)
Delayed payments (days)	40.23 (123.18)	36.77 (119.53)	50.01 (136.30)	40.62 (126.72)	114.64 (207.14)
House size (m^2)	77.20 (34.58)	90.46 (36.12)	114.92 (59.54)	82.61 (33.12)	78.38 (29.29)

Notes: Standard deviations are in parentheses.

2.1.3.2 Data Generation, Collection and Transformation

Protergia's tier-type contracts had a 2-year minimum contractual period, while the regular tariff contract had a one-year minimum contractual period. Social housing tariff contracts have no minimum contractual period as soon as households meet the eligibility criteria. The electricity meters are read every four months while consumers may get monthly bills based on projections regarding their electricity consumption. The four-month bill, that includes real consumption, adjusts the invoiced amount about any prepayments made. However, random deviations in the billing period may arise since this is human-based procedure. Thus, for reasons of comparison, both consumption and energy costs are converted into daily observations by dividing them with the number of days for each billing period. To estimate the electricity consumption in a specific period we consider only the meter reading and not any projections, while for estimating the amount paid, we sum up all payments made during that period.

The data generated from this process are safely stored in the company's central database. Every customer has access to a personalized profile within Protergia's application, where they can see their consumption, pay their bill, participate in questionnaires, receive conversation tips and be compared with neighbours[8].

[8] Neighbour comparison is available only to customers with smart meters.

2.2 EVIDENT Use Case 3 Datasets

2.2.1 Data Description, Usage and Schema

The dataset contains a variety of weather condition gathered from 615 cities in Sweden. Combined with CW's dataset presented in section 2.1.1, this dataset aims to investigate the complex interactions between energy measurements, consumers' demographics and weather conditions to develop advanced forecasting models that can accurately estimate energy consumption and production. These results will be crucial in formulating policies and strategies for sustainability and energy use. As a result of the insights gained from this extensive dataset, it will be possible to establish more sustainable energy policies and make well-informed judgments.

A crucial step in the pipeline is data pre-processing and cleaning, ensuring the dataset we use for energy consumption and production forecasting models is reliable, consistent, and appropriate for in-depth research.

During the pre-processing and cleansing of data, several crucial issues are addressed:

1. **Data Validation and Integrity**: Examining the dataset for errors, anomalies, and missing numbers is known as data validation and integrity. Information gaps are avoided by managing any missing data properly.
2. **Data Standardization**: Ensure uniform data types to facilitate the required analysis through data standardization. This includes converting variables to consistent units and resolving inconsistencies.
3. **Outlier Detection**: Outlier detection involves finding and managing outliers, or data points, so they don't impact the prediction models. Outliers are data points that considerably deviate from the average values in the dataset.
4. **Feature Engineering**: Feature engineering is the process of developing and improving pertinent characteristics to strengthen models' prediction ability. For deeper insights, this frequently entails the creation of new variables and transformations.
5. **Data Encoding:** The process of transforming category variables into numerical values to make them easier to incorporate into mathematical models. To preserve data integrity, techniques like one-hot encoding are used.
6. **Data Quality Assurance**: Data quality assurance is the process of using in-depth quality evaluations to ensure the accuracy and coherence of the data. This assurance mechanism's cross-validation and sanity tests are essential elements.

The effective completion of this analysis depends on this mindful phase of data pre-processing and cleansing. It sets the foundation for creating reliable forecasting models that are strong, trustworthy, and simple to understand, enabling conscious decision-making, improving energy management, and fostering a more sustainable future. Table 9 depicts the data schema related to the weather conditions dataset.

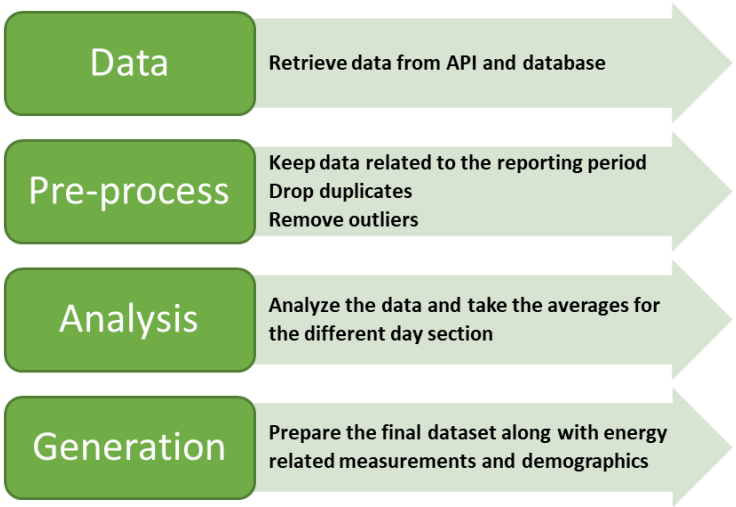
Table 9: Weather conditions dataset schema

Field Name	Data Type	Description	Constraints
country	String	The name of the country	-

city	String	The name of the city	-
temperature	Float	The temperature of the city	[-50, 50]
feels_like	Float	The real feel temperature of the city	[-50, 50]
temp_min	Float	The minimum temperature of the city	>-50
temp_max	Float	The maximum temperature of the city	<50
pressure	Float	The air pressure of the city	-
humidity	Float	The humidity of the city	[0, 100]
wind_speed	Float	The wind speed of the city	>0
wind_deg	Float	The wind degrees of the city	[0, 360]
sunrise	Date/Time	The sunrise time of the city	-
sunset	Date/Time	The sunset time of the city	-
weather_description	String	A description about the weather	

2.2.2 Data Generation, Collection and Transformation

Weather conditions from several external sources where the data is received through Application Programming Interfaces (APIs) and the CW dataset mentioned in section 2.1.1 were combined to create this dataset. The APIs from where the data are collected are the Open Weather Map[9] API and the World Weather API[10]. Using a Python agent, these APIs' information is collected and stored along with the data provided by the CW. This is the first step of the process to generate the dataset, depicted in Figure 7.



[9] https://openweathermap.org/api/
[10] https://world-weather.ru/

Figure 7: Steps followed to generate the dataset

The pre-processing steps, explained in section 2.1.1, are then carried out. Moreover, in the analysis step, it was important to aim at two objectives:

(a) Categorize the data per city and

(b) Process each city's data to create the averages.

Regarding the cities, it was important to have each city's data separately to create a panel data form. Since the data for each city are very regular, it was important to separate the timeline into three different day sections (morning, noon and evening). Finally, for each day section, the averages of each data were calculated separately to create the final dataset.

2.3 EVIDENT Use Case 4 Datasets

2.3.1 Data Description, Usage and Schema

EVIDENT's use case 4 aims to determine the impact of energy-related financial literacy, demographic factors, and behavioural intention/attitude on decisions to repair or replace household appliances across resident types. Specifically, this use case aims to answer the following research questions:

- How do financial, energy and environmental literacy impact the decision to repair or replace?
- How does the salience of financial information affect the decision to repair or replace?
- What type of information impacts willingness-to-pay (WTP) for a repair or replacement of an appliance? (financial, anticipated lifecycle or environmental)
- Does providing tips related to financial literacy enhance consumers' ability to make better choices?
- What barriers and facilitators do individuals encounter when deciding to repair or replace an appliance (Qualitative analysis)
- What is the impact of the serious game on real-life opportunities to make repair/replace decisions?

The research questions are investigated using a serious game whereby participants engage in a series of exercises using the online serious game, receiving valuable energy efficiency feedback on their behaviour within the game. As part of this, they complete a series of research tasks. The serious game[11], some of its parts are presented in Figure 8, is a life simulation type game within which players are tasked with maintaining a home over the passing of time. Participants are assigned a role in keeping with their residential status (i.e. landlord, tenant or homeowner) and will be given an avatar to represent themselves within the game. The participants then move this avatar around their virtual home and complete a series of actions, all aiming to maintain their avatars' comfort while also ensuring their environmental impact doesn't get too high. The participants' actions in the game will be guided by indicative gauges showing

[11] https://evident-h2020.eu/seriousgame/

their comfort, environmental impact (based on kWh of energy use within the game) and finances. Comfort ratings will reduce should an avatar's basic needs not be met (food, heat etc.) and are included as a means to motivate users to engage in actions in their home environment. At the end of the game, participants will be given a final score based on their environmental impact, comfort and finances and can see where they fall on a leaderboard.

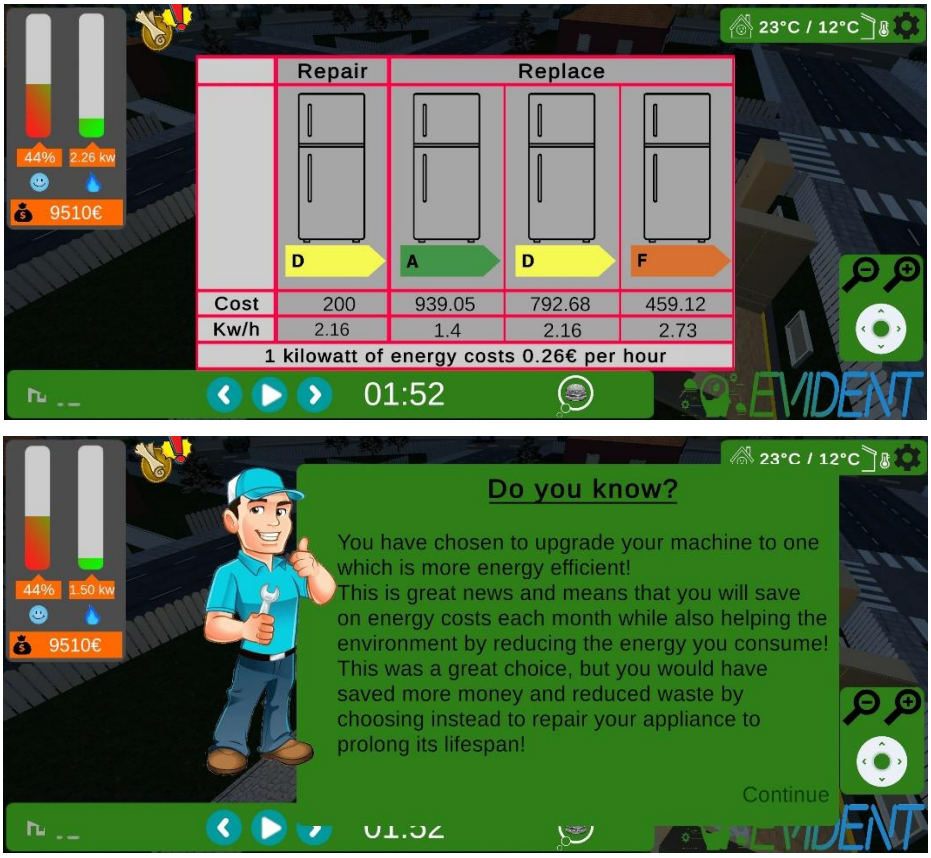


Figure 8: Screenshots from the EVIDENT Serious Game

Within the game:

- An appliance will break, and the user must decide whether they would like to repair or replace the appliance.
- A repairperson will show up in the house door and the user will enter into a discussion about whether they would like to repair the broken appliance or purchase a new appliance.
- For new appliances, differing levels of energy efficiency and cost will be available.
- Depending on the option selected, participants will then enter into a negotiation with the repairperson to determine their willingness-to-pay for a repair.
- For landlords or tenants, additional discussions will occur, with tenants given the option to pay more rent or a small fee in exchange for a better appliance, and landlords given the option to accept more rent from tenants in exchange for a better rated appliance.
- Once a final choice is made, users will be given feedback on their decision and on how to determine whether to repair or replace a broken appliance more easily.
- Users will then continue in the game, navigating more appliances that break.
- At the end of the game the points gained will be given as a total score and the user will be informed of their place on a leaderboard.
- Users will be advised also of where their score falls relative to the average.

Table 10 presents the dataset schema for the use case 4 participants' metadata.

Table 10: Use case 4 dataset schema for the participants' metadata

Field Name	Data Type	Description	Constraints
answers.demo_age	Integer	Participant's age	-
answers.demo_country	String	Participant's country of residence	-
answers.demo_income	Category	Participant's yearly income in thousand euros	-
answers.demo_employment_status	Categorical	Participant's employment status	-
answers.demo_role	Categorical	Participant's choice of role in the serious game	-
answers.demo_gender	Double	Participant's choice of gender of avatar for game	-
answers.demo_adults	Integer	Number of adults living in residence	-
answers.demo_children	integer	Number of children living in residence	

Table 11 presents the dataset schema for use case 4 participants' financial literacy responses.

Table 11: Use case 4 dataset schema for the participants' financial literacy responses

Field Name	Data Type	Description	Constraints
finance_financial_knowledge	Categorical	How would you assess your overall financial knowledge?	-
finance_borrow_100	Categorical	Suppose you need to borrow €100 for a year. Which is the lower amount to pay back: €105 or €100 plus three percent?	-
finance_inflation_qualitative	Categorical	Suppose over the next 10 years the prices of the things you buy double. If your income also doubles, will you be able to buy less than you can buy today, the same as you can buy today, or more than you can buy today?	-
finance_inflation_quantitative	Categorical	Suppose you had €100 in a savings account and the bank adds 10% per year to the account. How much money would you have in the account after five years if you did not remove any money from the account?	-
finance_interest_rate	Categorical	Suppose you put money in the bank for two years and the bank agrees to add 15% per year to your account. Will the bank add more money to your account the second year than it did the first year, or will it add the same amount of money both years?	-
finance_investments	Categorical	Is it safer to put your money into one business or investment, or to put your money into multiple businesses or investments?	-
finance_finance_self_assessment	Categorical	What do you think? How many of the last five questions did you answer correctly?	-
finance_blockchain	Boolean	Are you aware of blockchain technology?	-

finance_agree_or_disagree.finance_agree_or_disagree_invest_crypto	Categorical	I believe it is a good time for people to invest in crypto-assets or initial coin offering (ICOs).	-
finance_agree_or_disagree.finance_agree_or_disagree_close_watch	Categorical	I keep a close watch on my financial affairs	-
finance_agree_or_disagree.finance_agree_or_disagree_financial_goals		Please indicate how strongly you agree or disagree with the following: - I set long term financial goals and strive to achieve them	-
finance_agree_or_disagree.finance_agree_or_disagree_debt	Categorical	I have too much debt right now	-
finance_agree_or_disagree.finance_agree_or_disagree_long_term_savings	Categorical	I find it more satisfying to spend money than to save it for the long term	-
finance_agree_or_disagree.finance_agree_or_disagree_line_in_present	Categorical	I intend to live in the present more and do not consider the future	-
finance_agree_or_disagree.finance_agree_or_disagree_prepare_to_risk	Categorical	I am prepared to risk some of my own money when saving or making an investment	-
finance_agree_or_disagree.finance_agree_or_disagree_quick_money	Categorical	I make investments to make money quickly	-
finance_do_following.finance_do_following_mobile_payments	Categorical	I use my smartphone to make or receive payments	-
finance_do_following.finance_do_following_trading	Categorical	I am actively involved in trade activity	-
finance_risks	Categorical	Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?	-

Table 12 presents the dataset schema for use case 4 participants' environmental literacy responses.

Table 12: Use case 4 dataset schema for the participants' environmental literacy responses

Field Name	Data Type	Description	Constraints
enviroment_energy_saving	Categorical	Imagine there is an energy-saving measure that can reduce your energy bills, but may change your daily habits or lifestyle. Do you think you would keep this change?	-
enviroment_agree_or_disagree.enviroment_agree_or_disagree_spaceship	Categorical	The Earth is like a spaceship with very limited room and resources	-
enviroment_agree_or_disagree.enviroment_agree_or_disagree_humans	Categorical	Humans were meant to rule over the rest of nature	-

enviroment_agree_or_disagree.enviroment_agree_or_disagree_balance	Categorical	The balance of nature is very delicate and easily upset	-
enviroment_agree_or_disagree.enviroment_agree_or_disagree_modify	Categorical	Modifying the environment for human use seldom causes serious problems	-
enviroment_agree_or_disagree.enviroment_agree_or_disagree_protect_survival	Categorical	Protecting the environment is critically important for the survival of future generations	-
enviroment_agree_or_disagree.enviroment_agree_or_disagree_job_creation	Categorical	Protecting the environment is equally important as job creation	-
enviroment_agree_or_disagree.enviroment_agree_or_disagree_education	Categorical	Environmental education should be a necessary component of primary and secondary school education	-
enviroment_carbon_monoxide	Categorical	Carbon monoxide is a major contributor to air pollution in Europe. Which of the following is the biggest source of carbon monoxide?	-
enviroment_pollution_of_streams	Categorical	What is the most common cause of pollution of streams, rivers and oceans?	-
enviroment_renewable	Categorical	Which of the following is a renewable resource?	-
enviroment_ozone	Categorical	Ozone forms a protective layer in the earth's upper atmosphere. What does ozone protect us from?	-
enviroment_wastes	Categorical	Where does most of the waste in Europe end up?	-
enviroment_do_following.enviroment_do_following_turn_off_lights	Categorical	How often do you Turn off the lights at home when they are not in use	-
enviroment_do_following.enviroment_do_following_pick_up_litter	Categorical	How often do you Pick up litter that I find outside	-
enviroment_do_following.enviroment_do_following_turn_off_water	Categorical	How often do you Turn off the water when it is not in use (i.e. brushing teeth, etc.)	-
enviroment_do_following.enviroment_do_following_close_refrigerator	Categorical	How often do you Close the refrigerator door while I decide what to get out of it	-
enviroment_do_following.enviroment_do_following_recycle	Categorical	How often do you Recycle at home	-
enviroment_agree_or_disagree_extra.enviroment_agree_or_disagree_extra_get_information	Categorical	I get information on environmental issues from newspapers, magazines, or the Internet	-
enviroment_agree_or_disagree_extra.enviroment_agree_or_disagree_extra_identify_env_problems	Categorical	I can identify environmental problems and find personal solutions to them	-

enviroment_agree_or_disagree_extra.enviroment_agree_or_disagree_extra_integrate_viewpoints	Categorical	I can integrate different viewpoints on environmental issues and form my personal opinions.	-
enviroment_agree_or_disagree_extra.enviroment_agree_or_disagree_extra_communicate_info_to_others	Categorical	I can communicate relevant environmental-related information to others	-

Table 13 presents the dataset schema for use case 4 participants' serious game repair/replace responses.

Table 13: Use case 4 dataset schema for the participants' serious game repair/replace responses

Field Name	Data Type	Description	Constraints
sg_answer1	Categorical	Serious game 1st answer	-
sg_answer2	Categorical	Serious game 2nd answer	-
sg_appliance	Categorical	Serious game appliance that broke down	-
sg_negotiation_answer_1	Categorical	1st answer for the negotiation with the repairman	-
sg_negotiation_answer_2	Categorical	2nd answer for the negotiation with the repairman	-
sg_negotiation_answer_3	Categorical	3rd answer for the negotiation with the repairman	-
sg_negotiation_answer_4	Categorical	4th answer for the negotiation with the repairman	-
sg_negotiation_answer_5	Categorical	5th answer for the negotiation with the repairman	-
sg_negotiation_answer_6	Categorical	6th answer for the negotiation with the repairman	-
sg_negotiation_answer_7	Categorical	7th answer for the negotiation with the repairman	-
sg_negotiation_answer_8	Categorical	8th answer for the negotiation with the repairman	-
sg_negotiation_answer_9	Categorical	9th answer for the negotiation with the repairman	-
sg_negotiation_answer_10	Categorical	10th answer for the negotiation with the repairman	-
sg_negotiation_answer_11	Categorical	11th answer for the negotiation with the repairman	-
sg_negotiation_answer_12	Categorical	12th answer for the negotiation with the repairman	-

A total of 1056 individuals participated in this use case and played the EVIDENT Serious game. The descriptive statistics presented in Table 14 are based on the collected data from January to June of 2023 across Greece, Ireland and Sweden.

Table 14: Descriptive statistics of use case's 4 dataset

Demographics		Number of participants (N)	Percentage of participants (%)
Resident	Homeowner	502	47.9
	Homeowner and Landlord	24	2.3

	Homeowner and Tenant	8	0.8
	All three	2	0.2
	Landlord	11	1.1
	Tenant	494	47.2
	Landlord and Tenant	6	0.6
Income	5000	163	15.5
	17500	328	31.1
	37500	328	31.1
	67500	134	12.7
	85000	52	4.9
	150000	43	4.1
	250000	6	0.6
Employment	Unemployed	99	9.4
	Student	251	23.9
	Retired	7	0.7
	Prefer not to say	11	1.0
	Part-time employed	142	13.5
	Job-seeking	52	4.9
	Full-time employed	490	46.6
Country	Greece	284	26.9
	Italy	173	16.4
	Spain	123	11.7
	Ireland	111	10.5
	Estonia	72	6.8
	Norway	61	5.8
	Sweden	54	5.1
	Finland	44	4.2
	Latvia	39	3.7
	Czech Republic	34	3.2
	Austria	17	1.6
	Denmark	15	1.4
	Norway	12	1.1

	Switzerland	11	1.0
	United Kingdom of Great Britain and Northern Ireland	2	0.2
	Netherlands	1	0.1
	United States of America	1	0.1
Age bracket	<25	300	28.5
	26-40	611	58.0
	41-50	102	9.7
	51-60	32	3.0
	>60	8	0.8

2.3.2 Data generation, collection and transformation

As outlined in D3.2, the EVIDENT serious game is hosted on the EVIDENT platform, while recruitment of participants was done using Prolific[12] due to the requirement to achieve a high response rate and the time required to play the game (a minimum of 30 minutes), and to gather the date.

Prolific is an online platform that primarily serves as a tool for researchers to recruit and manage study participants for academic research, surveys, experiments, and other social science studies. It is increasingly popular in fields such as psychology, sociology, economics, and related disciplines. Prolific main features (related to this use case) are as follows:

1. **Participant Recruitment:** Prolific offers researchers access to a pool of pre-screened and diverse participants willing to participate in research studies. EVIDENT was able to specify the target demographic and criteria for participants, ensuring that the sample matched the study's requirements.
2. **Diverse Participant Pool:** The platform has a global user base, which means that participants could be recruited from various regions and backgrounds, increasing the geographical/climate diversity of the sample.
3. **Ease of Use:** Prolific provides an intuitive and user-friendly interface for both researchers and participants. The link to the EVIDENT platform was provided to participants, and they could quickly find and participate in the study.
4. **Participant Compensation:** Participants were compensated for their time and participation (according to a pro-rata average EU minimum wage).
5. **Data Quality:** Prolific is known for its emphasis on data quality. Participants are encouraged to provide accurate and thoughtful responses. Researchers can use various tools to ensure the

[12] https://www.prolific.com

quality of their data (in this use case, only users who had successfully completed 50 previous Prolific studies were able to participate in the EVIDENT serious game).

6. **Ethical Considerations:** The platform is designed to adhere to ethical guidelines in research. It provides for obtaining informed consent and protecting participant privacy, together with compensating participants for their time and ensuring there is no exploitation.
7. **Real-Time Data Collection:** Researchers can collect data in real time, track the progress of their studies, and monitor participant responses as they come in.

The data from the Serious Game is hosted on the EVIDENT platform. This online platform facilitates online experiments and quasi-experiments, allowing for the integration of a wide array of applications such as surveys and serious games within a user-friendly environment. Figure 9 outlines the step-by-step process undertaken to process the provided data, including the full process of analysis. More details can be found in *D4.4, "Analytical Usage Handbooks for Tools and Datasets"*.

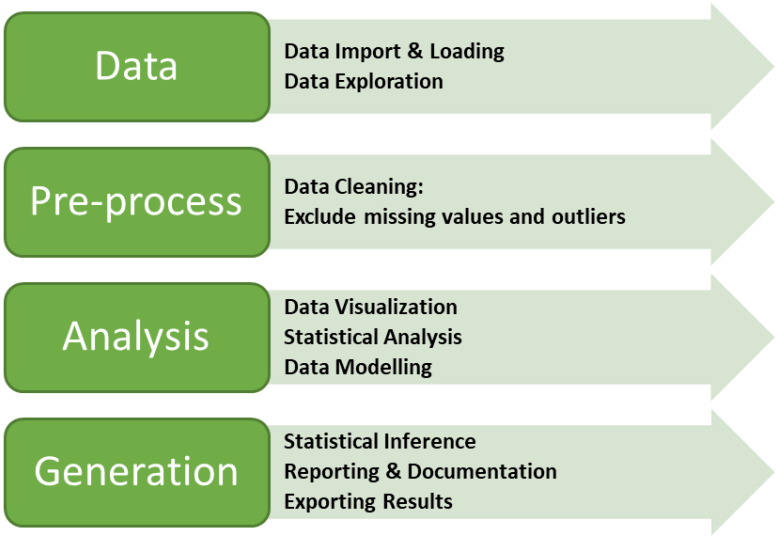


Figure 9: Use case 4, steps followed to prepare the data for analysis

2.4 EVIDENT Use Case 5 Datasets

2.4.1 Data Description, Usage and Schema

EVIDENT's use case 5 seeks to explore the impact of energy-related financial literacy, consumer motivation, point-of-sale information and demographic factors on discount rates and willingness-to-pay for efficient household appliances. Across a series of choice experiments, the impact of factors such as financial information (purchase price, operating cost, salience of financial information), risk reduction (i.e. extended warranty), and financial capacity (i.e. low-cost loans) on implicit discount rates for home appliances is examined. Further, the impact of direct rebound rates on efficient appliance selection is examined. Finally, through an additional quasi experiment, use case 5 seeks to explore the impact of average price biases on energy decision making.

Discrete Choice Experiment on willingness-to-pay for more efficient appliances

The experiment consists of gathering and analysing data on the following topics:

1. demographic information
2. current home appliance purchasing behaviour
3. financial literacy
4. environmental literacy
5. stated preference experiment consisting of four choice points
6. discount rates
7. discrete choice experiment consisting of ten choice points
8. questions examining direct rebound rates associated with the novel appliance selected.

Two choice experiments are included within the current experiment. The first is a stated preference experiment examining the impact of financial and energy framing on willingness-to-pay for energy-efficient appliances. Four choice points are presented within this experiment. Choice 1 presents five identical versions of an appliance, which differ only by one key feature and seeks to reduce hypothetical bias across the choice experiment. For example, for a washing machine, the key features are cost, capacity, spin speed, quick wash time and pause wash functionality (see Figure 10). Choice 2 consists of the participant's initial choice (at choice 1) alongside alternatives that differ only in the purchase price and energy rating, with a higher purchase price for more efficient appliances (e.g., A-rated appliances are most expensive; D-rated appliances are the least expensive). Choice 3 is similar to choice 2, however, operational costs per month are also presented in this instance. Again, operational costs are lower for more efficient appliances. Choice 4 is similar to choose 3, however, operational costs per year are presented in this instance.

	A	B	C	D	E
					
Cost	From €250	From €299	From €278	From €289	From €300
Capacity	7kg	8kg	7kg	7kg	7kg
Spin Speed	1400	1400	1600	1400	1400
Quick wash time	15 minutes for 2 kg	15 minutes for 2 kg	15 minutes for 2 kg	10 minutes for 2kg	15 minutes for 2 kg
Pause wash	No	No	No	No	Yes

Figure 10: Appliance presentation at choice 1

The second choice experiment is a Discrete Choice Experiment which explores the relative impacts of risk reduction (extended warranty) and financial supports (low-cost loan, loan term) on willingness to invest in more efficient energy appliances. Attributes were selected based on literature review, focus group analyses, cognitive walk-through and usability analyses. Once final attributes were determined, choice cards were developed using a fractional factorial design. A statistically efficient main-effects design with

ten choice sets was created in RStudio using the idefix package. As such, participants are presented with a series of ten choice points, each consisting of two appliances and a 'no preference' option.

Some question categories, such as demographics, financial and environmental literacy, are common between use cases 4 and 5 and have already been presented in section 2.3. Below, we present the data schema on participants' energy awareness responses, responses on the Discrete Choice Experiment. Table 15 presents the dataset schema for use case 5 participants' energy awareness responses.

Table 15: Use case 5 dataset schema for the participants' energy awareness responses

Field Name	Data Type	Description	Constraints
ReplacementFridge	Categorical	When would you would seek to replace the Refrigerator	-
ReplacementWashingMachine	Categorical	When would you would seek to replace the Washing machine	-
ReplacementOven	Categorical	When would you would seek to replace the Stove	-
EngRatingDishwasher	Categorical	What energy rating are your current Refrigerator	-
EngRatingWashingMachine	Categorical	What energy rating are your Washing machine	-
EngRatingOven	Categorical	What energy rating are your current Stove	-
EngAwarenessFridge	Categorical	If you are aware, what is the best energy rating available for Refrigerator	-
EngAwarenessWashingMachine	Categorical	If you are aware, what is the best energy rating available for Washing Machine	-
EngAwarenessOven	Categorical	If you are aware, what is the best energy rating available for Oven	-
Wash_setting	Categorical	What setting do you tend to use when using your washing machine?	-

Table 16 presents the dataset schema for use case 5 participants' Discrete Choice Experiment responses.

Table 16: Use case 5 dataset schema for the participants' Discrete Choice Experiment responses

Field Name	Data Type	Description	Constraints
Choice_1	Categorical	Which of these appliance options would you like to choose?	-
StatedPreference1	Categorical	Which of these appliance options would you like to choose?	-
StatedPreference2	Categorical	Which of these appliance options would you like to choose?	-
StatedPreference3	Categorical	Which of these appliance options would you like to choose?	-
StatedPreference4	Categorical	What setting would you use with your new washing machine?	-
DiscountRate	Categorical	Suppose you are buying new refrigerator with a life expectancy of 10 years. You have to choose between two models whose only difference is their energy efficiency level, a standard one and an efficient one that costs €100 more but can generate a reduction in your future electricity bills. How much would you have to save in your electricity bill	-

		approximately per year during the next 10 years in order to pay for the additional €100 Euros?	
DiscountRateRationale	Text/String	If you wrote "I don't know", why?	-
DCEChoice1	Categorical	Which of these appliance options would you like to choose?	-
DCEChoice2	Categorical	Which of these appliance options would you like to choose?	-
DCEChoice3	Categorical	Which of these appliance options would you like to choose?	-
DCEChoice4	Categorical	Which of these appliance options would you like to choose?	-
DCEChoice5	Categorical	Which of these appliance options would you like to choose?	-
DCEChoice6	Categorical	Which of these appliance options would you like to choose?	-
DCEChoice7	Categorical	Which of these appliance options would you like to choose?	-
DCEChoice8	Categorical	Which of these appliance options would you like to choose?	-
DCEChoice9	Categorical	Which of these appliance options would you like to choose?	-
DCEChoice10	Categorical	Which of these appliance options would you like to choose?	-

A total of 2220 individuals participated in the quasi-experiment, of whom 1751 provided complete responses (79% retention rate). The descriptive statistics presented in Table 17 are based on the collected data from January to June of 2023 across Greece, Ireland and Sweden.

Table 17: Use case 5 discrete choice experiment descriptive statistics

Demographics		Number of participants (N)	Percentage of participants (%)
Resident	Homeowner	1183	67.6
	Landlord	76	4.34
	Tenant	492	28.1
Income	Less than 25,000	223	12.9
	25-50,000	319	18.4
	51-75,000	351	20.20
	76-100,000	449	25.90
	101-200,000	344	19.83
	200,000+	49	2.82
Employment	Full-time	1352	77.12
	Part-time	158	9.01
	Job-seeking	18	1.03
	Student	53	3.02
	Unemployed	20	1.14
	Retired	135	7.7
	Prefer not to say	17	0.97

Country	USA	224	13.19
	UK	100	6.33
	Sweden	365	23.12
	Ireland	527	33.38
	Greece	283	17.92
	Germany	27	1.71
	France	19	1.2
	Canada	19	1.2
	Australia	15	0.95
	Other	172	9.82
Age bracket	Under 30	426	24.36
	31-40	617	35.28
	41-50	334	19.1
	51-64	190	10.86
	65 or older	182	10.41

Discrete choice experiment on willingness-to-pay for more efficient appliances

This quasi-experiment seeks to explore the impact of average price biases on energy decision making. While research has suggested the important role of behavioural biases and financial and environmental literacy on energy decision making, analysis of their combined impact is needed. One particularly challenging energy decision for consumers is nonlinear energy pricing wherein multiple marginal prices are applied.

The average price bias choice quasi-experiment is designed to elicit consumers' perceptions about different pricing schemes. The experiment aims to correlate the findings with participants' characteristics, potential behavioural biases and the participants' financial and environmental literacy levels.

The experiment consists of five discrete key sections:

1. a section about participant's demographic data
2. a small set of questions related to behavioural biases
3. a set of questions related to financial literacy
4. a section with questions related to environmental literacy
5. the choice experiment about price perceptions

Since some question categories, such as demographics, financial and environmental literacy, are common between use cases 4 and 5, we only present the data schema on participants' responses on the Average Price Bias Experiment. Table 18 presents the dataset schema for use case 5 participants' Average Price Bias responses. All 120 variables have the same data type and the same description ("Which tariff would you choose?") while there no constraints. Thus, the corresponding columns are missing from the table.

The name convention indicates a different case and a subcategory, as described in detail in *D3.2 "Implementation of preparatory actions for RCT, surveys and serious game"*.

Table 18: Use case 5 dataset schema for the participants' Average Price Bias responses

Case	Choice set	Field Names
Case 1 - Even distribution of consumption loads between tiers	1a - The first-tier price is fixed and the second-tier price changes	APB_option_1a_1, APB_option_1a_2, APB_option_1a_3, APB_option_1a_4, APB_option_1a_5, APB_option_1a_6
	1b - The second-tier price is fixed and the first-tier price changes	APB_option_1b_1, APB_option_1b_2, APB_option_1b_3, APB_option_1b_4
	1c - The first-tier price is fixed and the second-tier price changes	APB_option_1c_1, APB_option_1c_2, APB_option_1c_3, APB_option_1c_4, APB_option_1c_5
	1d - The second-tier price is fixed and the first-tier price changes	APB_option_1d_1, APB_option_1d_2, APB_option_1d_3
Case 2 - Uneven distribution of consumption loads between tiers	2a - The first-tier price is fixed and the second-tier price changes	APB_option_2a_1, APB_option_2a_2, APB_option_2a_3, APB_option_2a_4, APB_option_2a_5
	2b - The second-tier price is fixed and the first-tier price changes	APB_option_2b_1, APB_option_2b_2 APB_option_2b_3, APB_option_2b_4, APB_option_2b_5
	2c - The first-tier price is fixed and the second-tier price changes	APB_option_2c_1, APB_option_2c_2, APB_option_2c_3, APB_option_2c_4, APB_option_2c_5
	2d - The second-tier price is fixed and the first-tier price changes	APB_option_2d_1, APB_option_2d_2, APB_option_2d_3, APB_option_2d_4, APB_option_2d_5, APB_option_2d_6
Case 3 - Uneven distribution of consumption loads between tiers	3a - The first-tier price is fixed and the second-tier price changes	APB_option_3a_1, APB_option_3a_2, APB_option_3a_3, APB_option_3a_4, APB_option_3a_5
	3b - The second-tier price is fixed and the first-tier price changes	APB_option_3b_1, APB_option_3b_2, APB_option_3b_3, APB_option_3b_4
	3c - The first-tier price is fixed and the second-tier price changes	APB_option_3c_1, APB_option_3c_2, APB_option_3c_3, APB_option_3c_4, APB_option_3c_5, APB_option_3c_6
	3d - The second-tier price is fixed and the first-tier price changes	APB_option_3d_1, APB_option_3d_2, APB_option_3d_3, APB_option_3d_4, APB_option_3d_5
Case 4 - Even distribution of consumption loads between tiers	4a - The second-tier and the third-tier prices are fixed and the first-tier price changes	APB_option_4a_1, APB_option_4a_2, APB_option_4a_3, APB_option_4a_4
	4b - The first-tier and the second-tier prices are fixed and the third-tier price changes	APB_option_4b_1, APB_option_4b_2, APB_option_4b_3, APB_option_4b_4
	4c - The first-tier and second-tier prices are fixed and the third-tier price changes	APB_option_4c_1, APB_option_4c_2, APB_option_4c_3, APB_option_4c_4, APB_option_4c_5, APB_option_4c_6
	4d - The second-tier and third-tier prices are fixed and the first-tier price changes	APB_option_4d_1, APB_option_4d_2, APB_option_4d_3, APB_option_4d_4, APB_option_6d_4
Case 5 - Uneven distribution of consumption	5a - The first-tier and second-tier prices are fixed and the third-tier price changes	APB_option_5a_1, APB_option_5a_2, APB_option_5a_3, APB_option_5a_4, APB_option_5a_5

loads between tiers	5b - The second-tier and the third-tier prices are fixed and the first-tier price changes	APB_option_5b_1, APB_option_5b_2, APB_option_5b_3, APB_option_5b_4, APB_option_5b_5
	5c - The first-tier and the second-tier prices are fixed and the third-tier price changes	APB_option_5c_1, APB_option_5c_2, APB_option_5c_3, APB_option_5c_4, APB_option_5c_5, APB_option_5c_6
	5d - The second-tier and the third-tier prices are fixed and the first-tier price changes	APB_option_5d_1, APB_option_5d_2, APB_option_5d_3, APB_option_5d_4, APB_option_5d_5
Case 6 - Uneven distribution of consumption loads between tiers	6a - The first-tier and the second-tier prices are fixed and the third-tier price changes	APB_option_6a_1, APB_option_6a_2, APB_option_6a_3, APB_option_6a_4, APB_option_6a_5
	6b - The second-tier and the third-tier prices are fixed and the first-tier price changes	APB_option_6b_1, APB_option_6b_2, APB_option_6b_3, APB_option_6b_4, APB_option_6b_5, APB_option_6b_6, APB_option_6b_7
	6c - The first-tier and the second-tier prices are fixed and the third-tier price changes	APB_option_6c_1, APB_option_6c_2, APB_option_6c_3, APB_option_6c_4, APB_option_6c_5, APB_option_6c_6
	6d - The second-tier and the third-tier prices are fixed and the first-tier price changes	APB_option_6d_1, APB_option_6d_2, APB_option_6d_3, APB_option_6d_4

A total of 1030 individuals participated in the average price bias quasi-experiment. The descriptive statistics presented in Table 19 are based on the collected data from January to June of 2023 across Greece, Ireland and Sweden.

Table 19: Use case 5 average price bias experiment descriptive statistics

Demographics		Number of participants (N)	Percentage of participants (%)
Resident	Homeowner	668	64.85
	Landlord	25	2.43
	Tenant	337	32.72
Income	Less than 25,000	100	9.71
	25-50,000	203	19.71
	51-75,000	214	20.78
	76-100,000	272	26.41
	101-200,000	208	20.19
	200,000+	33	3.20
Employment	Full-time	837	81.26
	Part-time	97	9.42
	Job-seeking	8	0.78
	Student	32	3.11

	Unemployed	16	1.55
	Retired	33	3.20
	Prefer not to say	7	0.68
Country	Ireland	518	50.29
	Greece	114	11.07
	USA	97	9.42
	UK	96	9.32
	Italy	38	3.69
	Germany	21	2.04
	Netherlands	15	1.46
	France	14	1.36
	Sweden	11	1.07
	Other	106	10.29
Age bracket	Under 30	336	32.62
	31-40	334	32.43
	41-50	184	17.86
	51-64	133	12.91
	65 or older	43	4.17

2.4.2 Data generation, collection and transformation

The quasi-experiments from use case 5 were hosted on the EVIDENT platform. This online platform facilitates online experiments and quasi-experiments, allowing the integration of various applications, such as surveys and serious games within a user-friendly environment. To reach the widest possible audience, the studies were hosted on Qualtrics, an online survey tool. The recruitment process involved reaching out to potential participants through various channels, including social media platforms like LinkedIn and Twitter (now 'X'), the EVIDENT project's website, and email lists of EVIDENT consortium members. To ensure a diverse participant pool, special attention was given to groups that have been underrepresented in previous studies, such as the elderly, low-income households, renters, and landlords. Invitations to participate were distributed through local and regional community organizations that catered to these specific demographics. Additionally, to encourage participation from individuals who may not have a strong interest in environmental topics, an incentive was offered in the form of a chance to win a voucher from an online retailer.

Figure 11 outlines the step-by-step process undertaken to process the provided data, including the full process of analysis. More fulsome details can be found in *D4.4, "Analytical Usage Handbooks for Tools and Datasets"*.

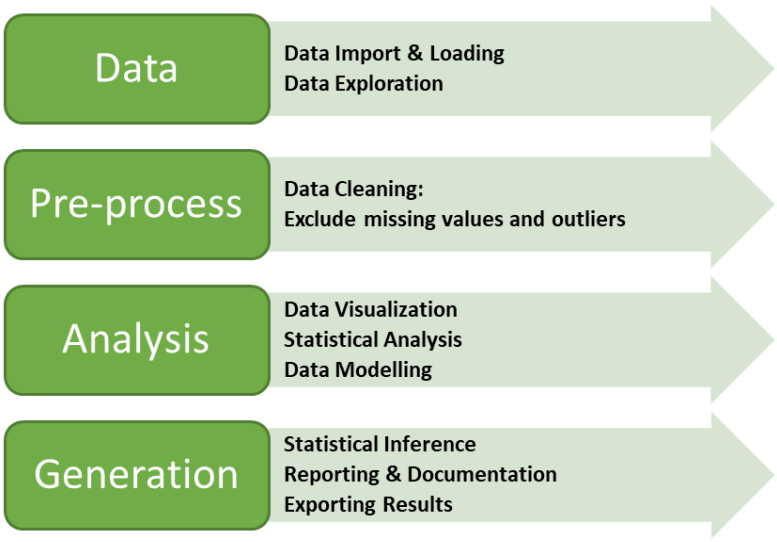


Figure 11: Use case 5, steps followed to prepare the data for analysis

3. EVIDENT Data Services

Till this point, several datasets and data sources have been utilized. For example, for designing and implementing the natural field experiment in use cases 1 and 2, the energy company, consortium partner CW, provides the data. In addition, the data provided by the PPC, the second energy company and also a consortium partner, to estimate how energy consumption is affected by the households' appliances and by consumers' consumption patterns. For the analyses in use case 3, the data provided by CW are leveraged while different aggregations are performed; for example, the hourly data are transformed into daily or weekly observations. Moreover, in these analyses, weather data for consumer locations are used. These data are collected through 3rd party services and data sources. For both use cases 4 and 5, two data collection platforms are used to collect the appropriate data. These platforms are the EVIDENT platform and the Qualtrics platform. The two experiments collect data as replies into three quasi-experiments designed to elicit consumer perspectives regarding energy usage and efficiency.

During the EVIDENT's project lifecycle, several consortium partners should cooperate and deliver solid and robust experimental results; thus, collecting and managing data in a centralized location is necessary. For better understanding and handling of the collected data, the EVIDENT consortium adopted a flexible architecture for scale-in/scale-out infrastructure, suitable for large-scale data collection, storage and processing. The architecture described in this section presents the methodological steps and the big data architecture agreed upon and followed by the corresponding EVIDENT consortium partners for the organized collection and retrieval of the data. The proposed architecture is also presented in this deliverable as a lesson learned and best practices for the proper collaboration between partners with different expertise and analytical skills. In addition, when it comes to data analysis, the researchers should consider their analytical tools as components of a wider modular business architecture.

Having said that and including all the previous information regarding the datasets, their schema, and their pre-processing and cleaning processes, Figure 12 presents the flexible architecture adopted during the project's lifecycle. This architecture presents an easy way for partners who provide data (e.g. CW and PPC) to upload data to the central database and for analytical partners to easily retrieve data when needed using the corresponding services.

Figure 12: EVIDENT consortium flexible architecture for scale-in/scale-out infrastructure

The proposed architecture's first step is identifying the different data sources. This also includes information about the corresponding datasets and the data schema (e.g. description of the data). The next step includes all these actions partners should follow to transform the provided data into a specific format agreed upon by the EVIDENT partners from the very beginning of the project. The data schema and the data transformation and cleaning processes have been presented for each dataset in section 2 of this deliverable.

The next step describes the central database adopted to host all collected and provided data. Since the data follows a strict schema, a MySQL database was developed. A unique table with all corresponding fields was created for each provided dataset while the corresponding partners undertook the population of these tables with data. Since needed each partner appended data several times during the project's lifecycle since the data became available over time. Figure 13 depicts a snapshot of the central database.

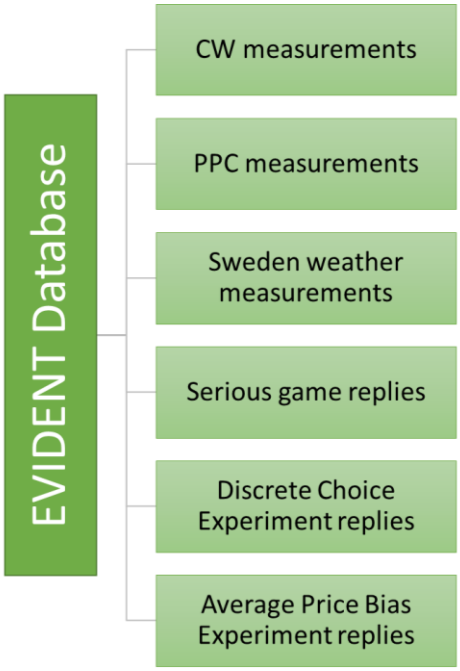


Figure 13: The EVIDENT central database

In the EVIDENT database, there are six different data tables, each corresponding to a different dataset. The tables have been organized based on the partner responsible for providing or collecting the data. Another distinction was made based on the use case each dataset is used for. For example, the data collected through the EVIDENT platform and the EVIDENT serious game are stored in the "Serious game replies" table.

The next step of the proposed architecture is creating the related data retrieval services. These services include creating a secure API developed to provide the corresponding data. For each table, two services are provided; the first provides all stored data of the table, while the second receives the date range for the retrieved data as a parameter. The data range refers to the measurement timestamp for the database tables that store either time-series data or panel data. In contrast, for the tables where quasi-experiment replies are stored, the date range refers to the date each reply was received.

The final step is the data analysis, where each partner has retrieved the data and can perform to deliver valuable results and insights. The analytical tools developed for each use case are presented in deliverables *D4.2 "Econometric analysis and robustness tests"*, *D4.3 "Updated econometric methodologies and robustness tests"*, and *D4.4 "Analytical usage handbooks for tools and datasets"*.

By adopting the architecture proposed above, the consortium partners managed to efficiently handle a large number of data and better organize different data sources. Finally, access to the data and the services were provided to the relative partners under a virtual private network (VPN) to ensure data security and protect against unauthorized access.

4. Support Multi-Domain Exchange of Information

The EVIDENT project seeks to propose policy measures that are effective, evidence-based, and tailored to the specific needs and circumstances of different user groups based on the insights gained from the different approaches, such as field trials, surveys and serious games leveraged during the project's lifespan. The project's ultimate goal is to accelerate the transition to a more sustainable energy future by helping consumers and businesses make informed decisions about their energy use and consumption. Through its combination of research methodologies and its focus on evidence-based policy measures, the EVIDENT project represents an important step forward in the global effort to address climate change and ensure a secure, affordable, and sustainable energy supply for all.

While the design and results of the five use cases remain innovative and novel, it is acknowledged that much of the research carried out today is built upon the foundations of previous studies. In this sense, the EVIDENT project represents a continuation of the work already conducted, while also contributing to the advancement of the field.

The opportunity to analyse and expand upon prior findings allows for a deeper understanding of the research area and enables research to build upon the knowledge that has already been established. The ultimate objective is to bring a new perspective, a new angle, and novel insights into the research area.

4.1 Data from Prior Relevant Surveys

In this sub-section we refer to a sub-set of the relevant literature to explore the relative studies that enhance the empirical findings of the EVIDENT project, that is, the studies that addressed similar research questions. Among other sources, this subsection is informed by EVIDENT *D1.1 'Analysis of best practices'* and *D3.1* '*Specifications of Preparatory Actions for RCT Surveys and Serious Game'*, where the reader can find reviews of the relevant literature. In this sub-section, we build upon these reviews and further explore the topic of **household energy consumption and conservation,** by presenting insights from studies not included in previous EVIDENT deliverables. These insights supplement and further support the original findings of the EVIDENT project.

4.1.1 Studies on household energy consumption

The issue of household energy consumption is becoming more popular in the academic literature. For example, (Han and Wei 2021) conducted a bibliometric and network analysis and found that in the period from 2010 till 2018, the number of publications in the last three years of the sample accounted for 85% of the total published papers. These trends are presented in Figure 14.

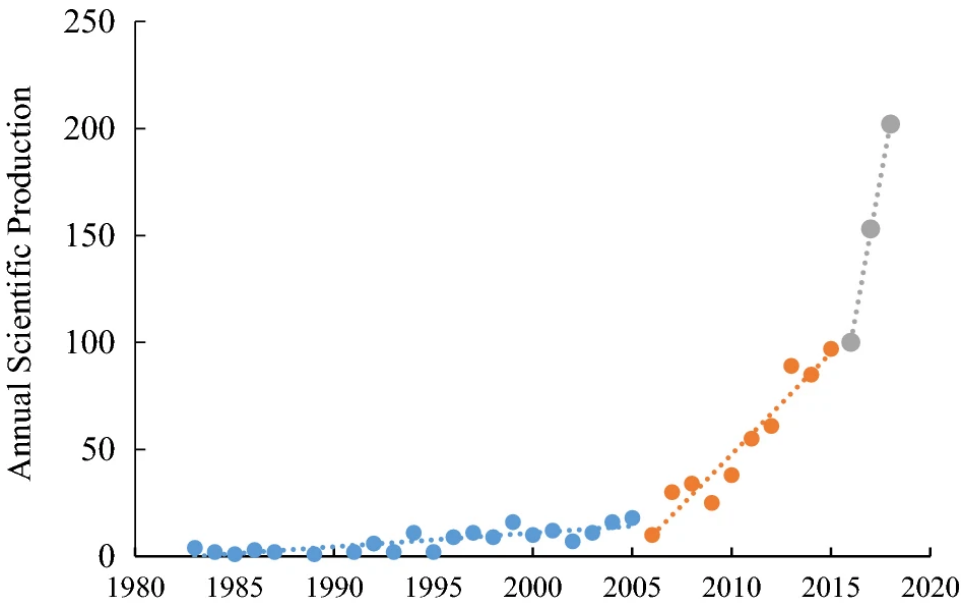


Figure 14: Publication trends in the area of household energy consumption[13]

Furthermore, the roles of social, cultural, and psychological factors are increasingly attracting scholars' attention. From the paper: *"In terms of specific research content, we have noticed that the papers related to these factors have high citations and tend to be published in leading journals."*, where *"a growing body of research has begun to use experimental intervention to study the impact of irrational factors on residents' energy use behaviour."*

Han and Wei (2021) also conducted a topic analysis of published articles in *Nature* journals during the period 2014–2018. In total, 39 articles were classified into 9 categories. The distribution per category is presented in Figure 15, where it is obvious that behaviour/intervention papers are prominent in the recent literature.

[13] Source: (Han and Wei 2021)

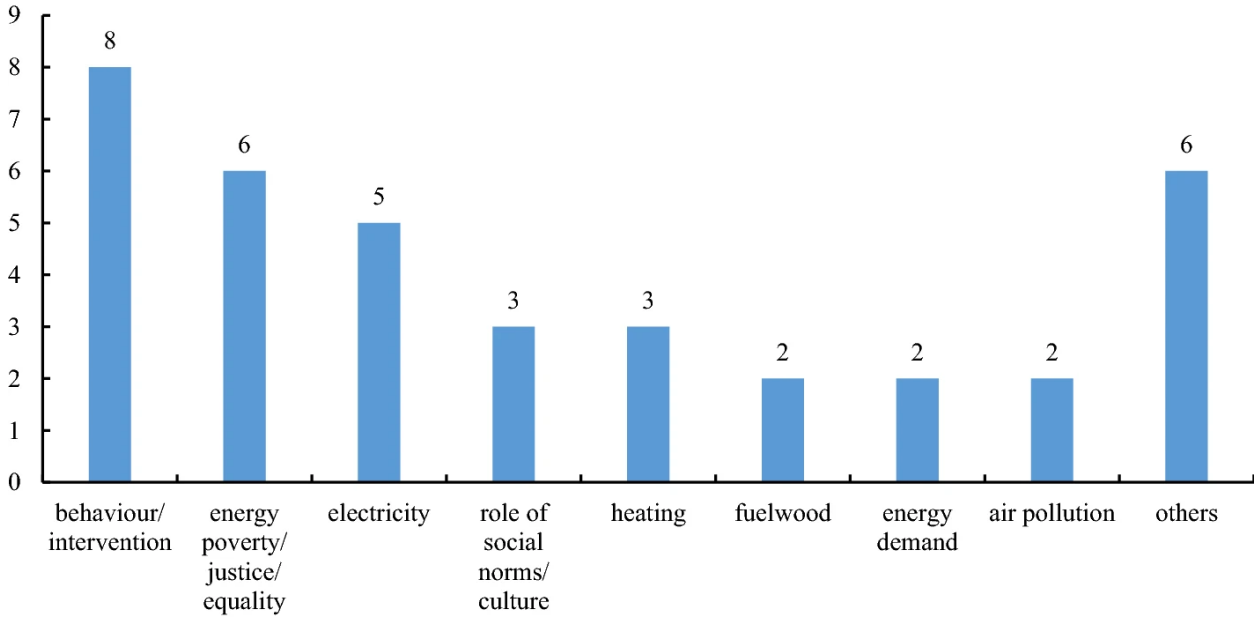


Figure 15: The number of articles in each category[14]

An example on behavioural interventions for energy efficiency regard **feedback systems such as home energy reports** designed and implemented in two EVIDENT use cases. Such feedback provides the opportunity to consumers to become active agents in the energy system. Feedback is an instrumental tool that empowers final energy consumers with significant knowledge and insights about their energy consumption. By thoroughly analysing the feedback received, energy consumers can gain a better understanding of their energy consumption patterns and take the necessary steps to modify their behaviour to maximize energy savings. In other words, final energy consumers can utilize the feedback to make more informed decisions about their energy usage.

The feedback mechanism helps final energy consumers to cultivate sustainable and efficient energy practices, while also providing them with an opportunity to interpret the results of their actions. This continuous feedback loop enables energy consumers to monitor and adjust their behaviour (sometimes even in real-time), thus, enabling them to make improvements and achieve their energy goals. Additionally, final energy consumers can also gain a better understanding of which changes lead to better energy outcomes, providing them with knowledge for future decision-making.

Moreover, feedback informs energy consumers of their progress in reducing their carbon footprint, resulting in an increased awareness of their impact on the environment. With the help of feedback, final energy consumers can better understand how their energy consumption patterns contribute to carbon emissions and work towards reducing their energy usage accordingly.

[14] Source: (Han and Wei 2021)

Feedback serves as a crucial mechanism for final energy consumers to not only adapt their behaviour according to energy-efficient practices but also to gain insights that enable a better understanding of the impact of their actions on energy consumption and the environment.

(Zangheri, Serrenho and Bertoldi 2019) reviewed more than 70 studies and found that while feedback mechanisms can bring results, they need to be constantly adapted to maintain the engagement of the consumers, and thus, the persistence of the savings. For example, the authors found from their examination of the aforementioned studies, that a higher frequency of feedback resulted in significantly increased energy conservation rates, with a marked 8% decrease in energy usage when feedback is given multiple times per week, compared to a meagre reduction of less than 3% when feedback is given every 2-6 months.

In addition to feedback mechanisms, another increasingly popular behavioural technique are serious games, which are games that have been created with an objective beyond just providing entertainment. They leverage the engaging and motivational aspects of game design to educate, train, or influence behaviour. In relation to energy efficiency, serious games can be utilized to inform and encourage individuals to modify their energy consumption habits.

(Delemere and Liston 2022) present a systematic review and content analysis of the use of behaviour change techniques (BCTs) within serious games for energy efficiency, defined as the use of gamification within interventions. The authors searched seven databases using search terms relating to serious games and energy efficiency and identified 21 serious games for energy efficiency. Across serious games, 36 individual BCTs were found with an average of 9.45 individual BCTs per game. Common BCTs included monitoring of behaviour by others without feedback, feedback on behaviour and self-monitoring of behaviour. The type and frequency of BCTs across the explored serious games are reported in Table 20.

Table 20: Number of game inclusions of each behaviour change techniques[15]

Behaviour change technique	Games included
1.1. Goal setting (behaviour)	4
1.2. Problem solving	6
1.3. Goal setting (outcome)	5
1.4. Action planning	3
1.6. Discrepancy between current behaviour and goal	9
1.7. Review outcome goal(s)	1
2.1. Monitoring of behaviour by others without feedback	18
2.2. Feedback on behaviour	17
2.3. Self-monitoring of behaviour	12

[15] Source: (Delemere and Liston 2022), Table 5

2.5. Monitoring of outcome(s) of behaviour without feedback	11
2.7. Feedback on outcome(s) of behaviour	8
3.1. Social support (unspecified)	8
3.2. Social support (practical)	7
4.1. Instruction on how to perform the behaviour	3
5.3. Information about social and environmental consequences	9
6.1. Demonstration of the behaviour	1
6.2. Social comparison	12
7.1. Prompts/cues	5
7.5. Remove aversive stimulus	1
8.2. Behaviour substitution	1
8.3. Habit formation	3
8.4. Habit reversal	1
9.3. Comparative imagining of future outcomes	1
10.1. Material incentive (behaviour)	6
10.2. Material reward (behaviour)	2
10.3. Non-specific reward	4
10.4. Social reward	4
10.5. Social incentive	2
10.6. Non-specific incentive	1
10.8. Incentive (outcome)	5
10.9. Self-reward	1
10.10. Reward (outcome)	5
10.11. Future punishment	1
13.1. Identification of self as role model	8
14.2. Punishment	3

From these 21 serious games for energy efficiency, the authors found that serious games accounted for average decreases in energy use of 15.2% in residential settings, 18.4% in commercial settings and 9.9% in educational and public settings. However, the authors also found that the majority of the studies did not conduct follow-up activities, thus, the persistence of the results (e.g., decrease of energy use) could not be assessed. In addition to the interest by the academic community, policy makers are also promoting energy efficiency activities including those based on behavioural interventions. For example, (Sussman and Chikumbo 2016) found that home energy reports were the most common utility-run energy efficiency behavioural change programmes in the United States, with *"robust savings ranging from 3% to 16%."* However, the authors could not indicate the upper limit of HER savings, nor the persistence of such savings, an issue explored by the EVIDENT project. The types of behavioural interventions in the authors'

report are illustrated in Figure 16. The International Energy Agency (IEA) also reports that competitions and games can lead to savings of around 14% for electricity and 10% for gas (Cornago 2021). However, one must note that competitions and games usually require significantly more resources than HERs.

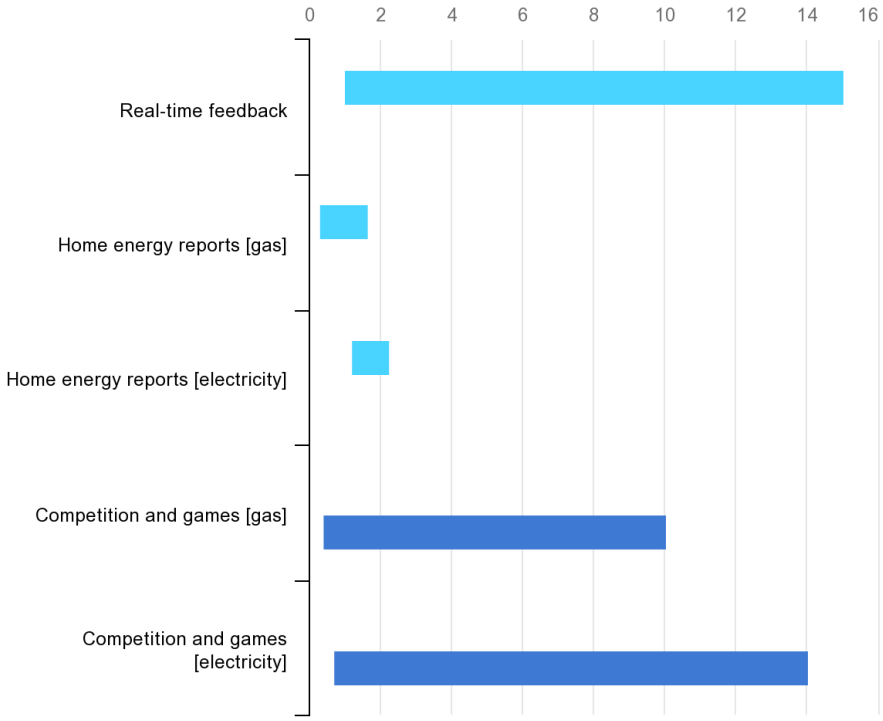


Figure 16: Energy savings across behavioural interventions implemented in the United States.[16]

In Europe, a cornerstone of EU policy on energy efficiency is the Energy Efficiency Directive or EED (2012/27/EU)[17]. The Directive sets binding energy savings targets for EU Member States, requires large companies to carry out energy audits every four years to identify energy-saving measures, promotes the use of energy-efficient technologies and practices, etc. An important element of the EED is that it recognizes the significant role of behavioural change in achieving energy savings targets, and it promotes awareness of energy efficiency and the provision of information to encourage behavioural change and support people in adopting energy-efficient practices.

Several measures under the EED are designed to encourage changes in behaviour. For example, large companies are required to carry out energy audits every four years to identify potential energy savings,

[16] Source: (IEA 2021)
[17] Source: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02012L0027-20230504

and the results must be communicated to employees and stakeholders. This measure promotes awareness and helps to establish energy-saving practices within the organization.

The EED also supports the implementation of energy-efficient technologies, such as smart grid technologies and appliances. The use of smart technologies can help to increase awareness and promote energy-efficient behaviour by providing real-time feedback on energy use and opportunities for energy savings.

More relevant for the EVIDENT project is the EED's empowerment of consumers to make informed decisions by providing them with transparent and clear information about their energy consumption and options for reducing it. This is achieved through several measures, including:

1. Regular billing information: Energy suppliers are required to provide consumers with regular billing information that includes details about their energy use and the associated costs, as well as tips and recommendations on how to reduce energy consumption and save money on their bills.
2. Information on energy efficiency: Energy suppliers are required to provide consumers with information on the energy efficiency of their products and services, as well as information on the availability and benefits of switching to more energy-efficient options.
3. National energy saving obligations schemes: EED encourages Member States to establish national energy saving obligations schemes to encourage households and businesses to take energy-efficient measures through the use of energy savings certificates and white certificates.
4. Energy performance certificates (EPCs): The EED promotes the use of EPCs to provide information to consumers about the energy efficiency of buildings. EPCs provide detailed information about the energy performance of a property and recommendations for improvements to increase energy efficiency. They can also be used to inform potential buyers or tenants about the energy performance of a property, allowing for informed decision-making and encouraging energy-efficient choices.

By empowering consumers with knowledge and resources, the EED can help to encourage behavioural change and support the adoption of energy-efficient practices. Consumers who are informed about their energy consumption and options for reducing it can make more conscious choices and take steps to reduce energy consumption, ultimately helping to reduce greenhouse gas emissions and promote sustainability.

The EED belongs to a group of European and national policies that promote energy savings. (Bertoldi and Mosconi 2020) found that without such policies energy consumption would have been 11% higher.

Several research initiatives inform the design and implementation of the EED and similar policies, including those funded by the Commission's FP and Horizon research and innovation programmes. Another initiative of note is the JRC's Competence Centre on Behavioural Insights, with a dedicated topic

on behavioural insights for energy (efficiency)[18]. An example of the Competence Centre's activities is the aforementioned study by (Zangheri, Serrenho and Bertoldi 2019). Additionally, in a dedicated report, JRC researchers (Nives and Paolo 2021) provide key conceptual and practical insights for encouraging decisions to invest in energy efficiency, with references and examples from four energy-related social sciences, namely economics, behavioural economics, psychology and sociology.

Another example where behavioural research into informed policy changes regarded the changing of the EU energy label classes for household projects. Since 2021 the scale's range is being gradually changed from A+++ to D to A to G, as researchers[19] found that consumers are more likely to choose the most energy efficient appliance when considering the new scale. Several research studies including informed this policy change. On the EVIDENT website one can find a dedicated blog post on the energy labelling system in the EU[20]. Another relevant study from the EU funded CONSEED project[21] found that including the financial costs over a product's lifetime in the labelling increases the consumers' demand for energy efficient across categories (buildings, automobiles, household appliances, etc.).

4.1.2 Project: Societal Appreciation of Security of Energy Supply

An example of a relevant project on energy consumption and consumers' stated preferences is the JRC project 'SASOS - Societal Appreciation of Security of Energy Supply'. In this project, the researchers focused on the security of energy supply while also investigating several aspects on energy use and attitudes. A discrete choice experiment (DCE) was designed and implemented using surveys targeting residential, commercial and industrial energy users.

The surveys were conducted in Estonia, Greece, the Netherlands and Portugal, with a focus on the value of loss load. As the methodology employed for Greece was slightly different than for the other countries, for now we focus only on the other countries. After conducting about 1000 interviews in each of the three Member States[22], the researchers estimated the participants' willingness-to-pay to avoid power outages (e.g., investments in storage solutions at the network level), and the willingness-to-accept (WTA) compensation when such outages were to occur.

It was found that consumers across countries and demographic groups do not share the same preferences[23]. For example, in the Netherlands, respondents who aim to decrease their energy usage have a higher likelihood of preferring the existing conditions (frequency and time length of power outages). On the other hand, in Estonia and Portugal, they tend to prefer the hypothetical situations that offer them

[18] Source: https://knowledge4policy.ec.europa.eu/behavioural-insights/topic/behavioural-insights-energy_en

[19] For examples see https://commission.europa.eu/system/files/2018-03/impact_of_energy_labels_on_consumer_behaviour_en.pdf and https://energy.ec.europa.eu/system/files/2018-07/energy_label_displays_-_final_report_0.pdf

[20] https://evident-h2020.eu/the-energy-labelling-system-in-the-eu/

[21] https://cordis.europa.eu/project/id/723741

[22] In total, the researchers collected 1043 completed interviews for Estonia, 1038 for the Netherlands and 1059 for Portugal.

[23] The results presented in the following paragraphs have been reported in detail in the following publications: (European Commission, Joint Research Centre; Bouman, T; Longo, A; Giaccaria, S; Efthimiadis, T 2019), (European Commission, Joint Research Centre; Efthimiadis, T; Giaccaria, S; Longo, A; Bouman, T 2019), (European Commission, Joint Research Centre; Efthimiadis, T; Giaccaria, S; Longo, A; Bouman, T 2019), (European Commission, Joint Research Centre; Giaccaria, S; Efthimiadis, T; Bouman, T; Longo, A 2019)

compensation for accepting more frequent and longer power cuts. Generally, individuals with significant 'selfless values' are inclined towards selecting the current conditions in all countries, while those with 'egoistic values' prefer a decline in their electricity expenses despite a decline in the quality of the power supply. Also, respondents value more a reduction in the number of unplanned than planned power outages in all countries.

On a scale of 1 to 7, with 1 being completely disagree and 7 being completely agree, respondents were **asked if they were willing to lower their energy usage.** On average, those in Estonia (M = 5.05, SD = 1.69) and the Netherlands (M = 5.13, SD = 1.62) were in agreement with this statement, while those in Portugal strongly agreed (M = 6.14, SD = 1.26). Analysis of data showed that typical demographic factors did not account for differences in energy conservation across the three countries. In Estonia and the Netherlands, the presence of 'biospheric values' exhibited a favourable correlation with a willingness to reduce energy consumption. Moreover, in the Netherlands, 'altruistic values' indicated a positive correlation with the willingness to reduce energy usage, whereas egoistic values did so negatively.

For different technologies respondents were asked to indicate whether they owned the technology or were planning to buy the technology for household use. **Except for smart meters, the adoption of all technologies is quite low across countries, less than 15%** (Figure 17), and the intention to buy such technologies in future was especially low in Estonia and in the Netherlands, but a bit higher in Portugal. Indeed, for almost every technology, the majority of the respondents indicated to not own the technology and also not having the intention to purchase it in the near future. One should note that the SASOS surveys were conducted mostly in 2017.

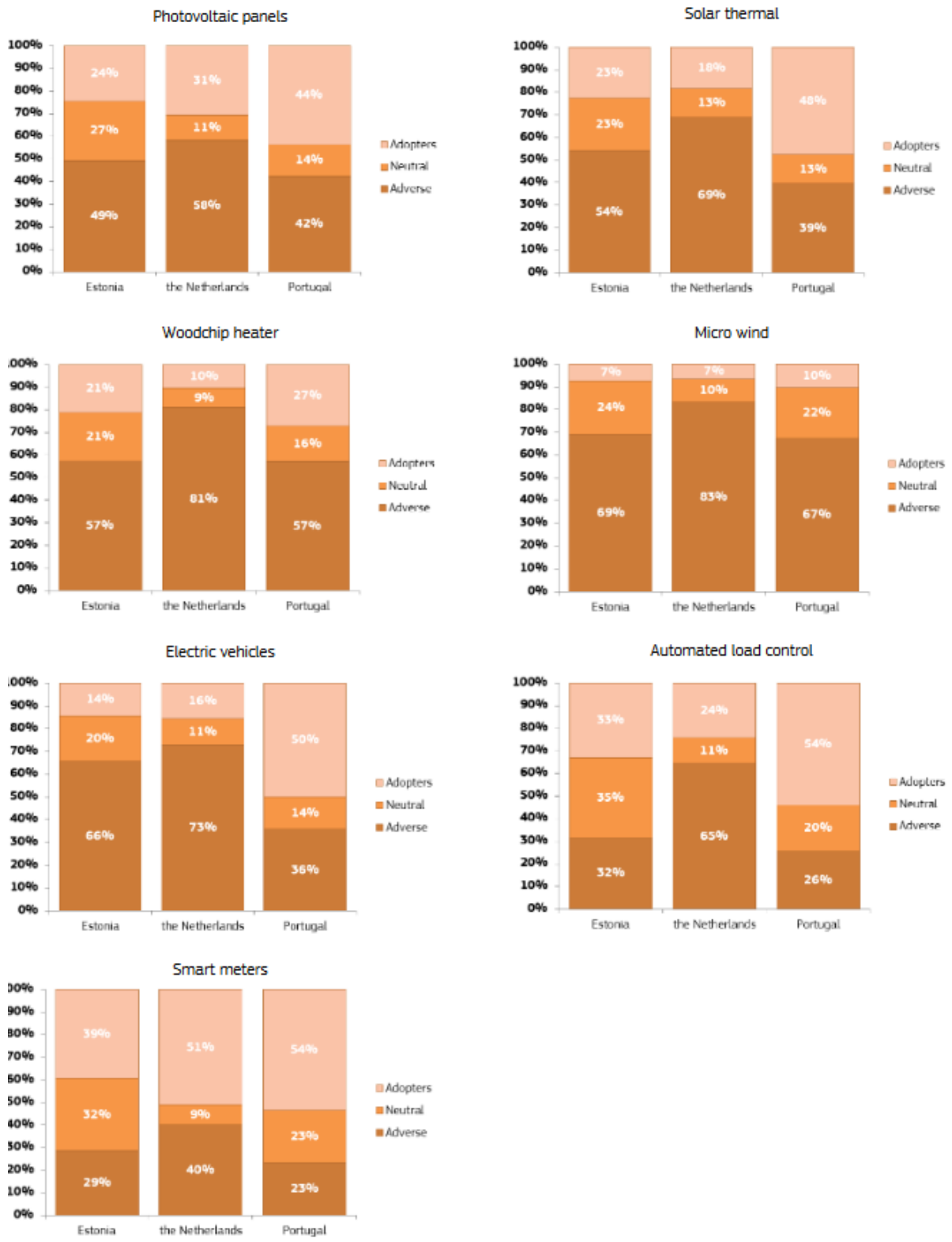


Figure 17: Attitudes of respondents toward the adoptions of energy technology options (SASOS project) [24]

As mentioned, similar surveys were conducted in Greece for the SASOS project. The first involved 1500 households completed via web and telephone interviews, and the second 750 non-residential customers included industrial, commercial and small medium enterprises participating in telephone interviews.

Households were asked to consider the most annoying inconveniences from a power cut. The most selected was damages to electrical devices, a significant result when elaborating policies for encouraging the replacement of household devices. The other inconveniences are reported in Figure 18.

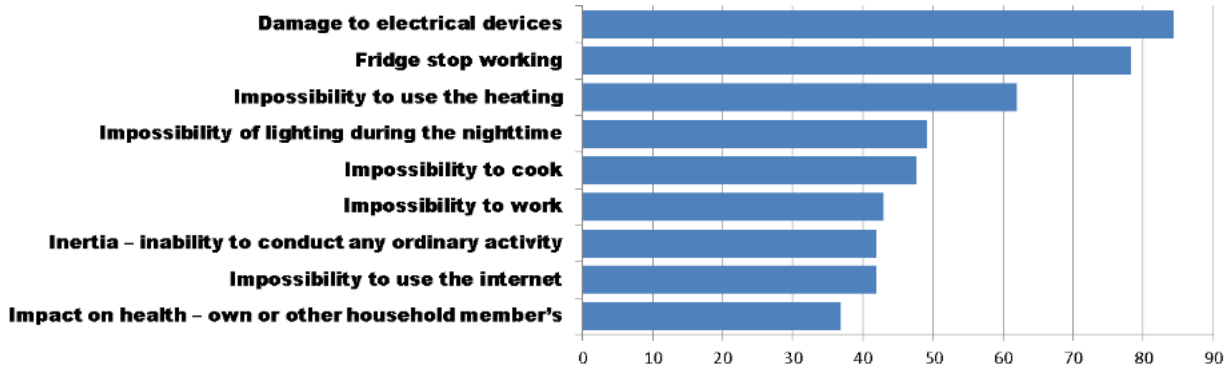


Figure 18: Worse consequences if the power cut would occur at home (% of respondents; (SASOS project)) [24]

Other interesting results of interest from the surveys conducted in Greece include:

- Over 80% of the participants in the survey support the implementation of energy-saving techniques, which include willingly reducing their own consumption, and promoting a greater utilization of renewable energy sources, as effective measures to enhance the electricity system's performance.
- Approximately 30% of the respondents expressed a positive inclination towards voluntary adjustment of their electricity consumption patterns by delaying the use of appliances during peak demand hours. In addition, a considerable percentage of approximately 50% of the participants were receptive towards automatic load control, albeit lower than expected.
- Nearly 50% of respondents agreed or tend to agree with buying energy saving technologies.
- Around 30% of survey participants exhibited a favourable inclination towards voluntarily modifying their electricity consumption patterns by deferring the usage of appliances during peak demand hours. In addition, a significant proportion of the sample, constituting approximately 50%, demonstrated a receptive attitude towards automatic load control.

Residential consumers were also asked about their potential changes in their use of electricity, which are reported in Figure 19. While the majority of the respondents agreed or completely agreed with all energy saving attitudes presented to them, one can note that the vast majority of the respondents indicated that

[24] Source: (European Commission, Joint Research Centre; Giaccaria, S; Efthimiadis, T; Bouman, T; Longo, A 2019)

they would like to reduce their energy consumption, use more renewable energies, **and adjust their electricity consumption during peak times through changes in their consumption behaviours.**

On the other hand, the least preferred attitudes regarded the (self) investment in energy storage and allowing technologies to automatically schedule the energy use (e.g., a washing machine that operates only during non-peak hours). In other words, among the options presented, the least favourable were those that would require monetary investment and installation of energy storage systems, and those where **control would be delegated to automated systems.**

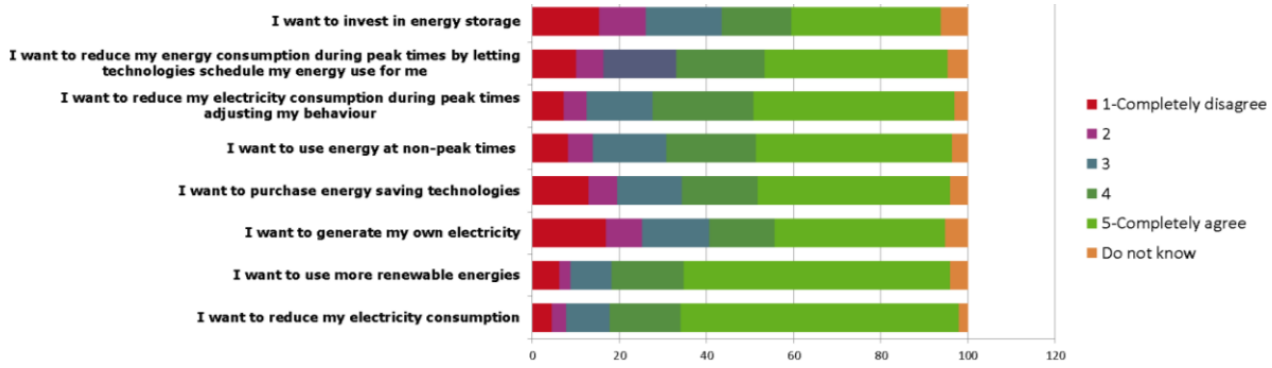


Figure 19: Attitudes toward potential changes in the use of electricity (SASOS project) [24]

4.1.3 Insights from the H2020 NUDGE project

The EVIDENT project is one of many Horizon 2020 and Horizon Europe projects studying energy behaviours and energy consumption. Some of the projects that are tackling similar issues as EVIDENT can be found on the EVIDENT website ('sister' projects, https://evident-h2020.eu/sister-projects/). An example is the project NUDGE – Nudging consumers towards energy efficiency through behavioural science (https://www.nudgeproject.eu). One activity of the NUDGE project was to conduct a survey with more than 3000 respondents from 29 European countries. The survey sample was sourced through several consumer organisations, and a Flemish panel.

From the results, the NUDGE researchers identified six profiles of energy consumers:

1. Environmentally conscious and well-informed energy consumers;
2. Concerned but comfort-oriented energy consumers;
3. Concerned but lacking awareness energy consumers;
4. Materialistic energy consumers escaping their personal responsibility;
5. Prone to social influence energy consumers; and
6. Indifferent energy consumers.

The survey revealed that the overwhelming majority of the respondents consider turning down the lights as the most common saving behaviour, with more than 90% of respondents stating that they often or always perform this action.

The NUDGE project further asked participants on their energy saving behaviours (NUDGE D1.1[25]). From the responses and given that many participants have air-conditioning (hence the high percentage of 'not applicable' in the relevant responses), the researchers conclude that the 'wearing more clothes is the most accessible behaviour in this category'. The results can be seen in Figure 20. The researchers implemented several control variables such as gender, country of residence, income, level of education, and others.

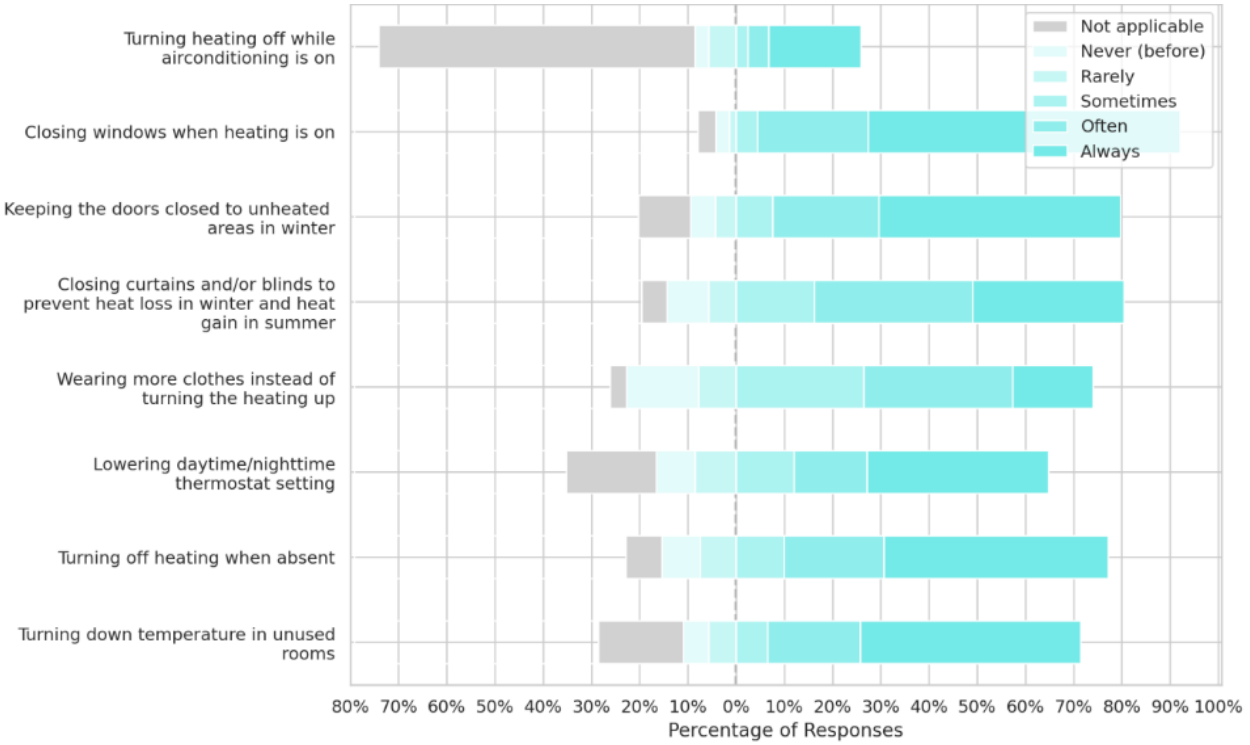


Figure 20: Saving behaviour: heating and cooling (NUDGE project)[25]

The survey also revealed that for reducing heating-related energy consumption, and developing effective measures, the two most important factors are the perceived ability to perform an activity ('perceived behavioural control') and the perceived social pressure to engage in an activity ('subjective norm') (Reinfandt and Brugger 2023). The impacts of these and other factors are presented in Figure 21.

[25] Source: NUDGE Project D1.1 https://www.nudgeproject.eu/wp-content/uploads/2021/10/D1.1-Profiling-of-energy-consumers-psychological-and-contextual-factors-of-energy-behavior-FINAL.pdf

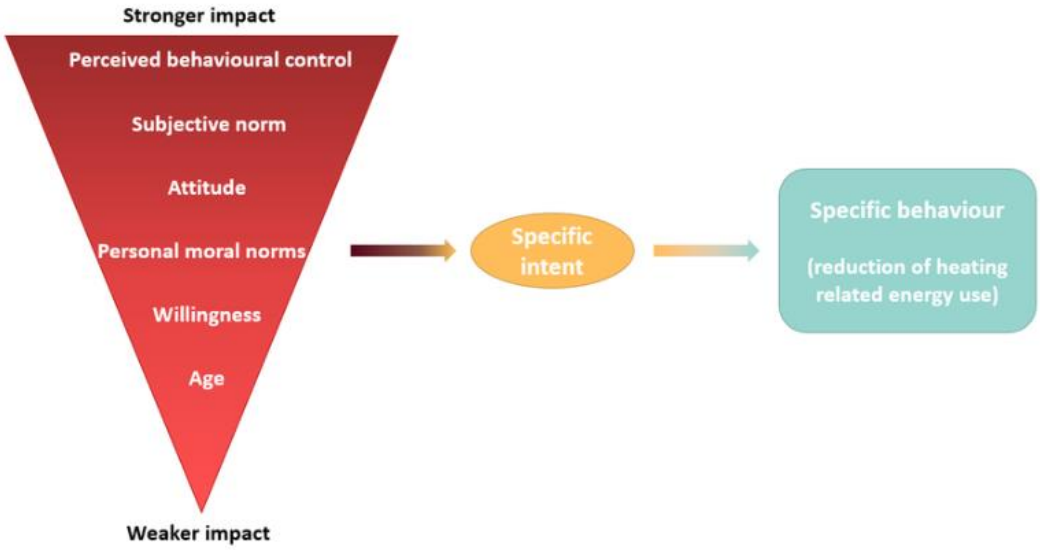


Figure 21: Factors determining people's intent to reduce heating-related energy consumption (NUDGE project)[26]

Other saving behaviours explored by the NUDGE researchers include 'water use' and 'kitchen', and the results are reported in NUDGE Deliverable 1.1.

Of particular relevance for EVIDENT's serious game[27] and the home energy reports[28], are the results from NUDGE on the use of general appliances, the willingness of respondents to share energy data with others, and their stated preferences to receive more insights on the energy consumption of their household.

For example, the respondents the majority of the respondents declared to use several energy saving behaviours such as using the washing machine only when fully loaded, using a clothes' line rather than a tumbler, etc. The reported behaviours are depicted in Figure 22.

[26] Source: (Reinfandt and Brugger 2023), URL: https://www.nudgeproject.eu/wp-content/uploads/2023/01/NUDGE-Policy-Brief-1-2.pdf
[27] Source: https://evident-h2020.eu/seriousgame/
[28] Source: https://evident-h2020.eu/the-evident-consortium-implements-its-first-field-experiment/

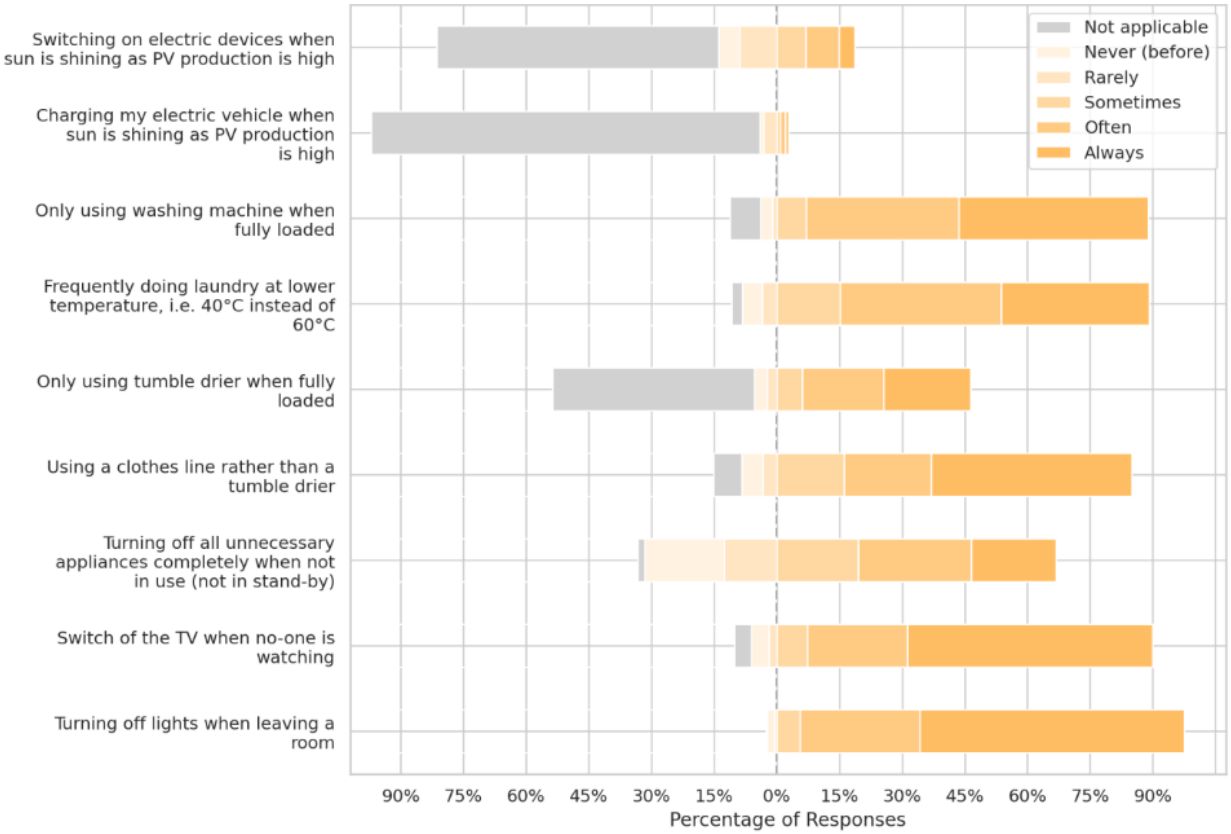


Figure 22: Saving behaviour: General appliances (NUDGE project)[25]

Regarding the willingness of the respondents to share their energy data with others, on average, 88.1% of respondents are willing to share their monthly energy data with their neighbours, energy providers and distributers, the government, and other third parties. Additionally, people seem to have the most diverse willingness to share their energy data with their neighbours. The willingness to share real-time data with neighbours decreases by almost 48% from monthly to real-time data.

The vast majority of the respondents also indicated their interest of gaining more insights on their household's energy consumption levels (Figure 23). These results are especially relevant for the EVIDENT use cases which employ home energy reports.

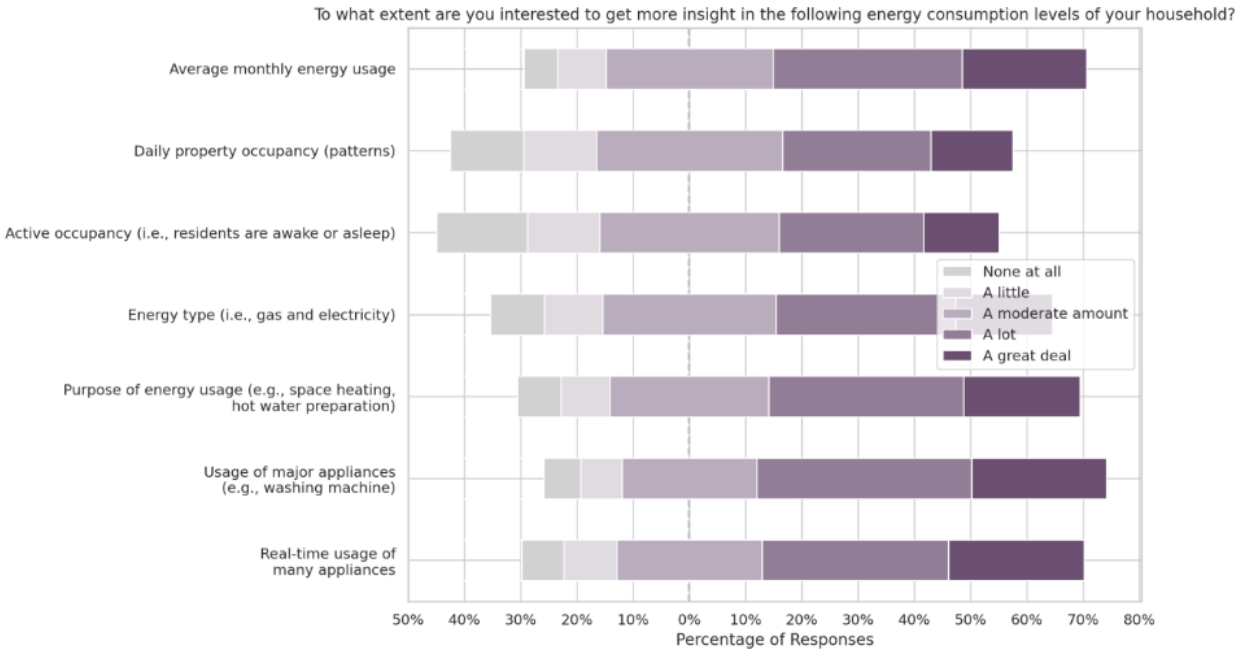


Figure 23: Stated interest to gain more insights on household energy consumption levels (NUDGE project)[29]

4.1.4 Insights from the H2020 ENCHANCE and WHY projects

Other Horizon 2020 sister projects of EVIDENT are WHY (https://www.why-h2020.eu/) and ENCHANT (https://enchant-project.eu/).

The ENCHANCT project focuses on intervention techniques to increase energy efficiency in EU households. The interventions being designed and used by the project include providing information and tips on energy conservation, delivering feedback to consumers on their energy usage habits, communicating social norms that encourage energy efficiency, employing commitment strategies, offering monetary incentives, differentiating between collective versus individual framing, or creating competition among households to encourage energy efficiency.

The WHY project also employs surveys and aims to understand how households invest resources towards the energy transition with the goal of improving energy demand modelling to forecast the domestic sector's energy consumption.

ENCHANT researchers conducted relevant surveys where they found that age often explains only a very small percentage of intent or behaviour compared to other factors[30]. For example, home ownership appears to significantly affect the intent to renovate a house regardless of the age of the home owner. However, younger persons were found to be more ambitious in the energy standards of their house

[29] Source: https://www.nudgeproject.eu/wp-content/uploads/2021/11/NUDGE-Project-presentation-Webinar-9Nov-1.pdf
[30] The results reported in this paragraph are from https://www.nudgeproject.eu/wp-content/uploads/2023/02/EUSEW-2022-NUDGE-et-al.pdf

renovations. Additionally, the researchers found that younger individuals have lower knowledge about energy efficiency (also partially confirmed by the WHY project), while they receive their information on sustainability mostly from friends, and less from TV or reading sources. On the other hand, the WHY project found that adults fear of losing comfort.

4.1.5 Energy consumption and life satisfaction

The characteristics of households' behaviours was also the central focus of a large study by (Piao and Managi 2023). The authors conducted a large-scale survey across 37 nations using a mixed approach of internet-based and face-to-face approaches. In total, the authors collected about 101,000 observations and found that energy is treated as a 'normal good', as and when household income increase, the household income spent on energy increases. Another interesting result is that energy curtailment behaviours and purchasing energy-saving electronic products will have 'a limited effect when household energy is mainly sourced from fossil fuels.' Furthermore, 'households show a positive association between household energy expenditure and life satisfaction in 27 out of 37 countries, including China, India, the United States, and Germany'.

4.1.6 Summary

In this subsection, we highlighted results from research projects that enhance those of the EVIDENT project. This overview is additional to those in other EVIDENT deliverables (e.g., D1.1 and D3.3). A common finding across research activities is that behavioural interventions can indeed lead to energy saving. However, they need to be customised to the particular audiences, one size does not fit all.

For example, in a systematic review of non-price interventions that target energy conservation behaviour in private households, (Andor and Fels 2018) found that social comparison, commitment devices, goal setting, and labelling have the potential to significantly reduce energy consumption of private households, but the effect sizes vary greatly. In particular, social comparison interventions showed the highest average effect size, followed by goal setting, commitment devices, and labelling.

The authors found that the effectiveness of each intervention depends on several factors, such as the design of the intervention, the specificity of the feedback, the delivery method, and the target audience. Therefore, the authors emphasize the importance of impact evaluations before implementing behavioural policy interventions. Also, one needs to be wary of the scalability of the different interventions, their cost-effectiveness, and their potential impact on different types of households. Another common concern regards the persistence of the savings gained, an issue that is also addressed in the EVIDENT project with follow-up surveys and other metrics.

4.2 Information Exchange and Knowledge Sharing onto other Domains

One of the main objectives of the EVIDENT project is to support the multi-domain exchange of information and link the project's empirical results and tools developed with applications to other policy domains such as water, waste and resource efficiency, food and transport. The open-code and open-data policy adopted by the EVIDENT consortium will support future research in the domain of energy efficiency and also in other domains since the analytical tools, part of the data and mainly knowledge and best practices can be easily replicated and adopted. For example, in the context of exploitation, DUTH has developed a smart

application for a gas provider in Greece. This application collects all technical knowledge of EVIDENT use cases 1 and 2, to develop a smart application that provides gas consumption feedback, peer comparison, personalized tips and an award system combined with a goal-setting system. Through this smart application, customers can be informed about their consumption and their neighbours' performance in terms of energy conservation, receive personalized tips and suggestions and win awards for being energy efficient.

Overall, to support multi-domain exchange of information the EVIDENT consortium adopted the following methodology as presented in Figure 24. In particular:

Figure 24: EVIDENT's information exchange and knowledge sharing methodology

1. **Identify the domains:** The first step is to identify the domains or areas of expertise relevant to the EVIDENT project. EVIDENT is an H2020 project that supports energy efficiency policy-making through different experiments and use cases. That means that the interventions designed and implemented in the project, the data, the results and the insights can contribute to future endeavours in different domains such as water conservation, waste management and others. This will enhance data-driven policy measures and protect the environment and the communities. For example, randomized controlled trials (RCTs) are commonly used when policy implications must be evaluated. Thus, RCTs are valuable tools that can be used in additional domains such as water waste, where behavioural interventions could play a crucial role in water conservation and raise consumers' awareness about the impact on the environment or in clinical trials where researchers evaluate the performance of a new drug. In addition, quasi-experiments play an important role in evaluating, measuring and verifying policy programs. They can be leveraged to create new links between individuals' behaviour biases and their actions upon a plethora of actions (e.g. how individuals respond to more liberal politics based on their beliefs and education, or what factors lead consumers to select reusable plastic bags in supermarkets). Another example could be the use of a serious game where participants could play and raise awareness about disabled people and their needs in means of public transportation. Finally, analytical models, such as forecasting models and causal inference models, can reveal hidden heterogeneous effects and contribute towards tailor-made policy implications. Overall, the EVIDENT project can inspire future research and enhance data-driven policymaking through its use cases, methodologies, data collected, and insights.
2. **Establish communication channels:** The second step is setting up effective communication channels to facilitate information exchange among stakeholders from different domains. This

includes establishing a communication kit[31] and participating in events and clustering activities with relevant stakeholders at national and EU levels. In the case of the EVIDENT project, this is achieved through the establishment of communication channels with sister projects[32], the participation in cross-domain events[33] and fora, the organisation of the EU Sustainable Energy Days and successful presence of the EVIDENT project in social media and other activities.

3. **Develop a common terminology space:** This step aims to develop a common terminology space where participants from different domains and expertise can use a common glossary to communicate effectively and avoid misunderstandings. For example, a dominant concept in the EVIDENT project is the concept of "nudge theory", a methodology extensively used in behavioural-based policy interventions but rather unknown to technical stakeholders. In the EVIDENT project, this approach was followed through the beginning of the project. In particular, *D1.1 "Analysis of best practices"*[34] was drafted to provide the reader with information regarding the current status of the behavioural-based interventions domain, presenting novel insights into the methods and toolkits to be developed within EVIDENT. In addition, several blog posts were written throughout the project's lifetime to present parts of EVIDENT's theory, methodologies and results in a simplified manner[35] and set a common ground for further communication and collaboration between the consortium partners and also external stakeholders, communities and policymakers.
4. **Encourage knowledge sharing:** This step fosters a culture of knowledge sharing. The EVIDENT project is a research and innovation action (RIA) that supports and encourages open science practices to enhance reproducible, transparent and credible research. In addition, through the adaptation of an open-data/open-code policy, the EVIDENT consortium aims to maximize public disclosure and engagement by giving the opportunity to policymakers, governmental bodies, regulators, citizens, academics and industry organizations and others to leverage our assessments, build upon them and reuse them to build additional agendas towards behavioural-based policy interventions. In addition, the EVIDENT project contributes towards open data research and raises awareness about open science practices by integrating the Zenodo API into the EVIDENT platform. The EVIDENT platform users are welcome to leverage the provided service to upload their research data to the Zenodo repository.
5. **Define data integration and interoperability standards:** This step aims to define and/or follow a data integration and interoperability standard to ensure that data from different domains can be combined, exchanged, and analysed efficiently. This involves establishing data formats, naming conventions, metadata standards, and data-sharing protocols. EVIDENT project's flexible architecture, presented in section 3 foresees these data integration and interoperability needs.

[31] More information about EVIDENT's communication kit can be found in D7.1, D7.2 and D7.3.
[32] More information about EVIDENT's sister projects can be found here.
[33] For example, the ELECTRON International event in Baku, 5-7 Dec. 2022 where the EVIDENT project was presented.
[34] EVIDENT D1.1 "Analysis of best practices" can be found here.
[35] Some examples of relevant EVIDENT blog posts: 1, 2, 3, 4, 5

Since the data used in the EVIDENT project are provided or collected by different partners (e.g. CW provides energy measurements, and SID undertakes the collection of use cases 4 and 5 replies through the EVIDENT platform), a common standard is adopted so the data can be easily exchanged between the project's consortium. The same principles are also followed when data sharing is of key importance. Project's data are stored in a way that can be efficiently retrieved by the EVIDENT partners and can be easily shared after removing any field that may contain personal and private information. The data schema for each dataset is presented in section 2.

6. **Facilitate cross-domain collaboration:** This step encourages collaboration and teamwork across different domains. This involves discussions and brainstorming sessions to explore innovative approaches and solutions and organising workshops with relevant stakeholders in the different aspects of policy-making domains. More specifically, through "exploratory workshops", ideas are analysed to gain a deeper understanding of a subject, including its associated issues, existing solutions, and future obstacles. In addition, "creating workshops" will bring together stakeholders to collaborate and foster cooperation within a common or intersecting interest to solve particular problems. The previous workshop brings closer professionals from different domains, such as scientists, policymakers and officials from different academic, industry, Nongovernmental Organizations (NGOs) and government organisations.
7. **Seek expert input:** Experts are encouraged to participate in the above activities to provide valuable insights, ensure accuracy, and enhance the quality of the project's outcomes. In the EVIDENT project case, the consortium has formulated an external advisory board consisting of members with different backgrounds. This background diversity will enable the EVIDENT consortium to receive constructive feedback and valuable remarks on the design of the econometric models and better define the processes for assessing the consumers' behaviour and the design and implementation of the EVIDENT use cases.
8. **Monitor and evaluate progress:** The monitoring and evaluating progress demands and holistic approach where the effectiveness of the information exchange processes is evaluated by members to identify any challenges or areas for improvement. Based on the challenges that may arise, the communication strategies or processes may update to ensure effective multi-domain information exchange. All monitoring and evaluation processes within the EVIDENT project are subject to two axes, the qualitative and the qualitative, to provide a comprehensive understanding of the potential action performance, outcome, and impact.

The EVIDENT project aims to facilitate the exchange of information across multiple domains and connect its empirical findings and tools with other policy areas. The project's open-code and open-data policy enables future research in energy efficiency and other domains, as the analytical tools, data, knowledge, and best practices can be easily replicated and adopted.

5. Conclusion

This deliverable, *D3.3 "Data collection and management"*, finalizes the tasks of WP3 "Intervention preparation and execution". The tasks of WP3 start with the preparation actions and the design of EVIDENT interventions and pilots. This includes the preparatory actions for the natural field experiments, the participant recruitment recommendations for the three quasi experiments and actions supporting the development of the EVIDENT serious game and its associated surveys. The progression involves the implementation of the randomized control trials, surveys and serious game, concluding with the data collection and management processes as the final step before the data analysis that takes place in WP4.

Apart from the three deliverables crafted under the umbrella of WP3, numerous out-of-papers actions (discussions, suggestions from external members, literature review, etc.) and processes (preparatory actions, recruitment recommendations, demographic questionnaires, etc.) took place from the beginning of this work package, fostering a successful collaboration among the EVIDENT partners and fulfilling the demanding activities of data collection.

This deliverable presents the schema of the corresponding datasets leveraged for the preparation and the implementation of all EVIDENT use cases. It also provides information about the data services developed to organise and assist the data management and data analysis from the analytical WP4. Additionally, the deliverable presents data and valuable insights from relevant previous research, projects and surveys and provides useful roadmap to achieve information exchange and knowledge sharing onto other domains.

6. References

Andor, Mark A.,, and Katja M. Fels. 2018. "Behavioral economics and energy conservation–a systematic review of non-price interventions and their causal effects." *Ecological economics* 148: 178-210. https://www.sciencedirect.com/science/article/pii/S092180091731039X?via%3Dihub.

Bertoldi, Paolo, and Rocco Mosconi. 2020. "Do energy efficiency policies save energy? A new approach based on energy policy indicators (in the EU Member States)." Edited by Elsevier. *Energy Policy* 139 (C). doi:https://doi.org/10.1016/j.enpol.2020.111320 .

Cornago, Elisabetta. 2021. *The Potential of Behavioural Interventions for Optimising Energy Use at Home.* 4 June. Accessed 2023. https://www.iea.org/articles/the-potential-of-behavioural-interventions-for-optimising-energy-use-at-home.

Delemere, Emma, and Paul Liston. 2022. "Exploring the Use of Behavioural Techniques in Serious Games for Energy Efficiency: a Systematic Review and Content Analysis." *Behavior and Social Issues* 31: 451-479. doi:10.1007/s42822-022-00103-4.

Della Vigna, Stefano, and Malmendier Ulrike. 2006. "Paying not to go to the gym." *American Economic Review* 96 (3): 694-719.

European Commission, Joint Research Centre; Bouman, T; Longo, A; Giaccaria, S; Efthimiadis, T. 2019. "Societal appreciation of energy security. Volume 1, Value of lost load – households (EE, NL and PT)." Publication Office. doi:https://data.europa.eu/doi/10.2760/139585.

European Commission, Joint Research Centre; Efthimiadis, T; Giaccaria, S; Longo, A; Bouman, T. 2019. "ocietal appreciation of energy security. Volume 3, Non-household actors (EE, NL and PT)." Publication Office. doi:https://data.europa.eu/doi/10.2760/672568.

European Commission, Joint Research Centre; Efthimiadis, T; Giaccaria, S; Longo, A; Bouman, T. 2019. "Societal appreciation of energy security. Volume 2, Long-term security (EE, NL and PT)." Publications Office. doi:https://data.europa.eu/doi/10.2760/05174.

European Commission, Joint Research Centre; Giaccaria, S; Efthimiadis, T; Bouman, T; Longo, A. 2019. "Societal appreciation of energy security. Volume 4, Value of lost load - Greece." Publication Office. doi:https://data.europa.eu/doi/10.2760/823702.

Han, Xiao, and Chu Wei. 2021. "Household energy consumption: state of the art, research gaps, and future prospects." *Environment Development and Sustainability.* 23: 12479–12504. doi:10.1007/s10668-020-01179-x.

IEA. 2021. *Energy savings across behavioural interventions implemented in the United States.* 4 June. Accessed 2023. https://www.iea.org/data-and-statistics/charts/energy-savings-across-behavioural-interventions-implemented-in-the-united-states.

Nives, Della Valle, and Bertoldi Paolo. 2021. "Mobilizing citizens to invest in energy efficiency." doi:10.2760/137315.

Piao, X., and S. Managi. 2023. "Household energy-saving behavior, its consumption, and life satisfaction in 37 countries." *Scientific Reports* 13. doi:https://doi.org/10.1038/s41598-023-28368-8 .

Reinfandt, N., and H. Brugger. 2023. "NUDGE Policy brief #1 – Profiling and nudging energy consumers to heat efficiently." https://www.nudgeproject.eu/wp-content/uploads/2023/01/NUDGE-Policy-Brief-1-2.pdf.

Sussman, Reuven, and Maxine Chikumbo. 2016. *Behavior Change Programs: Status and Impact.* Washington, DC: American Council for an Energy-Efficient Economy.

Zangheri, Paolo, Tiago Serrenho, and Paolo Bertoldi. 2019. "Energy Savings from Feedback Systems: A Meta-Studies' Review." *Energies* 12 (19): 3788. doi:https://doi.org/10.3390/en12193788.