

D3.2: Ethical and Data Protection Methodology of Dataset Management

WP3 - Re-analysis of Existing Data

Report No. D3.2 / Date: 30/09/2021



AUTHOR(S) NAME

Alim Nayum, Christian A. Klöckner, Anca Sinea, Melania Lese and with input from all consortium partners

Norwegian University of Science and Technology, Norway



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957115.

ENCHANT Report

D3.2: Ethical and Data Protection Methodology of Dataset Management

VERSION: 01 / DATE: 30.09.2021

AUTHOR(S)

Alim Nayum (NTNU), Christian A. Klöckner (NTNU), Anca Sinea (UBB), Melania Lese (UBB)
and with input from all consortium partners

Quality assurance: All project partners

PROJECT NO.: 957115 (H2020) / PAGES/ANNEXES: 14/1

ABSTRACT

The Ethical and Data Protection Methodology of Dataset Management provides methodological guidelines for handling already existing external data. It ensures that the outlined ethical requirements (see D1.3) and key regulatory aspects within the European and national legislation are followed also for these datasets.

REPORT NO.: D3.2

ISBN: NA

CLASSIFICATION: Public

CLASSIFICATION THIS PAGE: Public



DOCUMENT HISTORY:

VERSION	DATE	VERSION DESCRIPTION
1	25.09.2021	First version for a quality check sent to partners
2	29.09.2021	The final version sent to the European Commission



Table of contents

1.	Introduction and Overview	5
1.1	Collection and use of secondary datasets in ENCHANT	5
1.2	Purpose and scope of this document.....	5
2.	Procedure for collection of secondary data	6
2.1.	Data sources.....	6
2.2.	Data collection.....	7
2.3.	Ethical and privacy issues.....	7
3.	Secondary Database Management	8
3.1	Use of secondary data	8
3.2	Data documentation	8
3.3	Data storage and curation.....	8
4.	Confidentiality Agreements	9
4.1	Licensing	9
4.2	Protection of personal data.....	9
4.3	Confidentiality of secondary data.....	10
	ANNEXES.....	11
	Annex I: Data documentation templates.....	12



1. Introduction and Overview

1.1 Collection and use of secondary datasets in ENCHANT

ENCHANT's intervention tools aim to increase energy efficiency behaviour among millions of European residents through a contextual and multi-disciplinary approach. A deeper understanding of secondary data, which has already been collected, is therefore important for designing intervention packages aiming to bring desired outcomes. However, there are significant concerns regarding secondary data, such as to which extent it is reliable, accurate, statistically significant, or replicable, that need to be addressed. Although datasets from previous studies on people's energy behaviour do exist and can provide useful insights, methodological concerns can be raised about sources, categories, and objectives of data collection. Municipalities, NGOs, energy providers, and academic researchers have different purposes for similar data collection and hence differ largely in their deployment of data collection methods and strategy. Since the purpose/goal of the data collection steers data collection methodology, it is hard to use datasets compiled for different goals for comparative analytical purposes. Besides, while some data collection activities are systematically conducted at the national level, others are only done through local smaller-scale research projects. Moreover, data heterogeneity exists across secondary datasets, and this is particularly evident regarding behavioural analysis of energy efficiency. This data heterogeneity makes structured comparison difficult.

Adding to the methodological concerns, are the ethical and licensing aspects of existing external datasets. ENCHANT primarily makes use of secondary data sources that are provided by the user partners or are publicly available. All these secondary data sources contain already anonymized datasets, which will not allow identifying individuals directly or indirectly. Nevertheless, ENCHANT seeks to handle all external datasets and subsequent data analysis in accordance with the European and national legislation.

1.2 Purpose and scope of this document

This Ethical and Data Protection Methodology of Dataset Management (D3.2) aims to describe the procedure for collection of relevant secondary data through user-partners and from previous projects, establishing and managing a database of diverse sets of data, confidentiality agreements with the respective stakeholders, and confirmation that the processes comply with national and EU legislation. It should be noted that a comprehensive framework for collection and handling of the primary and secondary data used in the project, and relevant ethical requirements are described in ENCHANT's D1.2 and D1.3. Thus, this deliverable (i.e., D3.2) provides the most relevant aspects of management and ethical issues focusing on external secondary data used in the project.



2. Procedure for collection of secondary data

Supplemented with collecting and re-analysing already existing secondary data, ENCHANT will test the effects of the intervention packages in real-life settings with the objective to unlock an energy efficiency potential among the general public through behavioural change. Identification and sourcing of already existing secondary data are important for designing ENCHANT's intervention matrix and Randomised Control Trial (RCT) protocol.

2.1. Data sources

ENCHANT will conduct search and synthesis for already existing secondary data from various sources. Table 1 gives an overview of data sources and how the secondary data are identified and collected. In most cases, these already existing secondary data are open access unless the owners of the document/data have made explicit restrictions on them.

Table 1: Data sources and collections

Source of data	How data is collected
Published studies, reviews, meta-analyses from both user-partners and previous projects	Relevant bodies of literature are identified by conducting a comprehensive review of existing behavioural interventions through published studies, reviews, and meta-analytical syntheses. Search engines (e.g., Google Scholar) and databases (e.g., the European Commission's document database, Web of Science, Scopus) will be used to identify the relevant literature. Furthermore, user partners are asked for available internal reports.
Documents published by relevant stakeholders, such as policymakers and regulators, NGOs, and professional organizations	Relevant bodies of documents and data from previous studies and research projects about interventions, intervention evaluations/ assessments, plans of action and periodical reports implemented in interventions, and other relevant documents in previous interventions are located through direct contact with the user partners and search engines (e.g., Google Scholar) and databases (e.g., European Commission's document database, Web of Science, Scopus).
Data from both user-partners and previous projects on energy efficiency and consumer behaviour	Data on the topic of energy efficiency and consumer behaviour from user-partners and previous projects are collected through a chain referral method, based on enquiries of experts at the national and European levels.
Existing data sets, and studies on the topic of energy efficiency and consumer behaviour	An inventory of all relevant data sets is made, and a data review is conducted to integrate and process the disparate empirical sources



2.2. Data collection

To design and test intervention packages, ENCHANT will collect relevant data on the topic of energy efficiency and consumer behaviour from case cities and previous projects. Based on the consolidated data set of previous projects, the re-analysis of the secondary data will necessitate efforts to integrate and process the disparate empirical sources, and then to deploy a descriptive and inferential statistical analysis. To this end, it will build upon and feed into the literature review performed by WP2. The curation procedures of the data sets will follow two main steps: first, the data repository from previous studies and projects will be catalogued and processed to be operational and accessible. Second, the descriptive and inferential analysis will inform a data review of energy consumption behaviour to date presented as a meta-analysis report. In order to achieve a consolidated dataset from ENCHANT interventions, data collection within the consortium, data sorting and processing, and data analysis will be performed.

2.3. Ethical and privacy issues

No important ethical issues regarding secondary data collection were identified. In the process of data collection, personal data will not be collected without anonymization prior to accessing the datasets. Also, an important aspect of the secondary data collection is that most data accessed by the research team is open data provided in already anonymized form.



3. Secondary Database Management

ENCHANT is a project that also depends on already existing secondary data provided by case cities and other sources (e.g., NGOs, municipalities, energy providers). This makes it necessary to define procedures for how data access rights for secondary data are achieved and how that data is used and matched with primary data.

3.1 Use of secondary data

ENCHANT will not only collect new data but also reanalyse secondary data sources, such as datasets provided by the user partners. These secondary data sources either contain already anonymised datasets or the datasets will be anonymised by the user partners before analysis, in such a way that will not allow identifying individuals directly or indirectly as a default. In cases where past energy use data will be accessed on a personal level, the respective households need to give consent to that. Data will then be pseudonymised for matching with primary data recorded in the project.

3.2 Data documentation

All data files provided by ENCHANT include metadata on the content of the data file and the context in which the data was collected. It is important to ensure the usefulness of the data for researchers and analysts not foreseen in the data collection. This metafile shall include metadata concerning the kind of data included, involved personnel, date and duration of the data collection, variable names/labels, recruiting procedures, response rates, whether or not it is anonymised, related WPs and tasks, and finally a summary. Templates for such metafiles are provided in Annex I.

3.3 Data storage and curation

All collected data shall include a metafile when stored on ENCHANT's secure storage solution and/or ENCHANT's SharePoint server accessible through the Microsoft TEAMS® frontend. The file will after the end of ENCHANT be made available for external users of the data. Data will be stored and protected on ENCHANT's encrypted server space for secure data storage. WP1 (NTNU) and WP3 (UBB) are responsible for the curation of all data collected in ENCHANT and its safe storage. The storage solutions for personal raw data and anonymised data include daily backup routines to prevent data loss. (Aggregated) data underlying project publications will be made available as supplemental information, in a thematic repository or on the Zenodo platform, and will be mentioned in the main text of the publication.



4. Confidentiality Agreements

ENCHANT will use already existing secondary data that are either publicly available, obtained through the project's user partners or from a previous research project. In case of publicly available data, ENCHANT will provide details of the source(s) and confirm that the data are openly and publicly accessible and may be used for research purposes. Regarding the use of personal data that were collected from a previous research project, ENCHANT will provide details regarding the initial data collection, methodology and informed consent procedure. ENCHANT will also obtain permission from the owner/manager of the dataset(s) to use the data in the project. When using personal data provided by the project's user partners or a third party and the data subjects have not expressly consented to its use in research projects, ENCHANT will, in accordance with the GDPR, inform them that data are acquired for the research project and what ENCHANT will be using the data for¹.

4.1 Licensing

During collection, re-analysis, and storage of secondary data, ENCHANT makes sure that licenses and contracts for use of secondary data will be obtained, specifying their use in relation to the GDPR. All non-public data collected will be transferred only after the elaboration of GDPR procedures is implied. An agreement regarding the compliance of these GDPR procedures will be signed in every case. Before the transfer of data, the parties involved will make sure every possibility of data transfer with anonymization was considered.

4.2 Protection of personal data

ENCHANT considers the protection of personal data and privacy a priority and will fully safeguard the rights of our participants and data owners. All secondary data sources shall contain already anonymised datasets or will be anonymised by the user partners before analysis. For example, where past energy use data will be accessed on a personal / household level, the data will be pseudonymised for matching with primary data recorded in the project.

Energy consumption data from secondary data sets will be connected to survey data through a pseudonymised key table. Keys to pseudonyms are stored in a separate location from the data. In cases where past energy use data will be accessed on a personal level, the respective households need to give consent to that. Data will then be pseudonymised for matching with primary data recorded in the project. Each partner in the Consortium is responsible for declaring that all their actions are according to national

¹ European Commission (2018). Ethics and data protection. 14 November 2018. pp. 12-13.



and EU legislation. Only the researcher responsible for the specific ENCHANT dataset to be linked to other previous/existing datasets will have access to the key table.

Research in ENCHANT also involves observation and data processing methods such as audio, video, photos, and field notes but only after informed consent is provided by the participant. Prior to conducting methods that include such data, the Data Protection Officer for the task will be responsible for assessing the data protection impact according to Article 35 of the GDPR.

The informed consent form must contain terms regarding the specific data processing methodology that will be implemented. In cases where data is from a database, permission from the owner of the data set should be provided prior to data processing.

In the case of publicly available data, written permission by the owner of the data set will be obtained.

4.3 Confidentiality of secondary data

ENCHANT does not have any contribution concerning secondary data unless it involves data from partners or other third parties and they have certain commercial implications. In such a case, ENCHANT's respective research team signs agreements with the entities entrusting data to ENCHANT. In general, ENCHANT will make sure that the secondary data is also anonymized and handled on the principles for primary data described above.



ANNEXES



Annexe I: Data documentation templates

The following two templates shall be used to document the necessary background of the data files for internal and external use in ENCHANT.

- 1) Data documentation template for qualitative data in ENCHANT
- 2) Data documentation template for quantitative data in ENCHANT



1) Data documentation template for qualitative data in ENCHANT

Name of the data set: _____ Date the data set was finalized:

Date/time period the data was collected: _____ to _____.

Responsible partner for the collection of the data:

_____ (name) _____ (institution)

Data produced in WP: _____ Task: _____

Data anonymised on (date): _____ by _____

Information about the participants:

Number: _____ Age: _____ Sex: _____

Participants' background: _____

Recruitment procedure: _____

Original language of the material: _____

Data collected by (interviewer): _____

Transcribed by: _____

Transcription rules: _____

Translated to English by: _____

Ethically cleared by: _____ on (date): _____

Interview guidelines (or the like): _____

Size of the data (e.g., number of words): _____

Short summary: _____



2) Data documentation template for quantitative data in ENCHANT

Name of the data set: _____ Date the data set was finalized: _____

Date/time period the data was collected: _____ to _____.

Responsible partner for the collection of the data:

_____ (name) _____ (institution)

Data produced in WP: _____ Task: _____

Data anonymised on (date): _____ by _____

Information about the participants:

Number: _____ Age: _____ Sex: _____

Participants' representative for which population: _____

Recruitment procedure: _____

Response rate: _____

Original language of the material: _____

Translated to English by: _____

Ethically cleared by: _____ on (date): _____

Variables in the dataset:

Variable name	Variable type	Variable label	Answering format/value labels	Comments

Variable types:

- T = text
- D = date / time
- B = binary / dichotomous
- C = categorical
- O = ordered categorical / ordinal
- I = interval / ratio / Likert scales with 5 or more categories

Short summary: _____

