



DATA MANAGEMENT PLAN.v2

Alexandros Charalambides

March 2018



This project has received funding from the European Union's Horizon 2020 programme under Grant Agreement No 695873

EXECUTIVE SUMMARY

ENERFUND is a tool that will rate and score deep renovation opportunities – like a credit score used by banks to rate clients. The tool will be based on a methodology to be developed and on a set of parameters such as EPC data, number of certified installers, governmental schemes running, etc. By providing a rating for deep renovation opportunities – whether for private establishments or for public buildings – funding institutes can provide targeted loans, retrofit companies can identify sound opportunities, municipalities can promote targeted incentives and the public's trust for retrofitting will be enhanced. The strategic, long term aim is to provide a unified global tool for the promotion, not only of deep renovation of buildings, but of all energy related activities in the building sector.

ENERFUND will be designed with the user in mind and will comprise of (a) a database of building related data, (b) links to open-data sources, (c) a Geographic Information System (GIS)-based map and (d) dedicated login-page for specific type of users to host their own data on the ENERFUND tool.

The academic partners will contribute know-how on large-scale data processing, numerical methods and energy data analysis, the SME partners will primarily offer expertise in production-grade software development, on user experience assessment and large-scale data processing, while the remaining partners will provide inputs on the on the user experience assessment.

The aim of this Data Management Plan (DMP) is to specify which data will be openly accessible and to describe the data management life cycle for all datasets to be collected, processed or generated by ENERFUND.

Table of Contents

EXECUTIVE SUMMARY	2
DESCRIPTION OF THE DAT.....	4
DATA COLLECTION AND GENERATION	4
DATA MANAGEMENT, DOCUMENTATION AND CURATION.....	5
DATA SECURITY AND CONFIDENTIALITY OF POTENTIALLY DISCLOSIVE INFORMATION.....	6
DATA SHARING AND ACCESS.....	6
PROCESSING OF PERSONAL DATA.....	7
RESPONSIBILITIES	8
COPYRIGHT AND INTELLECTUAL PROPERTY	8
PROJECT DETAILS:.....	9

DESCRIPTION OF THE DAT

ENERFUND will leverage the latest advances in open-data databases, Geographic Information Systems and communication protocols to design and implementation of a decision making tool for the funding of energy retrofit of buildings. The ENERFUND user-centric architecture will consist of three distinct services, which can be combined to form an effective tool to be used by interested stakeholders:

- A decision making algorithm that will access the data available in the database and provide the user with intelligence on the opportunity or not to fund the retrofit of a specific building.
- An online map that will display energy related data of buildings, as well as other related data, both at an individual building level or at an aggregated level (i.e. district level).
- A data management tool that will enable users (i.e. municipalities) to monitor and analyse the energy performance of the buildings in their jurisdiction (i.e. statistical analysis).

Quantitative data will be collected and generated for and from the ENERFUND project. Initially, the project will try to identify the magnitude of the problem and to extract requirements for our solution, using publicly-available, easily-accessible data. Each data entry will contain information such as the size of the building, its energy rating, its heating system, its geolocation, etc. This dataset will serve as a starting point to create a machine learning algorithm, that based on the analysis of past projects, questionnaires, etc, will be able to identify buildings (or location) of sound funding opportunities for energy retrofitting. Finally, the tool will be assessed with user studies during workshops and piloting activities and will be released to the public.

DATA COLLECTION AND GENERATION

For the population of the ENERFUND database, there are two types of data to be collected and/or generated: (a) Energy Performance Certificate (EPC) data and (b) other related data to be identified through WP2 and WP3 that will affect the decision making process for the energy retrofitting of buildings. With regard to EPCs data, the initial database, will be populated from other external sources where the data can be extracted from online APIs (i.e. the Italian EPC repository), or where a complete online database exist (i.e. the Danish EPC repository), and where the data can, or is, geocoded so that it can be displayed on the

ENERFUND map. Secondly, EPC repositories that cannot automatically imported into the ENERFUND database, but are geocoded, will be accessed at a secondary level, while finally a functionality will enable users to input manually their own EPC data into the users ENERFUND database. In a similar manner other related data (i.e. noise level maps, economic activity, transparency index, etc) will also be used in a similar manner as with the EPC data. All inputted data will be filtered as to identify any erroneous data that might be present. A complete list of EPC external datasets and other open-data databases that will be used will be available through the WP4 deliverables.

The size of the database will depend upon the number of data inputted and the level of sophistication of the developed algorithm. Thus, for the proof of concept and the validation of the results, a program pilot will be performed in a specific region/country for a defined period of time. Such a developed database will enable the ENERFUND project and the consortium to build a large dataset of information, which will be invaluable for the current grant proposal, future proposals and for work with collaborators (i.e. for a pan-European EPC database across Europe).

Data quality and standards will be met through regular data quality checks, and automatic standardization checks. Data collection and generation will be standardised using appropriate software thus avoiding errors through manual input of data. Consistency and reproducibility of the data generation, the machine learning algorithms and any data collected will be maintained through peer review and cross checking the results with existing measurements.

DATA MANAGEMENT, DOCUMENTATION AND CURATION

During the lifetime of the project, when data collection or generation is performed, the data are straightaway backed-up onto ENERMAP's servers Any sensitive information stored will be encrypted using the Advanced Encryption Standard (AES) which is the algorithm trusted by numerous organizations. Any sensitive information that needs to be transmitted electronically will always be transmitted through a secure connection, and the data will always be encrypted using AES. Only authorised users will be able to access the sensitive information by using the appropriate software and decryption keys. No sensitive information will be made available to external users.

After the lifetime of the project, the Cyprus University of Technology will provide a long-term, reliable and secure repository for the data. The Cyprus University of Technology will provide sufficient backup of all data at multiple offsite locations. Data will be stored for at least 10 years after the completion of the project. Additionally, any data selected for publications or selected to be stored in a research repository will also be publicly available for 20 years.

A detailed documentation of the structure, annotation and description of data will be kept by all partners of the project. This will include detail representation of the methods used to collect, and analyse the data. Additionally, detailed documentation will be generated of all modules that lead to the creation and generation of the machine learning algorithms, and browser add-ons. The procedures for software development, data collection, and data generation will also be documented. Detailed documentation will also be produced for the public. All methodologies used to generate, collect and generate any data will be represented as modular operating procedures in an open file format.

DATA SECURITY AND CONFIDENTIALITY OF POTENTIALLY DISCLOSIVE INFORMATION

No personal data will be included in our data and therefore, there is no risk of disclosure of personal information. Furthermore, as the collected and curated data are going to be openly available and shared with the public via a dedicated repository through ENERMAP (during the duration of the project) and through Cyprus University of Technology (after the completion of the project), no data security risks exist. It should be noted that the Project partners will take all the necessary permission's and fail-safe procedures to ensure that data collected fully anonymised, and cannot be traced back to the user. All sensitive information will be encrypted using a state-of-the-art encryption algorithm. Finally, no sensitive information will be stored.

DATA SHARING AND ACCESS

All collected and generated anonymised data will be suitable for sharing, as prescribed by EU regulations and standards (i.e. the INSPIRE directive). No sensitive data will be stored or shared. The public and new users can discover about the ENERFUND collected and generated data through organised workshops, public awareness activities, Social Media posts, project website

and publications in Journal and conferences. Data will be publicly available through a dedicated repository, through ENERMAP published under a permissive re-use license like Open Data Commons Open Database License (ODbL). No unreasonable delays to data sharing are expected, as no Intellectual Property Rights (IPR) emerging from the data as such is foreseen.

However, no data will be shared with companies who do not abide with EU regulations on open data handling and that may exploit the data in unlawful ways.

PROCESSING OF PERSONAL DATA

Personal data will neither be stored on the ENERFUND database nor displayed on the ENERFUND map. However, all data sources (i.e. competent authorities, energy agencies, other open data repositories, will be identified on the website), with a disclaimer that ENERFUND partners do not interfere with the data provided with them, and that a user may request the removal of the data of a building under his/her ownership through the owner of the dataset. Since the ENERFUND dataset will be updated every 6months, all changes to the original EPC datasets will be also be reflected in the ENERFUND dataset. A “Remove/Alter” form will also be added so that users can request the removal or correction of their data, and the Project Coordinator, through the relevant partner, will inform the database owner for the request.

When collecting the relevant EPC data from the various sources, either the partner informed the database owner for the use of the data and their consent was provided (in case the data was not already easily publicly available) or the repository’s regulation was abided to (in case the data was already easily publicly available).

Furthermore, only EA and ENERMAP have access to the stored data on the website, and no other project partner may alter or interfere with the stored data,

RESPONSIBILITIES

The Project Coordinator (Dr Alexandros Charalambides) will have the overall responsibility for implementing the Data Management Plan with the support of the WP4 Task Leaders. Further, each partner of the ENERFUND project will be responsible for verifying that the data provided are accurate and records are up to date. The Technical Committee will ensure that all the appropriate protocols are in place, for storing and backing up the collected, generated and curated data. Furthermore, the lead partners of WP4 will be responsible that the appropriate permissions are assigned to each user.

COPYRIGHT AND INTELLECTUAL PROPERTY

The intellectual property will be jointly owned between the institutions of the ENERFUND partners. If a project partner decides to move institutions during the duration of the project, the Institution to which they move would not become a joint owner, and the ownership will remain with the institution at which partners were originally based.

PROJECT DETAILS:

Website: www.enerfund.eu

Twitter: [@enerfund](https://twitter.com/enerfund)

Facebook: [/enerfund](https://www.facebook.com/enerfund)

Disclaimer:

The sole responsibility for the content of this publication lies with the authors. It does not necessarily reflect the opinion of the European Union. Neither the EASME nor the European Commission are responsible for any use that may be made of the information contained therein.