DELIVERABLE 7.10: DATA MANAGEMENT PLAN

E. Wilczynski, S. Pezzutto, A. Novelli, P. Zambelli (EURAC), A. Bardi (OpenAIRE)

Revised by
D. von Gunten (CREM)

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

enermaps.eu

Table of Contents

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

List of figures

List of tables

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

List of abbreviations

FAIR	Findable, accessible, interoperable and re-usable
RIA	Research and Innovation Action
GDPR	General Data Protection Regulation
IPR	Intellectual property rights
DOI	Digital Object Identifier
EEA	European Environment Agency
ESPON	European Spatial Planning Observation Network
DMP	Data management plan
WP	Work package
D	Deliverable
CSV	Comma-separated values
EU	European Union
INSPIRE	Infrastructure for Spatial Information in the European
NUTS	Nomenclature of territorial units for statistics
MB	Megabyte
GB	Gigabyte
JSON	JavaScript Object Notation
R&I	Research and innovation
GUI	Graphical user interface
HRE4	Heat Roadmap Europe 4
ExcEED	European Energy Efficient Building & District Database
DoA	Description of the Action
KO	Key Output
DMT	Data Management Tool
EOSC	European Open Science Cloud
H2020	Horizon 2020
DOAJ	Directory of Open Access Journals
OpenDOAR	Directory of Open Access Repositories
re3data	Registry of Research Data Repositories

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

1. INTRODUCTION

EnerMaps is a Horizon 2020 (H2020) Coordination and Support Action (CSA) project that aims at improving data management practices in energy research and management. Currently, energy data is often difficult to find, mixed in different repositories, and fragmented, which can slow project progress, increase project costs, and create an overall lack of efficiency in the field of energy. EnerMaps will act as a quality-checked database of crucial energy data that will communicate and disseminate data efficiently using practices to make the data findable, accessible, interoperable and re-usable (FAIR) (1).

By dynamically linking different databases in the EnerMaps Data Management Tool, the EnerMaps project will create a common knowledge base that is user-friendly, promotes hard-to-find datasets, and includes additional visualization and analytical capabilities. Another layer of the EnerMaps project is the creation of an online energy research community. Users can use the EnerMaps Open Research Gateway to collaborate, share data, and provide feedback on the project. This community-led component will provide an exhaustive view of scientific data in the energy field.

The objectives of EnerMaps can be defined by four Key Outputs (KOs), summarized in Figure 1. Summary of the key outputs of EnerMaps and the related issues it will resolve:

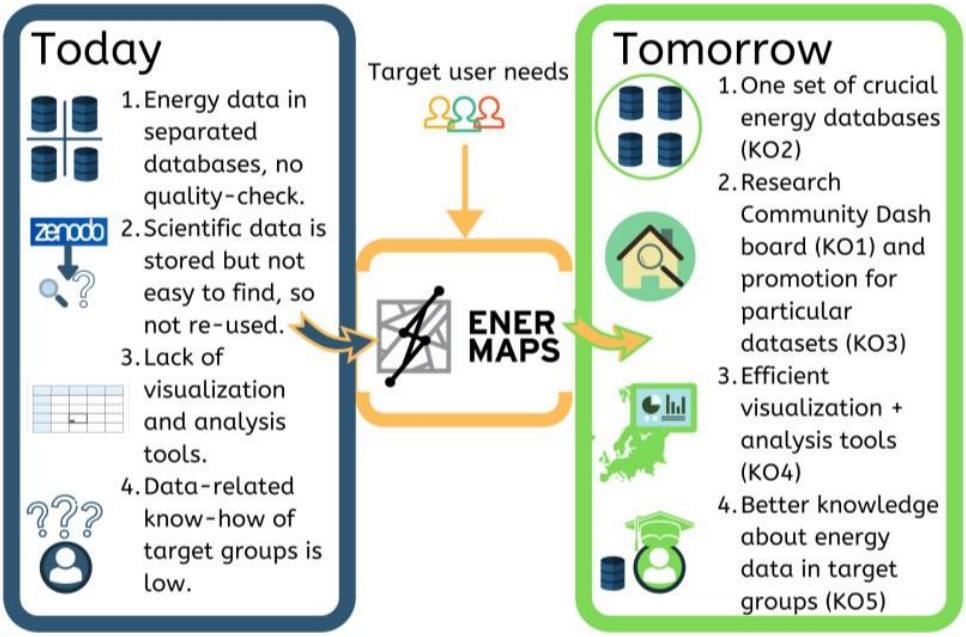

Figure 1. Summary of the key outputs of EnerMaps and the related issues it will resolve

EnerMaps is a collaborative project among seven institutions that make up the EnerMaps Consortium (from this point referred to as simply the Consortium). The Consortium includes experts in energy data management

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

(CREM, EURAC, e-think, and TU Wien), in data management and artificial intelligence (Idiap), in scientific data management (OpenAIRE), on the interactions between energy and civil society (e-think and EURAC), and on the dissemination and communication in the field of sustainability (REVOLVE Media).

EnerMaps is intended to outlive its project lifespan. The project will run in perpetuity with the guidance of an established EnerMaps research community and using the foundations laid out by the Consortium. Because of the long-term motivations for the project, EnerMaps will be vital for long-term planning and will act as a central hub for scientific research in the energy field long after its inception.

1.1. Document structure

Chapter 1 presents the objectives of the Data Management Plan and the logic used in creating this document.

Chapter 2 provides the substantial content of the Data Management Plan, including best practices.

Chapter 3 concludes the document.

1.2. Objectives

This document is the first version of the data management plan (DMP) for the EnerMaps project. The DMP provides descriptions of the data used in EnerMaps, and provides an outline of how the data is managed, shared, and preserved. Data management is a core function of the EnerMaps project, which harmonizes, manipulates, and displays a wide array of datasets of varying resolutions, with all of the data made publicly available. Datasets are procured for EnerMaps from both online, public sources as well as private sources with permissions to open the data.

This DMP describes the steps undertaken by the Consortium to make the data generated by EnerMaps to be findable, accessible, interoperable and re-usable (FAIR). The DMP is framed in accordance to the “Guidelines on FAIR Data Management in Horizon 2020” (1), which lists the following points that are to be addressed by a DMP that promotes FAIR data (see also Figure 2) (1):

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

The EnerMaps DMP was created using DMPOnline (2), an online tool developed by the Digital Curation Centre (DCC) to develop DMPs that meet specific regulations. In this case, the DMP was developed to satisfy the

The EnerMaps project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N°884161

requirements of the Horizon 2020 framework on DMPs by using information from the “Guidelines on FAIR Data Management in Horizon 2020” (1).

Figure 2. Documented phases of a Data Management Plan (3)

The European Commission’s “Guidelines on FAIR Data Management in Horizon 2020” provide a summary table with a detailed list of the specific issues that should be addressed in the Data Management Plan. This table has been reproduced in Table 1 and includes the location of where each issue is addressed within the DMP.

Table 1: Guidelines summary

DATA SUMMARY	
State the purpose of the data collection/generation.	Objectives, page 4
Explain the relation to the objectives of the project.	Data management guidelines, page 11

The EnerMaps project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N°884161

Specify the types and formats of data generated/collected.	Data formats, page 14
Specify if existing data is being re-used (if any).	Data from external sources, page 11
Specify the origin of the data.	Data from external sources, page 11
State the expected size of the data (if known).	Data preservation, page 18
Outline the data utility: to whom will it be useful.	Types of data managed in EnerMaps, page 10
FAIR DATA: MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA	
Outline the discoverability of data (metadata provision).	Data vocabulary definition and metadata, page 12
Outline the identifiability of data and refer to standard identification mechanisms. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?	Data identification, page 15
Outline naming conventions used.	Naming conventions, page 13
Outline the approach towards search keyword.	Data vocabulary definition and metadata, page 12
Outline the approach for clear versioning.	Data versioning, page 15
Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how.	Data vocabulary definition and metadata, page 12
FAIR DATA: MAKING DATA OPENLY ACCESSIBLE	

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so.	Data licensing, page 14
Specify how the data will be made available.	Data formats, page 14
Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?	Data formats, page 14
Specify where the data and associated metadata, documentation and code are deposited.	Data formats, page 14
Specify how access will be provided in case there are any restrictions.	Data formats, page 14
FAIR DATA: MAKING DATA INTEROPERABLE	
Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.	Data vocabulary definition and metadata, page 12
Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?	Data vocabulary definition and metadata, page 12
FAIR DATA: INCREASE DATA RE-USE (THROUGH CLARIFYING LICENSES)	

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

Specify how the data will be licenced to permit the widest reuse possible.	Data licensing, page 14
Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.	Data licensing, page 14
Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.	Data licensing, page 14
Describe data quality assurance processes.	Data quality process, page 12
Specify the length of time for which the data will remain re-usable.	Data preservation, page 18
ALLOCATION OF RESOURCES	
Estimate the costs for making your data FAIR. Describe how you intend to cover these costs.	Assessment of project costs for data management, page 18
Clearly identify responsibilities for data management in your project.	Project tasks and responsibilities related to data management, page 19
Describe costs and potential value of long-term preservation.	Assessment of project costs for data management, page 18
DATA SECURITY	
Address data recovery as well as secure storage and transfer of sensitive data.	Data access, page 17 Data preservation, page 18

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

ETHICAL ASPECTS	
To be covered in the context of the ethics review, ethics section of the Description of the Action (DoA) and ethics deliverables. Include references and related technical aspects if not covered by the former.	Sensitive data, page 16
OTHER	
Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any).	Data quality process, page 12

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

2. ENERMAPS DATA MANAGEMENT PLAN

2.1. Data summary

2.1.1. TYPES OF DATA MANAGED IN ENERMAPS

The data used in EnerMaps comes in the form of structured data (vector data, raster data, or attribute tables) with varied spatial extents and resolutions. The Consortium has identified the following target groups and end-users who will especially benefit from EnerMaps' data offerings:

- ***Energy researchers*** *(including staff in universities and non-university research organizations working in the field of energy) will be one of the main beneficiaries of the EnerMaps project through a) better tools to share data resulting from scientific projects, b) better access to energy data through a two-layer system, where the EnerMaps Open Research Gateway provides access to all available data without selection or quality-check and the EnerMaps Data Management Tool provides access to selected and curated databases as well as related insights.*
- ***Industry*** *(especially renewable technology industry), including energy utilities, energy managers, and planners in the energy field will have free and easy access to databases and related insights that they have selected as crucial for the development of their activities.*
- ***Energy consultants*** *have large data needs, which evolve very rapidly since the subject of consulting projects varies. It is central for them to access a large range of open data, which have been checked and curated before as it saves precious time to collect respective data/information. For them, the EnerMaps Data Management Tool will be a common gateway to find and access energy data as well as understand related insights.*
- ***Public administrations*** *(officers of public administrations operating within the energy sector) require reliable energy data as it is central to the development of coherent energy policy. Public authorities have a strong need for energy data to develop and implement efficient policies and instruments which support the energy transition. Hence, public administration officers have been identified as a lead-user and are part of the training program and will benefit from the availability of a European-wide curated database of energy data and related insights. Thanks to the EnerMaps Open Research Gateway, officers can monitor the uptake of Open Science practices in the energy field and follow the scientific progress (in terms of published research products) of research projects funded by more than 20 national and international funders worldwide.*
- ***Civil society*** *can take advantage of the valuable insights provided by the EnerMaps Data Management Tool and its user-friendly presentation.*
- ***Data providers*** *can use EnerMaps to promote and extend the use of the data they share. Data usage is central to data providers and tools to increase it are necessary. They will benefit from the data quality-check procedures and enrichment activities. In addition, data providers can contribute with the metadata of their datasets to the EnerMaps Open Research Gateway and benefit from different services offered by OpenAIRE (e.g. usage metrics, enrichment notification broker) by becoming OpenAIRE compliant (4).*
- ***Energy research communities*** *who need to disseminate data created during the implementation of their scientific project will have the ability to subcontract the EnerMaps team for dissemination and communication activities. By making their research outputs available in the EnerMaps Open Research Gateway (by depositing in Zenodo or in other OpenAIRE compliant data sources), they ensure their products enter the European Open Science Cloud (EOSC) via the OpenAIRE Research Graph.*
- ***Policy makers*** *will profit from the EnerMaps project indirectly. Officers of public administration working in the energy field will support policy makers in gaining a better understanding of EnerMaps, which can*

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

provide valuable insights (e.g. through the use of EnerMaps analysis tools) and facilitate the development of related policies.

The EnerMaps project, including the EnerMaps Data Management Tool, will be designed to operate beyond the lifecycle of the project, so that EnerMaps can be used as a long-term solution for energy data access and selection for end-users. In addition, an Open Research Gateway organized by OpenAIRE will be used to further engagement with relevant end-users and allow for improved feedback and collaboration among energy data users.

2.1.2. DATA FROM EXTERNAL SOURCES

Preliminary data selection was driven by low-carbon topics identified by the European Commission (5), including renewable energy sources (i.e. wind, solar, geothermal, hydro), electricity transmission grid, energy system integration, energy efficiency (including building stock), as well as relevant socio-economic data. Whenever possible, preferences are made for datasets that cover a large spatial extent (ideally, datasets comprising EU27+UK as this is the region of focus for the project). EU-wide data was selected from a variety of sources, including EUROSTAT, the European Environment Agency (EEA), and European Spatial Planning Observation Network (ESPON). However, there was also an emphasis to including energy data from recent, publicly financed R&I projects.

Metadata records about research products available via the EnerMaps Open Research Gateway are selected from the OpenAIRE Research Graph, an open metadata graph obtained by aggregating metadata records from more than 10,000 scholarly communication sources worldwide, including Crossref (6), the Directory of Open Access Journals (DOAJ) (7), the Directory of Open Access Repositories (OpenDOAR) (8), and the Registry of Research Data Repositories (re3data) (9), as well as project information from more than 20 national and international funders worldwide. In addition, users of the Gateway and the OpenAIRE EXPLORE portal (10) will have the possibility to enrich the aggregated metadata records.

2.2. Making data findable, accessible, interoperable, and re-usable (FAIR)

2.2.1. DATA MANAGEMENT GUIDELINES

There are several best practices the Consortium can follow to optimize the findability, accessibility, interoperability, and reusability (FAIR practices) of the data. In creating FAIR data, the Consortium should: create a common repository for open-data that is easily accessible by data consumers, apply open-data to verify the results in scientific research, and to manage other data appropriately and as designated.

Data outputs will be managed by one or more of the following web-based tools and repositories:

1. *EnerMaps Open Research Gateway: OpenAIRE-based tool that will act as the first layer of the data management system for energy research and will connect all publications, grey literature, datasets, software, and other types of research products available online.*
2. *EnerMaps Data Management Tool: A curated database of crucial energy data with efficient visualization and analysis tools that will act as the second layer of the data management system.*
3. *EnerMaps GitLab (11) repository: Git-based repository where all data, metadata, and documentation will be stored, and where data backups and versioning will be managed.*

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

In the European Commission's Guidelines on FAIR Data Management in Horizon 2020 (Version 3.0, 26 July 2016) (1), there are a number of additional recommendations for best practices that should be addressed in the DMP, including dataset naming convention and description, a standard schema for metadata, and processes for data sharing and preservation. All of these topics are explained in this section.

2.2.2. DATA QUALITY PROCESS

A data quality process was established in order to ensure the integrity of the EnerMaps project. The data quality process considers the following areas: i. Data inventory, ii. Data processing, iii. Data definition and comparability.

2.2.2.1. Data inventory

A data inventory is a repository where all project data that is to be made available for redistribution can be found. For the EnerMaps project, a GitLab repository will serve as the data inventory location.

To increase the interpretability of the data, each dataset will include metadata that will describe and document the data. Data documentation will be created with guidance from the W3C Proposed Recommendations on the Data on the Web Best Practices (12).

The H2020 EnerMaps project will be compliant with the Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) (13). The Open Geospatial Consortium has created a list of standards for geospatial interoperability to aid in the proper the implementation of the INSPIRE Directive.

2.2.2.2. Data processing

The purpose of data processing is to prepare the data according to the users' needs. This includes conversion of the data to a common format, adaptation to a specific spatial reference system, selection of indicators for graphical user interface (GUI) interactions and the transformation of data structures to fit users' needs. This will be done via dedicated open source data processing tools (and programming languages) and will be described in a report as part of WP2.

2.2.2.3. Data definition

One of the objectives of EnerMaps is to make energy data more comparable. This not only implies data standardization in terms of formats and units, but also utilizing a common time reference among datasets. To accomplish this, data from a common reference year will be collected for each dataset. In addition, data for the most recent year will also be collected.

2.2.3. DATA VOCABULARY DEFINITION AND METADATA

A key component of data interoperability is that the data and metadata use vocabularies that follow FAIR principles so that the data can be understood and used correctly by data consumers and members of the EnerMaps Consortium. Standard vocabulary can be used when available, but there is often not a reliable

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

definition that is unambiguous and adequately specific. In order to establish a suitable vocabulary, the Consortium will develop the community-supported schema.org standard (14) for semantic energy data in cooperation with other research and innovation (R&I) projects, including the Heat Roadmap Europe 4 (HR4) (15), European Energy Efficient Building & District Database (ExcEED) (16), and HotMaps (17), as well as open-source communities.

By referring to the schema.org standard, as well as best-use cases from other projects, the current state of the metadata standard includes many common fields to allow a user to identify the dataset and learn of its origin, creator, use conditions, when the data was published/updated.

There were two main guiding principles for further developing the current standard. First, the metadata needs to allow different kinds of users to assess whether a dataset would be appropriate for their work. There are numerous different types of data used in energy research, and the energy field is extremely broad in its scope. For this reason, several required and optional fields were considered to describe the dataset to provide a better understanding of the data. This includes required text metadata fields that provide brief descriptions of the dataset and descriptions of the methodology used in creating the dataset, as well as a field with a set of keywords for the dataset.

The second main principle that defined the fields and structure of the metadata was the prevalence of raster datasets. The data included in EnerMaps usually has a spatial component, and while many of the datasets simply provide data by EU member state (i.e. NUTS0 level) in a spreadsheet file, there are also raster datasets with images captured by satellites. Datasets that are spreadsheet data files and datasets that are raster image files are intrinsically two very different file types with differing metadata demands. This resulted in a Boolean metadata field specifically denoting whether the dataset is a raster or not. If the dataset is not a raster, there are a series of additional metadata fields to gather information on field names.

2.2.4. NAMING CONVENTIONS

An important aspect for making data findable is to attribute a suitable naming convention to project data and documentation, and the EnerMaps project will have separate naming conventions for each of these items.

For project documents, the naming convention will be based on the hierarchical structure of the project. The parent file will be the title of the work package, and each work package file will contain a minutes file as well as files for each task within the work package. The minutes file will contain the meeting minutes for the work package, while the task files will contain documents pertaining to that task's activities and may potentially contain an additional minutes file containing meeting minutes specific to that task.

Project data will principally be named using keywords that describe the data. The first component of each data repository name will describe the spatial resolution of the data (for example, "NUTS0" to indicate the nomenclature of territorial units for statistics at the country level). The subsequent components of the data repository name will include keywords specifying the data more acutely. These keywords will be derived from the repository pages found in the EnerMaps wiki that will be developed during WP5.

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

2.2.5. DATA FORMATS

EnerMaps collects and generates spatially explicit data. This includes vector data in the form of lines, points, and polygons at varying spatial extents (i.e. NUTS0, NUTS1, NUTS2, etc.), as well as raster data (i.e. satellite image data) and tables that can be georeferenced. The collected data will be processed by the EnerMaps Data Management Tool, where it will be possible to download the data onto a private computer. The data is useful for both manual and automatic use (i.e. machine evaluation); however, the latter will require a machine-readable data format that will be provided, as well.

EnerMaps will utilize the Data Package format that has been established by Frictionless Data (18). Each dataset and its associated metadata and documentation will be deposited in its own repository on GitLab. Each data repository will have a README file that provides standard information (i.e. title and description of the data) as well as a datapackage.json file that allows the information to be accessed by machines. Any scripts that are required to load data will be made available in the data repository in a file called scripts. Finally, data will be available in a text format (for example, in a CSV format) instead of shapefiles and other file types that require dedicated software to open.

2.2.6. PROJECT FEEDBACK

Training events on the EnerMaps Data Management Tool will be hosted during which feedback will be collected through a questionnaire. This feedback will be analysed and regularly shared with the Consortium during the cyber-meetings. In addition, EnerMaps will use a Kialo social network (19) where users can exchange ideas, network, and send feedback to Consortium members.

2.2.7. DATA ENHANCEMENTS

EnerMaps will enrich the raw data through various enhancements, including the harmonisation and correlation between groups of similar datasets. In addition, the development of new analysis tools that are fully integrated into the EnerMaps Data Management Tool will have the ability to be applied to any subset of the available data. Finally, a contextualisation of the data will provide hints to the users concerning which data could complement the data they are currently consulting.

2.2.8. DATA LICENSING

Each dataset used in the EnerMaps project will be accompanied by a data license that details the permissions associated with the use of that dataset. The licenses for each EnerMaps dataset will be stored in their respective GitLab repository, where they will be found in the respective dataset's README and metadata file.

EnerMaps is intended to be a fully open-source tool, meaning that all data can be used, shared, or altered. While the majority of the data to be included in EnerMaps will already be open source, there are some datasets from recent scientific projects that are not yet available for public access. The Consortium, with the support from EURAC's legal office, will explore opening these datasets in EnerMaps (with approval of the respective project leaders, creators of the datasets, and fully complying to related legal issues treated among others in the GDPR - Regulation (EU) 2016/679 (20)).

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

License restrictions will be assessed for each dataset to ensure that there are no license violations regarding the use of the original dataset. Datasets will be evaluated as to whether they are reused datasets in order to ensure the output dataset uses a license that is compatible with that of the reused dataset.

2.2.9. DATA PROVENANCE

Data provenance refers to the historical record of data and allows data consumers to trace a data record to its source of origin. EnerMaps will include the data author's contact information in the datapackage.json metadata file so that users can trace the data back to the original author and project page.

Metadata will also provide detailed information on the quality of a dataset, including any issues or feedback from the data authors. Data quality documentation is important in communicating the usefulness of data. Having a clear understanding of the data quality will let a data consumer know if the data will work for their needs, and increases the re-use and confidence in the data.

2.2.10. DATA VERSIONING

The data used in EnerMaps will be updated constantly and at varying rates, and clear data versioning is vital in communicating the changes in the data and allowing consumers to identify the specific iteration of data they are using, to discover more recent versions of the data, and to foster improved reproducibility and clarity in research. To accomplish this, clear version numbers that identify the iteration of the data is a necessity.

EnerMaps will utilize Git (21), GitLab's distributed version control system, to maintain data history and status changes. Each data file is assigned a unique hash string, and a change in status of the data file (i.e. a newer version of the data) will result in a new hash string. Additional tools that will be used in the EnerMaps project include Git Large File Storage extension (22), which will be used for large files (namely large raster files) and the Git Journal tool which automatically produces changelogs for each new data release.

2.2.11. DATA IDENTIFICATION

Data identification is an important aspect of creating FAIR data, as it allows data to be found and re-used easily. For EnerMaps, Zenodo (23) will be used to identify the data according to the Digital Object Identifier (DOI) standard. Zenodo assigns a unique DOI classifier to the GitLab repository for each stable release of the data set, allowing the data to be easily discoverable and citable.

2.2.12. CONCLUSION OF BEST PRACTICES AND GUIDELINES

The recommendations for best practices and guidelines from the European Commission and W3C provided the basis for the measures that the EnerMaps Consortium will implement to adhere to FAIR guidelines and ensure proper data licensing and documentation. The use of GitLab for data storage and backup and associated services like Git for data versioning are also key components to ensuring the integrity of the data quality process.

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

2.3. Ethical aspects

2.3.1. GENERAL DATA PROTECTION REGULATION (GDPR)

The EnerMaps project is fully compliant with the General Data Protection Regulation (GDPR) regulations laid out in Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (20) and respects regulations on intellectual property rights (IPR) (24).

Database access will be fully anonymous. If a user wishes to use EnerMaps, they will be required to register an account. The user must provide their email address and a password and grant their consent that their email address be processed and used for account-related communications and management. In addition, data on the type of user (i.e. industry, research, etc.) will be logged to provide the Consortium with key statistics on data access by final users. The user is free to revoke their consent at any time, which would lead to the deletion of their account.

2.3.2. SENSITIVE DATA

Sensitive data is data that is either private or confidential and includes personal user data. The proper management of sensitive data is imperative to maintain the individual privacy and remain in compliance with both EU and international regulations.

In order to ensure sensitive data is properly managed, data that is considered sensitive should first be identified. As part of the development process for EnerMaps in assessing the use-cases for different data and identifying new datasets, interviews with experts will be conducted to create user-stories describing user data needs and uses. Any response data captured via these stakeholder and expert interviews are considered sensitive data and will be treated responsibly and securely. In addition, feedback from users on the Data Management Tool and on any training event are also considered sensitive data. Thus, for the EnerMaps project, the main ethical and privacy issues with sensitive data arise from ensuring the data remains private and that proper consent is obtained before the data is shared or published in any way.

Sensitive data will be stored in a private GitLab repository. Measures to protect the privacy of individuals providing sensitive data will be taken in any instance where sensitive data will be collected and published. When possible, response data will be anonymized so that it cannot be directly attributed to the responder (for example, by delineating a numeric code to an individual). In addition, data will be reported in aggregated forms to further prevent any firm or individual from being identified through their response. If anonymization is not possible, then the explicit permission will be received prior to the publication of sensitive data. In any case, sensitive data will always remain confidential. Participation in interviews and surveys is completely optional, and individuals have the right to revoke the use of their responses at any time if they choose. The data collected from these interviews will only be used for the reasons specified during the interview process and will not be used for any other purpose.

Further insights will be provided in upcoming deliverables that will provide details of the initial data collection as well as of the informed consent procedure.

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

2.4. Data security

2.4.1. DATA ACCESS

Improving data access of energy data is a core function of the EnerMaps project, and so there are a variety of resources available for users to access the project data. EnerMaps will employ a two-layer system where users may access data, visualized in the following Figure 3:

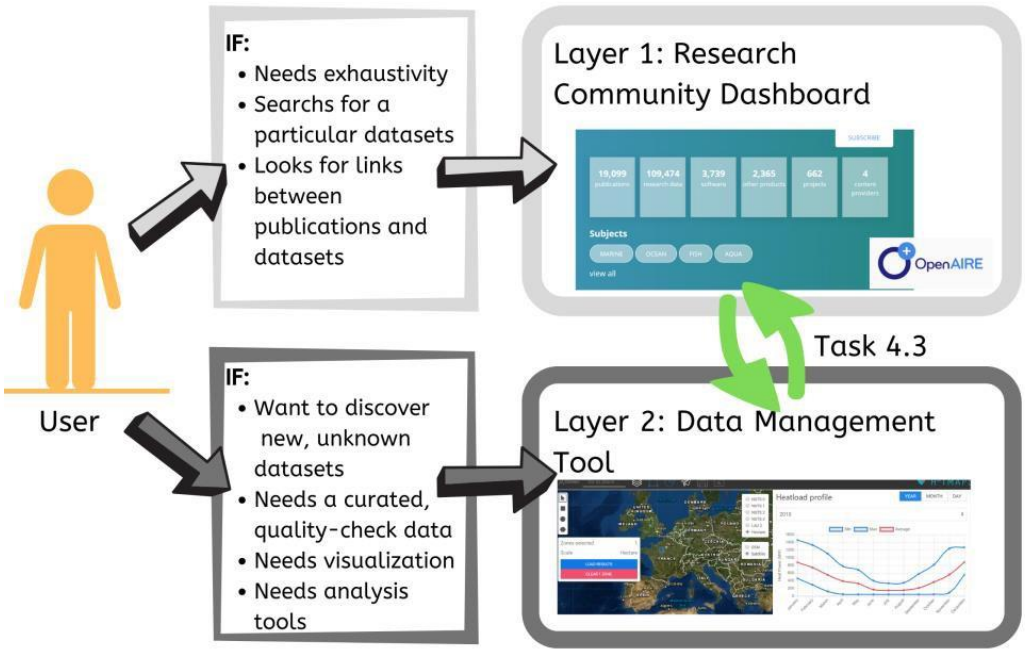

Figure 3. Visual representation of the two-layer EnerMaps data management system

As part of WP4, OpenAIRE will create an Open Research Gateway that will be linked with the Data Management Tool. The Gateway will act as the first layer of the EnerMaps data management system, and will allow community users to share relevant, qualifying data and other types of output of their research (e.g. protocols, methods, software, publications), and will provide an additional avenue for data dissemination. The Open Research Gateway is intended to be an exhaustive resource of energy data intended for users looking for a specific dataset or attempting to link a dataset to a publication (or vice versa) or to other types of objects of the research life cycle (e.g. software, project) to better contextualize the dataset and make it easier to interpret, reproduce, and re-use.

The EnerMaps Data Management Tool will act as the second layer of the EnerMaps data management system, and will provide curated, quality-checked data as well as visualization and analysis tools. Only select datasets will be included in the Data Management Tool from the Open Research Gateway.

Finally, data will be made available on the EnerMaps GitLab online repository where it is available to the public (however, editable data will only be available to members of the EnerMaps Research Community).

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

2.4.2. DATA PRESERVATION

Data preservation refers to data upkeep and maintenance to ensure that the integrity of the data is upheld in the future. This includes properly maintaining the data repository and data backups to ensure the long-term value of the data.

Data backups are expected to occur once a week, though this is subject to change based on the amount of data that will ultimately need to be backed up. Estimating the size of the data is difficult to accomplish at this time since the data that will be initially included in EnerMaps has not been fully identified yet. An initial inventory will be established as part of WP1 which will contain 50 datasets identified by the Consortium, as well as whatever datasets are identified through expert interviews. A better impression on the size of the data can be made after these datasets have been identified. However, EnerMaps is designed and intended to have useful energy data added on a consistent basis, so the size will likely change over time, as well.

The free use of GitLab has limits on the size of data (100 MB per file, 10 GB per repository). Raster data has the potential to be very large, and for large files it is possible that a GeoTiff compressed binary format will need to be used and managed using the Git Large File Storage tool that divides large files into more manageable pieces. Vector data, however, is usually much smaller in size, and in most cases these files can be stored in CSV format.

As EnerMaps data will be publicly available, the data can be downloaded and copied to personal computers without monitoring. In addition, data can be either imported from or exported to external energy models.

2.5. Allocation of resources

2.5.1. ASSESSMENT OF PROJECT COSTS FOR DATA MANAGEMENT

The EnerMaps project will take advantage of several free software and services that will be used to make the data of the project open and FAIR. Zenodo will be used as the data identifier for the project, and GitLab will be used as the data repository. Frictionless Data's Data Packages format, which will be used in the data format process, is also a free service.

However, there will be some expected capital costs associated with the EnerMaps project, mainly in the form of worker hours. Researchers and data managers will need to dedicate time in managing and developing EnerMaps and its various tools. In addition, there will be legal costs associated with maintaining legal compliance in reusing datasets for the project.

The use of free services for many of the core components of data management are essential in maintaining a long-term preservation of the data. Using GitLab to manage data repositories has the added benefit of allowing EnerMaps data to be re-assigned to a different git-host. The EnerMaps project is intended to be a long-term effort that functions after the project end with the support of an established community of energy researchers.

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

2.5.2. PROJECT TASKS AND RESPONSIBILITIES RELATED TO DATA MANAGEMENT

Tasks involving data management are present in all stages of the project. EURAC will lead the first substantive data management tasks that begin in WP1, including the identification and selection of the initial datasets and functionalities of the EnerMaps Data Management Tool. EURAC, TUW, and e-think will conduct the quality control process on the initial data inventory, as well.

CREM and Idiap will complete the substantive data management tasks for WP2, which aim to create the EnerMaps database. CREM will create the guidelines for the overall development framework and the description of input data integration workflow, while Idiap will create the guidelines for data integration process and database deployment.

Using tools that are currently available in the HotMaps Toolbox, five calculation modules will be created as part of WP3. EURAC will implement three calculation modules, while TUW and Idiap will contribute one module each. In addition, CREM will be responsible for the creation of the graphical access to the database via the HotMaps Toolbox and for the results and analysis of the harmonization process, while EURAC will create the guidelines for energy data visualization and TUW will create the methodology for the analysis tools.

The goal of WP4 is to integrate the EnerMaps Data Management Tool within EU-wide scientific data management initiatives. OpenAIRE will lead the creation of an Open Research Gateway, which EURAC will ensure functions coherently with the EnerMaps Data Management Tool. EURAC will lead the development of the schema.org standard for energy data to improve its discoverability. CREM will coordinate and internally review all material.

Finally, e-think will lead the development of guidelines for using HotMaps and other Open Source Databases for WP5, and will be responsible for the creation and the diffusion of the different training materials.

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

3. CONCLUSION

This is most recent version of the EnerMaps Data Management Plan and has been released as a public report. The EnerMaps DMP is intended to be a “living document,” meaning it will be revisited and revised as necessary throughout the project lifespan to reflect changes in project progression. A new DMP will be released for EnerMaps every six months, allowing for all major updates to be quickly captured, and for an up-to-date timetable for future updates.

The EnerMaps project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N°884161

4. REFERENCES

1. **European Commission.** Guidelines on FAIR Data Management in Horizon 2020. [Online] 2016. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

2. **Digital Curation Center.** *DMPonline.* [Online] 2020. https://dmponline.dcc.ac.uk/.

3. **ELRA.** Data Management Plan. [Online] 2020. http://www.elra.info/en/services-around-lrs/dmp/.

4. **OpenAIRE.** OpenAIRE Provide. [Online] 2020. https://provide.openaire.eu/home.

5. **Bastiaanssen, Veerle, et al.** Opportunities and barriers for opening of research databases in low-carbon energy research. [Online] 2018. https://op.europa.eu/en/publication-detail/-/publication/bb6a4096-0fe9-11e9-81b4-01aa75ed71a1/language-en/format-PDF/source-85008501.

6. **Crossref.** Crossref. [Online] 2020. https://www.crossref.org/.

7. **DOAJ.** DOAJ (Directory of Open Access Journals). [Online] 2020. https://doaj.org/.

8. **Jisc.** Directory of Open Access Repositories (OpenDOAR). [Online] 2020. https://v2.sherpa.ac.uk/opendoar/.

9. **re3data.** Registry of Research Data Repositories (re3data). [Online] 2020. https://www.re3data.org/.

10. **OpenAIRE.** OpenAIRE EXPLORE. [Online] 2020. https://explore.openaire.eu/.

11. **GitLab.** *GitLab.* [Online] 2020. https://about.gitlab.com/.

12. **W3C.** W3C Data on the Web Best Practices. [Online] 2017. https://www.w3.org/TR/dwbp/.

13. **European Parliament.** Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). [Online] 2007. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32007L0002.

14. **schema.org.** Energy. [Online] 2020. https://schema.org/Energy.

15. **Heat Roadmap Europe.** *Heat Roadmap Europe.* [Online] 2020. https://heatroadmap.eu/.

16. **ExcEED Project.** European Energy Efficient Building & District Database. [Online] 2020. http://www.exceedproject.eu/.

17. **HotMaps Project.** *HotMaps.* [Online] 2020. https://www.hotmaps-project.eu/.

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

18. **Frictionless Data.** Specifications for Packaging and Transporting Data. [Online] 2017. https://specs.frictionlessdata.io/.

19. **Kialo.** Kialo. [Online] 2020. https://www.kialo.com/social-network-26786.

20. **European Parliament.** Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. [Online] 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

21. **Git.** *Git - fast version control.* [Online] 2020. https://git-scm.com/.

22. **Git LFS.** *Git Large File Storage.* [Online] 2020. https://git-lfs.github.com/.

23. **Zenodo.** *Zenodo.* [Online] 2020. https://www.zenodo.org/.

24. **European Union.** Intellectual property rights. [Online] 2020. https://europa.eu/youreurope/business/running-business/intellectual-property/rights/index_en.htm.

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

Open source tools to share, compare, and reuse low-carbon energy data

WHAT IS ENERMAPS?

EnerMaps Open Data Management Tool aims to improve data management and accessibility in the field of energy research for the renewables industry.

EnerMaps tool accelerates and facilitates the energy transition offering a qualitative and user-friendly digital platform to the energy professionnals.

The project is based on the FAIR principle defining that data have to be Findable, Accessible, Interoperable and Reusable.

EnerMaps project coordinates and enriches existing energy databases to promote a trans-disciplinary research and to develop partnerships between researchers and the energy professionals.

Project Coordinator
Jakob Rager, CREM
jakob.rager@crem.ch

Communication Coordinator
Clémence Contant, REVOLVE
clemence@revolve.media

The EnerMaps project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°884161

enermaps.eu