ENER MAPS

THE OPEN DATA TOOL EMPOWERING
YOUR ENERGY TRANSITION.

DELIVERABLE 1.2: FOCUS GROUP REPORT

E. Wilczynski, S. Pezzutto, J. Balest (EURAC)

Revised by
D. von Gunten (CREM)

31 October 2020

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

enermaps.eu

Executive Summary

The Focus Group report discusses the results of the Focus Group meeting that was held on the 14th of October 2020 and presents the methods and insights gathered from the responses to two surveys that were administered during the meeting. The first survey focused on User Stories and Prioritization .It was framed to gather feedback from the stakeholder analysis process, gain insight in how the information gathered from Deliverable 1.1: User stories and prioritization (1) could be used to develop the EnerMaps Data Management Tool (EDMT), and assess potential of the EDMT in addressing the need for an online energy research community. The second survey focused on Dataset Identification and Quality Check Process (Task 1.2) and requested feedback on the methods in building the dataset list, including identifying additional metadata fields, and carrying out the various steps of the quality check process.

Based on the responses for the first survey, the consortium members voiced the need for capturing the specific needs of researchers and gather feedback from users of similar tools. With respect to the stakeholder analysis, the consortium members suggested more stakeholders and a greater variety of the types of stakeholders. The consortium recognized the importance of an online community dedicated to energy research. While the Kialo social network (2) can address most of the online community needs in relation to the Horizon 2020 (H2020) EnerMaps project, the EDMT can supplement these needs by providing a forum for discussion for individual datasets. Finally, the information gathered during the stakeholder analysis can be especially valuable for targeting stakeholders for capacity building events.

The feedback for the second survey revealed an overall satisfaction with the dataset list, but further metadata fields that had not been captured were suggested (e.g. dataset size, data format, and file type). The quality check process also received satisfactory responses, and there were suggestions that a similar protocol for future datasets would be of value, as well. However, some additional steps that could be added to the quality check process were identified, which were mainly enhancements to the completeness check and to note whether data was the subject of a peer-reviewed publication. Finally, the consortium members stressed the importance of being extra cautious during the quality check process, especially for the consistency analysis, so that inaccurate assumptions are not made when selecting datasets for comparison.

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

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

List of abbreviations

CSA	Coordination and Support Action
D	Deliverable
EDMT	EnerMaps Data Management Tool
EU	European Union
GA	Grant agreement
GDP	Gross domestic product
H2020	Horizon 2020
M	Month
QC	Quality check
R&I	Research and innovation

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 effectively and efficiently using practices to make the data findable, accessible, interoperable, and re-usable (FAIR) (3).

This document is Deliverable 1.2 (D1.2) of the EnerMaps project, which is the Focus Group report that has resulted from the feedback provided by all consortium members for the Focus Group meeting held on the 14th of October, 2020. The purpose of this report is to discuss the subjects of the Focus Group meeting, including the presentation of work to the EnerMaps consortium, and also the logic in choosing questions that were posed to the consortium members via two surveys. The results of the feedback gathered through these survey questions are also discussed. Finally, conclusions from the results are presented, as well as how these outcomes can be realized in future tasks of the project. The Focus Group report was submitted by month 7 (M7) of the project.

The insights provided by this report can be used in several future tasks for the project. The main objectives of this report are to present a summary of the work related to the collection of user needs and stories, the selection of datasets, and the methods for the quality check process, and to gather feedback from consortium members on the work that has been carried out so far.

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

2. FIRST SURVEY: USER STORIES AND PRIORITIZATION

2.1. Background

The partners that carried out Subtask 1.1.5: User Stories on assessing user needs of the EnerMaps project used a sociological approach in determining user needs. Sociology has methods aimed at collecting user needs and stories that can provide insights for the EnerMaps project and the developers of the EnerMaps Data Management Tool (EDMT). These methods are discussed in the following sections. It should be noted that this information is a brief summary for the information provided in Deliverable 1.1 on User Stories and Prioritization that was released in July 2020 (1).

2.1.1. WHO ARE THE USERS?

In order to assess user needs, the users themselves must first be identified. To identify the users, a stakeholder analysis was carried out. Table 1 (below) shows the different steps that were considered for the stakeholder analysis.

Table 1. Stakeholder analysis steps considered for the EnerMaps project (selected methodologies bolded).

STEPS	METHODS
1. Identification of stakeholders	Focus groups
	Semi-structured interviews
	Snow-ball sampling
2. Differentiation and categorization stakeholders	Analytical categorization (top-down)
	Reconstructive categorization (bottom-up)

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

3. Investigation of relationships between stakeholders	Actor-linkage matrices
	Social network analysis
	Knowledge mapping

Ultimately, the snow-ball sampling method was used for step 1, reconstructive categorization (bottom-up) for step 2, and social network analysis for step 3. Snow-ball sampling is a technique where stakeholders are identified from other stakeholders, which allows for a closer interaction between subjects (4). Reconstructive categorization (bottom-up) has stakeholders apply the categorization directly, which promotes the collection of different views from stakeholders (5). Social network analysis utilizes structures of networks to examine social structures, and was selected due to its efficiency (6). Using these methods, the stakeholder analysis identified 49 experts (Table 2).

Table 2. Identification and categorisation of stakeholders.

TARGET GROUP		
Lead User	**Number**	**Percent (%)**
Research and scientific	11	22.44
Industry	3	6.13
Energy managers	2	4.08
Energy planners	2	4.08
Energy utilities	4	8.16
Energy consultants	3	6.13
Public administration officers	6	12.25
End users		
Civil society	4	8.16
Data providers	3	6.13
Policy makers	11	22.4
People filling the format	8	
Total	57	

From the list of experts summarised in Table 2, ten were selected for interview based on geographic distribution, gender inclusiveness, and interdisciplinary expertise.

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

2.1.2. USER STORIES AND NEEDS

A total of eight identified experts were interviewed (four men and four women coming from six different countries). Interviews entailed 25-minute, semi-structured qualitative discussions administered over the phone. The questions were structured to gather insights for EnerMaps in the context of the experts' main professional activities to gather two aspects of the experts' work process: user stories and user needs. A user story is all of the formal and informal procedures related to the use of data, often presented in a narrative form. In the context of the EnerMaps project, these include common problems in data handling reported by the experts as well as the experts' previous experiences with software based on the aggregation and elaboration of different datasets.

Common problems that were identified by the experts included:

- *Difficulties in accessing databases (Legal and market-related)*
- *Low level of detail in the dataset (not sufficiently segregated data, low granularity data)*
- *Data normalization and data transparency*
- *Data representation*

Common experiences identified by the experts included:

- *Common experiences with calculation software*
- *Most of the companies/institutions perform calculation with internal software*
- *HotMaps (7) experiences: difficult to become an autonomous user and complicate interactions with other software*

The interviewed experts were also asked to provide their data- and software-related needs. The data-related needs identified by the experts included:

- *Energy consumption data (industrial/residential use, types of buildings using energy, …)*
- *Energy production data (renewable/fossil, …)*
- *Meteorological data (rainfall, solar insulation, sunshine, …)*
- *Decentralized energy production and consumption (prosumers, off-grid production, …)*
- *Socio-economic data (population, demographics, economic status, status of the building stock, …)*

The software-related needs included:

- *Harmonization of the measurement units and method of production of different datasets*
- *Presence of a glossary of all common use terms and an easily accessible wiki*
- *Easy tools to perform calculations and represent results*
- *Creation and support of an online community of users*

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

2.1.3. INSIGHTS FOR THE PROJECT AND THE DEVELOPERS

The data gathered from the stakeholder analysis presented in the previous section can provide numerous insights for the development of the EnerMaps project. Examples of these insights are summarised below.

For activities related to the creation of processes or procedures of the project, there is potential for the following insights to be acquired:

- *To promote where possible the open access, dealing both with legal and market issues*
- *To define a common procedure for creating datasets to be uploaded in the EDMT, in order to simplify the data acquisition and promote harmonization*
- *To promote the collection of data at middle-long term period, for producing datasets able to elaborate time trends*
- *To promote the data quality, preparing a list of dataset details sufficient (data segregation, granularity, normalization)*
- *To be transparent, within the tool and the trainings, on the mechanisms behind the calculations and representations*
- *To simplify, as much as possible, the procedure to download, analyse, and obtain representations with the EDMT*

In addition, the user needs and stories can provide an idea of communication activities for the project. For example, graphs and other kinds of representations based on graphical and communication expertise should be clear and understandable by all the actors of the energy transition. In addition, based on the results, a section with guidelines on how to transform data and data results into action guidelines should be included.

2.2. Survey Question Descriptions and Results

The results of the first survey administered during the Focus Group can be found below. Thirteen people participated in the Focus Group meeting with at least one representative from each of the consortium organisations. Survey responses were received from five of the consortium organisations (CREM, e-think, Idiap, REVOLVE Media, and TU Wien).

2.2.1. QUESTION 1: IS THERE FURTHER INFORMATION OR INPUT FOR THE DEVELOPMENT OF THE ENERMAPS DATA MANAGEMENT TOOL (EDMT) THAT YOU NEED OR EXPECT FROM THIS WORK OF USER NEEDS COLLECTION?

The EnerMaps Data Management Tool (EDMT) is a key outcome of the EnerMaps project. As it is intended to be a tool used by various types of researchers, its value and functionality is determined by the needs of the users. During the stakeholder interview process, unfortunately only a limited number of questions could be asked (the topics of focus that were asked during the stakeholder interviews are available in D1.1: User Stories

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

and Prioritization that was released in July 2020). However, future tasks will allow us to gather additional insights from stakeholders, which is the justification behind the opening question of the first survey.

Concerning Question 1, a common response among the consortium members was that it is important to capture the feedback from researchers operating outside of an academic setting (e.g. researchers in government and NGOs). In addition, it would be worth investigating the specific ways that researchers might use EnerMaps, and whether there is potential for it to become part of the regular workflow of a researcher.

Consortium members voiced the need for capturing the specific needs of researchers. For example, do stakeholders have preferences on the spatial granularity of the data? Asking specific questions like this will provide a greater depth of insight. In addition, asking stakeholders to prioritize different types of data (or specific datasets) would be useful in understanding which data types they demand the most.

Finally, an interesting insight to gather would come from interviews with users of similar tools. Since EnerMaps builds upon many of the features of the H2020 HotMaps project (7) and respective Toolbox (8), it would be worth gathering insights from HotMaps users on the downsides of HotMaps in order to try to remedy these deficiencies with EnerMaps.

2.2.2. QUESTION 2: WOULD YOU HAVE IMAGINED MORE OR FEWER STAKEHOLDERS? IF SO, WHICH ONES?

The total number and variety of the stakeholders that were identified and interviewed is important. A high number and variety will increase the confidence in the feedback and the overall value of the expert review process. Of course, time and resource constraints allow for feedback to be gathered from a limited number of stakeholders. Still, it is important to gather feedback on the work carried out for this subtask in order to make adjustments for future stakeholder outreach tasks (for example, Task 1.4: Final experts review and feedback).

The consensus among the consortium members is that the feedback process would be improved with an increase in both the number and variety of stakeholders. The consortium members recognize that while eight interviews may be appropriate for the qualitative research approach, this sample size likely does not capture a full array of viewpoints. In addition, the responses stress the importance of capturing a variety of viewpoints. The identified types of stakeholders include members of research, industry, energy managers, energy planners, energy utilities, energy consultants, public administration officers operating in the energy field, civil society, data providers, and policy makers. It is important to gather feedback from as many different stakeholder types, as possible.

2.2.3. QUESTION 3: WHAT DO YOU THINK ABOUT THE NEED FOR AN ONLINE COMMUNITY AND HOW THE DEVELOPMENT OF EDMT COULD ADDRESS THIS NEED?

A desired outcome of the EnerMaps project is to establish an online community for energy research, where researchers can collaborate and review each other's work and datasets. The question was posed to gather insights from the consortium members on two areas. The first part of the question requests opinions on the importance of an online community in energy research. The purpose of the second part of the question is to

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

determine the extent to which the EDMT (or, at least, the current, expected EDMT) fulfils the demands of an online energy research community.

The consensus among the consortium members was that establishing an online community is a valuable outcome of the EnerMaps project. However, there were differing opinions on the role that the EDMT plays in this outcome. For example, some believed that the Kialo social network (2) that will be developed as part of the project could handle all of the social community needs of the entire platform. However, others noted that it might be of additional value in including interactive components to the EDMT, as well. For example, adding a discussion platform (e.g. an online forum or a comment system for each dataset) to the EDMT would allow for a more focused discussion on specific datasets.

2.2.4. QUESTION 4: DO YOU HAVE ANY IDEAS ON HOW TO USE ONE OR MORE OF THE INPUTS PROVIDED BY THIS WORK IN THE DEVELOPMENT OF THE EDMT? PLEASE DESCRIBE YOUR IDEA.

The development of the EDMT will require contributions from a variety of sources. The work carried out and discussed in D1.1: User Stories and Prioritization (1) and summarized in this report may provide important insights in the development of the EDMT. The final question of the first survey was designed to be open-ended in order to gather a wide range of thoughts from the consortium members on where, specifically, these insights might apply to the EDMT development.

The consortium members agree that the user needs are captured by D1.1 (1), and this work can be used in selecting further datasets for the EDMT. Also, the work has illustrated the demands of the different types of stakeholders (e.g. the types of data they require for their work) and can provide an adequate basis for targeting stakeholders for capacity building events (e.g. capacity building workshop).

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

3. SECOND SURVEY: DATASETS IDENTIFICATION AND QUALITY CHECK PROCESS

3.1. Background

An essential component of the creation of the dataset inventory (Task 1.1: Inventory) is the identification of the datasets and to carry out the quality check process on the selected datasets. The following sections illustrate the steps applied in selecting the datasets as well as the methodologies used in conducting the quality check process.

3.1.1. DATASET IDENTIFICATION

Task 1.1: Inventory of the EnerMaps project entails the creation of the dataset inventory. To create the initial selection of 50 datasets, a dataset identification process was carried out. As described in Section 2 of this report, the expert interviews carried out in Subtask 1.1.4: Experts selection of datasets resulted in several datasets that was added to the dataset list. To complete Subtask 1.1.1: Dataset identification, a literature review was conducted. Literature from energy research, including from academic journal papers, conference publications, and similar articles, was reviewed to assess the data demands for energy research and identify any common datasets used in this research.

As is stated in the grant agreement, there was a focus on datasets for renewable energy sources, and also including socioeconomic and climatology data to ensure interdisciplinary research demands are satisfied. The literature review identified the following categories as being particularly important to current energy research demands:

- *Renewable energy source data (e.g. solar, wind)*
- *Energy efficiency data (including building stock)*
- *Environmental data (e.g. water, emissions)*
- *Climate data (e.g. temperature, wind speed, solar insolation)*
- *Socioeconomic data (e.g. GDP, energy prices, population)*

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

Figure 1 (below) illustrates the breakdown of the selected datasets by similar categories.

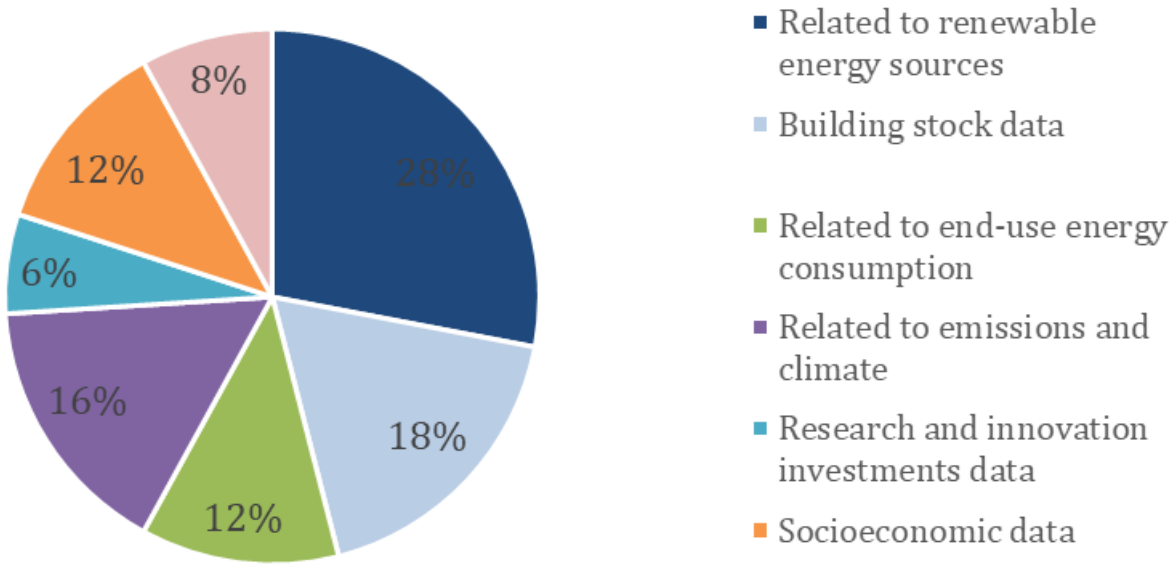

Figure 1. Current breakdown of categories of data in the EnerMaps dataset list.

The most well-represented category is data related to renewable energy sources (RES). However, categories for building stock data, data related to end-use energy consumption, data related to emissions and climate, and socioeconomic data each had over five datasets included in the list. Finally, three datasets related to research and innovation (R&I) investments were included while four were uncategorised.

3.1.2. METADATA ASSESSMENT

The basis for the metadata assessment was identifying the fields mentioned in the grant agreement for Subtask 1.1.2: Dataset description and metadata assessment. The following fields are those that are commonly found in metadata:

- *Creators*
- *Object*
- *Year*
- *Link*
- *Content*
- *Origin*
- *Geographical extension*
- *Granularity*
- *Time references*
- *Access conditions*
- *Terms of use*

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

The number of fields were extended by considering the fields that would be necessary to be compliant with DataCite (9) and schema.org (10) standards, as these standards will be used when creating the metadata for datasets that are processed by the EDMT. This added several additional, including:

- *Identifier*
- *Identifier type*
- *Publisher*
- *Resource type*

Further consultation with members of the consortium identified further necessary fields (e.g. providing a field on temporal granularity as well as spatial granularity).

3.1.3. QUALITY CHECK PROCESS

The purpose of the quality check (QC) process is to assess the accuracy and quality of the selected datasets. As of the creation of this report, the QC process has not been fully completed yet, although the methodologies for each step of the process have been established and presented during the Focus Group meeting to the other consortium members. Upon the completion of the QC process, the full results will be released in D1.6: Results of the quality-check process.

Before starting the quality check process, the selected datasets were categorized into three levels as defined in the grant agreement (GA). Level 1 includes 20 datasets, primarily those that were identified by experts as part of Subtask 1.1.4: Experts selection of datasets. Level 2 includes 20 datasets that have undergone most of the QC process. As the table below demonstrates, these datasets underwent the entire QC process except for the in-depth statistical comparison with related datasets. Level 3 includes 10 datasets that underwent the entirety of the QC process but, unlike Level 2 datasets, will also be statistically compared with similar datasets. The metadata analysis was performed for all 3 levels. The following table shows the steps that are applied to datasets at each level.

Table 3. Data levels applied to each quality check process step.

QUALITY CHECK PROCESS STEP	LEVEL 1 DATA	LEVEL 2 DATA	LEVEL 3 DATA
Collection of experts' feedbacks (Task 1.4)	X	X	X
Availability of users feed-back in the Kialo social network (2)	X	X	X
Existence check of relevant metadata		X	X
Methodology check of datasets		X	X
Completeness control of the dataset		X	X

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

Check of statistical accuracy		X	X
Consistency analysis of the datasets		X	X
Comparison with similar datasets			X

For Task 1.4, we will seek feedback from experts for all datasets from all levels. In addition, all levels will of available for review by users in the Kialo social network (2). The following five rows demonstrate the more formal quality check process, which is applied to only Level 2 and 3 datasets. But only Level 3 datasets will undergo an additional statistical comparison with similar datasets. The key components of the QC process are described in further detail in the following sections of the report.

3.1.3.1. Methodology check

For the methodology analysis, the presence of documentation associated with the dataset was checked for and, if it was available, examined. The methodology was summarized and added to the dataset list.

3.1.3.2. Completeness check

To check for missing data, each dataset was examined for the presence of blank or null values. For practical reasons, this was performed only on datasets that could be opened in spreadsheet software (i.e. Microsoft Excel). Cells that contained blank/null values were those that either had “empty” cells (i.e. no data), or those with a value that was defined to be representation for missing data (for example, a colon). The extent to the amount of missing data was reported in the dataset list as a percentage of the whole dataset. Finally, the documentation for the data was also reviewed for any mention on previously existing missing data and how resolving these blanks in data were handled (for example, through a method of extrapolation).

3.1.3.3. Accuracy check

The accuracy check was conducted similarly to the methodology check. The documentation for the data was examined for any elaboration on if the accuracy of the dataset was assessed. If this information was provided, a summary was added to the dataset list.

3.1.3.4. Consistency analysis

As shown in Table 3, the consistency analysis will be applied to Level 2 and 3 datasets. The approach used for this analysis is summarized as follows:

1. For each dataset that is to be assessed, another dataset that should logically be correlated with the assessed dataset is selected (the grant agreement gave the example of electricity consumption and population density of specific areas).

2. Many of the assessed datasets are panel data with a structure that does not always line up easily for simple analysis. To reduce the dimensionality of the data, a subset of each dataset was extracted from the whole

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

dataset. For example, the consideration of one year across multiple locations, or one aggregate location—like data for the entire European Union (EU) level—over numerous time periods.

3. By taking a subset of the data, a simple linear regression can be performed. The review the p-value statistic from the regression would indicate whether there is a statistically significant correlation between the two datasets in the model (the assessed dataset and the compared dataset). When testing the significance of correlation, the null hypothesis is that there is not a significant correlation (or linear relationship) between the two variables, while the alternative hypothesis is that there is significant correlation (or linear relationship) between the two variables. Since we are intentionally testing datasets that should be correlated, the expected outcome of the analysis is that the p-value will be suitably low (i.e. less than 0.05), resulting in a rejection of the null hypothesis and accepting the alternative.

4. Since we are only interested in a very basic level of correlation, the entire output from the linear regression analysis (for example the r-squared) was not considered, since the details or extent of this correlation was not of interest.

To further demonstrate the consistency analysis process, two examples were provided during the Focus Group meeting. These examples are described below.

3.1.3.4.1. EXAMPLE 1

To conduct the check for an energy efficiency indicator from the Global Tracking Framework (11), this data was compared on whether it was consistent with an energy efficiency indicator from Eurostat (12). Since these are both panel datasets that cover each individual EU country over a period of years, a longitudinal subset of each dataset was extracted which included data at the aggregated EU27 level over years where the datasets overlap (1990 to 2014). For visualization purposes, Figure 2 is a graph of these two variables, clearly demonstrating similarities in the linear trends for each dataset. The resulting regression analysis also suggested a statistically significant correlation between the datasets.

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

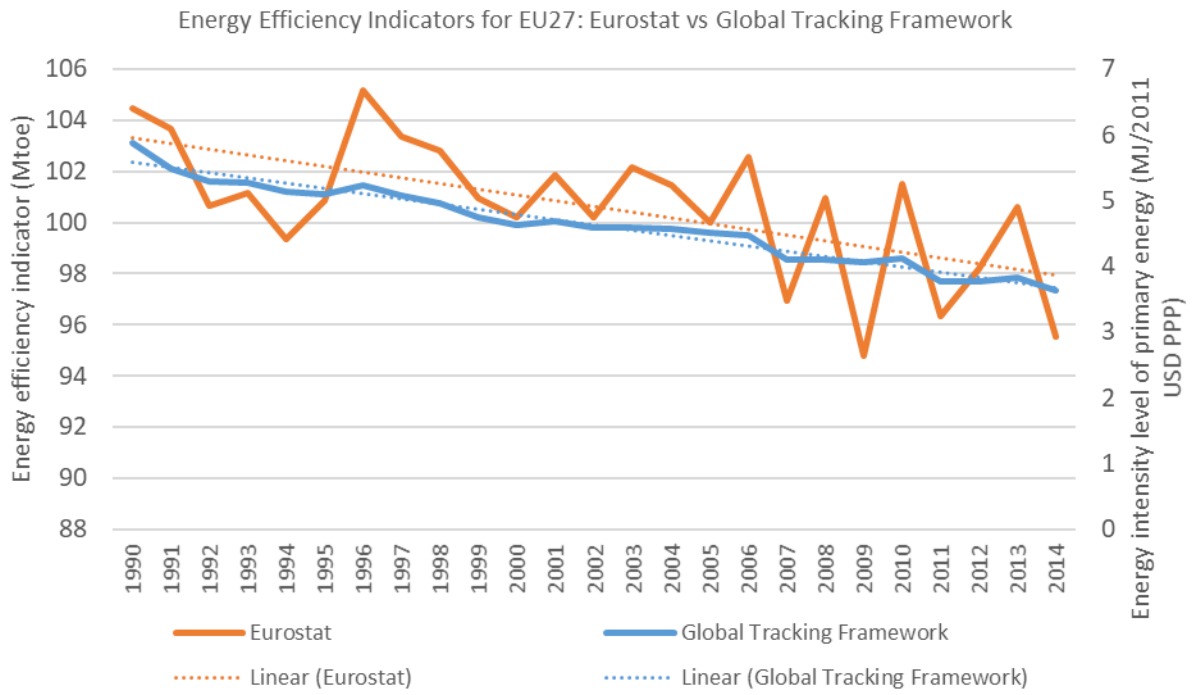

Figure 2. Line graph of dataset subsets for energy efficiency indicators from the Global Tracking Framework and Eurostat.

3.1.3.4.2. EXAMPLE 2

This second example differs in two ways from the first example. First, the dataset that we are assessing, a list of Horizon 2020 projects by country, does not have a similar dataset (13). To assess the consistency of this dataset, a dataset that is logically consistent with the assessed dataset must be identified. Logically, the number of projects should be consistent with the size of that country's economy, and it is expected that the number of projects by country are consistent with the gross domestic product (GDP) of each country (14). The second difference compared with the first example is that instead of focusing on one geographic extent over a period of time, a cross sectional subset was extracted to consider the number of projects for individual EU members state at one year (2019). Figure 3 below shows a scatterplot of the assessed dataset with the GDP by country.

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

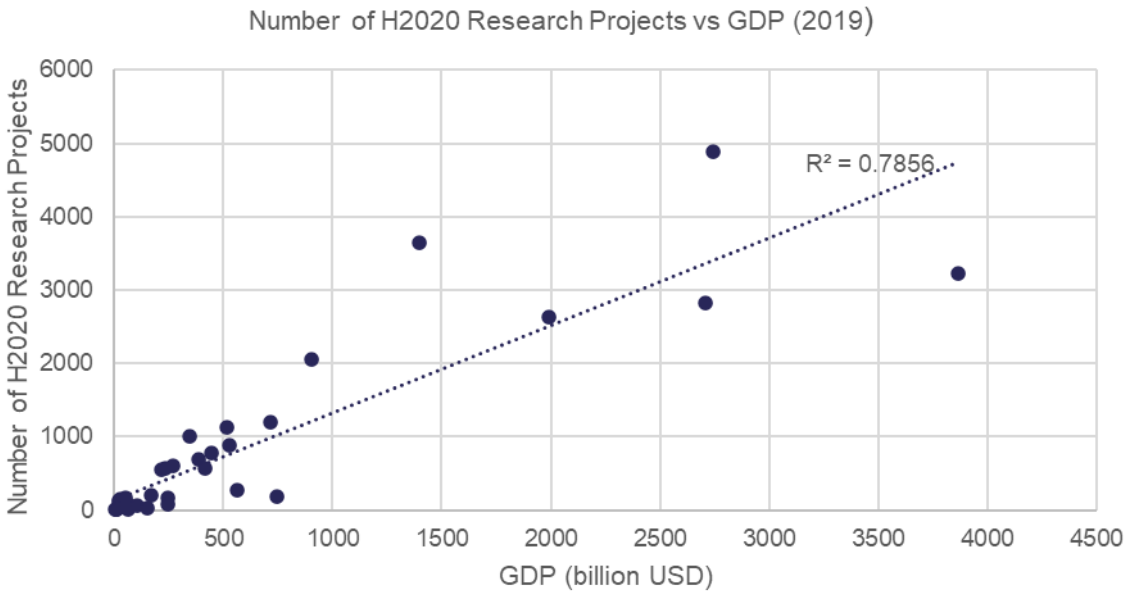

Figure 3. Scatterplot of the number of H2020 projects and GDP of EU member states.

The scatterplot and resulting the linear regression both suggest a statistically significant correlation between the datasets.

3.2. Survey Question Descriptions and Results

The results of the second survey administered during the Focus Group can be found below. Survey responses were received from five of the consortium organisations (CREM, e-think, Idiap, REVOLVE Media, and TU Wien).

3.2.1. QUESTION 1: WHAT METADATA FIELDS WERE NOT RECORDED THAT YOU BELIEVE SHOULD ALSO BE COLLECTED?

The metadata analysis was the most time-intensive activity that was described in this report. The purpose of this analysis is to conduct a rigorous assessment of the metadata for the selected datasets, and logically it is of increased benefit to assess additional relevant fields. The opening question of the second survey asks the other consortium members to consider the selected metadata fields. The members were provided with the current dataset list and the metadata fields that have been considered so far so that they can offer insights into other fields that should be considered.

While the consensus from the consortium members indicates an overall satisfaction with the metadata that was selected, a few additional fields were identified that could be of added use. For example, a “license” field that is distinguished from the “terms of use” field which clarifies the license in case the data is owned by someone

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

other than the creators of the data. Other fields that were identified include the size of the dataset, the data format, and the file type.

3.2.2. QUESTION 2: GIVEN THE ABSENCE OF DETAILED BASELINE MEASUREMENTS IN WHICH TO COMPARE THE DATA, DO YOU FEEL THE QUALITY CHECK PROCESS THAT WAS CARRIED OUT ADEQUATELY ASSESSES THE QUALITY OF THE DATA?

Comparing datasets is a key element for two steps of the quality check process. Level 2 and 3 datasets are compared with similar datasets as part of the consistency analysis on order to determine whether the assessed data is consistent and reliable with respect to similar data. In addition, Level 3 datasets will undergo a more rigorous statistical comparison with similar datasets. However, while it is logical that datasets that attempt to measure the same metric or are intrinsically correlated (like energy use and economic output), the process of using this comparison as a basis for determining consistency is not perfect. Two datasets can both be similarly inconsistent, either by coincidence or due to their reference of a common, inconsistent dataset. Ultimately, the results of these analyses only determine the consistency of each dataset with respect to the dataset it is being compared with, and not some scientific standard. For this reason, it is important to hear other opinions on whether these analyses were adequate. The purpose of this question was to gather any thoughts from the consortium members on the methodology of comparing datasets.

The responses were overall positive with regards to the methods for the quality check process. However, since the QC process was not completed and a full report on the consistency analysis and results from this analysis were not available during the Focus Group meeting, the entirety of the QC process could not be fully assessed. However, as noted in one of the responses, checking the quality of many datasets is not necessarily required if they come from highly reputable sources, like Eurostat.

An additional response mentioned the potential value in having the QC process available for future datasets uploaded to EnerMaps (however, the limitation and appropriateness of simple linear regression when comparing certain datasets).

3.2.3. QUESTION 3: WHAT STEPS WOULD YOU ADD TO THE QUALITY CHECK PROCESS IN ORDER TO MEASURE THE QUALITY OF THE DATA AND ENSURE ITS RELIABILITY?

The quality check process is a multi-step process and is not applied the same way for each level of dataset. For example, Level 3 datasets undergo the entire process while Level 1 datasets undergo only a few steps of the process. The purpose of the QC process is to ensure the reliability of the data, and so it is important that the process considers as many positions as possible in order to accomplish a rigorous assessment. Consortium members were therefore asked to review the steps of the QC process and determine if these steps were sufficient or if additional steps should be considered to enhance the process.

The consortium members identified further steps that could be added to the quality check process to enhance the overall process. One suggestion was to gather feedback from users of certain datasets (notably, datasets

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

resulted from research and innovation projects) and report on their use-cases. Other suggestions included including some additional detail for the metadata analysis, like the type of data format used and a more in-depth completeness check on the level of missing data (notably in more complex datasets, like highly-granular time-series data). Another useful step would be to make note of data that is the subject of a peer-reviewed publication. These datasets would be of increased value over other datasets.

3.2.4. QUESTION 4: PLEASE PROVIDE ANY GENERAL FEEDBACK ON THIS PRESENTATION.

The final question of the survey is more open-ended. Rather than focusing on any specific portion of the work presented in Section 2, the consortium members were asked to provide any general feedback or specific remarks to an individual topic of this section.

The concluding remarks by consortium members echoed some of the past responses. For example, the consistency analysis, while adequate, should be approached carefully to ensure inaccurate assumptions are not made when selecting datasets for comparison. Another point that was brought up was with regard to the confidentiality of the results. Of course, more transparency is better if full results can be presented provided privacy is prioritized.

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

4. CONCLUSIONS

Based on the results of the surveys from Sections 2 and 3 of this report, insights can be taken and potentially applied to future tasks within the project. These insights are further discussed in the following sections.

4.1. User Stories and Prioritization

Task 1.4 requires the finalisation of the dataset list, which includes a round of feedback from experts. The results from the Section 2 of this report can aid in the selection of experts (Subtasks 1.4.1) as well as the collection of feedback from experts (Subtask 1.4.2). Insights were provided on the number and variety of stakeholders that should be consulted from Question 2 (Section 2.2.2), where the consortium members suggested a higher number of stakeholder feedbacks than the eight gathered for Subtask 1.1.4. To this end, strategies for selecting an adequate number of experts from a sufficient variety of fields will discussed further among the consortium members. To increase the level of feedback gathered from users, REVOLVE can carry out a survey on the EnerMaps home page where users can submit answers to questions that will be developed by REVOLVE and the other consortium members.

4.2. Datasets Identification and Quality Check Process

The feedback from the second survey fall into one of two areas, either on the metadata analysis or the quality check process. With regards to the metadata analysis, several fields were suggested to be added to the dataset list (Section 3.2.1). Since the final version of the dataset list will be published as a deliverable (D1.4) during M10, these additional fields can be added so that they are reflected in the final version of the deliverable. As mentioned in Section 4.1, the expert review of the dataset list will potentially lead to further insights, as well as possible modifications to the datasets in the list.

The quality check process will be completed in anticipation of the creation of D1.6: Results of the quality-check process, which will be a confidential report on the results from this process. The final step of the quality check process, which is the statistical comparison of Level 3 datasets, will be carried out, along with any additional consistency analyses for Level 2 and 3 datasets that were exchanged for datasets suggested by the consortium members. The feedback for these processes stressed caution in selecting datasets in which to conduct the analyses.

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

5. REFERENCES

1. **Balest, Jessica, et al.** *EnerMaps Deliverable 1.1: User Stories and Prioritization.* 2020.

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

3. **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-datamgt_en.pdf.

4. **Ghaljaie, F., Naderifar, M. and Goli, H.** Snowball Sampling: A Purposeful Method of Sampling in Qualitative Research. *Strides in Development of Medical Education, 14(3).* [Online] 2017. 10.5812/sdme.67670.

5. *Stakeholder categorisation in participatory integrated assessment processes.* **Hare, Matt and Pahl-Wostl, Claudia.** 2002, Integrated Assessment, Vol. 3, pp. 50–62.

6. *Stakeholder analysis and social network analysis in natural resource management.* **Prell, Christina, Hubacek, Klaus and Reed, Mark.** 2009, Society and Natural Resources, pp. 501-518.

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

8. **HotMaps Toolbox.** [Online] 2020. https://www.hotmaps-project.eu/hotmaps-project/.

9. **DataCite Metadata Working Group.** DataCite Metadata Schema Documentation for the. [Online] 2019. https://doi.org/10.14454/7xq3-zf69.

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

11. **Global Tracking Framework.** Energy Efficiency Indicator. [Online] 2018. https://energydata.info/dataset/world-global-tracking-framework-2017/resource/5ed45e2a-0291-4338-aeda-46da78470aff.

12. **Eurostat.** Energy efficiency indicator. [Online] 2018. https://data.europa.eu/euodp/data/dataset/YIX54AYLew2DOmPqK8dRfQ.

13. **European Commission.** CORDIS EU research projects under Horizon 2020. [Online] 2018. https://data.europa.eu/euodp/en/data/dataset/cordisH2020projects.

14. **International Monetary Fund.** World Economic Outlook Database: GDP. [Online] 2019. https://www.imf.org/external/pubs/ft/weo/2019/02/weodata/index.aspx.

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

The Open Data Tool empowering
your energy transition.

WHAT IS ENERMAPS?

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

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

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