dialogues

Energy citizenship
for a sustainable future

D4.1

MILESTONE 4 - State of the energy citizenship data report

deliverable type
Report

work package
4

dissemination level
Public

month and date of delivery
Month 14 | 23. June 2022

leader
NSR

authors
Martin Inge Standal
Berit Therese Nilsen

programme H2020 | contract number 101022585 | duration 36 months | start May 1, 2021 | end April 30, 2024

www.dialoguesproject.eu
info@dialoguesproject.eu
/dialoguesh2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

dialogues
Energy citizenship for a sustainable future

Contributors

NAME	ORGANISATION
Martin Inge Standal	**NSR**
Berit Therese Nilsen	**NSR**

Peer Reviews

NAME	ORGANISATION
Jens Røyrvik	**NSR**
Andrea Kollman	**EI-JKU**
Candela Bravo	**LOBA**

Revision History

VERSION	DATE	REVIEWER	MODIFICATIONS
1	**20.06.2022**	**Jens Røyrvik**	**MODIFICATIONS**
2	**22.06.2022**	**Andrea Kollman**	**MODIFICATIONS**
3	**22.06.2022**	**Candela Bravo**	**MODIFICATIONS**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

Table of Abbreviations and Acronyms

Abbreviation	Meaning
CAL	Citizen Action Lab
OMICS-research	Omics is a rapidly evolving, multi-disciplinary, and emerging field that encompasses genomics, epigenomics, transcriptomics, proteomics, and metabolomics.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

Index of Contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

1 Executive Summary

Work Package 4 supports DIALOGUES' aim to operationalise the concept of energy citizenship through its contribution to the Knowledge Platform. The current report is DIALOGUES Deliverable 4.1 and presents the results of work performed in Task 4.1., "Assembling and curating existing data". In short, the aim of Task 4.1 is to identify and describe information about existing data relevant to energy citizenship in order to facilitate the future use of this data by researchers and policy-makers. The deliverable contributes in particular to two of DIALOGUES research objectives, namely objective 3: "Encompass the various strategic priorities and topical areas of citizen engagement and under the unifying concept of energy citizenship, including mapping probable pathways and the life-cycle timing between modes of engagement" and objective 4: "Analyse if energy citizenship is more likely to emerge locally, or at regional, national or supranational levels and for what reasons".

An important part of creating targeted impact in DIALOGUES, is to identify relevant datasets from projects shared by our consortium partners, as well as searching for accessible and relevant datasets that can be reused for DIALOGUES research purposes. The OpenAire database was created in 2018 to establish a communication infrastructure specifically to support openly available data in European research. Thus, we chose to focus our efforts on the OpenAire database as this task aims to include relevant data from a wide spectre of disciplines, topics, and perspectives.

We developed a procedure for how to implement our approach. In short, the strategy started by 1) defining our objectives, 2) deciding on what kind of data would be interesting, 3) how the data we decided to use ought be presented, and finally, 4) how we should curate these data to make them useful for ourselves and others. The OpenAire search strategy involved defining 51 search terms based on the DIALOGUES objectives, DIALOGUES deliverables 2.1 and 2.2, and input from DIALOGUES partners. In addition, we asked DIALOGUES partners to provide datasets they knew, as we consider it an asset to include datasets that are very well known to the partners, both for later analysis as well as to ensure relevance.

In addition, we examined the reference list of DIALOGUES deliverable 2.2 for potential datasets that were not already included. Furthermore, in cooperation with the other DIALOGUES partners, a meta-data template was established. In this template information about the identified datasets, such as their contents and other relevant meta-data, could be stored.

The OpenAire search yielded 5778 hits that were reduced to 31 datasets through assessment of relevance and availability. These 31 datasets were then curated and added to the meta-data template. Other DIALOGUES partners also provided information about 13 relevant data sources to the template, while the reference list of deliverable 2.2 yielded no additional hits. Thus, the data template was populated by 44 datasets in total.

Our findings reveal a heterogeneous sample of datasets that contain information on aspects such as individuals' sociodemographic factors, actions and behaviors, and

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

internal processes, which we believe will contribute to DIALOGUES' objectives. For instance by complementing the DIALOGUES data collection, identifying gaps in previous research, and reanalysing the datasets to answer some of the DIALOGUES project's research questions. However, our findings also show that there is great variability and much noise in the open data repositories, which contribute to a difficult and time-consuming effort in finding and curating relevant datasets. By further development of the FAIR-principles of open data for the energy citizenship domain, some of these issues could potentially be improved upon. The DIALOGUES project could further develop recommendations that have already been suggested by previous researchers (e.g., Schwanitz et al., 2022), of a data repository for energy research and by developing a vocabulary of concepts important for energy citizenship that can be used to describe future datasets.

One of the main contributions of DIALOGUES Task 4.1, is to collect and curate available energy citizenship data for use in Task 4.2, and Task 4.4. Task 4.2 will provide meta-analyses of the contents of the datasets collected in order to answer some of DIALOGUES' research questions. Task 4.4 will develop a knowledge platform that will describe the contents of-, and meta-data for the datasets collected in this task, for easy access for other researchers.

Furthermore, some of this deliverable's core contributions are to 1) identify the state of energy citizenship open data, 2) identify what kind of knowledge is missing, and 3) contribute to increasing DIALOGUES CALs and surveys' accuracy. In short, the present report reveals the shortcomings of energy citizenship open data in its current state where relevant data is overwhelmed by noise and lack data standards. We also identify three gaps of energy citizenship data for further consideration: First, data availability would be improved by considering and aligning the contents of terms and concepts we use for open data in energy citizenship research. Secondly, more data is needed to get a deeper understanding of underprivileged groups, beyond using nationally representative samples. And thirdly, identifying what can be assumed to be general traits and which are context-specific when it comes to facilitating energy citizenship, can be further investigated in the CALs.

2 Introduction

A critical aspect of efficiency in European research is the re-use of data instead of duplicating expensive data collections. DIALOGUES fully supports this view and attempts to improve efficiency in European research by focussing on the reuse and reanalysis of existing datasets and supplementing these with targeted new data collections when gaps in knowledge are identified. We have therefore also assessed the *availability* of the datasets we have been working with, inspired by the FAIR principles. This report presents the work done in Task 4.1 "Assembling and curating existing data", and the outcome of this work. It further elaborates on the aims and objectives of this task, and what we plan to use this data for. Proper data management, and the curation and re-use of research data are important for policy makers, funders, and researchers (Bahim

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

et al., 2020). Poorly curated data can even delay scientific advancement (Bahim et al., 2020). Thus, efforts to make data openly available are backed by several initiatives like the Open Data Directive, and the FAIR principles (Wilkinson et al., 2016), motivated by the same urge to make use of rich, relevant and accessible research data that has already been gathered. In addition to describing our work in these processes, we also aim to point to what we experience to be the main challenges with the way the open data is currently presented. An important output of this work will be to propose and present to stakeholders an approach to curating and facilitating easy access to and re-usage of research data that are already contained in open repositories. Suggestions as to how the challenges we have faced in this task might be overcome are some initial thoughts we hope to further develop.

2.1 Assembling accessible data through the lens of energy citizenship

Deliverable 4.1 is contributing to DIALOGUES' aim to operationalise the concept of energy citizenship through its contribution to the Knowledge Platform and data repository as online tools for exploring, visualizing, and understanding energy citizenship in a way that is relevant for policymakers and energy market actors. In addition, it investigates broad trends in citizen engagement within the sustainable energy transition, involving an interdisciplinary, multi-method, overarching research process that leverages past data, uses targeted, and minimal new data collection, and incorporates and transcends past work to contribute a more generalized understanding of citizen engagement in the energy transition.

As a part of this contribution, in Task 4.1 we have assembled and curated existing data, and in this deliverable, we provide insight into the state of energy citizenship data. Through this effort, Task 4.1 is targeting in particular two of DIALOGUES research objectives, namely objective 3: "Encompass the various strategic priorities and topical areas of citizen engagement and under the unifying concept of energy citizenship, including mapping probable pathways and the life-cycle timing between modes of engagement" and objective 4: "Analyse if energy citizenship is more likely to emerge locally, or at regional, national or supranational levels and for what reasons".

There are several general-purpose data repositories that can be used to access openly available data, such as Mendeley Data (https://data.mendeley.com/), and Zenodo (http://zenodo.org/), DANS (http://www.dans.knaw.nl/), and others. OpenAire was created in 2018 to establish an open communication infrastructure specifically to support European research (OpenAire, 2022). By searching the OpenAire database, data can be identified from other repositories such as those above. As the purpose of this task is to have a wide span to include relevant data from a spectre of disciplines, topics, and perspectives, we chose to focus our efforts on the OpenAire database.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

3 Methods

Prior to starting the data collection, we developed a strategy that could be implemented to identify datasets relevant to energy citizenship and assess the state of energy citizenship data. In short, the strategy started by defining what our objectives were. We then decided where data could be obtained (i.e., OpenAire and DIALOGUES partners). The next step was about deciding what kind of data would be interesting, and how the data we decided to use ought to be presented. The last step was to decide how we should curate these data to make them useful for ourselves and others. At its core, DIALOGUES is about developing and reaching a definition of the concept of energy citizenship, and it was, therefore, important for this task to avoid bias and reproduction of already established knowledge. By going through the DIALOGUES project description, and project deliverables, in particular deliverables 2.1 “DIALOGUES Integrated Research White Paper” (Biresseliolu et al., 2021a) and 2.2 “Comprehensive, interdisciplinary report on energy citizenship” (Biresseliolu et al., 2021b), we had a process where we iteratively suggested, discussed, retracted, and developed what would be fruitful categories to prioritize. As stated throughout the project, diversity and representation, contextual matters, and various types of behaviours and opinions linked to energy engagement are core themes in our approach to developing an understanding of the concept of energy citizenship. We, therefore, decided on a number of categories describing various populations along social dimensions, like gender, age, and socioeconomic status, in addition to demography and nationality. Actions and behaviours, as well as more internal processes like beliefs and motivations, were also included, in addition to what research question the various datasets mostly informed. Lastly, we added some dimensions describing the datasets, such as research area, sample size, and time of data collection.

3.1 Curating and describing meta-data

A template spreadsheet for describing dataset information meta-data was developed. As described above, categorical information to be included in this template was developed from DIALOGUES objectives and research questions. The template was forwarded to all project partners to suggest changes. The feedback from project partners resulted in minor changes in categories. A shortened version of the data curation temple can be seen in Table 1.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

Table 1: Extract from meta-data curation template

Gender	Age	Socioeconomic status	Race / Ethnicity	Demography	Nationality	Level of data collection	Actions and behaviors	Internal processes	Social and physical context / external factors	Other relevant information
E.g., Yes/no, which	Range	E.g., Income, education, occupation	E.g., Yes/no, which	E.g., Rural, urban, region	List of nations	E.g., Individual, household, other	E.g. voting, consumption, energy use, volunteerism	E.g., knowledge, attitudes, beliefs, values, motivations, perceptions	E.g. pollution, noise, crime, poverty	E.g. position and number in household

3.2 Data-search and dataset-extraction

Our data collection strategy involved: (1) searching the OpenAire database for relevant datasets using pre-determined keywords. (2) inviting DIALOGUES project partners to share information on relevant datasets. And (3) examining the references in the literature review conducted by DIALOGUES under deliverable D2.2, looking for relevant research where open datasets can be extracted for further use in DIALOGUES WP4.

3.2.1 OpenAire search strategy

An initial list of search terms used to extract data from OpenAire was developed from the DIALOGUES project description, DIALOGUES research questions, and from DIALOGUES deliverable 2.2 “Comprehensive, interdisciplinary report on energy citizenship”, which through a literature review describes important aspects of the energy citizenship concept (Biresseliolu et al., 2021b). This list was then discussed among a core group of DIALOGUES project members, and consecutively forwarded to all project partners for additional validation. A final list of 51 words was settled upon and can be found in Table 2:

Table 2: List of search terms

OR operator	AND operator				
energy	Access	co-ownership	engagement	participation	belief
citizen	Behavior	Culture	Ethics	policy	perception
citizenship	behaviour	Democracy	Gender	politicization	ethnicity
environment	Collective	Discourse	individualization	poverty	marginalization
environmental	community	Ecofeminism	Justice	services	class
sustainable	consumer	Ecological	knowledge	transition	race
sustainability	consumption	empowerment	Network	attitude	migrant
	Identity	Literacy	equity	inclusive	inclusion
	geography	geographical scale	geographical level	migration	

The search strategy involved combining all the first columns of words with an OR operator, and one of the words in the second column with an AND operator. E.g., “(citizen

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

OR citizenship OR energy OR environment OR environmental OR sustainable OR sustainability) AND access". This yielded 44 queries. The search was performed on the 3rd of November, 2021.

Exclusion criteria for datasets were those containing only non-EU/Canadian participants, lacking descriptions of the dataset (e.g., non-English language, no codebook), and non-individualized quantitative data (e.g., interview transcripts, smart-home meters, national-level data).

The report authors also reviewed the reference list of DIALOGUES report D2.2. However, this yielded no additional datasets for inclusion.

3.2.2 DIALOGUES partners' datasets

The template for meta-data was forwarded to all DIALOGUES project partners, which were asked to fill in relevant datasets that they knew of or worked on previously. This yielded 13 datasets being added by project partners.

4 Results

4.1 Data collection results

Our 44 search queries yielded 5778 database hits. After the removal of duplicates, 4640 remained. The report authors then screened the titles for relevance. Where there was doubt, the abstract was read to determine whether the dataset was relevant for DIALOGUES' research questions. After this initial screening, 228 datasets remained. Each of these was then investigated by four criteria: 1) relevance, 2) availability of the dataset, 3) availability of data explanation (e.g., codebook, data legend, or questionnaire), and 4) other exclusion criteria.

Relevance was assessed by a core group of project members in discussions critically examining whether the topic of research or data included in the dataset could provide relevant input to DIALOGUES topics. This left 30 datasets. Several of the hits in the search query also refer to various Eurobarometer surveys. A general description of the Eurobarometer was therefore included as a single entry. Thus, the search yielded 31 datasets in total.

As mentioned above DIALOGUES project partners also added 13 datasets. The relevance of these datasets for DIALOGUES research questions should be high, as they were included by the project partners' judgment. Sharing data across projects is also in line with the HORIZON 2020 Open Access guidelines (European Commission, 2022), and encourages collaboration, avoids duplication of effort, and enables researchers in DIALOGUES to build upon previous research.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

Finally, the review of DIALOGUES literature review report 2.2 yielded no additional open datasets. See Figure 1 for a flowchart of dataset inclusion.

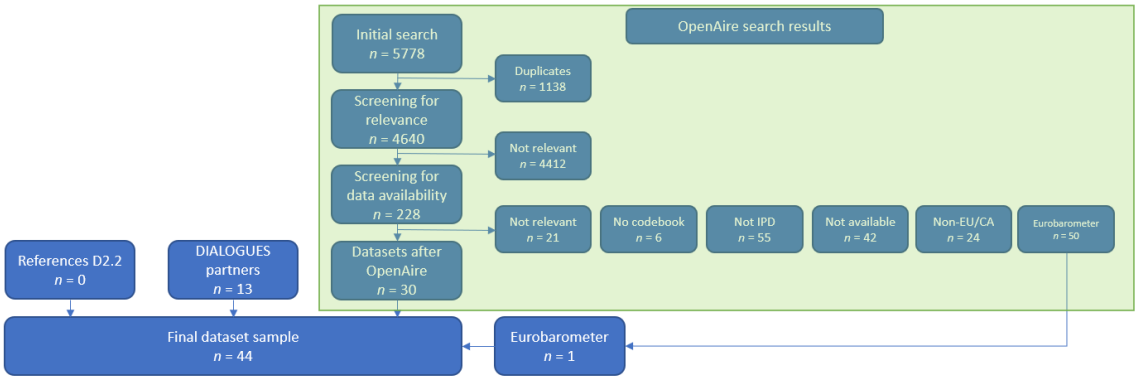

Figure 1: Flowchart of dataset inclusion. IPD = Individual participant data

4.2 Integration, quality, and reliability of the datasets

The results reveal a heterogeneous sample of datasets that include individual surveys, household-level data, and experimental data. Measurements of quality in open data datasets or register-based data involve assessments based on a standard for their particular purpose. For instance, the completeness and correctness of electronic health records (Weiskopf & Weng, 2013), data standards in OMICS-research (Field & Sansone, 2006), and clinical research (Richesson & Krischer, 2007). However, these methods are domain- and data-type-specific and not applicable for use in assessing the reliability and quality of the datasets identified as relevant for energy citizenship. This is due to the datasets showing large heterogeneity of question-wording, their scales, and the underlying constructs they attempt to measure.

As most open data repositories place few restrictions on the quality of the data descriptions and do not integrate or harmonize the data themselves, such spaces will become increasingly difficult to navigate and use for humans themselves or through computational data mining (Wilkinson et al., 2016). The FAIR principles (Wilkinson et al., 2016) were thus proposed to assess the availability of the datasets. Below we describe the human-readable state of the 228 datasets that were carefully examined for availability through the FAIR principles.

FAIR principles

(1) *To be Findable:* To a large degree data were provided with unique identifiers (usually doi). However, while meta-data or an abstract describing the contents of the datasets were available, 6 of the 228 final datasets lacked an English codebook or description of the variables.

(2) *To be Accessible:* A majority of the datasets were openly accessible, albeit not with any description of the appropriate software needed to access. Furthermore, 42 of the final 228 datasets were unavailable. The main reasons for not being

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

accessible (for the purposes of the present study) were that they only contained files that did not have individual participant data (e.g., graphs, images, summary statistics, statistical models, etc.), or that they were access restricted for data security reasons (e.g., confidential, embargoed, nationality-locked).

(3) *To be Interoperable:* Several datasets had descriptions of the contents in an abstract, short description, or provided keywords as to the contents. However, there was no formal, shared or broadly applicable language or vocabulary for representation in general. For instance, data formats that rely on specific software were often found.

(4) *To be Reusable:* To a large extent, datasets were uploaded with a clear and accessible data usage license and provenance.

Thus, for the purposes of the present study, few of the 228 datasets contained independent participant data that were assessed as relevant for energy citizenship. In many cases, it was impossible to assess the relevance or availability of the datasets by use of the accompanying titles, keywords, metadata, or descriptions alone. This resulted in a time-consuming effort to establish whether the dataset should be included.

Nonetheless, there are commonalities on a surface level across the final sample of 44 datasets identified in the present study. Most datasets include nationality and several measures of sociodemographic factors. Commonly the datasets also include some measure of actions, behaviors, and the study participants' internal processes. For instance, behavior on consumption, energy expenditure and saving, mobility and transportation, and civic participation were often surveyed. With regards to internal processes, surveys often included questions regarding participants' attitudes, values and beliefs with regard to climate, energy or civic issues, and knowledge regarding energy- or consumption-related themes. This will be further described below.

4.3 Contents of the datasets

By use of the template described above in Chapter 3 and seen in Table 1 , we curated the datasets into meta-data categories and short descriptions of which categories the datasets contained.

35 of the 44 datasets contained information on participants' age, gender and a measure of socioeconomic status. 3 contained information on two of these variables, and 4 contained information on only one. 2 of the datasets had no information on these sociodemographic factors. All datasets contained information on Nationality, and 10 of the datasets also had more detailed geographical information. This was often described as the region of the participants, or dichotomized as urban or rural.

We also categorized whether the datasets described actions and behaviors, where constructs such as consumption and presumption, diet, transport and mobility, energy use and saving, civic and community participation, were most frequently measured. Internal processes such as attitudes, beliefs, perceptions, identity, knowledge, or values were also noted. Most datasets contained one or more of these constructs.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

Besides the contents of the datasets, we also described the sample size, year of data collection, and the main area of research. The sample size varied from *n* = 37214 participants in the largest to *n* = 20 participants in the smallest. The median dataset sample size was *n* = 1677. In 38 of the datasets, we had information on the time of data collection. The oldest dataset started in 1971 and lasted until 2014, while the newest data was collected in 2021. The majority of the data had been collected in the last five years (52% of the 38 datasets were collected since 2017). The main area of research varied substantially. However, energy consumption, energy behavior, and environmental attitudes and beliefs were common research areas. The target population for the datasets to a large extent aimed at nationally representative samples or using convenience or opportunistic sampling strategies. Only two of the datasets had a special emphasis on targeting underprivileged or minority populations (e.g., by gender, ethnicity, financial resources, religion).

5 Discussion

Deliverable 4.1 aims, as the title suggests, to assemble and curate datasets, providing deliverable 4.2 “Meta-analysis of existing data in relation to DIALOGUES topics” with data needed to perform an analysis of the contents of the datasets. The discussion section will therefore be about the process of assembling and curating, rather than discussing the contents and implications of the datasets. There are however many interesting topics to discuss in relation to presenting, accessing and using open data, discussions that most likely will become increasingly relevant in the years to come.

5.1 Availability of data

Due to the extreme rise of available data, the meaning of availability and accessibility is changing. Accessing relevant data is now a matter of maneuvering through a jungle of irrelevant, overwhelming amounts of data that are presented and made available in a variety of ways and through a number of programs and formats. The way parts of these large amounts of data are made available without an agreement as to what kind of platform, language, program, system or taxonomy to apply, has reached a point where the noise from unwanted data to some extent is drowning the relevant data. Rather than the challenge being to find data, the challenge is now to find the relevant data through this noise. However, once a method for determining the relevant data can be established, the potential available and relevant data is greater than ever before.

5.2 Quality of data

Various research communities are currently looking for ways to implement the before-mentioned FAIR data initiatives in energy research (Schwanitz et al., 2022), and we do also indeed see them exist as a suitable framework to discuss the quality of data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

With regard to data being *Findable*, our approach to use OpenAire to a large extent dealt with this issue. We understand OpenAire's design and purpose to align with our interests in this task, and chose to use it as a tool for finding relevant data. We do however realise that it most likely did not provide us with all possible relevant datasets. After all, what is identified in the OpenAire search is 1) what is uploaded to the searched data repositories, and 2) what was identified by our search terms. A central point in our work was to identify gaps in the available data, and we consider the range of data covered by OpenAire to be extensive enough to give us a solid impression of what kind of data is available and what kind of data is missing.

The principle of *Accessibility* of data was as explained earlier challenged for parts of the datasets. This could be due to restricted access for security reasons, confidentiality, geographical restrictions or reasons related to the type of data and presentation, as we only wanted individual participant data, and not files for instance presenting findings or analyses. By mixing various types of data with analyses and illustrations, it can be argued that accessibility is fact reduced.

To what extent the data we encountered was *Interoperable* was not the most important point for our use in this task, but rather the point was to be able to access the data. To use the data for our metaanalysis in deliverable 4.2, however, the matter of it being interoperable is vital.

The fourth point of the FAIR data initiative, the aim of optimising data to be *Reusable*, is strongly supported by the efforts of this deliverable. By assembling and curating relevant data, this task has both utilised this principle through our search in the OpenAire repository as well as contributed to this principle by making these data accessible through deliverables 4.3 and 4.4.

5.3 Data and what can be done

The problem of overwhelming amounts of heterogeneous data as mentioned at the beginning of this section could be improved if we were able to follow and further advance the FAIR initiative in energy citizenship (or related) research. By elaborating and making further use of the FAIR principles, we could develop domain-specific guidance, e.g., for energy citizenship, or energy science more broadly (Schwanitz et al., 2022). This has already been started in other scientific domains such as earth and geosciences (Bailo et al., 2020; Stall et al., 2020), archaeology (Hiebel et al., 2020), oncology (Vesteghem et al., 2020), and agriculture (Ali & Dahlhaus, 2022). Such guidelines could include expectations for researchers, which data repositories to use (so conceptually similar data is collected in the same place), how to curate the data (by use of keywords and meta-data), which meta-data and vocabulary to use (with category examples e.g., for energy use, consumption, etc.), and have domain-specific examples for the FAIR principles themselves.

The DIALOGUES project could thus be in a position to build upon the work and recommendations by Schwanitz and colleagues (2022) in the energy research domain. They recommend an anchoring of meta-data and vocabulary that provide future possibilities to analyze and summarize data across projects and data collection efforts.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

DIALOGUES could contribute, for instance, by defining important constructs with keywords that can make relevant data findable for future energy citizenship research. That is, agreeing on a repository for data and vocabulary for meta-data that is in line with the conceptualizations of the energy citizenship research domain (see e.g., Biresseliolu et al., 2021b).

The benefits of clearer domain-specific guidelines for data-storage and meta-data tags could be large, as studies have indicated that only 20% of data is deposited in any repositories (Federer et al., 2018), let alone a repository easily accessible for researchers in a similar field. Furthermore, 60% of data scientists spend most of their time cleaning and labelling data (Crowdflower, 2016).

5.4 Data and what is missing

One aim of assessing and curating relevant datasets is to also reveal what kind of data is missing, that the DIALOGUES' CALs and surveys might provide. Before getting to know the datasets in more detail, what we can say on a more general level about the survey data contributing to our definition and understanding of energy citizenship is that they tend to ask for what we assume to be relevant social dimensions, background information and behaviour. What they are less able to provide, is what is meant by the concepts and terminology used, both from the researchers' end when they ask for instance what gender or ethnicity respondents identify with, as well as how respondents understand these concepts when filling out a survey.

Research has also shown that there are differences in survey response rates based on (non-)participant characteristics (Gustavson et al., 2019), and it is probably safe to assume that in many cases those responding to a survey are often more resourceful. Researchers often get a higher response rate by attempting to recruit as many as possible without consideration of underprivileged or at-risk groups. And even if researchers spend extra resources on recruiting such individuals, they may still be under-represented (Gustavson et al., 2019). The surveys are thus good at representing distribution and scope, and can say something about, e.g., the number of women and men that use public transport. The intentions behind asking for e.g. gender in a survey might however simply be to measure if there is a difference in how different genders respond, and not any intentions about understanding *why* there might be gender differences in behaviour. Potential differences that are difficult to uncover using survey methods.

The surveys are thus better at decontextualized data fitted for generalisations than an in-depth understanding of specific contexts. Especially since many datasets included attempts to get nationally representative samples, context-specific knowledge is less represented. Context-specific knowledge can potentially be enhanced by DIALOGUES' CALs and tailor-made surveys. Data contributing to our understanding of which limits and possibilities for active energy citizenship exist for underprivileged groups in contextual conditions is a potential gap that CALs and tailor-made surveys might provide.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

6 Conclusion

The aim of D4.1 was to provide the underlying information and data sources on existing energy citizenship research to be included in Task 4.4 – DIALOGUES knowledge platform. This will provide a possibility for scientists, policy makers, and other stakeholders to reuse this information to generate new knowledge.

Through a systematic process, our findings reveal a heterogeneous sample of 44 datasets that contain data potentially relevant for energy citizenship research. Data were curated on several dimensions such as sociodemographic factors, geography and nationality, actions and behaviors, and internal processes. However, our findings also show large variability and noise in open data repositories, which contribute to a time-consuming and difficult effort in finding and curating relevant datasets. By continuing the work of Schwanitz et al. (2022) and further developing domain-specific recommendations for FAIR-principles for energy citizenship, such as a commonly agreed-upon vocabulary for meta-data and a preferred data repository, collecting and curating energy citizenship data can be simpler going forward.

6.1 DIALOGUES' potential contributions

To sum up the most central contributions from this deliverable, for the project's internal use a core contribution is providing input for Task 4.2, a meta-analysis of energy citizenship data and for Task 4.4, DIALOGUES knowledge platform. As previously stated, although not all-encompassing in its scope, we believe that the datasets we have indeed identified will sufficiently represent the research fields contributing to understanding energy citizens' engagement, and also potential gaps in this research.

This leads to another part of the project D4.1 contributes to, which is to further elaborate on what kind of knowledge is missing or what DIALOGUES CALs and surveys might want to focus on. In short, we identify three aspects: First, as mentioned above, data availability would be improved by considering and aligning the contents of terms and concepts we use for open data in energy citizenship research. Secondly, more data is needed to get a deeper understanding of energy citizenship specifically for underprivileged groups. For instance, the reasons behind and consequences of gender differences. A third aspect is to provide insights into the field of energy citizenship as to what can be assumed to be general traits (i.e., survey associations) and which are context-specific (CAL findings).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

7 References

Ali, B. & Dahlhaus, P. The role of FAIR data towards sustainable agricultural performance: A systematic literature review. *Agriculture* **12**, 309 (2022).

Bahim, C., Casorrán-Amilburu, C., Dekkers, M., Herczog, E., Loozen, N., Repanas, K., ... & Stall, S. (2020). The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments.

Bailo D., Paciello R., Sbarra M., Rabissoni R., Vinciarelli V., & Cocco M. Perspectives on the Implementation of FAIR Principles in Solid Earth Research Infrastructures. Front. Earth Sci. **8**,3. https://doi.org/10.3389/feart.2020.00003 (2020).

Biresselioglu, M.E., Demir, M.H., Solak, B., Turan, U., Clément, G., Sahakian, M., Kollmann, A., Musina, D., Schibel, K-L. (2021a). *DIALOGUES Integrated Research White Paper – Version 1.*

Biresselioglu, M.E., Demir, M.H., Solak, B., Turan, U., Clément, G., Sahakian, M., Kollmann, A., Musina, D., Trenti, C., Santoro, M., Schibel, K-L., Massullo, C., Cardone, B., Carrus, G., Bogdanova, V., Yakova, L., Tzvetkova, G., Primova R. (2021b). *DIALOGUES Comprehensive, interdisciplinary report on energy citizenship.*

Crowdflower. (2016). *2016 Data science report.* Retrieved from https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. Accessed 21.06.2022.

European Commission. (2022). *Funding and tender opportunities. Open Access.* Retrieved from https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm. Accessed 21.06.2022.

Federer, L. M., Belter, C. W., Joubert, D. J., Livinski, A., Lu, Y. L., Snyders, L. N., & Thompson, H. (2018). Data sharing in PLOS ONE: an analysis of data availability statements. *PloS one*, *13*(5), e0194768.

Field, D., & Sansone, S. A. (2006). A special issue on data standards. *OMICS: a Journal of integrative Biology*, *10*(2), 84-93.

Gustavson, K., Røysamb, E. & Borren, I. Preventing bias from selective non-response in population-based survey studies: findings from a Monte Carlo simulation study. *BMC Med Res Methodol* **19,** 120 (2019). https://doi.org/10.1186/s12874-019-0757-1

Hiebel, G., Goldenberg, G. I., Grutsch, C. O., Hanke, K. & Staudt, M. FAIR data for prehistoric mining archaeology. *Int. J. Digit. Libr.* **2**, 139–150 (2020).

OpenAIRE A.M.K.E. (2022). *About us.* Retrieved from https://www.openaire.eu/about. Accessed 21.06.2022.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

Richesson, R. L., & Krischer, J. (2007). Data standards in clinical research: gaps, overlaps, challenges and future directions. *Journal of the American Medical Informatics Association : JAMIA*, *14*(6), 687–696. https://doi.org/10.1197/jamia.M2470

Schwanitz, V. J., Wierling, A., Biresselioglu, M. E., Celino, M., Demir, M. H., Bałazińska, M., ... & Suna, D. (2022). Current state and call for action to accomplish findability, accessibility, interoperability, and reusability of low carbon energy data. *Scientific reports*, *12*(1), 1-9.

Stall, S., McEwen, L., Wyborn, L., Hoebelheinrich, N. & Bruno, I. J. Growing the FAIR community at the intersection of the geosciences and pure and applied chemistry. *Data Intell.* **2**, 139–150 (2020).

Vesteghem, C. *et al.* Implementing the FAIR Data Principles in precision oncology: Review of supporting initiatives (2020). *Brief. Bioinform.* **21**, 936–945 (2020).

Weiskopf, N. G., & Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association : JAMIA*, *20*(1), 144–151. https://doi.org/10.1136/amiajnl-2011-000681

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*(1), 1-9.

8 Appendix

8.1 List of Tables

8.2 List of Figures

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

8.3 List of datasets

Partner datasets – OpenAire datasets

Dataset url/doi	Source/citation
Unpublished - https://greenfoot-project.com/	Greenfoot project
Home | ECHOES (echoes-project.eu); https://doi.org/10.5281/zenodo.3524917	Carrus, G., Chokrai, P., Fritsche, I., Klöckner, C. A., Masson, T., & Panno, A. (2020). Psychological factors in energy decisions: Results from experimental studies and a multinational survey. ECHOES report, (D4), 2.
https://doi.org/10.5281/zenodo.5617851	Patricia Albulescu, Irina Macsinga, & Laurențiu Gabriel Țîru. (2021). SMARTEES Timisoara Survey [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5617851
https://zenodo.org/record/3523916#.YrGrp0bP2Ul	ENABLE.EU team, & Galev, Todor. (2019). ENABLE.EU H2020 project dataset and questionnaire from a survey of households on energy use and energy choices (1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.3523916
https://ec.europa.eu/eurostat/databrowser/view/sdg_07_60/default/table?lang=en	Eurostat, 2021. Population unable to keep home adequately warm by poverty status. [online] Available at: <https://ec.europa.eu/eurostat/databrowser/view/sdg_07_60/default/table?lang=en> [Accessed 6 December 2021].
https://data.worldbank.org/indicator/EG.USE.PCAP.KG.OE	The World Bank, 2021. Energy use (kg of oil equivalent per capita) | Data. [online] Available at: <https://data.worldbank.org/indicator/EG.USE.PCAP.KG.OE> [Accessed 6 December 2021].
http://insulae-h2020.eu/deliverables/d-2-4-citizens-participation-common-conclusions-report/	INSULAE project, D2.4
ENERGISE openair datasets https://cordis.europa.eu/project/id/727642/results	ENERGISE project: survey and interview data
https://eige.europa.eu/gender-statistics/dgs/browse/ta	EIGE gender data sets that could be relevant when analysing barriers towards women's engagement in energy citizenship activities, such as gender pay gap, time use, share in STEM education, Internet access etc.use,

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

https://ariadneprojekt.de/nachhaltigkeitsbarometer-2021/ https://www-iass--potsdam-de.translate.goog/de/barometer?_x_tr_sl=de&_x_tr_tl=en&_x_tr_hl=en-US	Kopernikus-Ariadne
Unpublished - https://klimaneutral.berlin/	KliB Project
Unpublished - https://demokon.de/	Demokon Project
https://www.europeansocialsurvey.org/docs/findings/ESS8_toplines_issue_9_climatechange.pdf	Poortinga, W., Fisher, S., Böhm, G., Steg, L., Whitmarsh, L. and Ogunbode, C. (2018). European Attitudes to Climate Change and Energy: Topline Results from Round 8 of the European Social Survey. ESS Topline Results Series: Issue 9.
https://zenodo.org/record/5482382#.YrGquUbP2Ul	Lekavičius, Vidas. (2021). Lithuanian Household Energy Expenditure and Energy Poverty Data, 2019 (0.8) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5482382
https://zenodo.org/record/3346050#.YkFvWy8eNpQ	Nicholas Charron. (2017). PERCEIVE: WP1: Framework for comparative analysis of the perception of Cohesion Policy and identification with the European Union at citizen level in different European countries: Survey at citizen level and data relative to regional performance of the Cohesion Policy and institutional quality [Data set]. Zenodo. https://doi.org/10.5281/zenodo.3346050
https://zenodo.org/record/3712709#.YkF0ky8eNpQ	Troon, Marko. (2020). The public-private partnerships for environmental innovation and energy transition consultation of the GeoMode project (0.0.2) [Data set]. https://doi.org/10.15155/re-130
https://figshare.com/articles/dataset/CONSUMER_OR_CITIZEN_PROSOCIAL_BEHAVIORS_IN_MARKETS_AND_NON_MARKETS/1194811/1	Fosgaard, Toke R.. (2014): CONSUMER OR CITIZEN? PROSOCIAL BEHAVIORS IN MARKETS AND NON-MARKETS. figshare. Dataset. https://doi.org/10.6084/m9.figshare.1194811.v1
https://data.mendeley.com/datasets/w5t358925f/1	Soutter, Alistair; Mottus, Rene (2020), "Data for: "Global warming" vs. "Climate change": A replication on the relationship between political ideology, question wording, and environmental belief", Mendeley Data, V1, doi: 10.17632/w5t358925f.1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

https://data.mendeley.com/datasets/m9f43z5h6p/1	Peña-Vinces, Jesús; Solakis, Konstantinos; Guillen, Jorge (2020), "Data for: Environmental knowledge, collaborative economy and responsible consumption in the context of second-hand perinatal and infant clothes in Spain", Mendeley Data, V1, doi: 10.17632/m9f43z5h6p.1
https://data.mendeley.com/datasets/nswfpphxpn/1	Horgan, Graham; Kyle, Janet; Craig, Tony; Asvatourian, Vahé; Macdiarmid, Jennie (2018), "Data for: Relationship between pro-environmental attitudes, behaviour and dietary intake patterns", Mendeley Data, V1, doi: 10.17632/nswfpphxpn.1
https://data.mendeley.com/datasets/ffyrbsppkx/1	osman, magda; Thornton, Katie (2019), "Data for: Traffic light labelling of meals to promote sustainable consumption and healthy eating ", Mendeley Data, V1, doi: 10.17632/ffyrbsppkx.1
https://doi.org/10.15125/BATH-00193	Mogles, N., Gabe-Thomas, E., 2017. Energy behaviour change model validation. Bath: University of Bath Research Data Archive. Available from: https://doi.org/10.15125/BATH-00193
https://data.edmonton.ca/Surveys/Energy-Efficiency-Edmonton-Insight-Community/597p-bmf4	City of Edmonton (2017)
http://dx.doi.org/10.7939/DVN/10302	Comeau, Louise; Beckley, Thomas; Parkins, John; Stedman, Richard, 2015, "Energy Literacy Survey", https://doi.org/10.7939/DVN/10302, Scholars Portal Dataverse, V3, UNF:6:zfWecMfWmGQzVnlespxwGA== [fileUNF]
https://dx.doi.org/10.5255/ukda-sn-853189	Spaiser, Viktoria (2018). Environmental behaviour data collected through smartphones in a field-experimental setup. [Data Collection]. Colchester, Essex: UK Data Service. 10.5255/UKDA-SN-853189
https://dx.doi.org/10.5255/ukda-sn-854930	Echeverria-Arrondo, Ignacio and Wolfs, Bert (2021). Environmental Perception of Global Business Travel by Swiss Companies in the Zurich Airport Area, 2019-2020. [Data Collection]. Colchester, Essex: UK Data Service. 10.5255/UKDA-SN-854930

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

https://dx.doi.org/10.1594/pangaea.774535	Maiello, Antonella; Viegas, Cláudia Viviane; Ribeiro, José Luis Duarte; Frey, Marco (2013): Environmental policy and knowledge. Copying with complexity through inclusive management. PANGAEA, https://doi.org/10.1594/PANGAEA.774535,
https://dx.doi.org/10.5286/ukerc.edc.000007	McLachlan et al., 2017. Financing Community Energy Survey Dataset. Tyndall Centre for Climate Change Research, University of Manchester, and funded as part of the UKERC research programme
http://dx.doi.org/https://doi.org/10.11588/data/KPY585	Mertens, Alica; von Krause, Mischa; Denk, Alexandra; Heitz, Theresia, 2020, "Gender differences in meat-eating behavior and environmental attitudes – The mediating role of the Dark Triad", https://doi.org/10.11588/data/KPY585, heiDATA, V1, UNF:6:m/fJhuHUlo37Y+XZGy2i2w== [fileUNF]
https://dx.doi.org/10.5281/zenodo.3922367	Valerie van Niekerk. (2020). Influencing Dutch consumer behaviour with sustainable food products (Version 1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.3922368
http://dx.doi.org/10.17026/dans-zg8-9vbq	Ganzeboom, H.B.G. (Vrije Universiteit, Amsterdam) (2004): International Social Survey Programme NL - ISSP-NL 2003-2004: National Identity & Citizenship. DANS. https://doi.org/10.17026/dans-zg8-9vbq
https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:61887/tab/2	Ganzeboom, H.B.G. (VU University) (2014): ISSP-NL 2013 & 2014: Citizenship and National Identity. DANS. https://doi.org/10.17026/dans-zas-mmr7
https://dx.doi.org/10.5281/zenodo.820363	Kmetty, Zoltán, Bent, Caitlin, Shreeve, Greg, & Virág, Zsolt. (2017). NATCONSUMERS - main factors and attitudes behind energy consumption [Data set]. Zenodo. https://doi.org/10.5281/zenodo.820364

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

http://dx.doi.org/10.6092/unibo/amsacta/6219	Aiello, Valentina ; Reverberi, Pierre Maurice ; Brasili, Cristina (2019) PERCEIVE. WP1. Cohesion Policy and identification with the European Union at citizen level in different European countries. Survey at citizen level. European identity and citizens' support for the EU. University of Bologna. DOI 10.6092/unibo/amsacta/6219.
https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/BFWW4I	Kolcava, Dennis; Bernauer, Thomas, 2020, "Replication data for Kolcava, D., & Bernauer, T. (forthcoming). Greening the Economy through Voluntary Private Sector Initiatives or Government Regulation? A Public Opinion Perspective. Environmental Science & Policy.", https://doi.org/10.7910/DVN/BFWW4I, Harvard Dataverse, V1
https://dx.doi.org/10.5522/04/12808406.v1	Watson et al. (2020). Dataset accompanying: Two energy suppliers are better than one: Survey experiments on consumer engagement with local energy in GB. Energy policy.
https://data.mendeley.com/datasets/tzycxn3jxp/1	Emmerich, Philip (2020), "Data for: Public Acceptance of Emerging Energy Technologies in Context of the German Energy Transition", Mendeley Data, V1, doi: 10.17632/tzycxn3jxp.1
https://search.gesis.org/research_data/ZA4492	Andersen, Jørgen Goul, Deth, Jan W. van, Geurts, Peter, Viegas, José Manuel Leite, Badescu, Gabriel, Selle, Per, Teorell, Jan, . . . Armingeon, Klaus (2007). Citizenship, Involvement, Democracy. GESIS Data Archive, Cologne. ZA4492 Data file Version 1.0.0, https://doi.org/10.4232/1.4492.
https://search.gesis.org/research_data/ZA7633	European Commission, Brussels (2020). Flash Eurobarometer 485 (EU Citizenship and Democracy). GESIS Data Archive, Cologne. ZA7633 Data file Version 1.0.0, https://doi.org/10.4232/1.13550.
https://dx.doi.org/10.5255/ukda-sn-5714-1	Walker, G. (2007). Community Energy Initiatives: Embedding Sustainable Technology at the Local Level, 2005. [data collection]. UK Data Service. SN: 5714, DOI: 10.5255/UKDA-SN-5714-1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

https://dx.doi.org/10.5255/ukda-sn-7876-1	Waddams,C., Deller, D., Jeon, J.Y., Moniche, A., Giulietti, M., Loomes, G. (2016). Consumer Energy Supplier Switching: Data from 'The Big Switch' and Subsequent Surveys, 2012-2014. [data collection]. UK Data Service. SN: 7876, DOI: 10.5255/UKDA-SN-7876-1
https://dx.doi.org/10.5255/ukda-sn-5741-1	Department for Environment, Food and Rural Affairs. Environment Statistics and Indicators Division, BMRB. Social Research. (2007). Survey of Public Attitudes and Behaviours toward the Environment, 2007. [data collection]. UK Data Service. SN: 5741, DOI: 10.5255/UKDA-SN-5741-1
https://dx.doi.org/10.5255/ukda-sn-6366-1	Department for Environment, Food and Rural Affairs. Environment Statistics and Indicators Division, Energy Saving Trust, TNS Social Research. (2010). Survey of Public Attitudes and Behaviours toward the Environment, 2009. [data collection]. UK Data Service. SN: 6366, DOI: 10.5255/UKDA-SN-6366-1
https://www.gesis.org/en/eurobarometer-data-service/home	European Commission, Brussels (YEAR): Eurobarometer XX.X (YEAR). Kantar Public, Brussels [producer]. GESIS Data Archive, Cologne.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585

dialogues

Energy citizenship
for a sustainable future

www.dialoguesproject.eu

info@dialoguesproject.eu

/dialoguesh2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101022585