Solar Energy Advances 5 (2025) 100114

Contents lists available at ScienceDirect

Solar Energy Advances

journal homepage: www.elsevier.com/locate/seja

Open data in the solar thermal community: Status, barriers, and opportunities

Lukas Emberger [a], Gireesh Nair [b,*], Klaus Lichtenegger [c], Maria Moser [a], Thorsten Summ [d], Thomas Natiesta [e], Jianhua Fan [f], Pengcheng Wang [f], Viktor Unterberger [c], Philip Ohnewein [g]

[a] SOLID Solar Energy Systems, Graz, Austria
[b] Department of Applied Physics and Electronics, Umeå University, Umeå, Sweden
[c] BEST – Bioenergy and Sustainable Technologies, Graz, Austria
[d] Institute of New Energy Systems, Technische Hochschule Ingolstadt, Ingolstadt, Germany
[e] Austrian Institute of Technology, Vienna, Austria
[f] Technical University of Denmark, Kongens Lyngby, Denmark
[g] AEE INTEC, Gleisdorf, Austria

ARTICLE INFO

Keywords:
Open data
Solar thermal
Digitalization
Renewable energy

ABSTRACT

Open data in the renewable energy sector could contribute to a greener future and reduce carbon emissions. This study investigates existing open datasets in the solar thermal domain, challenges in publishing and using open data and its potential impact. The results support the common belief that open data is useful, offering considerable benefits for data re-users and the solar thermal community. While substantial open data in the solar thermal domain already exists, some datasets like plant statistics and cost data are hard to utilize due to licensing, accessibility, and quality issues. The study shows that data owners benefit less from publishing data compared to data re-users and service providers. While data sharing seems appealing to data owners, several barriers discourage data sharing. Suggestions to promote the use and publications of open data in solar thermal domain are discussed.

1. Introduction

The transition to renewable energy is direly needed to reach the climate goals and mitigate global warming. However - despite the efforts - a large part of the energy demand is still provided using fossil fuels [1]. With heat accounting for approximately half of the global energy demand, solar thermal might be a key resource in providing environmentally friendly and affordable heat. While the enormous potential of solar thermal is still largely overlooked by policymakers [1], it is nevertheless of utmost importance to further research and optimize solar thermal technology.

Harnessing data by learning from and utilizing it could support the adoption of solar thermal energy. Especially publicly accessible datasets, henceforth open data, might have a leverage effect as the data becomes accessible to more researchers and companies, which increases contributions to the field [2]. In addition, open data may also improve collaboration and transparency – as published results can be compared and reproduced more easily when relying on the same open data [2]. Especially in the field of renewable energy, an accelerated innovation might help to contribute to a greener future and reduce emissions. For example, Pfenninger et al. [2] pointed out that open data can help to develop better energy policies to mitigate climate change and promote energy sector transformation.

However, there are challenges to publishing and using open data, including legal, technical, or economic barriers. As a result, companies and organizations may be discouraged from using and publishing data, thereby reducing the availability of open data and hence its leveraging impact. This topic has already been researched by various authors, focusing on open data in general [3–9], government data [10–22], for private companies [23–29], and for specific domains [2,30–33]. However, to the best of the authors' knowledge, the topic of open data has not been analyzed in the context of the solar thermal domain yet. As noted by Janssen et al. [22], risks and barriers might be different depending on the specific datasets. It hence can be argued that it is imperative to understand the domain-specific barriers and opportunities for the solar-thermal community to facilitate effective promotion

* Corresponding author.
E-mail address: gireesh.nair@umu.se (G. Nair).

https://doi.org/10.1016/j.seja.2025.100114
Received 7 April 2025; Received in revised form 19 July 2025; Accepted 27 July 2025
Available online 27 July 2025

strategies for open data.

To address this research gap, this study analyzes the status of open data in the solar-thermal sector, aiming to understand its specific barriers and opportunities, and explore ways to promote data publication. Specifically, the study aims to address the following questions:

Q1. What are the most important barriers of using open data in the solar-thermal domain?
Q2. What are the most important barriers of publishing open data in the solar-thermal domain?
Q3. What are the most important benefits of publishing open data in the solar-thermal domain?

The study is based on information gathered through surveys and workshops with domain experts from the IEA SHC Task 68 group [34], which comprises of solar-thermal experts from 40 companies, universities, and research institutions. In addition to the surveys, barriers and benefits are further analyzed analytically by identifying and rating the usability of existing datasets related to the solar-thermal domain, and elaborating on the benefits of exemplary use cases that could be implemented with the open data.

For practitioners in the solar-thermal community, this work contributes a list of existing open datasets that are relevant for the solar-thermal domain, including a rating of their usability based on the FAIR principle (Findability, Availability, Interoperability, Reusability) [5,35]. Furthermore, it provides inspiration on how open data sharing could be beneficial, including ideas for re-uses, information about the main barriers and drivers which might also be used to motivate policies, and increases the visibility of the "open data" topic in the domain. For open data research, the work provides additional information to existing still comparably novel research on data sharing from the viewpoint of private companies and researchers, as given in [23].

The structure of this paper is as follows: Section 2 provides a synthesis of common barriers and benefits of open data based on a literature search, which informs the surveys, analysis, and interpretation of the results. In Section 3, the methodology of this work is presented, providing details about the survey, workshop, and analysis of activities that have been performed to address the research questions Q1-Q3. Finally, Section 4 shows the results, while in Section 5, a discussion on each of the research questions is provided.

2. Related work

This section provides a synthesis of the available work on open data, forming the baseline for the survey, analysis, and interpretation of the results. More specifically, related literature was screened to identify barriers and benefits for both using and publishing open data.

The synthesis is based on exploratory literature research (performed September 2024) that informed the surveys and workshops, and a semi-structured literature review (performed July 2025) that informed the analysis and interpretation of the results (see also Methodology Section 3). The exploratory literature search mainly relied on searches for *open data* combined with the keywords *benefits* as well as *barriers* and using snowballing, resulting in a total of 13 selected papers. The semi-structured literature search was performed scanning for titles with the search terms *open data AND (solar thermal OR barriers OR challenges OR benefits OR opportunities OR impact OR drivers OR private OR industry)* using u:search and Google Scholar in combination with filters (publication year > 2010, only English language, only articles). The results (2274 titles) were then screened for relevance based on their title, resulting in 472 unique articles (see Fig. 1). The articles have been additionally filtered by year (publication year > 2014) and by relevance - prioritizing review articles, and articles focusing on domains related to solar thermal, private sector, or research in general – to further reduce the number of titles (127 selected titles). After screening the abstract and checking for access, 25 articles were selected in the semi-structured search. Combining the exploratory and semi-structured literature search and removing duplicates, 34 articles were finally selected (see Fig. 2) to identify relevant barriers and benefits of open data.

The first title screening of the semi-structured search resulted in articles from several domains, however, no article dealing with open data in the solar thermal domain has been identified. Instead, as shown in Fig. 1, study focus is on open data in general, open government data, research and academia, the private sector, health and medicine, and various individual domains. Due to increasing pressure on public organizations after several open data initiatives in 2003–2013 [2,20,28] many early studies especially target public government data to better understand the barriers and benefits of open government data. The interest in researching government data is still prevalent today, with research shifting to facilitating the adoption of open data, categorization of barriers, and comparing experiences in different countries [36]. Stemming from government data, starting with 2015, many articles also cover research on public health – with a strong emphasis on privacy, ethics, and anonymity constraints. Notably, the SARS-CoV-2 pandemic led to several publications underscoring both the importance but also the challenges of open health data for the public [6]. The open government movement has also inspired the Open Science movement, promoting the use of open data in science in the past decade [6] in order to facilitate transparency, reproducibility, shared value-creation, and collaboration. As such, early work focuses on benefits and barriers of open science in general, while more recent work deals with challenges due to storing scientific data (e.g., [37]), the influence of open data on more transparent research and facilitating its reuse (e.g., [35]) and policies about sharing research data as part of scientific publications (e.g., [38]). In the case of the private sector, the number of articles about open data is limited, with most studies focusing on how companies might make use of government data. Notably, only four identified articles [23,24,26,39] cover data sharing among private companies, indicating that this topic is comparably novel [23].

While several studies have analyzed barriers and benefits of open

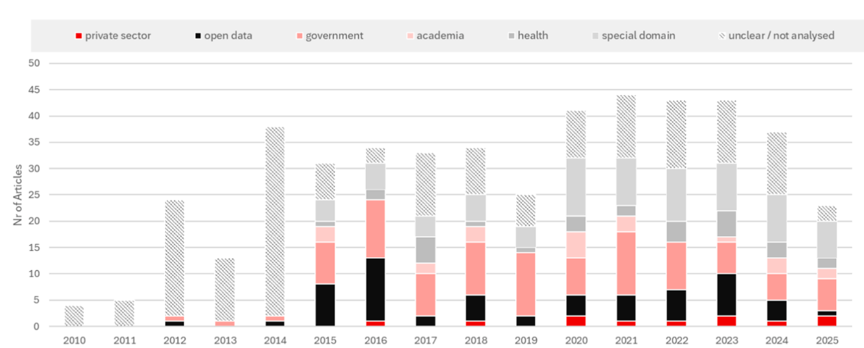

Fig. 1. Number, publication year, and focus of articles (472 articles).

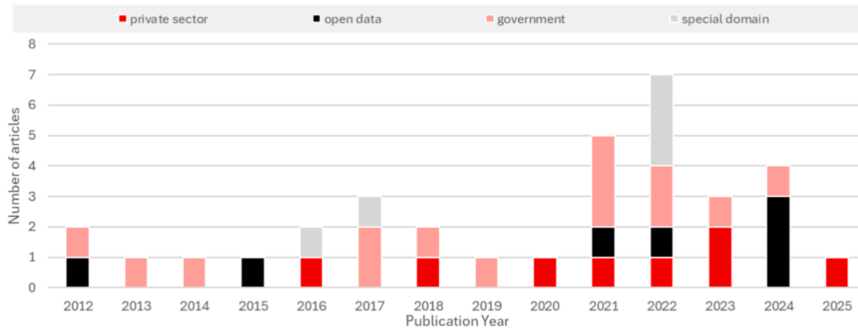

Fig. 2. Number, publication year, and focus of selected articles (34 articles).

data, no universally accepted standard for categorizing them has been established yet [32]. Instead, different categorization schemes exist. For example, barriers have been categorized based on the publication process [10], based on innovation resistance theory - focusing on what kind of barriers prevents people from embracing change [12], or based on categories derived from grouping similar barriers meaningfully [3–5,11,23,28,22]. Despite the differences, most authors clearly distinguish between barriers of publishing the data (i.e. addressing the impediments of the data owner publishing their data) and barriers for re-using the data (i.e. addressing a user that wants to use the open data). The lack of a common categorization is true for open data benefits as well, with different suggestions for example from Janssen et al. [22], Enders et al. [28] and Herala et al. [29]. Thus, barriers and benefits of open data were synthesized anew based on the selected literature, extracting statements that could be interpreted as either benefits or barriers. These statement snippets were then used to formulate distinct barriers and benefit categories, drawing on and adapting existing categorization schemes to group them meaningfully.

2.1. Benefits of open data

The use of open data is presumed to have both economic and societal benefits. For example, [27] note that studies predict that open data will unlock more than $3 trillion additional value worldwide, while societal benefits range from more transparency and accountability [3,5,10,28,29,22], more fairness due to a reduction of information asymmetry allowing small companies to access larger parts of the data [28,29], more efficient data co-creation reducing redundant data acquisition [5,29], and new insights and data service due to increased access and interconnectivity of the data [5,29].

However, the benefits are less clear from the viewpoint of the data owner - especially concerning private organizations - who might not get any value in return [27]. Instead, the benefits of data sharing are often perceived as unclear [32], or mainly speculative [29], while actual proof of benefits of publishing data is scarce [12]. This could be explained by the indirect nature of the identified benefits as shown in Table 1. Except for cases where *Payment in Return* (receiving money from publicly sharing data e.g., to receive funding in return, as part of advertisements, via donations/payments supporting the publishing) is possible, other benefits are harder to assess. For example, *Community Growth* describes the benefit that the domain might be gaining popularity through increased open data, potentially resulting in a bigger market, which in turn could benefit the data owner. Similarly, *Networking* (e.g. building contacts and increasing collaboration with others via data sharing and co-creation) could lead to new research opportunities, or contacts to customers. Another indirect benefit is that open data might be used to *Inform stakeholders* – for example allowing policy makers to perform more informed decisions and potentially leading to new business opportunities or more favourable incentives. Other benefits of data sharing could be *Feedback in return* (e.g. through hackathons, feedback to dataset quality and acquisition procedure, feature requests for new data), which could help address specific research questions or learn more about the demand of the domain, while getting *Data in return* addresses the hope that other people will also share their data in an attempt to decrease redundant data acquisition projects and make more data available. The data owner might also gain benefits from *New Services* emerging that are enabled by the open data. Although there is no guarantee that the data owner might receive them, some services might be offered back to the data owners or might create new business opportunities. The publication of sharing data might also lead to internal improvements, for example by allowing employees to develop *New Skills* or improving the *Business Culture* as employees might be more motivated to contribute to public data, might be more aware of data sharing and its possibilities, and as open data might allow breaking “communication-silos” between departments. Finally, another identified benefit is *Image and Visibility* describing an improved image of the company by showing innovation spirit and transparency, which could lead to increased media coverage, more job applications, or better image of the organization.

For categorization of benefits from publishing data, a combination of terminology from Enders et al. [28] Herala et al. [29] and Janssen et al. [22] was used - namely *Co-Creation (*entailing benefits that are derived by collaboration and benefiting from each other’s advances), *Internal Improvement* (addressing organisational improvements due to data sharing), *Compensation* (resulting in direct measurable income increase), and *Public Image* (improving others perception about the company).

2.2. Barriers for using data

The barriers for using open data synthesized based on the selected literature can be seen in Table 2, including a short description of the individual barriers. For categorizing, barriers were grouped based on what entity the barrier is addressing – namely the data-portal where the data is hosted, the data itself, the user who might reuse the data, or external conditions. For example, *Findability* (e.g. hard to find data, unclear dataset name), *Accessibility* (e.g., missing download options, bad data portal usability, registration requirements, language barrier) and *Support* (e.g., getting help, submitting feedback or annotations to the data) are typically dependent on the platform the data is hosted. In contrast, *Format* (e.g. uncommon data format, non-machine readable, requires proprietary software), *Quality* (e.g. missing data, incorrect data, out-of-date), *Documentation* (e.g. insufficient description of data structure, context, data acquisition procedure, metadata available) and *Licensing* (e.g., not available, reuse not permitted or limited, too complex to read) are connected the data itself and independent of the data-portal. Finally, *Missing Awareness* and *Missing Skills* for handling the data can be associated with the user, while *Missing Gain* for using the data and *Missing Open Data* are based on external conditions. The usage barriers for open data are further grouped based on the FAIR principle, describing that data needs to be Findable, Accessible, Interoperable, and Reusable to enable the benefits of open data [5,35].

Table 1
Benefits for publishing open data based on the selected literature.

Category	ID	Benefit	Description	Sources
Co-Creation	B1	Community Growth	Accelerate research and increase visibility of domain.	[3,5,10,18,22,24,25,28,30]
	B2	Networking	Increase Collaboration with others (e.g. Network building with industry, research, customers) potentially leading to new business opportunities	[8,10,15,23,24,28,30]
	B3	Inform stakeholder	Provide information to (external) stakeholder to enable more informed decisions	[2,5,15,18,24,28,22]
	B4	Data in Return	Allows to generate datasets together, creating potentially more valuable data	[2,5,23,24,28,22,39]
	B5	Feedback in Return	Get feedback in return to sharing data (e.g. input to data quality, feature requests, …)	[2,5,15,18,24,25,28,22]
	B6	Services in Return	Benefit from services that were enabled through the open data	[3,5,8,10,15,22–24,28,30]
Internal Improvements	B7	Business Culture	Working on open data could motivate employees and break communication-silos.	[22,24,28,30]
	B8	Skill development	Employees will acquire new skills	[28]
Public Image	B9	Image and Visibility	Improve image and visibility by showing innovation spirit and showing transparency and accountability	[3,5,8,10,15,18,24,28,30,39]
Compensation	B10	Payment in return	Receive money for publicly sharing data, e.g. as part of funding, one-time payments, or via advertisements.	[8,28]

Table 2
Barriers for using open data based on the selected literature.

Category	ID	Use Barrier	Short Description	Adressing	Sources
Findable	U1	Missing Data	Required data is not publicly available	External	[3,4,18,22]
	U2	Awareness	The user is unaware of open data	User	[3,10,15,18,22]
	U3	Findability	The data is hard to find	Data Portal	[3–5,7,9,10,13–16,19,31,32,22]
Accessible	U4	Accessibility	The data is hard to extract	Data Portal	[4,7,9,10,14–16,18,21,22,26,32]
Interoperable	U5	Missing Skills	The user lacks skills for using the data	User	[3,4,9,10,13,15,16,21,32,22]
	U6	Format	The data format is hard to use	Data	[4,5,14,18,22,26,31,32]
	U7	Quality	The quality of the data is insufficient	Data	[3,4,7,9,10,13–16,18,22,31,32]
	U8	Documentation	The documentation is insufficient	Data	[4,5,7,14–16,18,19,21,22,32]
Reusable	U9	Licensing	The licensing / terms of use are unclear	Data	[4,5,7,10,14,15,22,32]
	U10	Support	Does not allow user feedback / support	Data Portal	[4,7,10,14–16,18,22]
	U11	Missing Gain	Missing benefit or incentive for user	External	[3,4,9,13–16,19,21,22]

2.3. Barriers for publishing data

For categorizing barriers related to the publishing of data, we in large part follow the approach of Fassnacht et al. [23] with minor adaptions,[1] as their analysis covers most of the publishing barriers identified in this work's synthesis and offers a comprehensive framework for categorizing barriers: The *Technological* barriers describe barriers of IT-infrastructure or the data itself, the *Strategic* category groups barriers affecting the decision making process dealing with economic benefits and associated costs, the *Regulatory* barriers encompass privacy, ownership, and legal restrictions that need to be addressed before data can be published, while *Operational* includes barriers associated with the operational execution of the publishing and *Cultural* barriers are dealing with impediments connected to the believes of the data owners. The results are presented in Table 3, including a short description of each barrier.

3. Methodology

3.1. Data sources

The data gathered for this work is mainly based on workshops and surveys within the International Energy Agency Solar Heating and Cooling (IEA SHC) Task 68 group [34] focusing on efficient solar district heating. The group consists of approximately 70 professionals from 40 different organizations from international companies, universities, and research institutes working in the solar thermal field. The group is spanning 10 countries (Austria, China, Denmark, Germany, Netherlands, Poland, Spain, Sweden, Switzerland, UK), with experts from further countries (Estonia, France, Israel, Italy, South Africa) loosely associated via collaborations (see Fig. 3). IEA SHC Task 68 provides a well-suited expert base for this research due to its international composition and diversity of stakeholders, combining academic, industrial, and applied research perspectives. Its members are directly engaged in solar-thermal data handling, application, and research, and thus represent a comprehensive cross-section of the solar-thermal domain relevant for discussing open data issues. Their practical and domain-specific experience formed the basis for surveys, workshops, and expert evaluations carried out in the context of this study over the course of two years from 2023 to 2025. In addition to the surveys, data has been gathered through analysis and rating by the authors, who are also part of the IEA SHC Task 68 group and experts in the solar-thermal domain.

3.2. Activities

An overview of the activities carried out and their connection to the research questions Q1 to Q3 is depicted in Fig. 4, while the following text describes the activities in more detail.

3.2.1. Survey on perception of open data

To understand the perception of barriers and benefits within the solar-thermal community, a Mentimeter [40] survey was sent out by E-Mail via the Task Manager to all experts of the Task 68 group in September 2024. Participation was on a voluntary basis. In total, 32 of

[1] The following uses the terminology of Fassnacht et al.: Barriers S1 and C1 were merged, as were S2 with S6, R1 with R2, and T2 with T3. S4 was excluded as it focuses on data sharing rather than open data. O3 was moved to the category *cultural*. The barriers *fear of misinterpretation* and *fear of reputation damage* were added. The barriers names were adapted for easier interpretation by people working in the solar-thermal domain.

Table 3
Barriers for publishing open data based on the selected literature.

Category	ID	Publishing Barrier	Short Description	Sources
Techno-logical	T1	No Data to publish	There is no data available for publishing.	[7,10,11,23]
	T2	Insufficient Quality	Insufficient data quality like missing entries, or errors inside data.	[4–6,8,10–12,17,21–24,26]
	T3	Missing Infrastructure	Missing or inappropriate data structure for processing, storing, and publishing the data.	[4–6,10–12,14,17,22,23,26,32,39]
	T4	Security and Protection	e.g., fear of data breaches, unauthorized accesses, cyber-attacks.	[2,5,10,14,22–24,26]
Strategic	S1	Unclear Demand	Owned data is deemed as not valuable or demand for data is too unclear to publish.	[7,10–12,14,17,18,21,23–26]
	S2	Insufficient gain	No, unclear, or insufficient benefits of publishing the data expected.	[5,6,10,12,14,16,17,22,23,26,31,32]
	S3	Economic damage	Fear of economic damage or loosing competitive advantage.	[2,5,6,8,10,12,14,17,22–25,32]
	S4	Leaking critical information	Fear of disclosing competitive knowledge or unwanted data that should not be exposed.	[2,12,14,17,23–26,32]
	S5	Lack of resources for preparing	Lack of resources (time, money) to prepare and process the data prior to publishing.	[2,5,10,12,14,17,22–26,32]
	S6	Lack of resources for publishing	Lack of resources (time, money) to publish the data.	[2,5,10,11,14,16,22–24,26,32]
	S7	Lack of resources for maintenance	Lack of resources (time, money) to maintain data and publishing portal.	[2,5–7,10,12,14,16,17,22–24,26]
Regulatory	R1	Ownership and Usage Rights	No or unclear ownership of the data and who is allowed to use and redistribute it.	[5–7,10,14,21,23–25]
	R2	Legal Constraints	restrictions by law (e.g. EU general data protection regulation, cartel rights)	[4–6,8,10,12,14,16–18,22–26]
	R3	Missing legal frameworks	Missing standards and legal frameworks for handling sharing and publishing contracts.	[4,7,18,22,23,26]
	R4	Privacy Constraints	Fear of data privacy violations, or reidentification of personal data.	[5–7,10,11,16,18,21–23,30,32]
Opera-tional	O1	Missing Knowledge	Lack of skills and knowledge required for preparing and publishing data.	[5,10–12,14,16–18,21–26,32]
	O2	Institutional Barriers	Unclear decision process, organisation support, and organisation goals.	[2,6,11,12,14,17,18,23,25,26]
Cultural	C1	Lack of Interest / Awareness	Lack of interest or unawareness of open data, or reluctance to change, negative image of open data.	[2,6,7,10–12,14,16–18,23,24,26,32]
	C2	Fear of misuse	Fear of misuse of the data after publishing.	[6,10,12,14,16–18,23,24,26]
	C3	Fear of loss of control	Fear of losing control of what is done with the data.	[5,14,23,24,26,32]
	C4	Fear of misinterpretation	Fear of incorrect interpretation of the data.	[2,7,12,14,16,17,24,30,32]
	C5	Fear of Reputation Damage	Fear of reputation damage, for example because of publishing low-quality data.	[12,14,16,17,24,25,30,32]

the solar-thermal experts completed the survey. In addition to basic questions on the occupation of the participants and their years of experience, the survey included statements addressing open data barriers and benefits as based on the literature search. Using a Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree), the participants were asked to rate their agreement to the respective statement. In addition, open-ended questions were used to ask about the most important barriers and about suggestions for motivating the publication of open solar thermal data. A mapping of questions to the barriers and benefits of the related work section are shown in Tables 4–6. As the survey was based only on the initial exploratory literature search, some barriers and benefits were unfortunately merged or missed by the survey as noted in the tables.

3.2.2. Evaluation of barriers for using open data

In addition to the survey on the perception of solar-thermal experts, an evaluation was performed by the author team to provide further insights into the barriers of using solar-thermal open data. The evaluation was done in a two-step approach: First, existing open data relevant for the solar-thermal community was identified. Second, all the identified individual public datasets were rated by the authors based on their barriers of use. The following text describes these steps in more detail:

As a prerequisite of the evaluation, existing open data relevant to the solar-thermal community had to be identified first. This has been done via multiple activities: An online survey (see Activity 2 in Fig. 4) was conducted in March 2023 within the IEA SHC Task68 group, asking the survey participants about what public datasets they know and use for their work. In total, 49 participants took part in the Mentimeter survey anonymously, which was sent out to all the Task68 experts by email via the Task Manager. Answers were gathered using open-ended answers to the questions "*Which public datasets relevant to the solar thermal community do you know?*", allowing participants to fill in any known datasets and provide links. In addition, data search (see Activity 3 in Fig. 4) was carried out to further expand the list of existing open datasets. The search for open datasets was conducted using exploratory keyword search on prominent data hubs and search engines (Google, Zenodo, and GitHub). While the exact search terms were not systematically recorded at the time, they included general terms such as "solar thermal data", "collector database", "heat demand data", and similar topic-specific phrases chosen at the authors' discretion based on domain expertise. Zenodo and GitHub were selected because they were considered as prominent data hubs where data relevant to solar thermal might have been published. In addition, Google was included as a general-purpose search engine to ensure broader discovery, especially of datasets published on institutional or project-specific websites. Finally, a few additional datasets were identified during the workshop related to assessing the benefits of open data (see Activity 5 in Fig. 4) as a side effect.

Based on the existing public datasets, a detailed analysis (see Activity 4 in Fig. 4) was carried out to evaluate the barriers of use for each dataset. This was done relying on typical barriers for using open data as reported in the literature (see Table 7). For each dataset, the use was tested, and the experience was rated according to the instructions in Table 7. Results are reported using a scale from 1 (extreme barrier –

Fig. 3. Institutions active in IEA SHC Task 68. For more information, see https://task68.iea-shc.org/ [34].

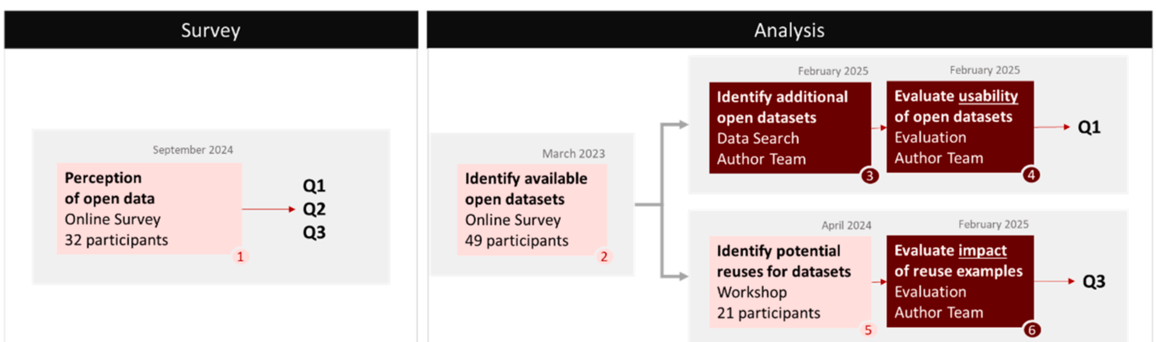

Fig. 4. Flowchart showing the activities and their relation to the research questions. *Online Surveys* and *Workshop* activities within the Task68 group are depicted in light red, while *Evaluations* are shown in dark red.

Table 4
Statements used to assess the barriers of using open data in the survey.

Category	ID	Usage Barrier	Statement
Findable	U1	Missing Data	I think the existing open-data for solar-thermal is sufficient.
	U2	Awareness	I have already used open data for solar-thermal
	U3	Findability	Open data is easy to find
Accessible	U4	Accessibility	It is easy to use the data (once found)
Inter-operable	U5	Missing Skills	— *Not covered in survey*
	U6	Format	The data is provided in a usable format
	U7	Quality	The quality of data is sufficient
	U8	Documentation	The datasets are well documented
Reusable	U9	Licensing	The terms of use were clear (licensing)
	U10	Support	— *Not covered in survey*
	U11	Missing Gain	Open data is beneficial to the solar-thermal community Open data could be a game-changer for the solar-thermal community

impossible to use) to 5 (not a barrier at all). The analysis has been carried out by three authors independently of each other to reduce bias, and the median value from their analysis is used for the final result. Note that only barriers addressing the data-portal and the dataset itself can be checked, as the other barriers (U1 Missing Data, U2 Awareness, U5 Missing Skills, and U11 Missing Gain) are dependent on the user or external factors (ref Table 2).

3.2.3. Evaluation of benefit of publishing open data

The potential benefit of open data was estimated in a two-step approach: In the first step, a workshop was conducted to come up with ideas on how the open data might be re-used in “use-cases” (see Activity 5 in Fig. 4). In the second step, the authors rated each of these re-use ideas in terms of their economic impact - assuming the idea would get implemented (see Activity 6 in Fig. 4). The following text describes the approach in more detail:

For the first step, a workshop was carried out within the Task68 group in a hybrid meeting in April 2024, as part of a regular IEA SHC Task 68 expert meeting. In total, 21 participants took part in the workshop with 12 people in person and 9 people online. To set the stage, the participants were presented with a typical lifecycle of a solar thermal plant (see Fig. 5) spanning from customer acquisition to design, construction, operation, and recycling. This was included to serve as a mental guideline to help the participants come up with interesting ideas for data utilization along every phase of a solar thermal plant. Next, the identified public datasets were presented to the participants, grouping together similar kinds of data. In total, 8 different dataset categories were presented (measurement data, plant statistics, partner database, irradiance data, component database, heat demand, and cost data). The participants were also provided with cards including an example image and a short description for each dataset category. After the introduction, the participants were split into 3 groups (2 in-person groups, and 1 online group) and asked to come up with ideas on how the open data could be utilized. This part of the workshop lasted 20 min, instructing the groups to write down the ideas using printed-out templates (see Fig. 5). During the process, participants were allowed to extend the list

Table 5
Statements used to assess the barriers for publishing open data.

Category	ID	Publishing Barrier	Statement
Techno-logical	T1	No Data to publish	Do you have access to data?
	T2	Insufficient Quality	— *Not covered in survey*
	T3	Missing Infrastructure	I don't know where to publish the data
	T4	Security and Protection	— *Not covered in survey*
Strategic	S1	Unclear Demand	I don't think my data is useful to others
	S2	Insufficient gain	I don't see my own benefit in sharing data
	S3	Economic damage	It would negatively affect the competitiveness of my organization
	S4	Leaking critical information	I am worried about leaking critical information
	S5	Lack of resources for preparing	It needs too much time to collect and prepare data
	S6	Lack of resources for publishing	It needs too much time/money to publish the data (e.g., selecting repository)
	S7	Lack of resources for maintenance	— *Not covered in survey*
Regulatory	R1	Ownership and Usage Rights	I am worried about privacy and legal constrains (allowed to share data?)
	R2	Legal Constraints	
	R3	Missing legal frameworks	
	R4	Privacy Constraints	
Operational	O1	Missing Knowledge	I don't have enough knowledge and experience to publish data
	O2	Institutional Barriers	— *Not covered in survey*
Cultural	C1	Lack of Interest / Awareness	The thought of sharing my data openly just never crossed my mind yet
	C2	Fear of misuse	— *Not covered in survey*
	C3	Fear of loss of control	I don't want to lose control about what is done with the data
	C4	Fear of misinterpretation	I am worried about other people mis-interpreting the data
	C5	Fear of Reputation Damage	— *Not covered in survey*

Table 6
Statements used to assess the benefits of open data in the survey.

Category	ID	Benefit	Statement
Co-Creation	B1	Community Growth	I would share data to stimulate the growth of solar-thermal community
	B2	Networking	— *Not covered in survey*
	B3	Inform stakeholder	I would share data to give decision-makers more information (potentially triggering investments)
	B4	Data in Return	I would share data to get data from others in return
	B5	Feedback in Return	I would share data to get feedback from users (e.g., via hackathons, crowd-wisdom, or new services)
	B6	Services in return	
Internal Improvements	B7	Business Culture	— *Not covered in survey*
	B8	Skill development	— *Not covered in survey*
Public Image	B9	Image and Visibility	I would share data to improve the visibility of my institution (e.g., improve image, increase citations)
Compensation	B10	Payment in return	I would share data if I got money for publishing it (e.g., per upload or per download) I would share data if required for funding

Table 7
Barriers reported for hindering the use of open datasets, which were used to rate the identified datasets, including instructions on how they are rated.

ID	Barrier	Instruction for Rating
U3	Findability	Try to find the dataset. First, only with the category description, then with a rough description of the dataset. If both does not work, try the name or other specific keywords.
U4	Accessibility	Once the data portal / source has been accessed, try to download the data.
U6	Format	Once downloaded, try to convert the data into a usable format and access the data.
U7	Quality	Inspect the data – how would you rate the quality of the data? Is it up to date? Is there considerable missing data?
U8	Documentation	Look for meta-data - does the data come with additional information? Is it sufficient to interpret the data?
U9	Licensing	Look for a licence – is it available? Is it clear? Are you allowed to use the dataset?
U10	Support	Did the web portal / source include any option to provide feedback? Does it allow to upload enhanced data?

of existing datasets with not yet published ones, in case they would provide additional leverage for a use case. The instruction was used to allow participants to refer to datasets even if current open data on this topic is limited or currently unusable due to other barriers.

In the second step, the authors qualitatively rated the ideas generated in the workshop, assessing the impact of the re-uses on four stakeholders groups: the data owner (who published the data), the service provider (who utilizes the data for the re-use), the end-user (who is targeted with the re-use) and the solar thermal community (which might benefit as a side effect). The rating was done using a custom 7-point ordinal scale from –1 (small negative impact – possible small disadvantage) to 0 (no impact – no gain, but also no disadvantage) to 5 (extreme positive impact – game-changing benefit). The rating has been performed by three authors independently of each other, using the median as the result. The asymmetric scale was deliberately chosen based on discussions among the authors about which impact could be expected in this domain. The core of the scale from 1 to 5 represents increasing levels of positive impact, while 0 was chosen to best represent a neutral outcome with neither gain nor disadvantage. In addition, −1 was added to the scale to represent the possibility of a harmful/negative impact to any stakeholder. While most reuses are expected to have positive outcomes, it was important to explicitly allow for neutral or even slightly adverse effects, as these may occasionally arise. At the same time, strong negative impacts were considered unlikely in this domain and thus not part of the scale.

3.2.4. Limitations

The barriers and benefits presented within the related work and used for the analysis are derived based on literature search (see Section 2). Due to the nature of the approach and the choice of keywords, important articles might have been missed and not included in the synthesis. Furthermore, the synthesis of the findings was inevitably influenced by the authors' perspective, as decisions on which items to merge and which to treat as distinct often reflects subjective or domain-specific judgements about the relevance and conceptual similarity. Additionally, the initial structure used to classify barriers and benefits used for the survey was originally developed as part of the exploratory literature search. While care was taken to avoid confirmation bias, it is still likely that the process of fitting new insights into the existing framework was influenced by a preference for consistency with the earlier structure.

The surveys presented within this work are based on responses from a limited number of respondents (between 20 and 50 people), which are part of the IEA SHC Task 68 community. Even though the group contains key players in the solar thermal field from many different universities, research centers and companies and from ten different countries, the results still might not be representative for the whole solar-thermal

Fig. 5. Documents provided to the participants during the workshop activity.

community. For the same reason, identified ideas for re-uses and datasets are by no means complete.

The authors of this paper participated themselves in the surveys as well. While participation of researchers would introduce considerable bias in studies of primarily psychological or sociological effects, the aim of the current article was to investigate the spectrum of opinions and experiences of solar thermal experts on the usefulness of open data and on potential barriers, with the goal of getting a picture of the situation as complete as possible. Since the solar thermal research community is small, the exclusion of the authors would have meant a significant reduction of the informative value of this article. Moreover, as the group of authors comprises of Task68 members who are particularly engaged in the topic of open data, excluding them would have led to a selection bias and disproportional representation of the Task68 expert group. In order to minimize the bias still introduced by this decision, results were analysed only after all surveys and workshops were completed.

A few of the authors who participated in the usability rating of the datasets have published datasets themselves. To limit bias, any authors affiliated with a dataset were not allowed to rate their own datasets.

The assessment of the use-case impact and the dataset barriers is only qualitative. To reduce individual bias in the rating and limit group influence, these ratings were performed by multiple authors independently. This approach increases the robustness of the results. Nevertheless, the ratings remain subjective and should be interpreted with caution.

Despite the limitations, we believe the results remain indicative and offer a relevant contribution to the knowledge in this field.

4. Results

This section shows the results of the analytic evaluations of the barriers for using open data (Section 4.1) and the benefits of publishing open data (Section 4.2), as well as the survey on the perception of open data (Section 4.3).

4.1. Evaluation: barriers of using open data

As described in the methodology section (Section 3) the barriers of using open data were derived in a two-step approach by first identifying existing open datasets relevant for the solar-thermal domain, and then performing a rating of the barriers of use for each individual dataset.

4.1.1. Available open data in the solar-thermal domain

The existing open datasets that were identified via a survey, data search, and as part of a workshop (see Methodology Section 3) are depicted in Table 8. In total, 42 datasets were identified, with 14 datasets as part of the survey, 26 additional datasets as part of the author search, and 2 additional datasets as part of the workshop.

The identified datasets are grouped into eight categories based on the type of data they contain: Datasets within the measurement data category (D1) contain sensor data as time series from multiple sensors, often including additional information about components and solar thermal plants. The plant statistics category (D2) contains datasets with various information like installation date, location, size, or other parameters of various plants around the world. The datasets of the partner database category (D3) provide information about contractors, financers, or other companies involved in solar-thermal. The cost data (D4) category refers to datasets providing information about the costs of solar-thermal plants and their components. Datasets within the Component Parameters category (D5) contain information about components that are critical for solar-thermal plants, like datasheets about solar-thermal collectors for example. The irradiation category (D6) encompasses data about irradiation and weather, including statistical, historical, and forecasting data. In the case of irradiation data, the list of datasets additionally found by the authors has been truncated due to the vast number of datasets, selecting the most important datasets based on the discretion of the authors. Hence, the datasets in Table 8 under this category just represent a small subset of available data, while a more complete list of datasets can be found in [41]. In addition, there exists a wide variety of APIs to collect and parse the data. One notable example is the Python library pvlib [42], which allows to collect data from multiple sources. The heat demand category (D7) provides information about heat demand. For example, heat demand densities of cities, existing heating networks, excess heat potentials, and sometimes even the potential for renewable energy technologies such as solar-thermal or PV. Finally, the base map category (D8) includes base map GIS data and satellite images like those provided by Google [43] or OpenStreetMap [44] that are often used in daily life (for example, to locate places or to find routes) but also in work-settings. Examples could be to identify the longitude and latitude of a location, to identify potential areas for placing collectors or storages, or planning logistics. This category was not identified during the survey, but during the use-case workshop (see Activity 3 in the Methodology section).

4.1.2. Barriers of use for the individual datasets

The results of the usability rating (see Section 3 for details) for the identified datasets are shown in Fig. 6. Due to unfamiliarity with specific data formats - particularly those involving geographic information as in the irradiation, heat demand and base map category - one author involved in the rating was unable to reliably assess a subset of the datasets. This observation coincides with the barrier for using data *U5 Missing Skills* (see U5 in Table 2) which emphasize that open data does not necessarily mean that everyone can use the data. To ensure fair and representative results, the ratings of this author were excluded for datasets with which they lacked the relevant data handling expertise. Consequently, the number of raters is reduced from three to two for a

Table 8
Identified public datasets including references. The source shows how the dataset was identified; "S" if the data was identified during the survey, "W" if it was identified during the workshop, and "L" if it was identified by the authors based on the literature.

ID	Category / Dataset Name	Short Description	Reference	Source
D1	**Measurement data**			
	Dronninglund	Measurement data of a solar district heating plant with a pit energy storage in Dronninglund (1 year, 10 min sampling interval).	[45]	S
	HojeTaastrup	Measurement data of a large-scale thermal pit storage in Hoje Taastrup connected to district heating but without a solar-thermal plant (1 year, 10 min sampling interval)	[46]	L
	FHW plant	Measurement data of a collector array installed in a large-scale solar district heating plant in Graz with high-accuracy measurements (1 year, 1 min sampling interval)	[47,48]	S
	PaSTS	Measurement data of multiple domestic solar heating plants in Germany. In addition to sensor data, it also contains flags on whether faults occurred at the plants (between 0 and 7 years, and different sampling intervals from seconds to minutes)	[49,50]	L
	PVT Eisenstadt	Measurement data of PVT collector at a test facility in Germany. (one summer period, 5 s sampling interval)	[51]	L
D2	**Plant Statistics**			
	SHIP plants	Interactive world map showing solar heat in industrial processes (SHIP) plants all around the world, displaying general information, technical parameters, information about the process and economic parameters of plants	[52]	S
	Solarheatdata.eu	Interactive world map listing solar district heating plants in Denmark and France. In addition to plant statistics (manufacturing year, manufacturer, collector area, etc.) it also allows to download the measurement data of selected plants in up to hourly resolution.	[53]	S
	Solvarmedata.dk	The Danish version of solarheatdata.eu.	[54]	S
	SDH EU	Plant database from Solar District Heating EU, focusing on solar district heating plants and providing basic data about the plants.	[55]	S
	Solar Wärmenetze	Interactive world map showing solar district heating plants in Germany, also including projects in development.	[56]	S
	NREL CSP	List of concentrating solar power (CSP) plants including technical, economic, financial and industrial data.	[57,58]	L
	SolarHeatWorldwide	Solar Heat Worldwide report showing global market development and trends of solar-thermal.	[59,60]	L
	REN21 global status report	Renewable global status report from REN21 containing statistics about market growth and installed capacity.	[1]	L
	Solaratlas	Statistics on installed solar thermal plants (collector area, number) in Germany based on subsidiary applications from the market incentive program (MAP).	[61]	L
D3	**Partner database (Contractor, Advisor, Financier, Manufacturer)**			
	RTC database	Webpage listing service providers and financers of renewable energy technologies including solar-thermal	[62]	S
	MCS database	Webpage listing contractors in the UK holding an MCS certification	[63]	S
	Solar Payback	An interactive map showing suppliers of solar process heating systems (SHIP) worldwide.	[64]	L
	SDH EU professionals	Webpage from the Solar District Heating SDH initiative, listing business partners, suppliers, consultants, and service providers in the field of solar district heating.	[65]	L
D4	**Cost data**			
	EU Study	EU study on long-term (2050) projections of technical and economic performance of large-scale heating and cooling, including cost functions for multiple energy supply technologies including solar-thermal.	[66]	L
D5	**Component properties**			
	ESTIF Database	Webpage listing solar keymark certificates for collectors, tanks, systems and controls hosted by Solar Heat Europe (ESTIF)	[67]	S
	CEN Database	Webpage listing any keymark certificates including solar-keymark certificates hosted by the Keymark Management Organization owned by CEN.	[68]	L
	TÜV Rheinland	Webpage listing certificates issued by DIN CERTO Germany including some solar-keymark certificates.	[69]	L
D6	**Irradiation data/forecasts**			
	Global Solar Atlas	Webpage including an interactive world map showing DNI, Global horizontal irradiation, temperature, and PV potential. (TMY data, global)	[70]	S
	Digitaler Atlas Steiermark	Interactive map of Styria including yearly global irradiation, ambient temperature, potential for solar but also other data like available district heating networks, land use designation, bodies of water, etc. (TMY data, Styria)	[71]	S
	PV-GIS	Interactive world map showing irradiation and ambient data based on different satellite datasets (e.g. ERA5, CM SAF) allowing to export up to hourly values. (historic data, global)	[72]	S
	CM SAF	Satellite records for climate and irradiation data covering EU, Africa and the Atlantic Ocean in high resolution and for many years (historic data, global but focus on EU & Africa).	[73]	L
	NSRDB	Satellite data primarily focusing on America including climate data and solar radiation. An interactive world map also allows access to other datasets like METEOSTAT and Himawari (historic data, global)	[74]	L
	ERA5	Reanalysis data combining satellite data and observations from across the world into more complete datasets including climate and irradiation data (historic data, global).	[75]	L
	CAMS	Satellite data including aerosol optical properties for high-resolution climate information and forecasting (historic & forecasting data, EU).	[76]	L
	Solarkataster NRW	Multiple datasets for Nordrhein-Westfalen including solar potential and yearly irradiation (TMY, Nordrhein-Westfalen)	[77]	L
	DWD Open Data Server	Open data portal of the Deutscher Wetterdienst providing historic irradiation data from German weather stations (historic, Germany)	[78]	L
	KNMI Open Dataplatform	Open data portal of Koninklijk Nederlands Meteorologisch Instituut providing climate data from weather stations in the Netherlands (historic, Netherlands)	[79]	L
D7	**Heat demand data**			
	Peta 5	Pan-European Thermal Atlas (PETA) is an interactive world map showing heat demand densities, excess heat potential, prospective district heating networks, and the availability of renewable energy sources such as geothermal, solar radiation, and biomass. Data can be downloaded at the sEEnergies Open Data hub. (EU)	[80,81]	S

(*continued on next page*)

Table 8 (*continued*)

ID	Category / Dataset Name	Short Description	Reference	Source
	Hot Maps	Interactive world map showing heat density, cooling density, building information, excess industry heat, and renewable potential including solar radiation on buildings and solar-thermal potential (EU)	[82]	L
	THERMOS	Software for district heating network optimization allowing to estimate of heat demand on building level for selected regions (global)	[83]	L
	Digitaler Atlas Steiermark	Interactive map of Styria (Austria) allowing to visualise multiple datasets including land register, land use designation, water bodies information, geological information, but also information about existing district heating networks, pipelines, electricity production and networks as well as a dedicated layer for solar potential on roofs. (Styria)	[71]	L
	Fernwärme Atlas	Interactive map of Germany showing district heating networks including heat supply, and temperature. Furthermore, it allows to show the area needed for solar-thermal fraction up to 15 %, the potential of industry waste heat (Germany)	[84,85]	L
	Thermische Netze	Interactive map of Switzerland showing multiple datasets including heating and cooling demand, existing thermal networks and energy sources. It also provides horizontal and inclined irradiation. (Switzerland)	[86]	L
	Austrian Heatmap	Interactive map of Austria, focusing on district heating. The map shows heat demand for 2021 and extrapolations for 2030 and 2050. In addition, it shows heating grids, energy sources (not including solar-thermal) and industry waste heat potential. (Austria)	[87]	L
	Open modelling of electricity and heat demand curves (DE)	Data estimating electricity and heat demand curves for buildings in Germany based on OpenStreetMap, ERA5 irradiance data, Peta5 heat demand, and based on statistical demand profiles for different types of households (VDI 4655) as input. (Germany)	[88]	L
D8	**Base Maps**			
	Google Maps	Provides (restricted free) API use for satellite images and base map data such as roads, buildings, navigation, place search etc.	[43]	W
	Open Street Maps	Provides community-driven and free API for satellite images and base maps.	[44]	W

few datasets. In addition, ratings were excluded if an author was involved in publishing the respective dataset (see Section 3.3 on limitations) affecting the rating of the FHW plant dataset. The exact number of authors rating the respective datasets is indicated in Fig. 6.

The results in Fig. 6 show that different barriers apply to different datasets. However, there are also distinctive patterns for some data categories, as described in the following: For measurement data (D1), the scores are high with almost no barriers to use. Most of the datasets are available on Zenodo, which supports findability and ensures that a licence is selected. With most datasets being covered by a publication or includes readme files the documentation is sufficient. The only two exceptions receiving moderate barriers are the PVT dataset which is provided in a MATLAB format[2] which is not easy to parse without the fee-based tool and the PaSTS dataset which does not provide an email address or other directions to receive support.

On the other hand, the plant statistics (D2) received relatively low scores indicating that data is *almost impossible to use*. The most evident issue is the *U9 Licencing* of the data, with most of the datasets not specifying if re-use of the data is permitted. Furthermore, the *U4 Accessibility* of the data also received low scores, as the data is often represented as part of a web interface, with no or only hidden options to download the data. The only exceptions with *somewhat of a barrier* are ship-plant.info and the NREL database, which provide rather clear licences. A similar observation is also made for partner data (D3) and for solar-thermal component data (D5) which also received low scores, especially for *U9 Licencing* and *U4 Accessibility*. In the case of the solar Keymark databases, the main issue with *U4 Accessibility* is that component data can only be exported via PDF but not in a machine-readable format. In the case of the cost category (D4) with only one entry, moderate scores were achieved for most barriers. However, it seems this dataset is hard to find resulting in a particularly low score for *U3 Findability*. Finally, the irradiation datasets (D6), the heat demand data (D7), and map data (D8) received relatively high scores – possibly influenced by the fact that the data is relevant to multiple disciplines and often hosted by large public organizations. However, there are exceptions for some datasets. For example, the local Digitaler Atlas Steiermark [71] providing data about Austrian solar potential and irradiation, during the study period, only allows to use the data for research and prohibit commercial use.

In summary, the evaluation indicates that the use barriers of the open data differ based on the data category. While a few categories show generally high usability (measurement data, irradiation data, heat demand, base maps), other categories (plant statistics, partner data, costs, and component data) are hardly usable, with the most evident barriers being *U9 Licensing* and *U4 Accessibility*. In addition, the economic aspects are underrepresented (*U1*) based on the number of datasets found, or hard to find (*U3*), which can be interpreted as a hindering factor for substantial techno-economic evaluations such as feasibility studies.

4.2. Evaluation: benefit of open data

As described in the methodology section (Section 3) the benefit of open data was evaluated by a two-step approach, first identifying potential interesting use-cases for reuse in a workshop, and then performing a rating of the benefits of use for each individual reuse example.

The results of the workshop for identifying potential re-use ideas based on the available datasets are shown in Fig. 7, while a detailed description of the ideas is provided in Table 9. In total, 21 re-use ideas were contrived by the participants during the workshop. By grouping together similar ideas, the list was reduced to 11 distinct re-use ideas (UC1-UC11). Most of them were found for the acquisition stage, the design & planning stage, and the operation stage, while only one idea was generated for the manufacturing & construction stage and no re-use was contrived for the recycling stage. The most often used dataset categories that are required for re-uses are the measurement data (D1), plant statistics (D2) and cost (D4) categories.

The impact rating of each re-use idea (see Section 3 on methodology for details) is shown in Fig. 8. The scores range from 0 (no impact) to 4 (considerable impact). For most use-cases, the impact of the re-uses on at least one stakeholder is at the level of 3 (moderate positive impact) or higher.

There are notable differences in the impact of the re-use ideas on the different stakeholders. More specifically, the re-uses seem to have the most impact on the service provider (median 3) while the end-users benefit to a lesser degree (median 2). This might be explained as the service provider can offer their solution to many users and can adjust

[2] During the proofing stage of this article, the dataset was additionally made available in CSV format.

their prices according to the benefit of the end user. Hence, the impact on service providers is often rated equal or a bit higher compared to the end user. Furthermore, most re-uses were estimated to have a small to considerable impact on the community (median 2). Especially the use-cases advertising (UC1) and identifying high-potential locations (UC3) were rated to have a high impact as they might influence stakeholders, policy creators, and investors on a larger level. While cost estimates are seen as very useful (UC6), there are doubts that they can be obtained from open data. In contrast to the other stakeholders, the data owner often has low or even no benefit from the open data (median 1.0). The exceptions are especially the advertisement of solar-thermal plants (UC1), showing similar plants (UC2) and identifying potential partners (UC7), in which cases the data owner might have a small impact due to increased visibility. Apart from that, the data owner mainly has only small benefits due to feedback on the data (UC4, UC6, UC8, UC9, UC10) or no benefit at all (UC3, UC5, UC11).

In summary, the estimates by the expert group further support the hypothesis that the open datasets could have a considerable (but not game-changing) impact and could lead to new services and re-uses. In particular, the impact is not only high for the end-user and service provider but also could lead to considerable benefits for the solar-thermal community. On the other hand, it may also indicate that data owners do not necessarily benefit as much from sharing the data.

4.3. Survey on open data

In total, 32 experts from the IEA SHC Task 68 working group took part in the survey (see Section 3 for details), with most of them being experienced with solar thermal (25 out of 32 people having >6 years of experience) and a majority of respondents are researchers (21 out of 31 people). Approximately half of the survey participants (14 out of 32) reported that they have published open data. The results are summarized in Fig. 9.

4.3.1. Benefits of open data

Focusing on publishing open data, the results in Fig. 9 indicate that the benefits for sharing data are convincing to the audience, with many statements receiving high agreement (5 – strongly agree). A majority of the respondents (87 %) considered *if needed for funding* [B10] a motivation for publishing data publicly. In addition, there are also less direct benefits like *helping the community grow* [B1], *informing stakeholders* [B3], and *increasing Image and Visibility* [B9] which received high scores. Interestingly, the lowest (but still modest) scores were obtained for *receiving data* [B4], *feedback* [B5-B6], or *monetary gain* [B10] in return. The options *required for funding* and *get money in return* received not only positive or neutral feedback, but both also received strong negative ratings (1 – Strongly disagree) from a few participants. This could indicate that offering money could even prevent some people from sharing data. This has also been reported in a use case in Switzerland where people were less motivated to store toxic waste after being offered a compensation compared to receiving no money [89].

4.3.2. Barriers for using open data

The survey results on barriers for using open data are mixed, with an average score of 2.9, with low values indicating strong barriers (see Fig. 9). The most evident barrier with the lowest score is *Missing Data* [U1] (average 1.8) with almost 85 % of the participants believing that existing data is not sufficient. The second most critical barrier, as reported by respondents, for using open data seems to be *Findability* [U2] (average 2.3). On the other hand, *Missing Gain* [U11] and *Missing Awareness* [U2] seem to be less relevant barriers since almost all respondents believe that open data is beneficial, and half of them agree that it is a game-changer, and approximately 40 % have already used open data. Most other barriers received both positive and negative scores with average values between 2.6 and 3.0. The results show no obvious correlation between the barriers to the occupation (researcher versus non-researcher), experience of the participants in the solar-thermal domain, or the publishing experience of the participants (already published open data versus not yet published open data).

4.3.3. Barriers for publishing open data

The results for the barriers to publishing open data are also mixed: A few respondents perceive them as strong barriers (5 – Strongly agree), while others do not perceive them as a barrier at all (1 – strongly disagree). Looking at the scores per participant (see Fig. 11), the results show that almost every survey participant (94 %) rated at least one barrier with a score of 4 or higher. This indicates that the barriers to publish data could differ based on a group- or on an individual level. In particular, Fig. 10 compares the scores for researcher and non-researcher – while researchers are mostly wary about the effort in *preparation* [S5] and *publishing* [S6], the non-researcher (operator, designer, manufacturer) are rather concerned about *leaking information* [S4], *losing control [C3]*, or *misinterpretation [C4]* of their data. However, both groups seem to be most wary of *privacy and legal constraints* [R1-R4]. Hence, in summary, while motivation to share data is high, there seem to be considerable barriers preventing publishing data.

As part of the survey, participants were also asked how to overcome the most critical barriers to publishing open data. The participants argued that best-practice examples and showing proof of the benefit (5 occurrences) might motivate data owners to publish their data. In addition, a few participants suggested that guidelines (1 occurrence) and a common data portal for publishing (5 occurrences) might help to reduce the time and effort required for publishing. Another frequent comment (5 occurrences) was that requiring open data to receive funding could help to increase the willingness to share data. Finally, a few participants suggested providing data owners with specific incentives for sharing data, for example providing money or access to data in return (1 occurrence each).

5. Discussion

The findings of the activities provide the basis for addressing the research questions as stated in Section 1:

> Q1 – What are the most important barriers for using open data in the solar-thermal domain?

Based on the survey with the solar-thermal experts (see Section 4.3), the most critical barrier to using open data is *Missing data* [U1]. Combined with the fact that *Missing Gain* [U10] seems not to be a barrier as per the solar-thermal experts, this indicates that the experts would appreciate more open data in the domain. This highlights the importance of understanding the barriers and benefits to publishing open data, as covered in Q2 and Q3.

As a prerequisite of the evaluation, the study identified 42 existing open data sets that are relevant to the solar-thermal community, which could be grouped into 8 different categories (namely measurement data, plant statistics, partner database, irradiance data, component database, heat demand, and cost data). This indicate that a considerable amount of open data relevant to the solar-thermal domain is already available. The survey results also show that about 40 % of the participants already use open data in their daily work, indicating that the community is largely aware of open data. This is in contrast to the finding mentioned before, stating that open data is missing. However, it might be explained by the fact that some datasets that are crucial for the solar-thermal community are either missing, or not usable due to other barriers. In particular, both the impact rating (see Section 4.2) and cost data (D4) could be of exceptional value to the solar-thermal community and policymakers by showing the potential of solar-thermal and increasing its visibility. However, we could only identify one dataset specifically focusing on costs, which however faces severe barriers in terms of use. In combination, this could explain the call for more open data at least in the case

Category	Dataset	Minimum Score	Findability	Accessibility	Format	Quality	Documentation	Licensing	Support	Nr of reviewer
			Barriers (Average Rating)							
D1 - Measurement data	Dronninglund	4	5	5	5	4	5	5	4	3
	HojeTaastrup	4	4	5	5	5	5	5	4	3
	FHW plant	4	4	4	4	4	5	5	4	2
	PaSTS	3	4	4	5	4	5	4	3	3
	PVT Eisenstadt	2	5	5	2	5	5	5	4	3
D2 - Plant Statistics	SHIP plants	3	4	4	5	3	3	3	4	3
	Solarheatdata.eu	2	4	2	4	4	4	3	5	3
	Solvarmedata.dk	2	4	2	3	4	3	2	4	3
	SDH EU	3	4	5	5	3	5	3	4	3
	Solar Wärmenetze	2	2	2	2	4	3	2	3	3
	NREL CSP	3	3	5	5	4	4	5	5	3
	SolarHeatWorldwide	2	4	3	3	4	4	2	3	3
	REN21 global status report	1	4	4	5	3	4	1	4	3
	Solaratlas	3	3	3	4	3	4	3	4	3
D3 - Partner Database	RTC database	1	3	3	3	4	4	1	4	3
	SDH EU professionals	1	5	3	3	4	4	1	4	3
	MCS database	1	4	4	3	4	4	1	4	3
	Solar Payback	1	4	4	3	4	4	1	4	3
D4 - Cost Data	EU Study	1	2	5	4	3	4	1	4	3
D5 - Component Properties	ESTIF Database	1	5	3	3	5	4	1	4	3
	CEN Database	1	5	3	4	5	4	1	4	3
	TÜV Rheinland	1	3	3	4	3	4	1	4	3
D6 Irradiation Data	Global Solar Atlas	4	5	5	4	5	5	5	5	2
	PV-GIS	4	5	5	5	5	5	4	4	2
	Digitaler Atlas Steiermark	2	3	3	3	4	3	2	4	2
	CM SAF	3	4	3	3	5	5	5	5	2
	NSRDB	3	5	5	5	5	5	3	4	2
	ERA5	3	4	3	4	5	5	5	4	2
	CAMS	4	4	4	5	5	5	5	4	2
	Solarkataster NRW	4	4	5	5	5	5	4	4	2
	DWD Open Data Server	3	4	3	4	4	4	4	4	2
	KNMI Open Dataplatform	4	4	4	4	4	5	5	4	2
D7 - Heat Demand Data	Peta 5	3	4	4	5	5	4	3	4	2
	Hot Maps	4	5	5	5	4	5	5	5	2
	THERMOS	3	4	3	5	4	5	3	5	2
	Digitaler Atlas Steiermark	2	3	3	3	4	3	2	4	2
	Fernwärme Atlas	2	3	2	3	3	3	5	4	2
	Thermische Netze	4	4	4	4	4	5	4	4	2
	Austrian Heatmap	2	4	5	4	5	4	2	4	2
	Open modeling of [...]	4	5	5	4	5	5	5	4	2
D8 - Base Maps	Google Maps	3	5	4	4	5	3	4	4	2
	Open Streen Maps	4	5	5	5	4	5	5	5	2

Color Legend	
1	Extreme Barrier (not usable)
2	Serious Barrier (almost not usable)
3	Moderate Barrier (difficult to use)
4	Somewhat a Barrier (minor difficulties)
5	Not a barrier (no difficulties)

Fig. 6. Result of the rating of the datasets in terms of their usability by the authors. The column on the outer right shows how many authors rated the dataset. The evaluation was conducted in February 2025 - and the usability of datasets may have changed since then.

of plant-statistics and cost data.

Other important barriers as per the survey are *Findability* [U3], *Documentation* [U8], and *Data Quality* [U7]. However, most of the data-related barriers received similar scores with mixed results. The reasons for the mixed results might be because the barriers are either perceived differently by individual respondents or because the statements were difficult to rate due to the barriers being different for different types of datasets. Compared with the rating by the authors (see Section 4.1), the barriers for using the datasets vary significantly for individual datasets. This may hence explain why the perceived barriers showed inconclusive results during the survey, as it addressed data usage in general rather than evaluating specific datasets. This observation also supports the argument of Janssen et al. [22] who suggest that barriers should be analyzed on a case-by-case basis.

Nevertheless, while barriers of using open data varied significantly for the different data sets based on the author assessment, specific patterns emerged across different categories of data. While some categories show generally high usability (measurement data, irradiation data, heat demand, base maps), other categories (plant statistics, partner data, costs, and component data) are hardly usable, with the most evident barriers being *Licencing* [U9] and *Accessibility* [U3].

While overcoming the barrier of *Missing Data* requires that more data is published (Q2, Q3), existing literature covers recommendations for data portals, ecosystems, and what guidelines and measures could help

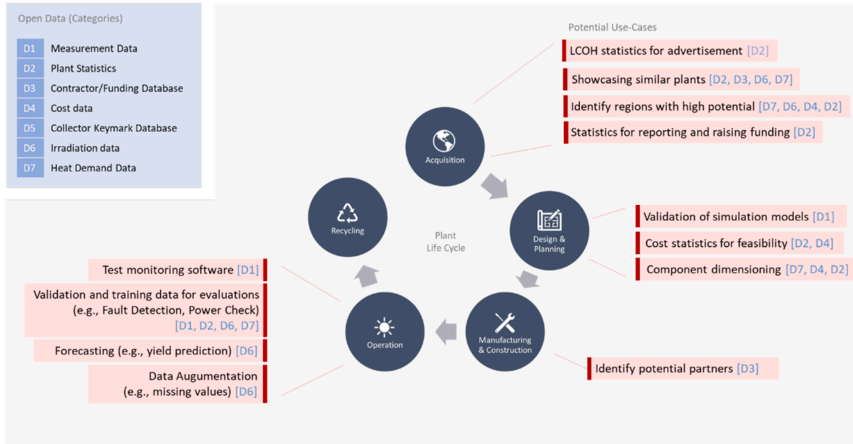

Fig. 7. Overview of identified reuse ideas (red blocks) and their required datasets (blue notes) along the lifecycle of a solar-thermal plant.

to mitigate them [5,19,35,90], for example by applying the FAIR principle (Findable, Accessible, Interoperable, Usable) as part of publication. In the case of specific data sets, correction might be comparatively easy by adding licenses and clarifying terms of use (e.g. addressing partner database (D3), costs (D4) and component database(D5) and plant statistics (D2)) and supporting the export of data (addressing mainly plant statistics (D2)).

Q2 – What are the most important barriers for publishing open data in the solar-thermal domain?

Based on the survey (see Section 4.3), the barriers for publishing data are mixed and vary for each participant. However, almost all respondents (94 %) strongly agree to at least one barrier, which indicates that the barriers very likely have a considerable influence in hindering the publishing of open data.

The survey results also show a difference between barriers as perceived by researchers and non-researchers (see also Fig. 11): While for both groups *privacy and legal constraints* [R1-R4] are perceived as the most evident barrier, researchers seems to be concerned about the *effort for preparation* [S5] and *publishing open data* [S6], while private organizations and companies are additionally concerned about *leaking information* [S4], *incorrect interpretation* [C4], *loss of control* [C3], and *no-own benefit* [S2]. The results also indicate more barriers for private companies compared to researchers.

The survey results also include suggestions by the solar-thermal experts on how to overcome the most important barriers and promote the sharing of open data, as also depicted in Table 10:

- **Guidelines for publishing data** - One suggestion is that guidelines and best practice examples could help to mitigate barriers. For example, the effort for preparation [S5] and publishing [S6] might be reduced due to instructions and tips. Guidelines might also increase the confidence that no major (e.g. legal) topic has been missed [R1-R4], speeding up the publishing process and mitigating fears.
- **Common Data Portal –** A common data portal was also suggested by the survey participants as it might improve the ease of publishing [S6].
- **Require for Funding -** Some participants in the survey suggested that publishing open data should be facilitated or even required for receiving funding. This might present a clear benefit for data owners who are willing to share their data, potentially mitigating fears associated with data sharing. In addition, it could also encourage collaboration within project consortiums, exchanging knowledge on data preparation and publication and might allow both research and industry to specifically designate resources for publishing data. However, as one participant noted in the survey, such a requirement might also lead to bad-quality datasets being published to avoid unintended re-use of the data.
- **Show proof of benefits -** Another suggestion based on the survey was to show the benefits of open data to the data owners. Similar to receiving funding, this might help to justify resources to be spent for publishing and balance out fears of data sharing. However – as shown by the impact rating – the economic impact of open data on data owners is low. While this is less critical for government organizations, it could be a major barrier for industry and research. To circumvent this problem, a recommendation could be to deliberately leverage the effects of more indirect benefits: For example, by combining data-sharing with a specific data re-use (e.g., leveraging benefit *Service in Return* [B6]), data owners would not only benefit from sharing the data but could also act directly as end-users or even service providers themselves, resulting in a higher economic impact. Similarly, datasets could be published as part of challenges, actively asking for collaboration on specific topics that are of special interest to the solar community (e.g., leveraging benefit *Feedback in return* [B5]), to collaboratively gather data [B2, B4] or used deliberately to inform stakeholder [B3] and increase the growth of the community [B1]. While this facilitates the utilization of open data, it could also largely increase the visibility and image of the involved data owners [B9].

Q3 – What are the most important benefits for publishing open data in the solar-thermal domain?

Based on the impact rating of the authors (see Section 4.2), open data might yield considerable benefit for the solar-thermal sector – especially to data re-user and service providers, but also to the solar-thermal community as a whole. This coincides both with claims in literature [9] as well as with the survey results on the barriers of using of open data (see Section 4.3) which indicates that open data is beneficial but not game-changing. However, the impact rating also shows that data-owners, who need to publish the data, receive fewer benefits

Table 9
Reuse examples identified in a Task 68 workshop including a short description for each idea, the number of original ideas that were merged together (denoted # in the table header), and the used dataset categories which are utilized within the idea. Data Categories that are only secondary or optional to the re-use idea are marked with brackets.

Project Stage / Re-Use Idea	Description	#	Used Dataset Categories
Acquisition stage			
Plant statistics for advertisement	This re-use aims to use plant statistics data (D2) and costs (D4) especially to investigate the levelized cost of heat (LCOH) of implemented plants. This information is highly critical for assessing the feasibility of solar-thermal plants and essential to compare with other energy production technologies. Showing multiple realized plants might also help to prove the maturity of the technology to the customers. Hence, using this data might help to advertise the use of solar thermal, and raise the awareness of potential customers and policymakers. However, care must be taken while comparing the data as different local conditions might influence LCOH and lead to high variations and uncertainty of the results.	2	D2, D4
Showcasing similar plants	In this use case, the plant statistic data (D2) is used together with irradiation data (D6), and heat demand data (D7) to identify plants that are similar to the demands of a new potential customer. By showcasing already implemented plants, the economic feasibility and technological readiness of solar thermal can be easily proven to the customer. In addition, analysing already implemented similar plants might help to identify opportunities, hurdles and how they can be overcome. Cost data (D4) might be of help to doublecheck and enhance the results.	2	D2, (D6, D7, D4)
Identify high-potential locations	This re-use utilizes many datasets that are mostly connected to spatial data (heat demand D7, irradiation data D6, cost data D4, geographic data D8) and aims to identify locations that are of particular interest for solar-thermal plants. For example, spatial data might be used to identify potential locations for a Big-Solar project with large-scale pit storage, based on geological parameters, irradiation, performance, costs, and checking for existing heating networks. Similarly, the data might also be used to quickly verify if a plant can be feasible at a potential location – for example letting the customer use a web application. Such examples have already been implemented for photovoltaics (e.g., https://energy3000.com/), and also by one solar-thermal collector manufacturer (https://www.absolicon.com/fs/).	1	D7, D6, D4, D8
Design & Planning stage			
Validation of simulation models	In this re-use, the plant operational/measurement data (D1) is a crucial asset for enhanced reliability in the design and planning process. Despite the existence of numerous design tools, it is imperative to assess the predictive accuracy of models to mitigate risks in the planning and development process. In this regard, it is essential to ascertain the plausibility of the models and to validate the calculation methods and their results based on measurement data from real systems.	7	D1
Plant Dimensioning	Sizing plant components and choosing the ideal technologies to provide energy is not a trivial task. For this task time plays a large role in generation (e.g. solar following irradiation), demand (e.g. heating demand following ambient temperature), and prices (e.g., fluctuating electricity prices) which influences technologies used and storage used. In this re-use example, the heat demand (D7) and weather data (D6) are collected to automatically fill in this information and dimension the plant accordingly. Data from similar plants (D2) and costs (D4) may help to size also in terms of techno-economic optimum.	2	D7, D6, (D4, D2)
Estimating system costs	This re-use utilizes plant statistics (D2) and costs (D4) to estimate the cost of a plant. This could be useful for example in the early stages of planning, to quickly check the feasibility of a plant, based on similar existing plants (D2), or based on estimated system or component prices (D4).	3	D2, D4
Manufacturing & Construction stage			
Identify potential partners	This re-use aims to find partners for constructing/planning solar-thermal plants. For example, a company focusing on solar-thermal collectors might want to cooperate with a local storage manufacturer or an energy planner. Using a partner database (D3) would allow to scan for potential partners.	1	D3
Operation stage			
Validating Monitoring Software	Measurement data is the main input for monitoring tools. During the development of the software, hence, it is important to ensure that the data import works fine and that algorithms can work on the data correctly – which is done by testing the software. While artificial data and data from other domains might be helpful for testing this kind of functionality, domain data often contains special characteristics or standards. Hence, solar-thermal measurement data (D1) can be of great help for testing monitoring software and analysing the special needs of solar-thermal data.	1	D1
Data Imputation / Completion	For plant controlling, sensor information is important to adapt the control accordingly. However, if sensors break this can lead to problems, where crucial information is missing to the control. Using measurement data (D1) one can look for relations in the data so values can be estimated ("imputed") by different sensors in case of a sensor break. In addition, it may allow to install less measurement equipment in case relations can be approximated sufficiently from other sensors ("soft sensors", "sensor fusion"). Open weather data (D6) might be useful to impute irradiance in case it is not part of the measurement data.	1	D1, (D6)
Anomaly Detection	To ensure the performance of solar-thermal plants over their long lifetime, fault detection or anomaly detection might be used to monitor the plant and identify any issues. Public measurement data (D1) might allow the development of such methods and testing them on real plants. Similar to the use case above, open weather data (D6) might be useful to augment irradiance data	1	D1, (D6)
Yield Prediction	Measurement data (D1) and plant statistics (D2) contain reliable information about solar thermal yield. Hence, it can be used to develop and test solar-thermal yield prediction algorithms. Open weather data (D6) might be useful to augment irradiance in case it is not part of the measurement data, while collector parameters (D4) might be used for estimates or to compare the results.	4	D1, (D2, D4, D6)

ID	Use Case	Maximum Score	Minimum Score	Impact on: End User	Service Provider	Data Owner	Community
UC1	LCOH statistics for advertisement	3	2	2	3	2	3
UC2	Showcasing similar plants	2	2	2	2	2	2
UC3	Identify high-potential locations	4	0	2	1	0	4
UC4	Validation of Simulation Models	3	1	2	3	1	3
UC5	Plant Dimensioning	3	0	2	3	0	2
UC6	Estimating system costs	3	0	2	3	0	2
UC7	Identify potential partners	2	1	1	2	2	2
UC8	Validation of Monitoring Software	3	1	2	3	1	3
UC9	Data Imputation	2	1	2	2	1	2
UC10	Implementation of Algorithms	3	1	3	3	1	3
UC11	Forecasting	3	0	3	3	0	2
	Average	2.8	0.8	2.1	2.5	0.9	2.5
	Median	3.0	1.0	2.0	3.0	1.0	2.0
	Standard Deviation	0.6	0.8	0.5	0.7	0.8	0.7

Color Legend	
5	Extreme Impact
4	Considerable Impact
3	Moderate impact
2	Small impact
1	very small impact
0	No Impact
-1	Negativ Impact

Fig. 8. Results of rating on the impact of the use cases on different stakeholders of the open data. The values show the median values of the three authors who participated in the rating.

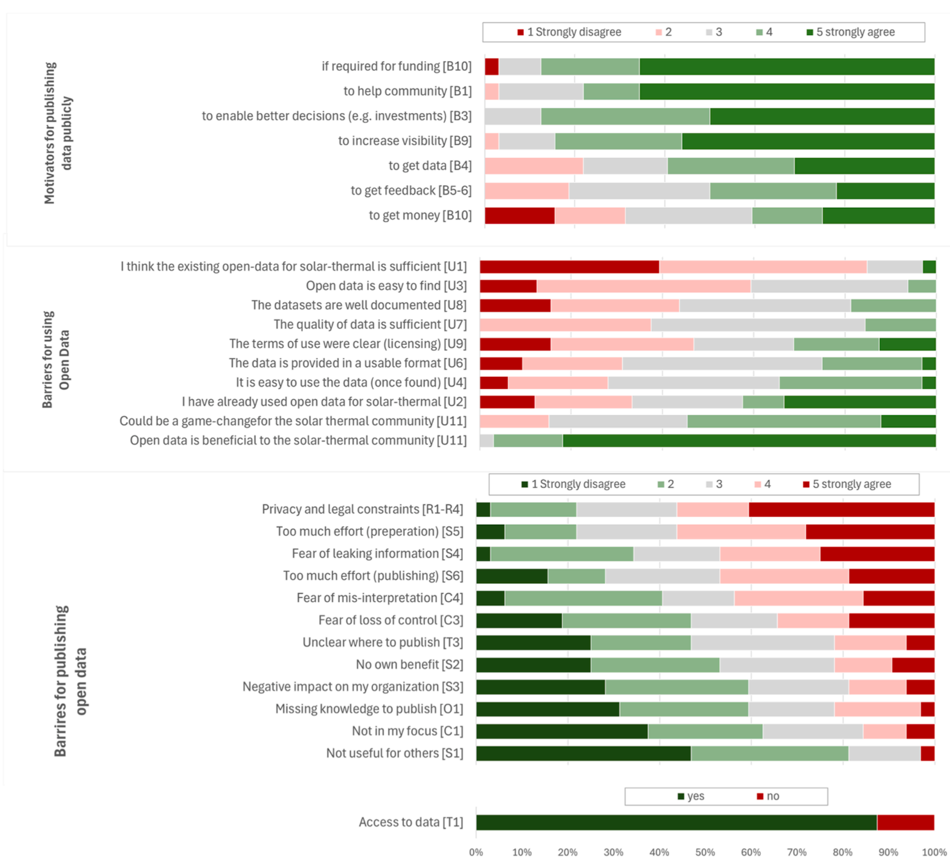

Fig. 9. Results of the survey about the perception on open data.

compared to the other stakeholders.

In contrast, the survey results (see Section 4.3) on the perception of open data indicate that the benefits of open data are evident to the solar-thermal community. Receiving funding [B10], improving visibility and transparency [B9], and supporting the solar-thermal community in general [B1] were rated as the most important motivators for the survey participants. In contrast, more direct benefits like getting data [B4] feedback [B5], or monetary return [B10] seem to be less relevant in

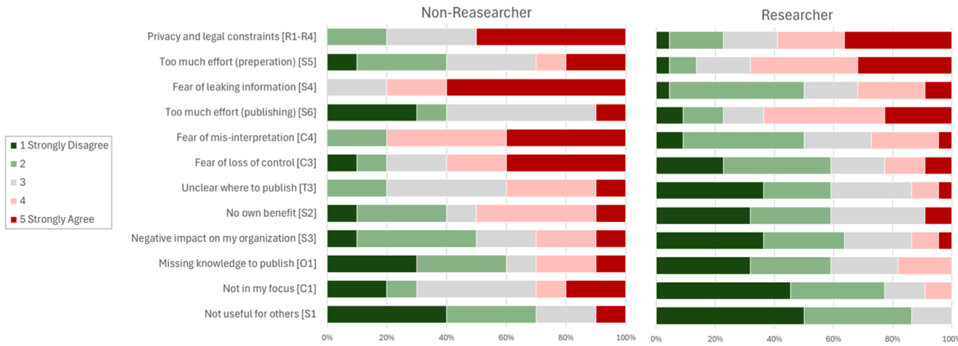

Fig. 10. Results for "What keeps you from sharing data?" with the bars representing the percent of participants picking the corresponding answer and distinguishing between researchers (21 respondents) and non-researchers (11 respondents).

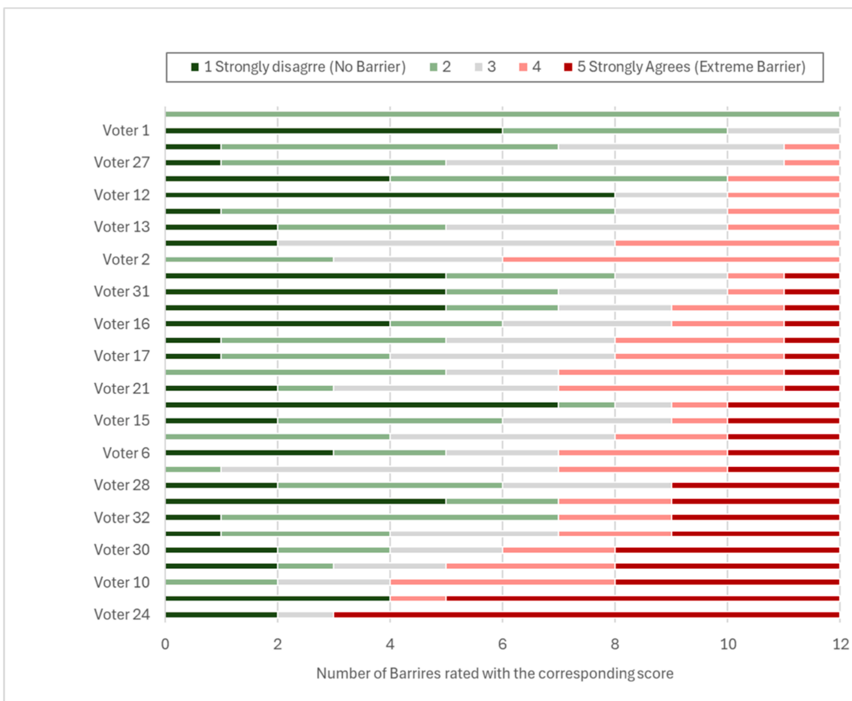

Fig. 11. Individual rating of barriers for each participant. The x-axis shows how many barriers were rated with the respective score from 1 (not a barrier) to 5 (extreme barrier).

Table 10
Important barriers identified for publishing data and corresponding mitigation strategy ideas.

Barrier	Average scores for researchers	Average scores for non-researchers	Direct Mitigation Strategies	Indirect Mitigation Strategies
Effort for publishing [S6]	3.8	–	data portal, guidelines	Proof of benefits, funding
Privacy and legal constraints [R1-R4]	3.7	3.7	Guidelines	
Effort for preparation [S5]	3.6	3.1	data portal, guidelines	Proof of benefits, funding
Fear of leaking Information [S4]	–	4.2	Guidelines	Proof of benefits, funding
Fear of misinterpretation [C4]	–	3.8	Guidelines	Proof of benefits, funding
Fear of loss of control [C3]	–	3.6	Guidelines	Proof of benefits, funding
No own benefit [S2]	–	3.0	Proof of benefits, funding	

comparison.

Receiving funding and direct money in return are perceived differently, although both are related to a direct economic impact [B10]. A possible reason might be that participants would not believe that selling the data directly would yield as much revenue compared to receiving funding as part of research calls. Another possibility is that the benefit of funding is rated higher because this will allow data-owners to increase their network and enable them to participate in funded research programs, hence addressing the benefit Networking [B2] instead of monetary gain [B10].

6. Conclusion

This study aimed to analyze the status of open data in the solar-thermal sector - analyzing barriers of using open data, as well as assessing the barriers and benefits of publishing open data. Data has been gathered both via surveys with domain experts from the IEA SHC Task 68 group, as well as through evaluations by the authors who are also part of the Task 68 expert group.

The results support the common belief that open data is useful, offering considerable benefits to data re-users and the solar thermal community. The study also showed that considerable open data relevant to the solar-thermal community already exists. However, some datasets like cost data and plant statistics, which might be of exceptional importance for re-use, are hard to utilize due to licensing, accessibility, and quality issues. This could be addressed by raising the awareness of service providers and data owners - improving the usability of the datasets by adjusting licenses and formats as well as facilitating cooperation with end users and applying guidelines derived from research on open data portals and reusability of datasets.

In addition, crucial datasets seem to be missing, based on the survey results, especially concerning data about costs – which might be of high strategic value for the whole solar-thermal community due to its impact on the adoption of solar-thermal energy.

While the analysis shows that data owners benefit less from publishing data in terms of direct economic income, the less direct benefits of open data nevertheless seem appealing. However, severe and diverse barriers exist for every participant of the survey, hindering the publication of open data. For example, concerns over leaking information, misinterpretation, loss of control, or no own benefit are especially prominent among private companies, while researcher are more concerned about resources for preparing and publishing. For both groups, the main barriers were privacy and legal constraints.

To address these challenges, providing clear guidelines and establishing dedicated repositories could improve the quality of open data and boost data owners' confidence. Additionally, facilitating open data in funding schemas or publishing data with a specific re-use in mind might help promote publishing open data, as it makes the benefits more tangible to the data owners.

While some mitigation strategies have been suggested in this work, the measures still need to be applied and tested to prove their usefulness. Hence, future work might provide an overview of existing measures and stimuli to facilitate the sharing of open data, utilize some of the strategies and analyze their effectiveness for the solar-thermal domain. In addition, future work might also investigate the motivations behind publishing the identified open datasets – particularly the ones from researchers and private companies – and assess the decision to share data in hindsight. This could include evaluating whether measurable positive outcomes (or unintended negative impacts) resulted from data sharing.

Moreover, future initiatives might work on collaborative data collection and sharing efforts within expert groups such as the IEA SHC Task groups – for example targeting cost data due to its high relevance. Such coordinated activities may help to address domain-specific challenges, including the underrepresentation of solar-thermal in energy planning, while increasing its visibility. Structured collaboration could create mutual benefits for both data providers and users, strengthening the solar-thermal community and expanding the market. The categorization schemas of benefits and barriers from literature could further support these efforts, by transparently assessing the risks and benefits of data publication.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT in order to improve the clarity and readability of selected sections of the work. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

CRediT authorship contribution statement

Lukas Emberger: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Gireesh Nair:** Writing – review & editing, Supervision, Methodology. **Klaus Lichtenegger:** Writing – review & editing, Methodology, Formal analysis, Conceptualization. **Maria Moser:** Writing – review & editing, Methodology, Formal analysis. **Thorsten Summ:** Writing – review & editing, Formal analysis. **Thomas Natiesta:** Writing – review & editing, Formal analysis. **Jianhua Fan:** Writing – review & editing, Investigation. **Pengcheng Wang:** Writing – review & editing, Investigation. **Viktor Unterberger:** Writing – review & editing, Investigation, Conceptualization. **Philip Ohnewein:** Writing – review & editing, Investigation.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The work has been made possible by contributions of the International Energy Agency (IEA) Technology Collaboration Platform (TCP) Solar Heating and Cooling (SHC) community active in Task 68, Efficient Solar District Heating.

This work was supported by the Swedish Energy Agency project number 52686-1 "RESILIENTa Energisystem Kompetenscentrum" and by the Austrian Ministry of Climate Action and Energy [grant number FO999890460]. It also builds on insights and results obtained in the projects DigiSol [grant number FO999891071], Offgrid Solar Cooling [grant number FO999898771] and BEST 4.0 [grant number FO999892426], funded by BMK, BMIMI, BMWET, the federal states of Styria and Lower Austria and the city of Vienna. All Austrian projects were operated by the Austrian Research Promotion Agency (FFG).

References

[1] REN21, Renewables 2024 global status, 2024. https://www.ren21.net/gsr-2024/.
[2] S. Pfenninger, J. DeCarolis, L. Hirth, S. Quoilin, I. Staffell, The importance of open data and software: is energy research lagging behind? Energy Policy 101 (2017) 211–215, https://doi.org/10.1016/j.enpol.2016.11.046.
[3] D. Sari, M. Gasco-Hernandez, The use of open data: challenges and good practices from a user's perspective, in: Proc. Annu. Hawaii Int. Conf. Syst. Sci., Hawaii International Conference on System Sciences, 2024, https://doi.org/10.24251/hicss.2024.269.
[4] A. Zuiderwijk, M. Janssen, S. Choenni, R. Meijer, R.S. Alibaks, Socio-technical impediments of open data, 10 (2012).
[5] A.I. Ugochukwu, P.W.B. Phillips, Open data ownership and sharing: challenges and opportunities for application of FAIR principles and a checklist for data managers, J. Agric. Food Res. 16 (2024) 101157, https://doi.org/10.1016/j.jafr.2024.101157.
[6] N. Lichtenauer, L. Schmidbauer, S. Wilhelm, F. Wahl, A scoping review on analysis of the barriers and support factors of open data, Information 15 (2023) 5, https://doi.org/10.3390/info15010005.
[7] M.K. Natvig, S. Jiang, E. Stav, Using open data for digital innovation: barriers for use and recommendations for publishers, JeDEM - EJ. EDemocracy Open Gov. 13 (2021) 28–57, https://doi.org/10.29379/jedem.v13i2.666.
[8] S. Temiz, M. Holgersson, J. Björkdahl, M.W. Wallin, Open data: lost opportunity or unrealized potential? Technovation 114 (2022) 102535 https://doi.org/10.1016/j.technovation.2022.102535.
[9] A. Zuiderwijk, M. Janssen, K. Poulis, G. Van De Kaa, Open data for competitive advantage: insights from open data use by companies, in: Proc. 16th Annu. Int. Conf. Digit, Gov. Res., ACM, Phoenix, Arizona, 2015, pp. 79–88, https://doi.org/10.1145/2757401.2757411.
[10] Investigating open government data barriers: a literature review and conceptualization, Lect. Notes Comput. Sci., Springer International Publishing, Cham, 2018, pp. 169–183, https://doi.org/10.1007/978-3-319-98690-6_15.

[11] W.S. Wibowo, D.I. Sensuse, S. Lusa, P.A. Wibowo Putro, A. Yulfitri, A systematic literature review on open government data: challenges and mapped solutions, J. Theor. Appl. Inf. Technol. 101 (2023) 1818.
[12] A. Nikiforova, Innovation resistance theory in action: unveiling barriers to open government data adoption by public organizations to unlock open data innovation, (n.d.).
[13] A. Zuiderwijk, M. Janssen, S. Choenni, R. Meijer, Design principles for improving the process of publishing open data, Transform. Gov. People Process Policy 8 (2014) 185–204, https://doi.org/10.1108/TG-07-2013-0024.
[14] M. Beno, K. Figl, J. Umbrich, A. Polleres, Open data hopes and fears: determining the barriers of Open data, in: 2017 Conf. E-Democr. Open Gov. CeDEM, IEEE, Krems, Austria, 2017, pp. 69–81, https://doi.org/10.1109/CeDEM.2017.22.
[15] I. Kawashita, A.A. Baptista, D. Soares, Open government data use by the public sector - an overview of its benefits, barriers, drivers, and enablers, in: 2022. https://doi.org/10.24251/HICSS.2022.315.
[16] A. Zuiderwijk, M.D. Reuver, Why open government data initiatives fail to achieve their objectives: categorizing and prioritizing barriers through a global survey, Transform. Gov. People Process Policy 15 (2021) 377–395, https://doi.org/10.1108/TG-09-2020-0271.
[17] A. Nikiforova, A. Zuiderwijk, Barriers to openly sharing government data: towards an open data-adapted innovation resistance theory, in: Proc. 15th Int. Conf. Theory Pract, Electron. Gov., ACM, Guimarães Portugal, 2022, pp. 215–220, https://doi.org/10.1145/3560107.3560143.
[18] A. Ibrahim, S.D. Abdullah, A. Arief, Benefits and barriers of open and one government data: a systematic review, IOP Conf. Ser. Mater. Sci. Eng. 1125 (2021) 012026, https://doi.org/10.1088/1757-899x/1125/1/012026.
[19] F. Xiao, D. He, Y. Chi, W. Jeng, C. Tomer, Challenges and supports for accessing open government datasets: data guide for better open data access and uses, in: Proc. 2019 Conf. Hum, Inf. Interact. Retr., ACM, Glasgow Scotland UK, 2019, pp. 313–317, https://doi.org/10.1145/3295750.3298958.
[20] M. Beno, K. Figl, J. Umbrich, A. Polleres, Perception of key barriers in using and publishing open data, (2017).
[21] S. Martin, M. Foulonneau, S. Turki, M. Ihadjadene, Risk analysis to overcome barriers to open data, Electron. J. E-Gov. 11 (2013) 348–359.
[22] M. Janssen, Y. Charalabidis, A. Zuiderwijk, Benefits, adoption barriers and myths of open data and open government, Inf. Syst. Manag. 29 (2012) 258–268, https://doi.org/10.1080/10580530.2012.716740.
[23] M. Fassnacht, C. Benz, D. Heinz, J. Leimstoll, G. Satzger, Barriers to data sharing among private sector organizations, in: 2023. https://doi.org/10.24251/HICSS.2023.453.
[24] A. Herala, E. Vanhala, J. Porras, T. Krri, Experiences about opening data in private sector: a systematic literature review, in: 2016 SAI Comput. Conf. SAI,, IEEE, London, 2016, pp. 715–724, https://doi.org/10.1109/SAI.2016.7556060.
[25] P. Runeson, T. Olsson, J. Linåker, Open data ecosystems—An empirical investigation into an emerging industry collaboration concept, J. Syst. Softw. 182 (2021) 111088.
[26] M.T. Çaldağ, E. Gökalp, Understanding barriers affecting the adoption and usage of open access data in the context of organizations, Data Inf. Manag. 9 (2025) 100049, https://doi.org/10.1016/j.dim.2023.100049.
[27] N. Kazantsev, N. Islam, J. Zwiegelaar, A. Brown, R. Maull, Data sharing for business model innovation in platform ecosystems: from private data to public good, Technol. Forecast. Soc. Change 192 (2023) 122515, https://doi.org/10.1016/j.techfore.2023.122515.
[28] T. Enders, G. Satzger, M. Fassnacht, C. Wolff, Why should I share? Exploring benefits of open data for private sector organizations, (2022).
[29] A. Herala, Benefits from open data: barriers to supply and demand of open data in private organizations, (n.d.).
[30] F.W. Donker, B. Van Loenen, A. Bregt, Open data and beyond, ISPRS Int. J. Geo-Inf. 5 (2016) 48, https://doi.org/10.3390/ijgi5040048.
[31] J. Conde, A. Munoz-Arcentales, J. Choque, G. Huecas, A. Alonso, Overcoming the barriers of using linked open data in smart City applications, Computer 55 (2022) 109–118, https://doi.org/10.1109/MC.2022.3206144.
[32] Z. Sugg, Social barriers to open (water) data, WIREs Water 9 (2022) e1564, https://doi.org/10.1002/wat2.1564.
[33] C. Fischer, S.D. Hirsbrunner, V. Teckentrup, Producing open data, Res. Ideas Outcomes 8 (2022) e86384, https://doi.org/10.3897/rio.8.e86384.
[34] International Energy Agency Solar Heating and Cooling Programme, IEA SHC Task 68, (n.d.). https://task68.iea-shc.org/(accessed 2025).
[35] G. Umbach, Open Science and the impact of open access, open data, and FAIR publishing principles on data-driven academic research: towards ever more transparent, accessible, and reproducible academic output? Stat. J. IAOS J. Int. Assoc. Off. Stat. 40 (2024) 59–70, https://doi.org/10.3233/SJI-240021.
[36] Y. Gao, M. Janssen, C. Zhang, Understanding the evolution of open government data research: towards open data sustainability and smartness, Int. Rev. Adm. Sci. 89 (2023) 59–75, https://doi.org/10.1177/00208523211009955.
[37] K. Lassila-Perini, C. Lange, E. Carrera Jarrin, M. Bellis, Using CMS open data in research – challenges and directions, EPJ Web Conf. (2021) 251, https://doi.org/10.1051/epjconf/202125101004.
[38] L. Wang, Y. Zhan, Does cross-policy mix work on digital innovation: analyzing by the impact of open government data and low-carbon city pilot policies in China, J. Clean. Prod. 491 (2025), https://doi.org/10.1016/j.jclepro.2025.144842.
[39] P. Runeson, T. Olsson, Challenges and opportunities in open data collaboration – a focus group study, in: 2020 46th Euromicro Conf. Softw. Eng. Adv. Appl. SEAA, Portoroz, Slovenia, IEEE, 2020, pp. 205–212, https://doi.org/10.1109/seaa51224.2020.00044.
[40] Mentimeter, (n.d.). https://www.mentimeter.com.
[41] M. Sengupta, A. Habte, S. Wilbert, C. Gueymard, J. Remund, E. Lorenz, W. van Sark, A.R. Jensen, Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications: Fourth Edition, 2024. https://www.nrel.gov/docs/fy24osti/88300.pdf.
[42] K.S. Anderson, C.W. Hansen, W.F. Holmgren, A.R. Jensen, M.A. Mikofski, A. Driesse, pvlib python: 2023 project update, J. Open Source Softw. 8 (2023) 5994, https://doi.org/10.21105/joss.05994.
[43] Google Maps, Google (n.d.). https://www.google.com/maps (accessed March 31, 2025).
[44] OpenStreetMap contributors, OpenStreetMap, (n.d.). https://www.openstreetmap.org (accessed March 31, 2025).
[45] I. Sifnaios, G. Gauthier, D. Trier, J. Fan, A.R. Jensen, Dronninglund water pit thermal energy storage dataset, Sol. Energy 251 (2023) 68–76, https://doi.org/10.1016/j.solener.2022.12.046.
[46] I. Sifnaios, S. Furbo, A.R. Jensen, Monitoring data of the Høje Taastrup water pit thermal energy storage, Data Brief 58 (2025) 111305, https://doi.org/10.1016/j.dib.2025.111305.
[47] D. Tschopp, P. Ohnewein, R. Stelzer, L. Feierl, M. Hamilton-Jones, M. Moser, C. Holter, One year of high-precision operational data including measurement uncertainties from a large-scale solar thermal collector array with flat plate collectors, located in Graz, Austria, Data Brief (2023) 109224, https://doi.org/10.1016/j.dib.2023.109224.
[48] D. Tschopp, P. Ohnewein, R. Stelzer, L. Feierl, M. Hamilton-Jones, M. Moser, C. Holter, One year of high-precision operational data including measurement uncertainties from a large-scale solar thermal collector array with flat plate collectors, located in Graz, Austria, (2023). https://doi.org/10.5281/ZENODO.7741083.
[49] F. Ebmeier, N. Ludwig, G. Martius, V.H. Franz, PaSTS an operational dataset for domestic solar thermal systems, (2024). https://doi.org/10.5281/ZENODO.11093493.
[50] F. Ebmeier, N. Ludwig, G. Martius, V.H. Franz, PaSTS an operational dataset for domestic solar thermal systems, in: Proc. 15th ACM Int. Conf. Future Sustain. Energy Syst, Association for Computing Machinery, New York, NY, USA, 2024, pp. 529–534, https://doi.org/10.1145/3632775.3662159.
[51] F. Veynandt, F. Inschlag, C. Seidl, C. Heschl, Measurement data from real operation of a hybrid photovoltaic-thermal solar collectors, used for the development of a data-driven model, Data Brief 49 (2023) 109417, https://doi.org/10.1016/j.dib.2023.109417.
[52] AEE INTEC, SHIP Plants Database, (2023). https://ship-plants.info.
[53] solarheatdata.eu, SOLARHEATDATA.EU, (2025). https://solarheatdata.eu/.
[54] solvarmedata.dk, SOLVARMEDATA.DK, (2025). https://solvarmedata.dk.
[55] J.-O. Dalenbäck, Ranking list of European large scale solar heating plants, (2017). https://www.solar-district-heating.eu/en/plant-database/.
[56] Projektlandkarte solare Wärmenetze, (2024). https://www.solare-waermenetze.de/projektbeispiele/projektlandkarte-solare-waermenetze/.
[57] R. Thonig, A. Gilmanova, J. Lilliestam, CSP.guru 2023-07-01, (2023). https://doi.org/10.5281/ZENODO.8191855.
[58] J. Lilliestam, L. Ollier, M. Labordena, S. Pfenninger, R. Thonig, The near- to mid-term outlook for concentrating solar power: mostly cloudy, chance of sun, Energy Sourc. Part B Econ. Plan. Policy 16 (2021) 23–41, https://doi.org/10.1080/15567249.2020.1773580.
[59] IEA Solar Heating and Cooling Programm, Solar Heat Worldwide, (2024). https://www.iea-shc.org/solar-heat-worldwide (accessed March 30, 2025).
[60] W. Weiss, M. Spörk-Dür, Solar heat worldwide 2024, IEA SHC (2024), https://doi.org/10.18777/ieashc-shww-2024-0001.
[61] BSW – Bundesverband Solarwirtschaft e.V., Solaratlas - Der Vertriebskompass für die Solarbranche, (n.d.). https://www.solaratlas.de (accessed March 30, 2025).
[62] Renewable Thermal Collaborative, Partner Locator, (n.d.). https://www.renewablethermal.org/partner/(accessed 2025).
[63] MCS, Find A contractor, Mcscertified (n.d.). https://mcscertified.com/find-an-installer/(accessed 2025).
[64] Solar Payback, Suppliers of Turnkey Solar Process Heat Systems, (n.d.). https://www.solar-payback.com/suppliers/ (accessed March 30, 2025).
[65] Solar District Heating EU, SDH - Find Professionals, (n.d.). https://www.solar-district-heating.eu/en/find-professionals/ (accessed March 30, 2025).
[66] European Commission. Joint Research Centre., ILF Consulting Engineers Austria GmbH., AIT Austrian Institute of Technology GmbH., Long term (2050) projections of techno-economic performance of large-scale heating and cooling in the EU., Publications Office, LU, 2017. https://data.europa.eu/doi/10.2760/24422 (accessed February 12, 2025).
[67] ESTIF – European Solar Thermal Industry Federation, Solar Keymark Database, (n.d.). https://solarkeymark.eu/database/(accessed 2025).
[68] CEN KEYMARK, CEN Keymark Database, (n.d.). https://keymark.eu/en/certificates/certificates-data-base (accessed March 30, 2025).
[69] T.Ü.V. Rheinland, DIN CERTCO - Zertifikate und Registrierungen, (n.d.). https://www.dincertco.tuv.com/ (accessed March 30, 2025).
[70] Global Solar Atlas, Global Solar Atlas, (n.d.). https://globalsolaratlas.info/download/world (accessed March 30, 2025).
[71] L. Steiermark, G.I.S. Steiermark, Digitaler Atlas Steiermark, (n.d.). https://gis.stmk.gv.at/wgportal/atlasmobile/map/Versorgung%20-%20Entsorgung/Energiewirtschaft%20und%20-planung?snapshot=SolarTool&tool=webgis.tools.custom.solartool&presentation=dv_int_rdata-solarpotenzial,dv_int_rdata-solarstrahlung_jahressumme,dvg_geb%C3%A4ude_aus_dom,dv_datenstand_globalstrahlung=off&append-services=solarstrahlung_jahressumme,globalstrahlung_solartool,als,gebaeudebauwerke (accessed March 30, 2025).

[72] European Commission, PVGIS, (2024). https://re.jrc.ec.europa.eu/pvg_tools/en/(accessed March 30, 2025).
[73] Deutscher Wetterdienst, CM SAF Web User Interface, (n.d.). https://wui.cmsaf.eu/safira/action/viewProduktList?id=2 (accessed March 30, 2025).
[74] National Renewable Energy Laboratory, NSRDB Natioanl Solar Radiation Database, (n.d.). https://nsrdb.nrel.gov/data-viewer (accessed March 30, 2025).
[75] C3S, ERA5 hourly data on single levels from 1940 to present, (2018). https://doi.org/10.24381/CDS.ADBB2D47.
[76] Copernicus Climate Change Service (C3S), Atmosphere Data Store, (n.d.). https://ads.atmosphere.copernicus.eu/datasets (accessed March 30, 2025).
[77] Landesamt für Natur, Umwelt und Verbraucherschutz NRW, Energieatlas NRW, (n.d.). https://www.energieatlas.nrw.de/site/karte_solarkataster (accessed March 30, 2025).
[78] Deutscher Wetterdienst, Hourly station observations of solar incoming (total/diffuse) and longwave downward radiation for Germany, (2024). https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/hourly/solar/DESCRIPTION_obsgermany_climate_hourly_solar_en.pdf.
[79] Royal Netherlands Meteorological Institute, KNMI Data Platform, (n.d.). https://english.knmidata.nl/(accessed 2025).
[80] sEEnergies, Pan-European Thermal Atlas PETA v5.2, (2022). https://www.seenergies.eu/peta5/(accessed March 30, 2025).
[81] sEEnergies, sEEnergies Open Data Hub, (n.d.). https://s-eenergies-open-data-euf.hub.arcgis.com/(accessed March 30, 2025).
[82] Hotmaps Project, Hotmaps, (n.d.). https://www.hotmaps.eu/map (accessed March 30, 2025).
[83] THERMOS, Thermos Tool, (n.d.). https://www.thermos-project.eu/thermos-tool/tool-access/(accessed March 30, 2025).
[84] Hochschule für angewandte Wissenschaft und Kunst Hildesheim/Holzminden/Göttingen, Der Fernwärmeatlas, (n.d.). https://fernwaerme-atlas.hawk.de/.
[85] J. Pelda, S. Holler, U. Persson, District heating atlas - analysis of the German district heating sector, Energy 233 (2021) 121018, https://doi.org/10.1016/j.energy.2021.121018.
[86] Bundesamt für Energie BFE, Thermische Netze (Nahwärme, Fernwärme, Fernkälte), (2025). https://opendata.swiss/de/dataset/thermische-netze-nahwarme-fernwarme-fernkalte.
[87] e-think Zentrum für Energiewirtschaft und Umwelt, Austrian Heat Map, (n.d.). https://austrian-heatmap.gv.at/karte/.
[88] C. Büttner, J. Amme, J. Endres, A. Malla, B. Schachler, I. Cußmann, Supplementary data: "open modeling of electricity and heat demand curves for all residential buildings in Germany," (2022). https://doi.org/10.5281/ZENODO.7071795.
[89] B.S. Frey, F. Oberholzer-Gee, The cost of price incentives: an empirical analysis of motivation crowding-out, (n.d.).
[90] M. Lnenicka, A. Nikiforova, M. Luterek, P. Milic, D. Rudmark, S. Neumaier, K. Kević, A. Zuiderwijk, M.P. Rodríguez Bolívar, Understanding the development of public data ecosystems: from a conceptual model to a six-generation model of the evolution of public data ecosystems, Telemat. Inform. 94 (2024) 102190, https://doi.org/10.1016/j.tele.2024.102190.