Efficient Data Management and Validation

IEA SHC TASK 68 | Efficient Solar District Heating Systems

Efficient Gathering, Storing, Distributing and Validation of Data

This is a report from SHC Task 68: Efficient Solar District Heating and work performed in Subtask B: Data Preparation & Utilization

Lukas Feierl, Sabine Putz, Viktor Unterberger,
Mohamad S. Mortada, Yuvaraj Pandian, Adam R. Jensen,
Thomas Natiesta, Stefan Mehnert, Daniel Tschopp, Philip Ohnewein
29.01.2024
Report number, DOI: 10.18777/ieashc-task68-2024-0001

Solar Heating & Cooling Technology Collaboration Programme (IEA SHC)

The Solar Heating and Cooling Technology Collaboration Programme was founded in 1977 as one of the first multilateral technology initiatives ("Implementing Agreements") of the International Energy Agency.

Our mission is *to bring the latest solar heating and cooling research and information to the forefront of the global energy transition.*

IEA SHC members carry out cooperative research, development, demonstrations, and exchanges of information through Tasks (projects) on solar heating and cooling components and systems and their application to advance the deployment and research and development activities in the field of solar heating and cooling.

Our focus areas, with the associated Tasks in parenthesis, include:

- Solar Space Heating and Water Heating (Tasks 14, 19, 26, 44, 54, 69)
- Solar Cooling (Tasks 25, 38, 48, 53, 65)
- Solar Heat for Industrial and Agricultural Processes (Tasks 29, 33, 49, 62, 64)
- Solar District Heating (Tasks 7, 45, 55, 68)
- Solar Buildings/Architecture/Urban Planning (Tasks 8, 11, 12, 13, 20, 22, 23, 28, 37, 40, 41, 47, 51, 52, 56, 59, 63, 66)
- Solar Thermal & PV (Tasks 16, 35, 60)
- Daylighting/Lighting (Tasks 21, 31, 50, 61, 70)
- Materials/Components for Solar Heating and Cooling (Tasks 2, 3, 6, 10, 18, 27, 39)
- Standards, Certification, Test Methods and LCA/LCoH (Tasks 14, 24, 34, 43, 57, 71)
- Resource Assessment (Tasks 1, 4, 5, 9, 17, 36, 46)
- Storage of Solar Heat (Tasks 7, 32, 42, 58, 67)

In addition to our Task work, other activities of the IEA SHC include our:

- SHC Solar Academy
- *Solar Heat Worldwide*, annual statistics report
- SHC International Conference

Our members

Australia	France	Slovakia
Austria	Germany	South Africa
Belgium	International Solar Energy Society	Spain
Canada	Italy	Sweden
CCREEE	Netherlands	Switzerland
China	Norway	Türkiye
Denmark	Portugal	United Kingdom
EACREEE	RCREEE	
ECREEE	SACREEE	
European Commission	SICREEE	

For more information on the IEA SHC work, including many free publications, please visit www.iea-shc.org.

Contents

1 Executive Summary

Solar thermal plants have proven to be a successful player in providing heat for district heating networks. However, to ensure the efficient operation of such plants and to achieve optimal coordination with other heat generation units, thorough monitoring, quality control, and system control are required. These tasks strongly depend on accessible and reliable measurement data, which are often unavailable.

Thus, this report focuses on efficient **data gathering, storage, distribution, and validation**, covering data management topics- from sensor selection to permanent data storage. The report is mainly targeted at system designers and plant operators, aiming to provide checklists and recommendations on these topics.

The report considers a general solar district heating plant, as depicted by IEA SHC Task 55 (see Figure 1) – including a collector field, heat storage, and heating center (including a biomass boiler, heat pump, and other auxiliary heating) up to the interface to the district heating network. The topics are described on a summary level of detail while referring the reader to individual articles in case more information is needed. In addition, research groups may use this report to get an overview of data management in the solar-thermal field and identify related work.

The work consists of five sections: The Required Data section lists recommended measurements and discusses meta information required to interpret the data successfully. The Data Gathering section provides recommendations for data logging – e.g., sampling rate, encoding, and formatting. The Data-Distribution section shows proven examples of architectures for collecting and distributing data. Furthermore, the Data Storage section describes what data storage technologies (e.g., CSV files or relational databases) are currently used. The section also discusses the experiences, advantages, and challenges of the respective technologies. Finally, the Data Validation section lists common data-validation procedures that can be applied to solar-thermal data and links to open-source implementations where available.

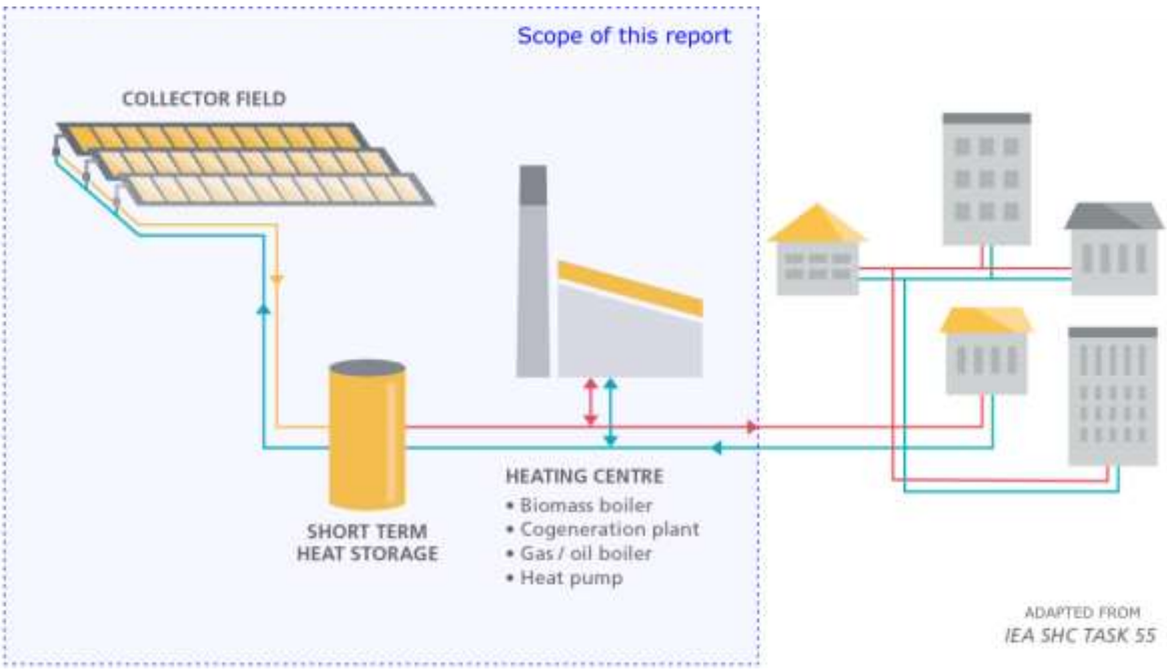

Figure 1: Illustration of a solar district heating plant with a grey box depicting the scope of this report. Source: adapted from IEA SHC Task 55.

2 Required Data

This section is targeted at plant operators and provides recommendations about what data must be gathered to ensure that the most common evaluations of solar thermal plants can be performed. This includes not only measurements but also supporting information (i.e., metadata), which is necessary for properly analyzing solar thermal plants. Hence, the first subsection deals with recommended measurements, while the second subsection deals with required metadata.

2.1 Recommended Measurements

This subsection provides a list of recommended sensors that should be installed at a plant to gather the most crucial measurements. This topic has also previously been covered by multiple authors, see Table 1. Hence, this section aims to summarize the results of these reports and adds new recommendations based on the authors' experiences, especially considering concentrating collectors. For each of the recommended measurements, the text discusses why the measurement is important and which evaluation it is relevant for.

Table 1: Related literature on recommended measurements

Reference	Description
IEA SHC Task 45 [1], [2]	A fact sheet about monitoring, listing recommendations for monitoring solar-thermal plants. The report contains a list of recommended measurements, installation conditions, and sensor specifications (written in German).
IEA SHC Task 55 Fact sheet BD.3 [3]	A fact sheet about automated monitoring, listing recommendations for monitoring solar-thermal plants. The report contains a list of recommended measurements, installation conditions, and sensor specifications.
Faure et al. [4]	An article reviewing dysfunctions at solar-thermal plants, containing a description of typical measurements as part of the State-of-the-art section.
Kramer et al. [5]	An article comparing different collector performance tests (ICC, DCAT, PC), containing a brief overview of required measurements.
Zirkel-Hofer et al. [6]	An article discussing the selection and installation of measurement devices based on uncertainty examination of different sensors (focusing on concentrating collectors).
Effertz et al. [7]	A research report on measurement- and performance testing of solar-thermal plants, containing recommendations to sensor specifications and installation.
ISO 9806 [8]	An international standard for testing solar collectors, containing requirements for measurement devices and their installation.
ISO 24194 [9]	An international standard for collector performance guarantees, containing requirements for measurement devices and their installation.
Guide to ISO 24194[10]	Guide for international standard ISO 24194, providing tips on how to apply the corresponding ISO norm. It also contains a chapter about which measurements are required to apply the ISO and extends the ISO to multiple collector arrays.
Guide to ISO 9806 [11]	Guide for international standard ISO 9806, providing tips on how to apply the corresponding ISO norm. It contains information about required sensors, installation conditions, and sensor specifications for testing solar thermal collectors.
Solar Heat Data Website [12]	Website showing solar thermal plants and collecting measurement data. Required measurements are listed as part of the registration process.

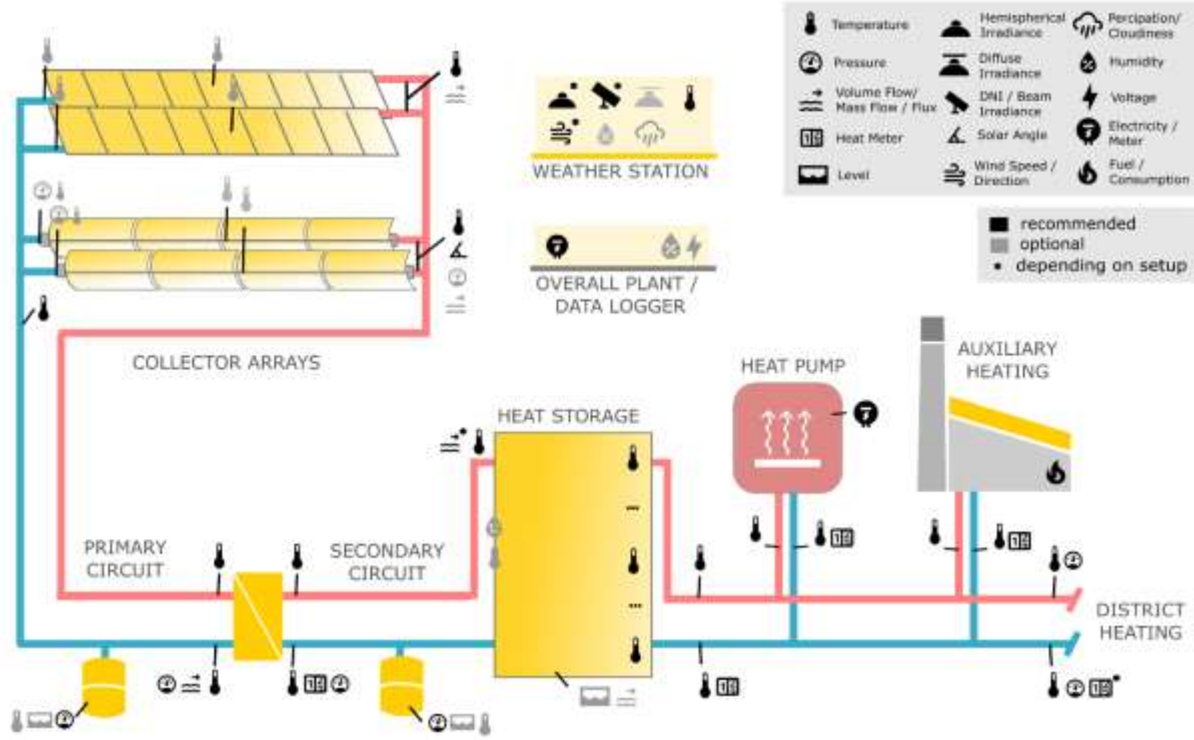

Figure 2: Illustration of a solar district heating plant displaying recommended sensors.

For a better overview, the recommended measurements are described in individual subsections based on the position of the sensors. More specifically, the following categories are used to loosely group the sensors (see Figure 2): The “Weather Station” groups all sensors related to solar radiation and weather. The “Collector Array” corresponds to measurements at the collector arrays. The “Primary Circuit” contains all sensors in the primary circuit that are not associated with individual collector loops. The “Secondary Circuit” contains all measurements in the secondary circuit up to the heat storage. The “Heat Storage” encompasses all sensors related to the heat storage. The “District Heating” accounts for sensors close to the interface of the district heating grid. In contrast, the “Heat Pump” and “Auxiliary Heating” contain all sensors close to a heat pump and (any/general) auxiliary heating, respectively.

A summary of the recommended sensors can also be seen in s 2 to as well as in Figure 2. Please note that the installation requirements and uncertainty considerations are not in the scope of this report. Instead, the interested reader is referred to the cited literature in Table 1.

1.1.1 Ambient Measurements (Weather Station)

This part comprises measurements related to the weather. Since solar irradiance is the energy source of solar-thermal plants, its measurement data is crucial for monitoring the collectors' efficiency. For example, irradiance data is required to calculate the effective collector efficiency , estimate thermal power according to the collector Keymark equation (ISO 9806) , and carry out the Thermal Power Check as defined by ISO 24194 [9] and the Dynamic Collector Array Test [13]. For non-concentrating collectors, it is often sufficient to measure the *hemispherical solar irradiance* on the collector surface [3,8,9]. In contrast, *direct normal irradiance (DNI)* is required for concentrating collectors [8,9]. For more accurate results, information about the *diffuse solar irradiance* is helpful [8,9]. However, this quantity is best determined by subtracting measured direct irradiance from measured hemispherical in-plane irradiance since measuring in-plane diffuse irradiance directly is often impractical. Irradiance sensors must be maintained well to prevent systematic errors. As shown, for example, by Tschopp et al. [14] and Ohnewein et al. [13], incorrect measurements might have a high influence on efficiency estimates. Based on their results, dirt on the sensors or insufficient sensor accuracy can lead to high systematic errors of 10% or higher. Hence, carefully choosing a proper radiation sensor and periodically cleaning and calibrating the sensor are recommended. In addition, an air shield may help to avoid dirt on a pyrheliometer [15].

The *ambient temperature* is also a crucial measurement, as it allows for estimating heat losses of collectors [4–6], storages [16], and pipes [16]. Please note that multiple sensors might be needed if the temperature is different at different places of interest (e.g., collector arrays in the field versus pipes inside the control room). As an example, ISO 24194 requires the ambient temperature sensor used for collector evaluations to be shielded from direct sunlight and placed on the field one meter above the ground and not further than 100 meters from the collectors [9]. Shields for ambient temperature sensors should ideally also be equipped with a ventilator, as stagnant air inside the shield enclosure may get heated up and not reflect the actual ambient temperature.

Additional recommended measurements are *wind speed* and the *wind direction*, which affect heat losses of the collectors (especially relevant for uncovered collectors). Hence, this data is required, for example, for ISO 24194 [9] and D-CAT [13]. Furthermore, the information about wind is especially important for tracking collectors, for example, for estimating torsion losses, for checking whether safety positions must be applied, and for fault detection in general.

Furthermore, operating experience from the RB1 authors suggests that measuring the *relative humidity and precipitation* and using *sky cameras* might lead to very useful insights and risk management strategies. For example, knowledge about precipitation might be used to estimate the need for collector cleaning or to indicate whether a draining system is required due to heavy rain. Similarly, information about snow precipitation can be helpful to estimate whether snow covers the collectors – in which case tracking collectors might get rid of snow covers by tilting the collectors down. In general, these measurements can help maintain optimal system efficiency but can also be used for forecasting district heating demand or applying short-term predictive control strategies accounting for cloud movements.

Table 2: Recommended measurements for ambient/weather conditions. Measurements are marked either recommended (R), optional (O), or depending on the setup (*).

Ambient Measurements (Weather Station)		
Measurement	Required for	
Hemispherical solar irradiance	Monitoring collector performance (for non-concentrating collectors)	*
Direct normal irradiance (DNI)	Monitoring collector performance (for concentrating collectors)	*
Diffuse horizontal irradiance	Monitoring collector performance	O
Ambient temperature	Calculating heat losses	R
Wind speed	Calculating heat losses, estimating torsion losses, checking the time for safety position	R
Wind direction	Estimating torsion losses, checking the time for safety position (for tracking collectors)	*
Relative humidity	Monitoring condensation, checking plausibility of other weather measurements	O
Precipitation	Estimate collector soiling, evaluate the need for drainage	O
Sky camera	Check cloudiness, short-term forecasting	O

2.1.1 Collector Arrays

This part comprises the collector arrays (or collector loops/rows) installed at the plant (see Figure 2). Required measurements may vary depending on the type of thermal collectors used. However, measuring *flow temperatures* for each collector array is common at most plants. For example, the flow temperature allows for detecting "freezing," "stagnation," and "sub-field stagnation"[1] events and for checking that all arrays deliver a similar flow temperature. Similarly, the *return temperature* is also measured at most solar thermal plants, for example, to calculate pipe losses and analyze the temperature increase due to the collectors. Depending on the aim of the measurement, it may suffice to place only one temperature sensor for all arrays. However, pipe losses and pressure drops may lead to a slightly unequal temperature distribution for each loop.

If *volume-* or *mass flows* are measured as well, this even allows the application of the Thermal Power Check (ISO 24194), the D-CAT, and similar methods for each individual loop - permitting a more in-depth evaluation of the

[1] I.e., only a part of the plant is in stagnation.

collector performance of each loop. Due to more focus on quality assurance, recently built plants often install a reference loop for measuring flow- and return temperatures, irradiation, and volume flow for one specific collector loop. This enables a more accurate estimation of collector performance, as pipe distribution losses are limited, and modeling is more straightforward (compared to modeling the whole collector field). The measured collector loop then often serves as a representative of all the collectors of the respective type for defining collector performance guarantees.

Depending on the type of thermal collectors, it is also recommended to measure the *flow-* and *return pressure* for each row, especially in the case of concentrating collectors with high pressure. This allows the monitoring of the pressure drop and can also be used to detect leaks and blockages. For tracking collectors, the *tracking angle* must either be calculated or measured to adjust the tracking actuators for performance optimization and to detect downtime where the collectors are defocused. For large-scale systems with many collectors in a row, it is also recommended to install temperature sensors within the row *(intermediary temperature)* to allow for faster flow control and detection of overheating.

Table 3: Recommended measurements for the collector arrays. Measurements are marked either recommended (R) or optional (O).

Collector Arrays (each)		
Measurement	Required for	
Flow temperature	Monitoring collector row performance, detecting “subfield stagnation”, anti-freezing, matching flow temperature of different loops, modelling pipe losses.	R
Return temperature	Monitoring collector row performance, anti-freezing, modelling pipe losses.	O
Flow/heat meter	Monitoring collector row performance	O
Return pressure	Pressure-drop monitoring	O
Flow pressure	Pressure-drop monitoring	O
Tracking angle	Adjustment of tracking actuators, performance optimization, detection of downtime/defocusing (for tracking collectors)	*
Intermediary temperature	Faster flow control, detection of overheating	O

2.1.2 Primary Circuit

For the primary circuit (see Figure 2), the *flow and return temperatures* should always be measured, as they are often used for system control and can also be used for freezing and stagnation detection. Combined with a *volume flow* sensor or with a *heat meter*, it allows for monitoring the performance of the whole collector field (e.g., according to ISO 24194) and calculating energy balances. Measuring the *pressure* (e.g., before and after the pump) allows for pump monitoring, leakage detection, and general safety monitoring. In addition, measuring the *expansion vessel level* and *pressure* is recommended for leakage detection and safety monitoring, especially in the case of high-pressure systems. To rule out damages at the expansion vessel due to high fluid temperatures, a *temperature* sensor might be installed at the expansion vessel as well.

Table 4: Recommended measurements for the primary circuit. Measurements are marked either recommended (R) or optional (O).

Primary Circuit		
Measurement	Required for	
Flow temperature	Performance monitoring, stagnation detection, check system control	R
Return temperature	Performance monitoring, detect “too little extraction”	R
Flow/heat meter	Performance monitoring, pump monitoring	R
Pressure	Leakage detection, pump monitoring	R
Expansion-vessel pressure	Safety monitoring, process check	R
Expansion-vessel level	Safety monitoring, leakage detection	O

Expansion vessel temperature	Monitoring expansion vessel temperature.	O

2.1.3 Secondary Circuit

The recommended measurements for the secondary circuit are similar to those for the primary circuit. However, it is recommended that heat metering is done at the secondary circuit. The reason is that water (typically used in the secondary circuit) has very well-known thermal properties compared to the fluid mixtures (typically anti-freeze glycol mixtures) often used in the primary circuit. Hence, heat calculation is more accurate for the secondary circuit. Apart from that, the recommended measurements are the same.

Table 5: Recommended measurements for the secondary circuit. Measurements are marked either recommended (R) or optional (O).

Secondary Circuit		
Measurement	Required for	
Flow temperature	Performance monitoring, stagnation detection, check system control	R
Return temperature	Performance monitoring, detect “too little extraction”	R
Heat meter (incl. flow)	Performance monitoring, pump monitoring	R
Pressure	Leakage detection, pump monitoring	R
Expansion-vessel pressure	Safety monitoring, process check	R
Expansion-vessel level	Safety monitoring, leakage detection	O
Expansion vessel temperature	Monitoring expansion vessel temperature.	O

2.1.4 Heat Storage

The “Heat-Storage” part comprises all measurements regarding heat storages. Hence, evaluations concerning energy content, heat losses, and stratification are of interest. It is highly recommended to measure the *temperature of the storage* at several positions (depending on the storage size, but a minimum of 5 measurements is recommended). In addition*, charging and discharging temperatures* and *flows* should be measured to calculate energy balances. The measurement of the *fluid level* in the tank can help to detect leakages, while information about the *heat flux*, *insulation humidity,* and *insulation temperature* might help to analyze and monitor the heat storage in more detail.

Table 6: Recommended measurements for the Heat Storage. Measurements are marked either recommended (R) or optional (O).

Heat Storage		
Measurement	Required for	
Temperature (top, middle, bottom)	Calculation of storage energy content, avoiding overheating, monitoring stratification,	R
Charging/discharging temperature	Calculating storage charging/discharging rate	R
Charging/discharging flow	Calculating storage charging/discharging rate	R
Tank level	Detect leakages	O
Heat flux	Monitor lid/insulation performance and degradation	O
Insulation humidity and temperature	Monitor lid performance and conditions, detection of critical situations requiring maintenance intervention	O

2.1.5 Auxiliary Heating

Data regarding auxiliary heating can also be of great interest in evaluations. Measuring *flow temperature, return temperature,* and *volume flow* allows using the auxiliary heating in energy balance calculations. In addition, knowing the *fuel consumption* allows calculating the efficiency and emissions of the auxiliary heating.

Table 7: Recommended measurements for the Auxiliary Heating. Measurements are marked either recommended (R) or optional (O).

Auxiliary Heating		
Measurement	Required for	
Flow temperature	System-control monitoring, efficiency of heat generation	R
Return temperature	System-control monitoring, efficiency of heat generation	R
Flow/heat meter	Energy balance evaluation, efficiency of heat generation	R
Fuel consumption	Calculation of emissions, efficiency of heat generation	R

2.1.6 Heat Pump

Heat Pumps can supply heat to the district heating network directly or be used as a booster when the storage temperature is too low. To monitor performance, *flow temperature, return temperature, volume flow*, and ideally, the *thermal power* should be measured at both the evaporator (source) and condenser circuit (sink). Measuring the *electric consumption* also allows the computation of the coefficient of performance (COP) and the energy efficiency ratio (EER) of the heat pump.

Table 8: Recommended measurements for the Heat Pump. Measurements are marked either recommended (R) or optional (O).

Heat Pump		
Measurement	Required for	
Flow temperature (source)	Performance monitoring, system-control monitoring	R
Return temperature (source)	Performance monitoring, system-control monitoring	R
Flow/heat meter (source)	Performance monitoring. Energy balance, COP, EER,	R
Flow temperature (sink)	Performance monitoring, system-control monitoring	R
Return temperature (sink)	Performance monitoring, system-control monitoring	R
Flow/heat meter (sink)	Performance monitoring. Energy balance, COP, EER.	R
Electric consumption	Monitoring electricity consumption, energy balance, COP, EER, and system-control monitoring	R

2.1.7 District Heating Interface

Especially with high solar shares, exchanging information with the district heating network is critical for operation. Experiences from district heating plants show that pressure readings and information about the storage content are essential to optimize the district heating load, especially in the case of a high solar-thermal share. Hence, measuring the *pressure* at the district heating interface is important for safe operation and monitoring system control. In addition, a *heat meter* is typically required to measure how much heat was transferred to the district heating network for monitoring and accounting purposes.

Table 9: Recommended measurements for the District Heating Interface. Measurements are marked either recommended (R) or optional (O).

District Heating Interface		
Measurement	Required for	
Pressure	Leakage detection, system control monitoring	R
Flow temperature	System-control monitoring	R
Return temperature	System-control monitoring	R
Flow/heat meter	Energy balance evaluation, accounting	R

2.1.8 Other Measurements

This part lists all measurements that cannot be directly assigned to one of the parts above. For example, it is strongly recommended to measure the *electricity consumption* of the plant – either in total or for specific parts of the system. This is needed for an energy balance evaluation, performance monitoring, and accounting. In addition, *webcams* can be installed during construction for security reasons and to check construction progress. During operation, they can also be used to check collector and mirror positions (in the case of tracking collectors), evaluate shading, or detect visible damage or dirt on the collectors. In addition, the *data logger voltage* and *temperature* might be helpful to monitor the data logger. Furthermore, *safety devices* like gas or smoke detectors might also be required for the system for safety reasons - providing measurement data as well.

Table 10: Recommended measurements not connected with a specific part of the plant. Measurements are marked either recommended (R) or optional (O).

Other Measurements		
Measurement	Required for	
Electricity meter	Energy balance evaluation, electricity consumption	R
Webcam	Check construction progress, collector positions, and mirrors, visible damage, dirt on collectors, shading	O
Data logger voltage	Feedback on data logger performance	O
Data logger temperature	Feedback on data logger performance	O

2.2 Meta Data

While measurement data is critical for evaluations, no analysis can be done without meta-information. For example, the position of the sensors and the measurement unit must be known to interpret the data correctly. As another example, knowledge about events influencing the system (e.g., services or faults) is also required to understand the behavior of a plant. Hence, this subsection provides a list of important meta-information that should be systematically recorded, stored, and made available along with the measurement data.

- **Plant Basics** - many calculations require basic information about the plant. For example, the location, i.e., longitude and latitude, are required to calculate the solar position, ideal tracking angles, and incidence angles. Also, the collector area of each collector field and its orientation are important parameters required for the thermal power check according to ISO 24194, for example. Additional recommendations are provided in Table 11.

- **Sensor Details -** Similar to the “Plant Basics,” specific information regarding the sensors, including the unit of measurement, accuracy, sensor type, and installation (e.g., orientation in case of irradiation sensors), is required to interpret the data correctly.

- **Plant Layout** – The measurement data can also not be interpreted without knowing the structure of the plant and at which position the sensors are installed. Hence, a piping and instrumentation diagram (PID) or a simplified representation of the plant layout should be available with the corresponding sensor location and tag numbers.

- **System-Control** – The valve positions, the rotation speed of pumps, setpoints, or control signals are critical for understanding the behavior of a plant. For these parameters, it also makes sense to store them as timestamp-value pairs together with the measurement data as they change frequently. In contrast, for some parameters that rarely change (e.g., control strategies and P-I setpoints), it makes more sense to gather the data on change. Nevertheless, both types of system-control data can be critical to understanding plant behavior.

- **Events (services, faults, etc.)** – Finally, knowledge about unusual events (services, faults, etc.) [17] is required as well to interpret the data correctly. In case of an event, the behavior of the plant might be compromised, which should be taken into account for calculations. On the other hand, events might also explain other phenomena like a sudden increase in performance after a service or abnormal data due to a fault. This type of information should be recorded in a logbook, ideally digital, e.g., a spreadsheet or survey style, or via ticketing.

Two good practice examples are the open-source Dronninglund dataset about a pit thermal energy storage [18] and the open-source FHW dataset about solar-thermal collectors installed at a solar district heating plant [19]. Both papers provide all essential information to interpret the data, listing most of the contents described in this section.

Table 11: Example datasets for "Plant Basics" metadata that should be stored to allow evaluation of solar-thermal plants.

Parameter	Description
Location	Longitude, latitude, and elevation of the plant, to determine solar angles, tracking angles, local timezone, and shading.
Collector-Arrays	Layout and orientation (azimuth, tilt, ground tilt, row-spacing, etc.) of each collector field to compute solar incidence angles and shading.
Collector-Types	Solar Keymark parameters of collectors to enable calculation of expected yield and performance check according to ISO 24194.
Piping	Length, diameter, insulation, insulation type, and –thickness (preferably a datasheet), for estimating heat losses.
Heat Storages	Geometry, sizing, and insulation to estimate heat losses, energy content, and stratification.
Pumps	Nominal power to estimate electric consumption.
Heat Pumps	COP, EER, setpoints, etc. for comparison with measurements, ideally datasheet from manufacturer.
Heat transfer fluids	Heat capacity, density, etc. for comparison with estimated thermal power and other calculations, ideally a datasheet from the manufacturer.
Nominal yield	Expected solar yield based on plant design for comparison to actual yield.
Accounting information	Number of collectors/gross area, to enable calculation of specific yield. Associated price/savings and CO_2-emissions/savings to calculate economical key performance indicators (KPIs). Information about replacement/counter resets to correctly calculate heat generation.

3 Data Gathering

This section provides recommendations regarding data logging – transforming the input of the measurement sensors into digital output and recording the results. The content of this section is mainly based on the IEA SHC Task 45 fact sheet [1, 2], Task 55 fact sheet BD.3 [3], and lessons learned from developing the SunPeek software [20]. A summary is provided in Table 12.

Table 12: Summary of recommendations for Data Gathering.

Topic	Recommendation
Sampling Rate	Use ≤ 1-minute intervals for storing measurement data
Time-zones	Encode timestamps (preferably in UTC format)
Datetime Format	ISO 8601
Encoding	UTF-8 encoding
Language	English if possible

3.1 Recommendations

3.1.1 Sampling Rate

The sampling rate describes how often data is acquired and stored. While the Task 45 fact sheet recommended a sampling rate of at least 5 minutes [1, 2] the newly introduced ISO 24194 Thermal Power Check method [9] requires a sampling rate of at least 1 minute. Hence, to support the ISO evaluation, a sampling rate of at least 1 minute is recommended by the authors. Even though some methods require even higher sampling rates (e.g., spectral methods from Räber [21] and Grossenbacher [22]), their use is relatively rare. However, if quasi-dynamic collector tests according to ISO 9806 are desired, a sampling rate between 1 and 10 seconds is needed.

3.1.2 Time Zones & Datetime Format

Measurement data always needs information about when the measurement took place to correctly interpret the data (based on the time dependency due to weather and demand). For further processing, the time zone of the timestamps must be correct, for example, to compute the angle of incidence or run plausibility checks like "no irradiation at night." A lesson learned from the SunPeek project [20] is that storing data in UTC is recommended, as time zones can change[2] and can often lead to bugs[3], especially in the case of timestamps with daylight saving. Thus, the recommendation is to use UTC for storing the data.

In any case, it is recommended to record timestamps with their associated timezone information. It is strongly urged to follow the timestamp conventions specified in ISO 8601. An example time stamp might be 2023-07-18T11:12:27+00:00, where the +00:00 denotes that it is in UTC.

3.1.3 Encoding

Similar to the time zones, another lesson learned from SunPeek [20] is that encoding can be a problem for storing and sharing data. The standard for most applications is the use of UTF-8, whereas less common encodings like "latin1," for example, can sometimes lead to errors and malformatted labels[4]. Hence, we highly recommend using UTF-8.

3.1.4 Language

Ideally, meta-data and sensor names should be in English, as it is the most used language worldwide. In addition, the characters can be easily encoded with UTF-8.

[2] E.g. https://www.timeanddate.com/news/time/south-sudan-time-zone-2021.html
[3] E.g., https://stackoverflow.com/questions/49777178
[4] Also see: https://en.wikipedia.org/wiki/Mojibake

4 Data-Distribution

Data is rarely stored only at the Programmable Logic Controllers (PLC). In contrast, data needs to be available to the operators both on-site and remotely to allow monitoring and performance evaluation. Hence, it is important to implement data distribution, allowing the transfer of information from the sensors to a permanent (final) storage and making the data accessible for evaluation. This is often a major obstacle in the evaluation of district heating plants. Thus, this section deals with data distribution and shows some proven examples of how data logging and distribution is implemented by some researchers and companies.

4.1 Example Architectures

4.1.1 Example 1

The basic flow of the data can be seen in Figure 3. The sensors transmit the measurement signals to the PLC, which transforms the signal into interpretable data. The data is stored every minute by appending it to a comma-separated-value (CSV) file using the current measurement values (no averaging). The file is accessible to a local computer, which serves as a host for the SCADA system. This setup is used so the operator can access the data on-site even in case of connection problems. The data is also sent to an FTPS server in daily batches, where an application automatically accesses the data and stores it in a database. At this point, the data is regularized to 1-minute intervals and validated. After this step, additional evaluations like fault detection and computation of key performance indicators are done. Finally, the measurement data can be accessed by a user using a web interface. The interface can also be used to set up new plants and add metadata (e.g., service events) to the database.

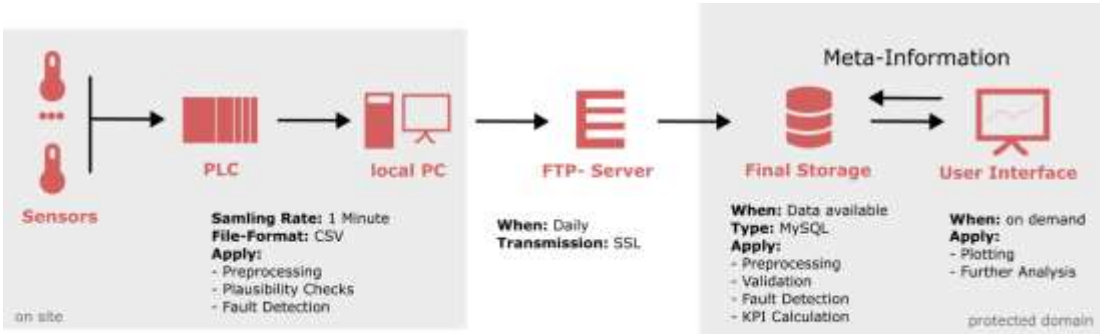

Figure 3: Data flow of example architecture 1.

4.1.2 Example 2

The sensors transmit the measurement signals to the PLC, which transforms the signal into interpretable data. The data is stored in the RAM of the PLC. Every day before 1:00 a.m., the monitoring data of the previous day (0:00 to 23:59) is sent as a CSV file to an FTPS server. The monitoring data of all monitoring devices are assigned to timestamps with a step-with of typically 1 minute. For example, all measured values from January 1st, 2023, go into the file "MSXXXX_202301010000_PLC-query.txt". "XXXX" stands for the sequence number of the respective solar thermal plant. In addition, the monitoring data is also stored locally.
When the monitoring data is analyzed, a data analyst runs a Python script that downloads all the daily files not yet downloaded and stores them in the database. The same script is used to perform several data analysis tasks, e.g., data cleaning, data transforming, and plotting. The data visualization is done via the software tool Visplore.

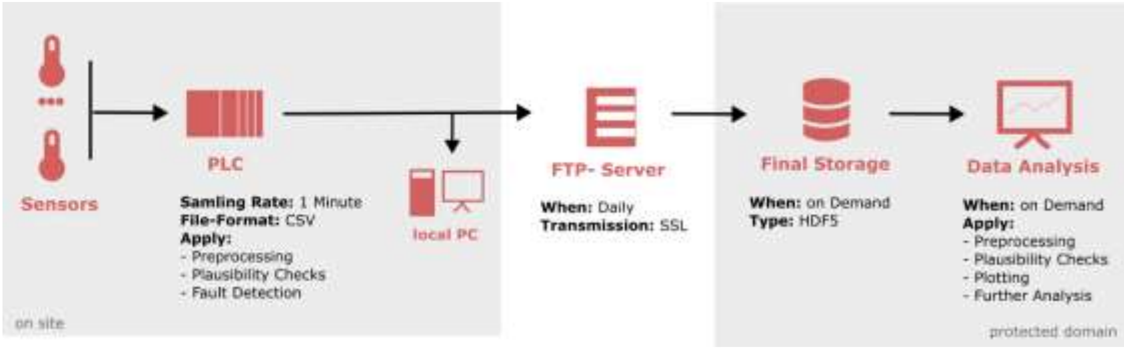

Figure 4: Data flow of example architecture 2.

4.1.3 Example 3

The sensor data is transmitted to remote I/O (Input/Output) islands, which are specialized units associated with each section of the solar-thermal plant. This arrangement allows efficient data aggregation and segregation based on specific equipment and processes. Multiple PLCs then handle the data. These PLCs serve as the control centers for their respective domains, enabling coordinated control and response mechanisms based on the incoming sensor data.
All the PLCs are seamlessly linked to a SCADA system with a centralized server, and for security purposes, they are also connected to a backup server. The servers are physically located on-site and function as repositories of crucial operational data. In addition to real-time signals from sensors, these servers also store calculated values derived from the collected data. This database forms the backbone of the monitoring system, allowing operators and maintenance personnel to gain real-time insights into the plant's behavior using an onsite supervision PC. The SCADA interface also allows viewing, analyzing, and exporting the measurement data for more in-depth evaluations. The SCADA system can also be accessed remotely to export data and can also be connected to an app.

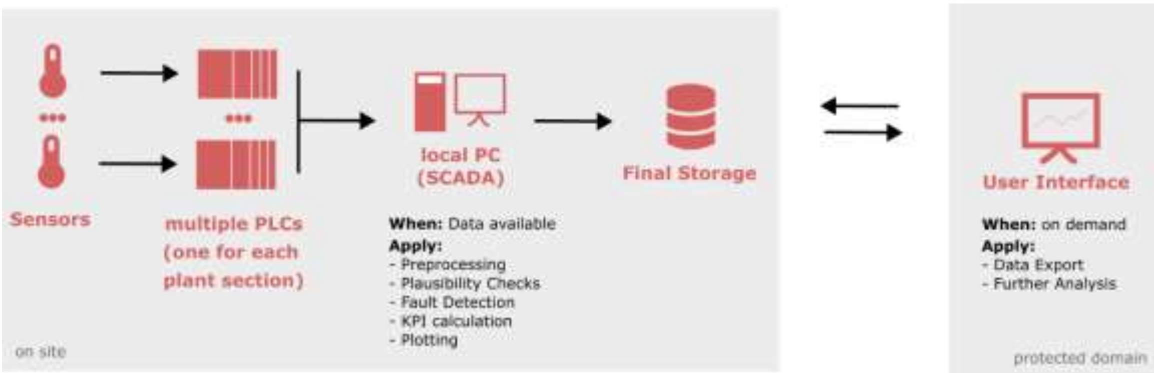

Figure 5: Data flow of example architecture 3.

5 Data Storage

Data must be storable and accessible to enable long-term evaluations. Different data storage technologies exist to support the permanent storage of data and allow the fetching of large portions of plant data for evaluations. This section aims to report what technologies are currently used by Task 68 experts and provide examples of different approaches to permanent data storage.

5.1 Overview

Figure 6 shows which data storage technologies are currently used by the experts of Task 68. These results were obtained by a survey during the 3rd task meeting with 20 participants (multiple answers allowed). The results show that comma-separated value (CSV) text files and Microsoft Excel spreadsheets (.xlsx) are the most common formats for storing measurement data. Both are standalone file formats, whereas it is also possible to store data in a dedicated database, e.g., PostgreSQL or MySQL. Additionally, only a few participants reported using file types or databases specifically dedicated to time-series data (e.g., TimescaleDB and Parquet).

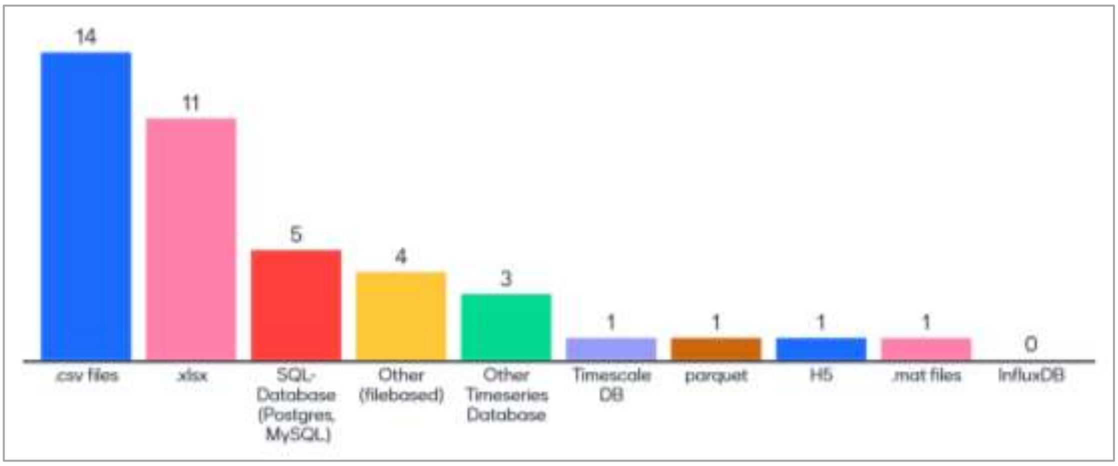

Figure 6: Survey conducted at the 3rd Task 68 Meeting about which storage technologies are currently in use.

5.2 Example Use Cases

The following use cases will show more insights into the various advantages and disadvantages of some of the database technologies. The aim is to aid researchers and companies in choosing the best technology for their use case (without guaranteeing completeness). The texts below are written by researchers/companies that work with solar-thermal data and use the corresponding database technology. To prevent any bias, the paragraphs are anonymized.

5.2.1 Comma-Separated Value (CSV) Files

Technology Description:
Please provide a short description of how the technology works:

CSV files store the measurement data in text format using a symbol (often "," or ";") to distinguish columns and another symbol (typically "line-breaks") to distinguish rows. Usually, the file contains a header row with the sensor names as columns. The following rows typically contain measurements - with one row corresponding to a specific timestamp. Hence, CSV files encode the data in a table-like style.

Context:
Please describe the context the technology is used in:

In this use case (offering monitoring as a service), the data is acquired from different operators. While the original plant data is stored in CSV, the processing is done via Python using HDF5 files, and the visualization and analysis are done with Visplore. Metering reports and calculated values are then stored again as CSV files.

What are the advantages? Why do you use this technology?
The following advantages are the main reasons for using this technology:

- **Common Data Format –** In the present use case, data might be provided by different operators with different data acquisition systems. As CSV is a very common data format, it is easy for the operators to prepare the data accordingly. No file-format conversions are necessary, and errors in data processing are thus mitigated.

- **Usability -** Most people use CSV files because they are a very common data format. Hence, managing the data does not require a software developer or specialized software for handling the data.

- **Simplicity –** In the present use case, the plant structure typically stays the same for the project duration. Only the initial configuration must be done for each plant. By using CSV files, the monitoring procedure can be applied for several similar plants with only minor adaptions of the processing Python script. Given that the plant operator strictly obeys the obligatory guidelines for installing and configuring the monitoring system, these efforts are minimized. Hence, in summary, the handling of the CSV files can be done very efficiently without much overhead due to its simplicity.

Any challenges? How were they tackled?
The following challenges were experienced:

- **Renaming of Sensors / New Sensors** - In some cases, plant operators have changed settings in the PLC without notification. This can result in errors or in the creation of new columns that are created and hence cause additional data processing efforts.

- **Limited Query Speed –** It is slower to query and filter large amounts of data with CSV files compared to databases. However, in this use case, the higher time effort for loading all new daily files is justified, as data from only one year and with a time step of typically one minute is analyzed. In addition, the analysis is performed only every few weeks or even monthly.

Any final thoughts?

Using CSV files for the storage of monitoring data was suitable in the use case presented above. The presented approach, developed around 2010, has been in successful operation for more than a decade with only minor adaptations. As explained above, the plants and requirements on their monitoring system were uniform. However, this solution would be less suitable in case more data must be processed. In future projects, time-series databases will be the method of choice. The monitoring team is building comprehensive knowledge and skills in data processing technology applying time-series databases.

5.2.2 MySQL Database

Technology Description:
Please provide a short description of how the technology works:

Relational Databases store data in multiple tables using a row-based format. In case new data is stored, additional rows of data can be inserted into the tables. In contrast to CSV files, special data structures (e.g., B-Trees) are used to allow fast access to the data (using indexing). Users can interact with the tables using the Structured Query Language (SQL) to fetch and store data efficiently.

What are the advantages? Why do you use this technology?
The following advantages are the main reasons for using this technology:

- **Query Speed -** Compared to CSV files, it is faster to query and filter large amounts of data. While fetching all data points for certain periods is faster with CSV, SQL databases are much faster for fetching single data points for larger timespans.
- **Storage Flexibility -** Another main advantage compared to CSV data is that meta-data can be stored alongside the measurement data. Through links between table elements, it is possible to store all data relevant to a respective plant in the database.
- **Integrability -** Relational databases put a high emphasis on security and consistency – so each transaction (write/fetch/update) leaves the database in a valid state. Relational databases have been used for decades and are well supported by programming interfaces and software solutions. Hence, it is fast to implement software for accessing the data, and easy to get support in case of questions. It also allows “live” updates and multithreading.

Any challenges? How were they tackled?
The following challenges were experienced:

One difficulty in relational databases is that the structure of the database should not change during runtime. Adding columns or changing tables at a later point in time is possible but typically requires some effort and time.
For example, our initial database design used a simple numeric entry in the “plants” table to store the total collector area of the plant. However, in some cases, new collectors are installed. To correctly calculate the specific solar yield, for example, the change in collector area and the time of the change must be known. Hence, there was some effort to change the database structure and add a new table, “plant-details” and “plant-detail-values,” which allows the storage of details (e.g., the collector area) with a start- and end date.
Similarly, initial designs for storing the measurement data consisted of a table with a [timestamp, datapoint_A, datapoint_B, …] structure similar to a CSV file. However, adding new data points to the plant means that a new column has to be created. Due to the row-wise storage of relational databases, this is a very resource-heavy task. Instead, the design was changed to a [datapoint, timestamp, value] table. With this design, adding new data points is not a problem anymore. As a drawback, however, the effort to query and format the data has now increased.

Any final thoughts?

Relational databases are a good choice for permanent data storage due to their usability and flexibility in storing all required data. It is also easy to integrate it into applications. Even though the read/write speed is better for Time-Series databases, the speed of relational databases is still sufficient for our use cases. Should the need arise, we will switch to the Time-Series database for the measurement data, while CSV data is still a good interface for sharing measurement data.

5.2.3 Parquet Files

Technology Description:
Please provide a short description of how the technology works:

Parquet is a file-based data storage that uses a column-oriented approach. While a row-oriented approach (e.g., CSV files, relational databases) stores the data based on the timestamp (e.g., time -> values for each sensor), the column-oriented approach stores the data based on the sensor (e.g., sensor -> values for each time). This can be very helpful in compressing data, as the measurements for a sensor are typically very similar to each other. Another advantage is if only the data of some sensors is needed: While the row-oriented approach requires reading the whole row, the column-oriented approach can parse only the desired sensor columns directly.

What are the advantages? Why do you use this technology?
The following advantages are the main reasons for using this technology:

- **Compression –** As discussed above, the data storage technology allows very efficient compression and hence reduces the required amount of disc space to store the data. This can be helpful, especially for on-site storage with limited disc space, but it also speeds up the reading process.
- **Query Speed –** In addition to compression, the sensor-oriented storage of data also leads to an increased query speed compared to CSV and relational databases: In cases when only some sensors are required, the necessary data can be fetched without loading unnecessary data from the other sensors.

Any challenges? How were they tackled?
The following challenges were experienced:

- **Limited Multi-Threading** – Parquet is a file-based storage technology. Difficulties arose when multiple users tried to write on the same data file. While multiprocessing is possible for reading and writing different sensors, only one sensor write can be done at a time.
- **Integration and Security** – Compared to relational databases, there is no concept of transactions that ensures that data is always up to date. Hence, data might get corrupted if a write fails. In addition, updating data might be more challenging compared to relational databases, as it requires rewriting the file.

Any final thoughts?

Working with parquet data is very fast and well supported by many libraries and software. However, implementing the read-and-write operations is more involved compared to CSV and relational databases and hence requires more knowledge. In summary, there is a trade-off between query speed and usability.

5.2.4 HDF Files

Technology Description:
Please provide a short description of how the technology works:

The Hierarchical data file (HDF) format is a file-based data storage (as are CSV and Parquet). HDF allows storing measurement data in “datasets” but also allows adding meta information and connections between datasets inside of “groups.” Both “datasets” and “groups” are indexed using B-trees and hence allow fast access and filtering. HDF5 is the latest version of the HDF format.

What are the advantages? Why do you use this technology?
The following advantages are the main reasons for using this technology:

- **Query Speed -** It is faster to query and filter large amounts of data with HDF5 files compared to CSV files. Compared to CSV files, the HDF5 format provides compression and indexing to speed up fetching data.
- **Compression –** Compared to CSV files and relational databases, the setup of the HDF5 files allows for efficient data compression. With the given data sizes, no drawbacks were experienced in processing and storing the data permanently in this single HDF5 file. The data size of the HDF5 files in this monitoring use case was usually below 1 GB.
- **Usability –** Having the data stored in one file can be beneficial for exchanging data. In addition, HDF5 is well integrated into many software libraries like Python, which is used for data analysis. It also allows the organization of multiple groups and adding metadata to the datasets.

Any challenges? How were they tackled?
The following challenges were experienced:

- **Fixed Dataset Structure** – Similar to the CSV files, the structure of HDF5 datasets is fixed. This can lead to problems when new columns are added or renamed by the operators. However, in the current use case, such events that justify manual corrections or rewriting “datasets” are rare.

Any final thoughts?

Using HDF5 files to store monitoring data is suitable for our use case. The approach has been used for over a decade and worked well with only minor adaptions. The improvement in speed compared to CSV files is evident, and the integration with Python and Visplore works very well.

6 Data Validation

For performing analysis, it is essential to work with correct data. Otherwise, the results will not reflect the actual behavior of the plant. Hence, data must be validated, and automated plausibility checks can greatly support users in doing so. Therefore, this section provides commonly used validation and pre-processing algorithms in the following two subsections. In addition, Table 13 lists available open-source software that contains some of the algorithms described in this section.

Table 13: Software implementing quality and data preprocessing algorithms.

Software	Description
PVAnalytics	Python library for analyzing PV plant data, including many useful functions for quality checks for solar irradiance, weather, and general data quality issues. Link: https://pvanalytics.readthedocs.io
SunPeek	Python library and open-source tool implementing the Performance Check method ISO 24194, including some data quality checks as part of data processing. Link: https://gitlab.com/sunpeek/sunpeek
Dronninglund Pit Storage Data [18]	Open-source data about a water pit thermal energy storage, located in Dronninglund, Denmark. The GitHub repository includes a public Python notebook that was used for the data treatment. Link: https://doi.org/10.1016/j.solener.2022.12.046 Link: https://github.com/PitStorages/DronninglundData
FHW Arcon South dataset [19]	Open-source measurement data from a collector array located in Graz, Austria. The document includes a description of the preprocessing algorithms used. Link: https://doi.org/10.1016/j.dib.2023.109224 Link: https://gitlab.com/sunpeek/zenodo-fhw-arconsouth-dataset-2017

6.1 Validation Algorithms

Validation algorithms try to analyze whether measurement data is reliable, for example, by spotting statistical outliers or comparing the measurements with physically plausible ranges. Discarding invalid data ensures that the results of further evaluations are correct. A list of common validation algorithms can be seen in Table 14.

Table 14: Summary of Plausibility Check Algorithms.

Plausibility Checks	
Algorithm	Short description
Min-Max-Check	Tests if the sensor value exceeds the min-max boundary and replaces values with NAN or meaningful values [3]. Open-source implementations are available in SunPeek [20].
Sensor-Hangs	Test if sensor values are constant for too long, i.e., stuck values, and replace them with NAN values [3]. Opens source implementation available in pvanalytics[23].
Irradiation at night	Irradiation during the night can imply that there is a wrong offset in the measurement. Idea from Visplause [24].
Ambient Temperature Offset Detection	Ambient Temperature typically does not change rapidly but increases and decreases continuously. Drastic changes can be signs of signal transmission problems or data-logging issues for example. Implementation available in pvanalytics [23].
Heat meter value shift	Heat meter counters are often cumulative, i.e., monotonically increasing. In case the count goes down or experiences large changes (compared with maximum expected yield for example), this can indicate a replacement of the heat meter or other issues.

Step change	Detect unfeasible drastic change in measured values, e.g., irradiance should not change faster than 1000 W/m²/min. [23]
Maximum and minimum feasible values (model-based)	Detect unrealistic or physically impossible values. Pvanalytics[23] contains algorithms for checking the plausibility of solar irradiance measurements (e.g., maximum value based on extraterrestrial irradiance) and humidity and temperature (e.g., by comparing with irradiation).
Outlier detection via statistics	For example, by checking for univariate outliers (e.g., single points compared to their typical distribution; too drastic changes of one value to the previous one). Idea from [24] but also implemented by Pvanalytics [23]
Ignored ranges	Exclude user-defined time periods from the data, in case events (e.g., services or faults) have occurred which falsify the data. Typically sets the values to NAN during such periods [18], [20].

6.2 Preprocessing Algorithms

Preprocessing algorithms are supposed to prepare the data for further evaluation. Common examples of preprocessing include regularization, which transforms the data to equidistant timestamps, and interpolation, dealing with missing data. This might help to simplify algorithms or reduce unexpected behavior, while some algorithms even require specific data properties like equidistant timestamps [25]. A list of pre-processing algorithms can be found in Table 15.

Table 15: Summary of preprocessing algorithms.

Pre-Processing Algorithms	
Algorithm	Short description
Drop duplicates	Discards duplicated timestamps, by using either (i) the first occurring value, (ii) an aggregated value like the average for example, or (iii) by replacing with NAN [20], [24].
Regularization	Resampling to equidistant timestamps for easier computation of key performance indicators and for merging data from multiple files. Typically done with linear spline interpolation [3], [18], [26].
Interpolation (short-term)	Replaces short periods of NAN data with linearly interpolated values. The period for which this is acceptable depends on the measured quantity [3], [26].
Interpolation (model-based)	Replaces periods with NAN data by using machine learning models or physical models relying on different sensors. For example, replaces NAN values for storage temperatures by interpolating adjacent temperature measurements [18]. Another example is using meteorological data from a nearby station for checking and interpolating measurement data on site (e.g., ambient temperature) [18]. Only recommended if a serially complete dataset is absolutely necessary.

7 Literature

[1] S. Knabl, I. C. Fink, and I. W. Wagner, 'ERSTELLT IM RAHMEN DES IEA TASK 45 „LARGE SYSTEMS“ UND IM RAHMEN DES NATIONALEN BEGLEITFORSCHUNGSPROGRAMMS ZUR FÖRDERAKTION DES KLIMA- UND ENERGIEFONDS "SOLARTHERMIE - SOLARE GROẞANLAGEN"', p. 16, 2012.

[2] AEE INTEC and FH Oberösterreich, 'Leitfaden zum Monitoringkonzept im Rahmen des Begleitforschungsprogramms zur Förderaktion des Klima- und Energiefonds "Solarthermie - Solare Großanlagen"', Klima- und Energiefonds, Gleisdorf, 2021.

[3] L. Feierl and P. Luidolt, 'IEA SHC FACT SHEET 55.B-D3.2. Automated monitoring, failure detection of key components, control strategies and self-learning controls of key components.' 2020. [Online]. Available: https://task55.iea-shc.org/Data/Sites/1/publications/IEA-SHC-T55-B-D.3.2-FACT-SHEET-Automated-Monitoring.pdf

[4] G. Faure, M. Vallée, C. Paulus, and T. Q. Tran, 'Fault detection and diagnosis for large solar thermal systems: A review of fault types and applicable methods', *Sol. Energy*, vol. 197, pp. 472–484, Feb. 2020, doi: 10.1016/j.solener.2020.01.027.

[5] K. Kramer, D. Tschopp, J. E. Nielsen, K. Kramer, and P. Ohnewein, 'Review and Outlook on Methods for Product Certification, Energy Yield Measurement, Power Output Determination and Commissioning Test for Large Solar Thermal Installations', in *Proceedings of the ISES Solar World Congress 2019*, Santiago, Chile: International Solar Energy Society, 2019, pp. 1–14. doi: 10.18086/swc.2019.06.02.

[6] A. Zirkel-Hofer *et al.*, 'Improved in situ performance testing of line-concentrating solar collectors: Comprehensive uncertainty analysis for the selection of measurement instrumentation', *Appl. Energy*, vol. 184, pp. 298–312, Dec. 2016, doi: 10.1016/j.apenergy.2016.09.089.

[7] T. Effertz *et al.*, 'StaMep - Entwicklung von standardisierten Messmethoden und Prüf- ständen für den Leistungsnachweis von Bauteilen solar- thermischer Kraftwerke Abschlussbericht zum Forschungsvorhaben'.

[8] *EN ISO 9806 Solar energy — Solar thermal collectors — Test methods*. Geneva: ISO, 2017.

[9] *ISO 24194: Solar energy — Collector fields — Check of performance*. Geneva: ISO, 2022.

[10] D. Tschopp *et al.*, 'GUIDE TO STANDARD ISO 24194:2022 - Power Check [unpublished manuscript]'.

[11] K. Kramer *et al.*, 'GUIDE TO STANDARD ISO 9806:2017 A Resource for Manufacturers, Testing Laboratories, Certification Bodies and Regulatory Agencies', Fraunhofer ISE, Freiburg, 2017. Accessed: Feb. 22, 2020. [Online]. Available: http://rgdoi.net/10.13140/RG.2.2.30241.30562

[12] 'Connect your system to solarheatdata.eu', solarheatdata.eu. Accessed: Nov. 24, 2023. [Online]. Available: www.solarheatdata.eu

[13] P. Ohnewein, D. Tschopp, R. Hausner, and W. Doll, 'Dynamic Collector Array Test (D-CAT). Final Report FFG Project 848766 - MeQuSo. Development of methods for quality assessment of large-scale solar thermal plants under real operating conditions.', AEE INTEC, Gleisdorf, 2020.

[14] D. Tschopp *et al.*, 'IEA SHC FACT SHEET 55 B-D1.1 - Application of Performance Check (PC) Method to Large Collector Arrays', IEA SHC Task 55, 2021. [Online]. Available: http://files.iea-shc.org/public/rba/b-d1-1-application-pc-test-methods-large-collector-arrays.pdf

[15] J. Jaus, 'Automated sensor operation systems', presented at the 11TH INTERNATIONAL CONFERENCE ON CONCENTRATOR PHOTOVOLTAIC SYSTEMS: CPV-11, Aix-les-Bains, France, 2015, p. 120002. doi: 10.1063/1.4931559.

[16] V. Quaschning, *Regenerative Energiesysteme: Technologie - Berechnung - Simulation*, 9., Aktualisierte und erweiterte Auflage. München: Hanser, 2015.

[17] L. Feierl, 'SunScreen: visual fault detection and diagnosis for solar-thermal systems', 2023, doi: 10.25365/THESIS.73521.

[18] I. Sifnaios, G. Gauthier, D. Trier, J. Fan, and A. R. Jensen, 'Dronninglund water pit thermal energy storage dataset', *Sol. Energy*, vol. 251, pp. 68–76, Feb. 2023, doi: 10.1016/j.solener.2022.12.046.

[19] D. Tschopp *et al.*, '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*, vol. 48, p. 109224, 2023, doi: 10.1016/j.dib.2023.109224.

[20] M. Hamilton-Jones, P. Ohnewein, L. Feierl, and D. Tschopp, 'SunPeek'. AEE-Intec, SOLID Solar Energy Systems, 2023. [Online]. Available: https://gitlab.com/sunpeek/

[21] R. Räber, 'Spektralmethode zur Fehlerfrüherkennung in wärmetechnischen Anlagen', ETH Zurich, 1997. doi: 10.3929/ETHZ-A-001845425.

[22] U. Grossenbacher, ‘Qualitätssicherungssystem für Solaranlagen - Methode zur permanenten Funktionskontrolle thermischer Solaranlagen’, EnergieBüro Grossenbacher, 77269, 2003. Accessed: Jul. 19, 2022. [Online]. Available: https://www.aramis.admin.ch/Default?DocumentID=64322&Load=true

[23] ‘PVAnalytics’. pvlib, 2022. [Online]. Available: https://pvanalytics.readthedocs.io/en/stable/

[24] C. Arbesser, F. Spechtenhauser, T. Mühlbacher, and H. Piringer, ‘Visplause: Visual Data Quality Assessment of Many Time Series Using Plausibility Checks’, *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 641–650, Jan. 2017, doi: 10.1109/TVCG.2016.2598592.

[25] L. Feierl, V. Unterberger, C. Rossi, B. Gerardts, and M. Gaetani, ‘Fault detective: Automatic fault-detection for solar thermal systems based on artificial intelligence’, *Sol. Energy Adv.*, vol. 3, p. 100033, 2023, doi: 10.1016/j.seja.2023.100033.

[26] P. Ohnewein *et al.*, ‘METHODIQA - Development of a Quality Assurance Methodology for Renewable Heat Systems Based on Intelligent Operational Monitoring’, *Energy Procedia*, vol. 91, pp. 376–383, Jun. 2016, doi: 10.1016/j.egypro.2016.06.285.