



HEAT CHECK

Deliverable 2: Full documentation on
access, handling, privacy

Written by Dennis Lange & Tore Kjeld
November 2024



center**denmark**
intelligent energy

 **nnovationsfonden**

The logo icon for Innovationsfonden, consisting of a dark blue diagonal slash mark.

Table of content

Contents

Table of content	2
1 Introduction	4
2 The cloud platform	5
2.1 Physical infrastructure	5
2.2 Software infrastructure	5
2.2.1 Data streaming layer.....	5
2.2.2 Data access layer	6
2.2.3 Data deep storage layer	6
2.3 Data infrastructure.....	6
2.3.1 Input sources	6
2.3.2 Data Lake.....	7
2.3.3 Output Sources	7
3 Privacy	9
3.1 Data controller.....	9
3.2 Contact information.....	9
3.3 Purposes and legal basis for processing your personal data	9
3.4 Categories of personal data	10
3.5 Recipients or categories of recipients	10
3.6 Transfer to recipients in third countries, including international organizations.....	10
3.7 Source of your personal data	10
3.8 Agreement on joint data responsibility	10
3.9 Retention of your personal data.....	10
3.10 Right to withdraw consent	11
3.11 Your rights	11
3.12 Complaint to the Danish Data Protection Agency	11
4 Data handling.....	12
5 Data access	15
5.1 Create user.....	15
5.2 Explore documentation	16
5.3 Explore available data sets	16
5.4 Explore a specific data set.....	16
5.5 Query the data	17
5.6 Download the data.....	18

5.7 Connect to API.....	19
5.7.1 Create token	19
5.7.2 Access data through the API	20

1 Introduction

This report aims to describe the cloud platform used in the HeatCheck project.

It will highlight subjects such as privacy and data handling, to document how these topics and concerns are addressed. Starting from a generalized stance, it will become increasingly specific, to end up showing exactly how data is accessed by partners in HeatCheck.

2 The cloud platform

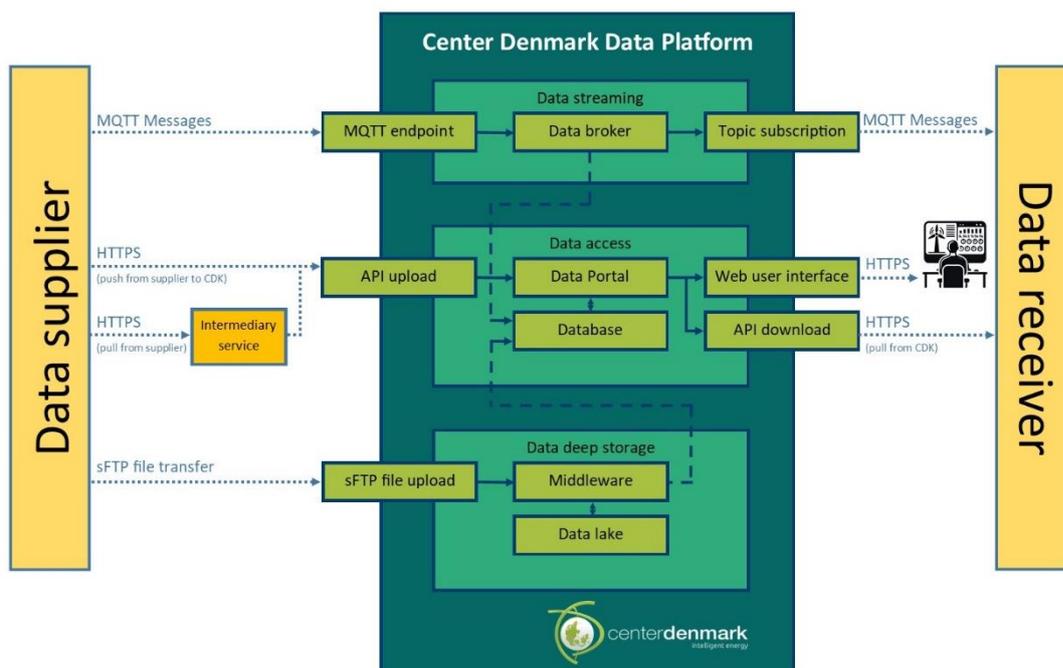
The Center Denmark Data Platform, or the HeatCheck Cloud Platform, is built for security, safety, reliability and scalability.

2.1 Physical infrastructure

The physical infrastructure is built around a redundant system, with servers in several countries. This ensures resilience against almost any physical disruption, while also providing the opportunity to connect to the nearest server.

2.2 Software infrastructure

The software is built on several opensource components, connected and expanded by custom software, to enable seamless interaction between components. To ensure a high degree of connectivity, the platform has several layers for both data ingestion, storage and extraction. This is in part to ensure flexibility to be able to receive data in many different ways. In part to have different levels, depending on the needed data speed, from when it is pushed to us, until it is available to partners.



2.2.1 Data streaming layer

For the highest data speed, the streaming section of the platform is used.

The MQTT protocol is built to update when a value changes, instead of every xx seconds. This also means that it is unpredictable when a value will be updated. This is handled by subscribing to the message topic. Then you receive the new value whenever it is updated.

In most cases, the time from when the message is sent from a data supplier to our MQTT endpoint, having passed through the Center Denmark platform, to when it has arrived at the subscriber, is less than a second.

At Center Denmark we further persist the data in the databases of the access layer as well.

2.2.2 Data access layer

For another high data speed connection, the data access layer can be used directly.

Data is then pushed directly from the data supplier, conforming to the data model that it will be exposed in, and directly connected to the database and portal, where it is available to data receivers. It is available to receivers both through our web interface and directly from the API. A common use case is to use the user interface to explore the data and define the particular queries for the API calls, which are then used in an automated setup. This will be described further in the chapter “Data access”.

In most cases, the time from when the message is sent from a data supplier to our API, to when it is available to data receivers on our portal, is less than a second.

Selected data from the access layer is persisted further to the data lake in the deep storage layer.

2.2.3 Data deep storage layer

For larger amounts of data in bulk, the deep storage layer is used directly, mostly through sFTP.

Here files are uploaded, to then be split and transformed through middleware services, before it is stored in our data lake. To extract data, other middleware services are defined, to extract specific data and transform it to an agreed data model, before making it available in the data access layer. This will be described further in the chapter “Data handling”.

2.3 Data infrastructure

Flow of Data (Data Lake perspective)

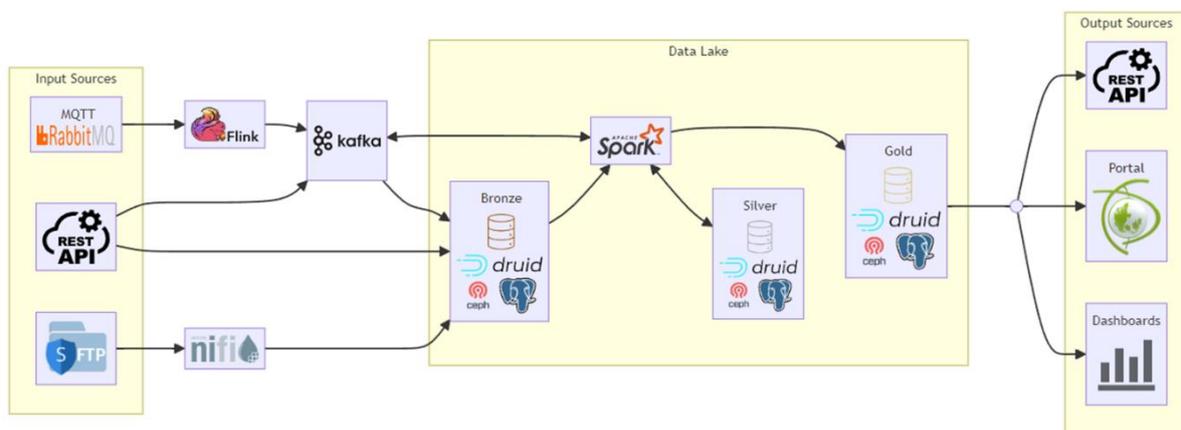


Diagram ? shows the infrastructure of Center Denmark from a Data Lake perspective. The input-sources, the data lake for data processing, and the output sources.

2.3.1 Input sources

Message Brokers:

- MQTT and RabbitMQ facilitate real-time message streaming.
- Flink acts as a stream processing framework for processing these messages before passing them to Kafka.

REST APIs: External APIs provide data in a structured or semi-structured format.

File-based Inputs:

- sFTP servers serve as a source of files that may contain batch data.
- For some occasions, Azure Blob storage is used instead of the Center Denmark's sFTP-server for transferring data.
- NiFi: Acts as a data ingestion and integration tool, routing files from sFTP or other sources, like Blob storage, into the system.

From here, Kafka acts as a central hub for ingesting data from Flink, API's and NiFi.

2.3.2 Data Lake

Data is deposited into the Data Lake as raw as possible and processed through a medallion-structure (bronze, silver and gold) to provide cleaned, uniform and enriched data for several applications. Here, Spark is the core data processing engine in the Data Lake.

- **Bronze layer:** Here, the raw data is formatted correctly, i.e. from .json, .csv or .parquet to the Delta format. This is due to the Delta format being more effective for performance, scalability and more uniform data schemas.
- **Silver layer:** Here, data is cleaned and formatted in two possible variations:
 - **Uncleaned:** Here, data is put on a generic data model and is not cleaned.
 - **Cleaned:** Here, data is put on a generic data model and some universally requested cleanings are performed (outlier detection)
- **Gold layer:** Here, data is formatted for specific applications – i.e. dashboards, PowerBI reports, specific requested datasets

For each layer, data can be pushed to relevant databases, essentially constituting a Data Lake House.

For master data, Postgres is utilized as a relational database due to its strong relational schemas and rich query capabilities.

For time-series data, Druid is utilized as a time-series database due to its optimization for handling large-scale quantity of data and high-performance analytics for time-series data.

2.3.3 Output Sources

Data from the databases are pushed to various products for exposing data – i.e. API for users to pull enriched data for their analytics.

The Center Denmark Portal is utilized as a common place for data users to explore datasets, find relevant meta-data, query data, and information on how to use the API to access data.

This Portal supports role-based access, i.e. the following data-restrictions apply, however login on the site will always be needed to access data.

- Public dataset

This is available for everyone, either fully anonymized data or open-data sources.

- Visible dataset

This is available for specific accounts, but the existence of the data set is available for everyone. Thus, it can be requested if you meet the criteria – either legally, commercially or both.

- Private dataset

This is available for specific accounts, and the awareness of this data set is only available for the same accounts. This can be due to restrictions from the data owners, or the fact that it is irrelevant to the regular user of the platform.

The role-based access can be for entire data sets or for limited number of rows in the datasets – i.e. filters for accessibility for specific meters.

Center Denmark also make data accessible in other forms – for example on **dashboards** to illustrate the value of cleaned and enriched data. This could be consumption data, building data and geo-data to show relevant dynamics in the energy- and utility sector.

3 Privacy

This is the complete information on the Center Denmark privacy policy and processing of personal data. For the HeatCheck project, individual agreements are made with each data owner, ensuring that all legal requirements are met across the project. For this purpose, several DPOs and legal entities have been involved, to further ensure full legal compliance.

As this is a legal document, it should be presented in full, to not take anything out of context. However, for this project, most data owners are working with Center Denmark through an “Agreement on joint data responsibility”.

Data is shared with DTU and DTI as research institutes under “research basis” and using individual “Data transfer Agreements”.

3.1 Data controller

Center Denmark is the data controller for the processing of the personal data we have received about you.

Center Denmark
Vendersgade 74
DK-7000 Fredericia
Company Reg. No.: 40868399
Phone: +45 23 40 86 12
E-mail: info@centerdenmark.com

3.2 Contact information

Should you have any questions regarding the processing of your personal data, you are always welcome to contact us via:

E-mail: info@centerdenmark.com
Phone: +45 23 40 86 12

3.3 Purposes and legal basis for processing your personal data

We process your personal data for the following purposes:

- The personal data is used to identify consumption profiles and analyze loads on specific lines, etc. The personal data will be correlated with other data sources such as weather data, BBR data, price data, and production data, among others, to identify flexibility potentials.
- This processing enables detailed analyses of the supply network concerning production and consumption to identify and utilize flexibility potential, as well as to visualize the energy balance and possible sector coupling.
- Additionally, we may process your personal data to respond to your inquiries.

The legal basis for our processing of your personal data is:

- The rules in the General Data Protection Regulation (GDPR), Chapter II, specifically Article 6 (1) (a), (b), and (f).
- The national data protection law, § 6, subsection 1, cf. particularly GDPR Article 6(1)(f), regarding processing necessary to pursue a legitimate interest not overridden by the

interests or fundamental rights and freedoms of the data subject - taking into account the categories of personal data processed. The legitimate interest justifying the processing is the consideration of the green transition and the environment and, more generally, the promotion of the development of an optimized pattern of energy consumption, which will also benefit energy consumers.

3.4 Categories of personal data

We process general personal data about you, including basic meter data, e.g. meter number and installation address, as well as consumption and production data as stated on the meter. If you contact us, we also process your name, contact details, and possibly information about your workplace. Personal data is anonymized prior to any form of visualization for others.

3.5 Recipients or categories of recipients

We disclose or assign your personal data to the following recipients:

- We disclose your personal data to partners in research and development projects.
- Data processors who process your personal data on behalf of Center Denmark. Center Denmark continuously monitors that the processing by data processors complies with legal requirements.

3.6 Transfer to recipients in third countries, including international organizations

We do not transfer your personal data to recipients outside the EU and EEA.

3.7 Source of your personal data

Data is collected in collaboration with utility companies for heat, electricity, water, and gas. Information is collected from utility companies covering your local area. If you contact us, your information may also come directly from you.

3.8 Agreement on joint data responsibility

We have entered into an agreement on joint data responsibility with a number of utility companies in connection with the preparation of detailed analyses of the supply network around production and consumption in order to identify and exploit flexibility potential and visualize the energy balance and possible sector coupling. According to the agreement, we are responsible for ensuring the security of processing, including determining the tools used for processing of personal data and ensuring that personal data is anonymized before publication of the analyses. You can, at any time, contact us or the utility company to exercise the rights under the General Data Protection Regulation (as listed below in point 11). We handle your request/inquiry in collaboration with the utility company.

3.9 Retention of your personal data

We retain your personal data for up to 10 years, but usually delete it before this deadline. The necessity of the specified retention period is continuously assessed, so that only what is necessary for fulfilling the purpose is retained. The retention period is reviewed at least once a year.

3.10 Right to withdraw consent

As a basic principle, we do not use consent as a legal basis for processing. You can contact us at any time if you do not want your data to be part of our work. You can do this by contacting us using the contact information provided above in section 2. 3.

3.11 Your rights

Under the GDPR, you have several rights regarding our processing of personal data about you. If you wish to exercise your rights, you can contact us using the contact details provided above in section 2. These rights include:

- Right of access: You have the right to access the personal data that we process about you, as well as a range of additional details.
- Right to rectification (correction): You have the right to have incorrect personal data about yourself corrected.
- Right to erasure: In certain cases, you have the right to have personal data about yourself deleted before the time of our regular general deletion occurs.
- Right to restriction of processing: In certain cases, you have the right to have the processing of your personal data restricted. If you have the right to restrict processing, we may only process the information in the future – apart from storage – with your consent or for the purpose of establishing, exercising, or defending legal claims, or to protect a person or important societal interests.
- Right to object: In certain cases, you have the right to object to our or lawful processing of your personal data. You can also object to the processing of your information for direct marketing purposes.
- Right to data portability: In certain cases, you have the right to receive your personal data in a structured, commonly used, and machine-readable format, and to have these personal data transmitted from one data controller to another without hindrance. You can read more about your rights in the Danish Data Protection Agency's guide on data subjects' rights, which you can find at www.datatilsynet.dk.

3.12 Complaint to the Danish Data Protection Agency

You have the right to lodge a complaint with the Danish Data Protection Agency if you are dissatisfied with the way we process your personal data. You can find the contact information for the Danish Data Protection Agency at www.datatilsynet.dk.

4 Data handling

Section 2.3 explains the general setup of Center Denmark's infrastructure from a data perspective. Here, the focus is on the specifics for Heatcheck.

Revisiting the input sources, Center Denmark handles the following:

External API's

- Bolig Selskab Sjælland
- Brunata
- Techem
- Aalborg Forsyning – To be determined

Common for these solutions are that each of these companies provide an external API to access their data. Here, Center Denmark extracts the data and load it into the Data Lake House. For Aalborg Forsyning, discussions are still ongoing for how to receive data in the most optimal manner, API's or sFTP.

sFTP:

- Horsens Varme
- HOFOR
- Ista – To be determined

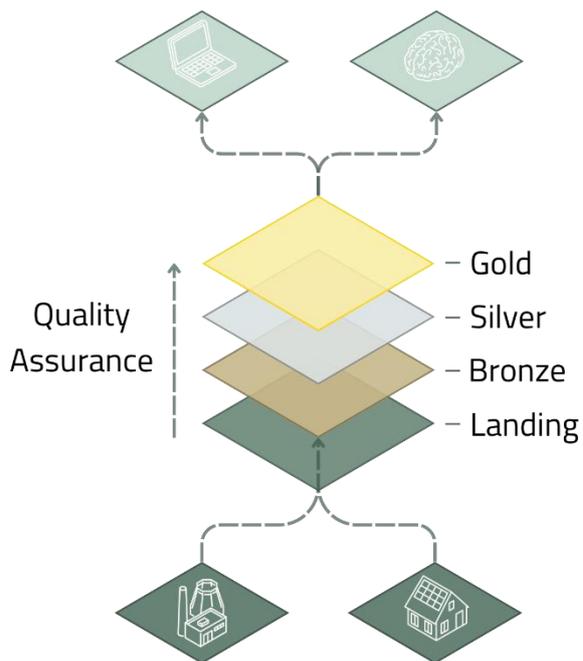
Here, Horsens Varme pushes data to Center Denmark's sFTP-server by a requested name-standard, thus automatically allowing NiFi to extract and load the data to the Data Lake House.

In the HOFOR-case, Center Denmark has access to their sFTP-server and NiFi is manually configured to extract and load the data.

For Ista, it is not determined how data will be facilitated but is under investigation.

In the project, no streaming brokers are utilized (**MQTT, RabbitMQ**), as it has not been relevant for the type of data that is facilitated. The quantity and the frequency of data are sufficiently low to handle everything as batch-processing, hence favoring API's and sFTP.

Figure ?? shows a more simplified figure for how the data is handled in the Data Lake House.



- **Landing Layer**

Here, data from the API's or NiFi, are loaded into the Landing layer. This is completely raw, and in this specific case, it is either .json files or .csv files.

- **Bronze layer**

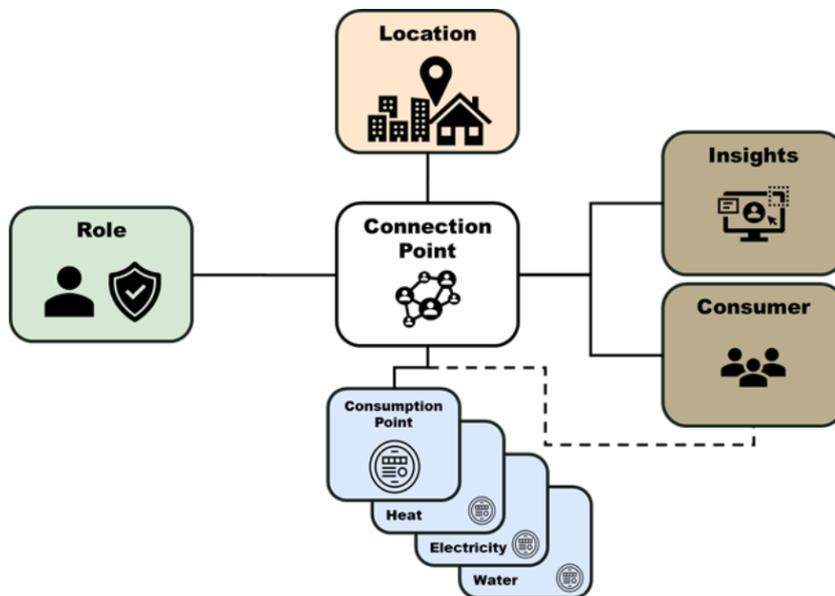
Here, the .csv or .json files are converted to the Delta format. This is the only transformation which happens at this layer.

- **Silver layer**

The silver layer in Heatcheck expands on the work done by Center Denmark in the Generic Data model for district heating consumption, published here:

<https://www.centerdenmark.com/media/n0ih03m4/generisk-datamodel-for-fjernvarmedata-v2-0.pdf>

The generic data model as is, is shown in Figure ??



This model has 3 important domains: **Role, Location and Data (Insights and Consumer).**

Hence, it revolves around the possibility of connecting district heating consumption data with:

- a. Location – this domain enables the possibility to verify the exact address matching the consumption, which allows for enriching data with additional building data, more precise weather data, price data, demographic data and much more.
- b. Role – this domain enables the possibility to provide role-based access to datasets depending on the ownership of data and subsequently read permissions.

The main driver behind this is the consumption point, which is an abstraction from the meter ID. This allows for a physical location-based entity and can map several metering devices to it. In addition, the connection point serves as an abstraction to include multiple consumption points, to allow the generic data model to handle consumption points for electricity, heat and water. Here, role relates to who has relevant access to the data for the consumption point, i.e. data suppliers specify the access, and the data receivers can obtain read permission for the data.

The extension in Heatcheck is to include the heat-cost allocators in this model, as it connects with the metering device as an IoT-device.

The point of the generic data model is to provide data to the users (in this case, DTU and DTI) in a generic form so the Heatcheck product can reach proper scalability, and the users can expect compatible data from any buildings with similar equipment (i.e. district heating meters and heat-cost allocators) as well as use the correct building- and weather data.

Center Denmark provides no additional cleaning other than ensuring there are no duplicates – this is due to the need from the users to have data as raw as possible but enriched with building and weather data.

- **Gold layer**

The gold layer serves as an application-based data processing layer. Here, specific operations can be done at the request of the data users.

Here, a common schema will be utilized and relevant data from the silver layer will be structured in this schema to create a common dataset for consumption, building information and weather data.

In Heatcheck, the gold-layer will be utilized in two variations, for the given heating seasons:

- **Heating season 2:** Data will be directly processed from the Bronze-layer, to provide a fast-track for DTU and DTI to investigate data needs.
- **Heating season 3:** Data will be processed from the Silver-layer, to provide a more scalable solution for the Heatcheck product.

As data is processed, it is pushed to the relevant databases as outlined in section 2.3, and in this case, made available on the Portal for DTU and DTI to investigate and query data, and eventually access it by API. This will be elaborated further in section 5.

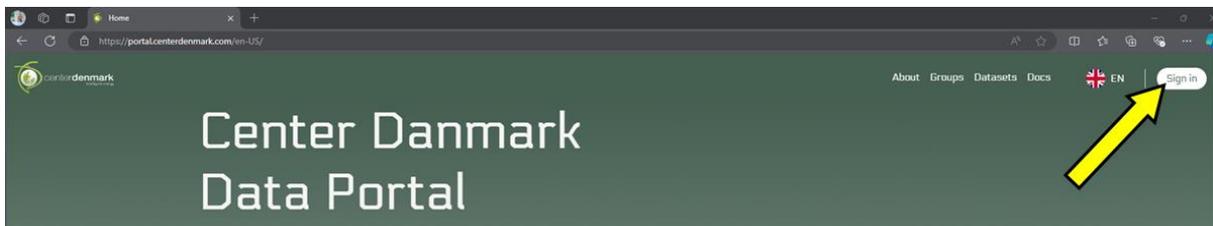
5 Data access

At Center Denmark, the data can be accessed in two main ways. The first way is through our online portal, which is a web user interface. This is a great place to explore data and investigate the different content of the data sets. The other way is through a REST API, which is mainly used for machine-to-machine interactions. A common use is to use the portal for exploration, to then define the code to interact with the API. As the direct use of the web UI can be the basis of an API integration, this chapter will follow that route.

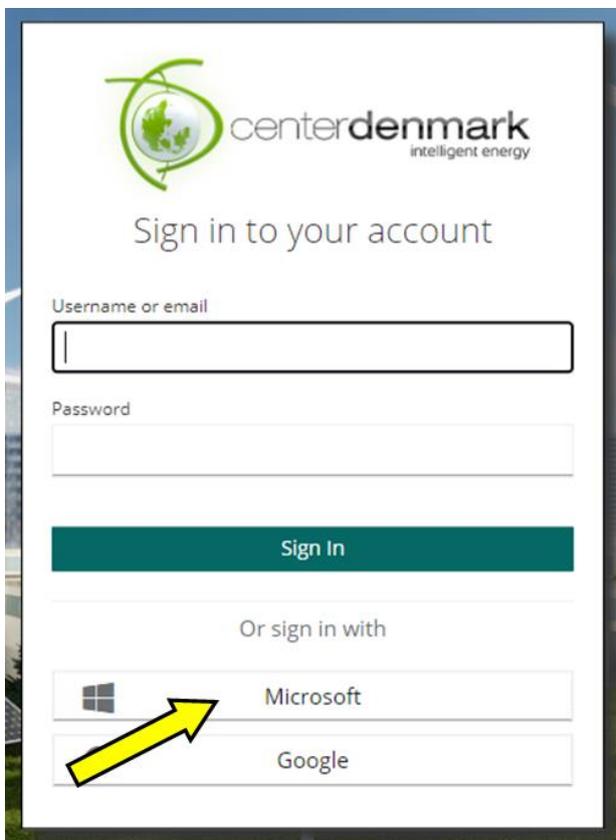
5.1 Create user

First, head to our online data portal here: <https://portal.centerdenmark.com/>

To create your user, there is no “sign up”. You just use “sign in”.



The most common way to create your user is, to sign in with a Microsoft account or a Google account.



Your user now exists, and we can grant you access to otherwise closed data sets.

Until then, you can explore our open data sets that are always available.

5.2 Explore documentation

The documentation can be found in the top bar menu “Docs”.



This wiki page will answer the most common questions.

5.3 Explore available data sets

To explore data sets, you can either use the search field on the front page or browse using the top bar menu “Datasets”.



Tip:

If you only have access to public data sets, for instance if it is your first login, it can be beneficial to deselect “private” and “visible” in the Dataset visibility section. This ensures that you will only see public data sets, which you can view, query and download.

The “private” box shows you datasets that you have access to, that are not public.

The “visible” box shows you data sets that you can ask for access to.



You can then click on a data set that you want to explore.

5.4 Explore a specific data set

Each data set has its own top bar.

The “Description” page describes the data set in common text.

The “Metadata” page describes the data set in a more programmatic way.

The “Field Metadata” page contains the data types and a description for each column type.

5.5 Query the data

The “Query” page allows you to set different filters for the data shown.

If nothing is done, the entire data set is shown.

By clicking the column title, you can sort the content by this. If you click it again, it will toggle between ascending and descending sorting.

Please note that if the data set has a time column, this is the only column the data can be sorted by.

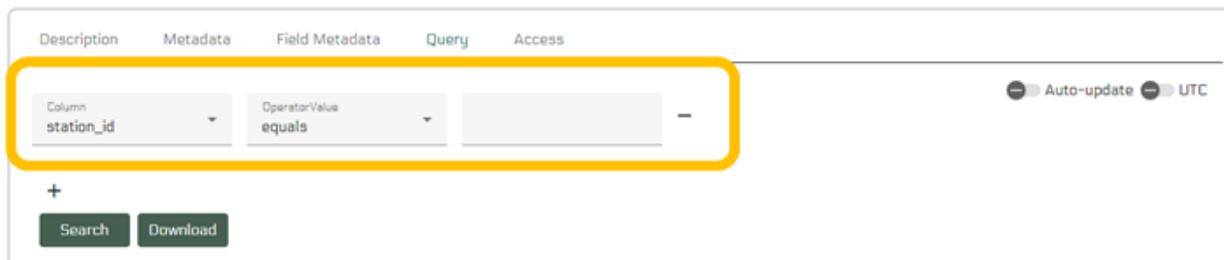
The “+” on the left, just below the data set top bar, lets you add a restriction.



You can then choose the column you want to filter by in the left field.

In the middle field the operator is chosen (equals, greater than, before (for time), and so on). The available operators are specific to the data type of that column.

In the right field you type the limiting value to complete the query.



After this, you can click search, to have the data shown updated to fit your query.

You can also click “+” again, to add another restriction.

If you want to remove one of your restrictions, you can click the “-“ to the right of it.

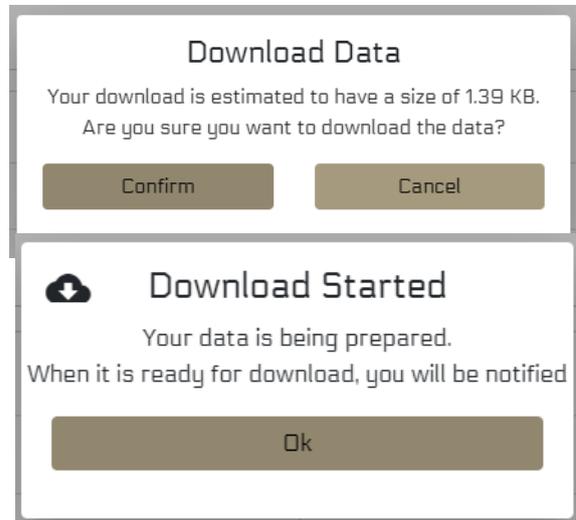
You can then click search again to update the data shown.

5.6 Download the data

If you prefer accessing the data through our API, jump to the next chapter.

Once you have restricted the data to fit your needs, you can retrieve the data in several ways. The simplest way is to click “download”, which starts a process of generating a .csv file for you.

A popup will ask you to confirm the download, while showing you an estimated size of the data set. You do not need to worry about clicking without knowing the size of the file.



When you click “confirm” you will see a new message, to tell you that your custom data set is being prepared.

Once the data set is ready, you will see a “ready” message resembling this image.



By clicking “View” on the right side of the message, you arrive in the “Downloads” page, where you can see your generated data sets.

Tip:

The “Ready” message disappears after about 5 seconds. If you missed it or want to find your data at a later time, you can find the “Downloads” page in the mini menu, placed to the far right in the top bar menu.



From the “Downloads” page, you can see a list of the data sets that you have generated.

To the right, the last two columns have the icons to either download or remove data sets.

Dataset ID	Dataset Name	Started At	Size	Status	Download	Remove
106cf0f7-b39d-4daf-a457-488956fa61b0	Mapping: Building Cluster/Weather - Sector-Coupled Clustered Data	Oct 31, 2024, 10:46:53 AM	3.52 KB	SUCCESS		
106cf0f7-b39d-4daf-a457-488956fa61b0	Mapping: Building Cluster/Weather - Sector-Coupled Clustered Data	Oct 31, 2024, 1:27:33 PM	1.12 KB	SUCCESS		
106cf0f7-b39d-4daf-a457-488956fa61b0	Mapping: Building Cluster/Weather - Sector-Coupled Clustered Data	Oct 31, 2024, 1:32:51 PM	1.12 KB	SUCCESS		

We kindly ask you to remove data sets, when you are done with them, to free up server space.

Also, data sets are automatically removed after 7 days.

5.7 Connect to API

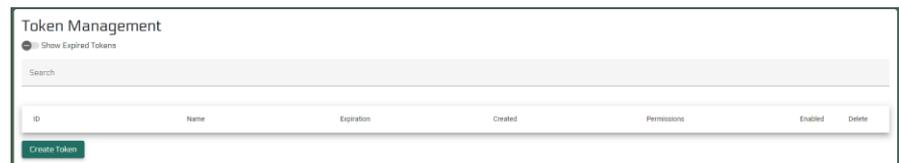
To access data through our API, you need to first create a token.

5.7.1 Create token

This is done via the “Tokens” button, which is found in the mini-menu in the far right part of the top bar menu.

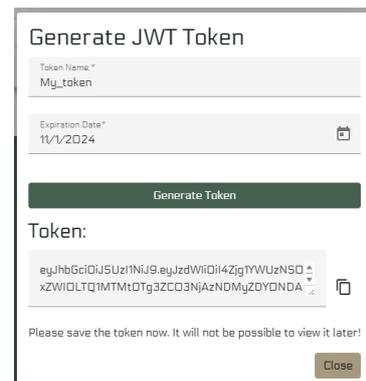
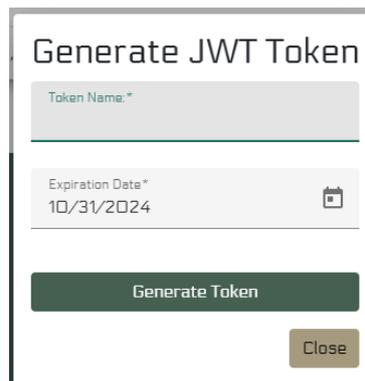


This brings you to the “Token Management” page.



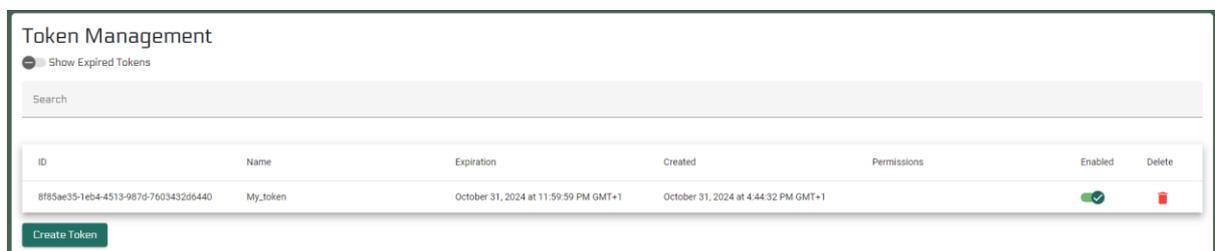
When you click “Create Token”, you can generate a JWT token, after setting a name and an expiration date.

As the dialog states, please save the token right away, as it will not be possible to see it again. If the token is lost, you will need to generate a new one.



Once you have copied and saved your token, you can close the dialog box.

Your new token should now be visible in the list on the token management page.

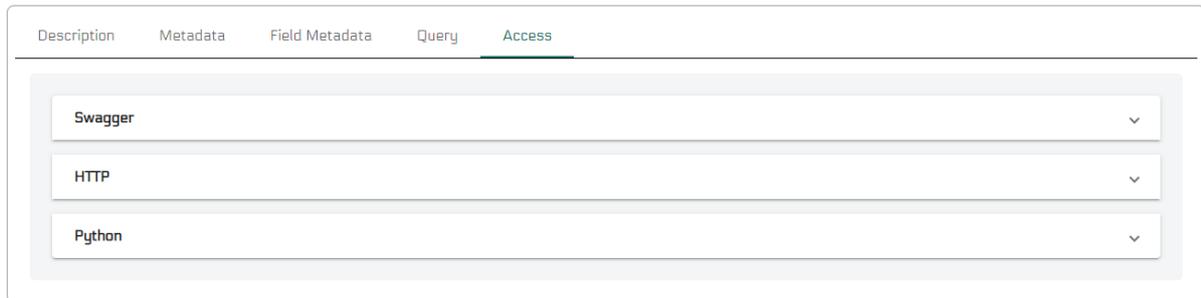


As is shown, this is also where you can disable and delete tokens.

5.7.2 Access data through the API

With your token ready, you can now define a data set, the same way that it was done in the “Query the data” section. Once you have defined your data set, instead of clicking download, in the top bar menu for the data set, select “Access”.

Here you can select one of three options, to download the data set you created in the query section.



Each section contains a step-by-step guide, which tells you how to download the data set you created in the query section.

