Template: Dataset Description

Before uploading data to the energydata.dk website, data owners must prepare specific information about their data. This information is crucial for ensuring smooth data usage by both users and data owners. Most of the information, unless otherwise specified, is visible to all users.

Stages of Adding Information:

1. Setting Up the Dataset
   a. Dataset name
   b. MQTT topic prefix
   c. Description
   d. Picture

2. Setting Up the Datastream
   a. Datastream name
   b. MQTT topic suffix
   c. Data type
   d. Datastream properties:
      - Comment,
      - Data license,
      - GDPR classification,
      - geo tag,
      - location,
      - organization,
      - project tag,
      - theme tag,
      - unit.

Setting up the dataset – General Data Information

Dataset Name

Name the dataset using the project name, company/lab name, and information type. The name should be intuitive for both the data owner, internal and external users.

Examples: Project name/Wind data; Master thesis/X data; Company name/Project name

MQTT Topic Prefix

Prefix, together with MQTT topic suffix, is used to identify a datastream. Topic prefix is mandatory if you want to connect to MQTT broker. Topic prefix is only visible to dataset owners. You can read more about its use in the API description.

The topic prefix is one string of alphanumerics

- Example: my-topic-prefix

Description

To ensure smooth data usage by both users and data owners, it is crucial to provide comprehensive information about the dataset. Include the following:

- General description of the data,
  (Describe the type of data, its source, etc.),
- Data granularity in the data set,
- Period covered by the dataset,
- Information about missing data or irregularities for specific periods,
- Usage Restrictions,
- Contact Person.

Adding a contact person is especially important when you plan to publish the dataset. Users who wish to use the dataset need to know who to contact in order to get permission to use the dataset.

The dataset description can be edited by data owners after datasets creation and shall be updated as soon as possible when the above information is available.

EXAMPLE DATASET DESCRIPTION

The dataset consists of synthetic data generated for demonstration purposes. It contains randomly generated records representing different data types commonly used in structured datasets.

Dataset Structure:

- *Alphanumeric Data: 2 independent datastreams containing randomly generated text strings.*
- *Integer Data: 3 independent datastreams with randomly generated numbers.*
- *Boolean Values: 1 datastream representing true/false values.*

The dataset includes 100 records spanning from April 1st, 2023, to April 5th, 2023. There are missing values between 14:00 and 18:00 on the 4th of April due to the server maintenance carried out at that time. Data is recorded at hourly intervals.

To use the dataset, the user needs to sign an NDA.

For further details regarding the dataset and the NDA, please contact: example@email.com

Picture

Adding a picture is an optional feature, but it makes it easier to navigate between different datasets. If you have many datasets, avoid using the same picture for all, as it can lead to confusion. The picture should be intuitive both for the data owner and users with access to the data.

Setting up the Datastream - Datastream Description

Datastream Name

While you can choose any name for your datastream, remember that others might use your data if you allow them. Therefore, the name needs to be intuitive and understandable for everyone.

MQTT Topic Suffix

Suffix, together with MQTT topic prefix, is used to identify a datastream. Topic suffix is visible only to dataset owner and users with "read" permission type to the dataset.

MQTT topic suffix can be several alphanumeric strings separated by "/".

- Examples: my/topic/suffix; currect/Lyngby/qwe; PriceLowEUR

Data Type

You can choose your datastream data type from the drop-down menu. The choices are:

- integer,
- double,
- string.

Properties

1. Comment

Any comment you might have for the datastream.

2. Data License

The data should have an assigned license indicating any restrictions on sharing, storing, or handling. If no license is specified, the default is "Do not share unless accepted by the data owner." According to DTU standards, data sharing should follow the principle of "as open as possible, as closed as necessary."

If your data is not licensed, but you wish it to be, you can take a look at the common license in the research spaces: About CC Licenses - Creative Commons.

Note that the license cannot be revoked and that you must own or control copyright in the work in order to add license.

3. GDPR Classification

When handling personal data, specify whether the dataset is freely available or requires special access permissions, such as an NDA. Clearly indicate that by writing "Irrelevant" or "NDA needed". Ensure all personal data handling complies with relevant legal and organizational policies.

4. Geo Tag

Geographic coordinates of the data collection place. The geo tag is used as a search parameter to help users find relevant datastreams.

5. Location

Name of the data collection location.

6. Organization

Your organization's name.

7. Project Tag

Tag the datastream with the project it is related to.

E.g. Smart Meter Data, Voltage Data, Electrical Measurements

8. Theme Tag

Tag the datastream with a theme to categorize it by its main topic or subject.

It should be clear and intuitive, as it is used as a search parameter to help users find relevant datastreams. By providing a well-defined topic or theme, you enhance the discoverability of your datastream, ensuring that users can efficiently find the data they need.

Examples: smart house; solar energy; switch status; CO2, indoor climate; District heating, building; building management data.

9. Unit

Unit of measurement for data in the datastream.

10. Customized

As a data owner you have an option to add other metadata you find relevant to the datastream.

Examples: Scaling; Precision; Transformer Number;