

Delivery no.: D2.7b
**System and data validation and warning tool
manual**

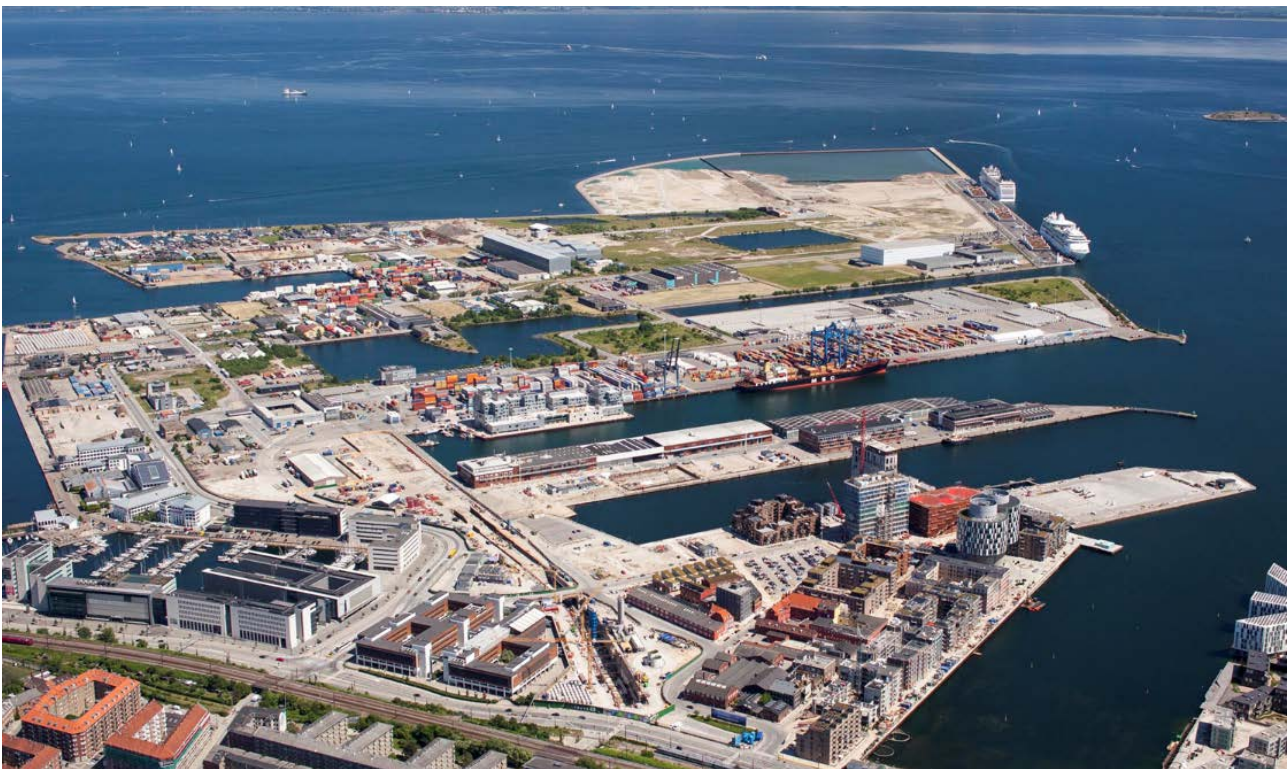


Photo: By & Havn / Ole Malling

DTU
Author, Anders Laage Kragh
Date 26-03-2019

Public deliverable

Confidential deliverable

Preface

EnergyLab Nordhavn – New Urban Energy Infrastructures is an exciting project which will continue until the year of 2019. The project will use Copenhagen's Nordhavn as a full-scale smart city energy lab, which main purpose is to do research and to develop and demonstrate future energy solutions of renewable energy.

The goal is to identify the most cost-effective smart energy system, which can contribute to the major climate challenges the world are facing.

Budget: The project has a total budget of DKK 143 m (€ 19 m), of this DKK84 m (€ 11 m) funded in two rounds by the Danish Energy Technology Development and Demonstration Programme (EUDP).

Forord

EnergyLab Nordhavn er et spændende projekt der løber til og med 2019. Projektet vil foregå i Københavns Nordhavn, og vil fungere som et fuldskala storbylaboratorium, der skal undersøge, udvikle og demonstrerer løsninger for fremtidens energisystem.

Målet er at finde fremtidens mest omkostningseffektive energisystem, der desuden kan bidrage til en løsning på de store klimaudfordringer verden står overfor nu og i fremtiden.

Budget: Projektets totale budget er DKK 143 mio. (EUR 19 mio.), hvoraf DKK 84 mio. (EUR 11 mio.) er blevet finansieret af Energiteknologisk Udviklings- og Demonstrationsprogram, EUDP.

Disclaimer

None

Project Information

Deliverable no.: D2.7b

Deliverable title: System and data validation and warning tool manual

WP title: Data and Measurements

Task Leader: Benny Stougaard Hansen

WP Leader: Benny Stougaard Hansen

Comment Period: 23 March 2019 to 17 April 2019

For further information on this specific deliverable, please contact:

Anders Laage Kragh, alkragh@elektro.dtu.dk

For other information regarding EnergyLab Nordhavn, please contact:

EnergyLab Nordhavn Secretariat

Center for Electric Power and Energy, DTU Electrical Engineering

Elektrovej

Building 325

DK-2800 Kgs. Lyngby

Denmark

E-mail eln@dtu.dk

Tlf. +45 45 25 35 54

www.energylabnordhavn.dk

Table of Contents

| | |
|--|-----------|
| 1. INTRODUCTION | 7 |
| 2. DATA VALIDATION | 7 |
| 2.1 Data from reliable and reproducible sources | 7 |
| 2.2 Data from non-reproducible but validated sources | 8 |
| 2.3 Data from non-reproducible and not validated sources | 8 |
| 2.4 Data validation | 8 |
| 3. SUPERVISION OF DATA CONNECTIONS | 10 |
| 3.1 The Meraki dashbord | 11 |
| 3.2 Monitoring in Energydata.dk | 13 |
| 3.3 Outstanding monitoring | 15 |
| 4. SYSTEM STABILITY AND FAULT TOLERANCE | 15 |
| 5. BACKUP | 17 |
| 5.1 Backup of metadata | 17 |
| 5.2 Backup of measurements | 17 |
| 6. CONCLUSION | 17 |

Resumé

In this document the data validation and supervision in Energydata.dk is presented. Further the Data Management System (DMS) vulnerability for hardware and software faults is described and the current data back-up procedures are explained.

In the conclusion recommendation for the coming development is given.

Version Control

| Version | Date | Author | Description of Changes |
|---------|------------|------------------------|------------------------|
| 1.0 | 18-03-2019 | Anders Laage Kragh | Initial version |
| 1.1 | 25-03-2019 | Benny Stougaard Hansen | First review |
| 1.2 | 26-03-2019 | Benny Stougaard Hansen | Final |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

Quality Assurance

| Author | Reviewer | Approver |
|--------------------|------------------------|-----------|
| Anders Laage Kragh | Benny Stougaard Hansen | WPL group |

| Status of deliverable | | |
|-----------------------|------------------------|-------------------|
| Action | By | Date/Initials |
| Sent for review | Anders Laage Kragh | 18-03-2019 |
| Reviewed | Benny Stougaard Hansen | 26-03-2019 |
| Verified | | |
| Approved | WPL group | 18-04-2019 |

1. Introduction

In this document the data validation and supervision in Energydata.dk is discussed. Further the Data Management System (DMS) vulnerability for hardware and software faults is described and finally the current data back-up procedures are explained.

Data is the measurements send to the Data Management System (DMS) from installed sensors in the Nordhavn living lab houses or entities as e.g. the battery in the parking garage. Further data can be from other data sources subscribed to by the project, e.g. weather forecast, power prices. The data definition does not include control signals relayed from control algorithms connected to the Data Management System (DMS) nor data export from the data warehouse, using either the web interface or the API. The supervision of the data flow is described and it is described how this can be done in two different monitoring systems. The challenges with data validation is discussed.

Further the architecture of the Data Management System (DMS) covering the servers and how the applications are deployed is described and discussed. The deployment of data acquisition equipment e.g. sensors and locally deployed aggregators and gateways is not in scope of this document.

The stability, fault tolerance, backup solution, calibration and MTBF for sensors and actuators is not discussed in this document.

2. Data validation

In the EnergyLab Nordhavn project data is coming from three types of sources:

- Data from reliable and reproducible sources
- Data from non-reproducible but validated sources
- Data from non-reproducible and not validated sources

In the following these three different kind of data is described:

2.1 Data from reliable and reproducible sources

These are data characterized by the fact that they are supplied from a source that it self performs validation and data can be retransmitted if data is lost for various reasons.

Examples of such data in the project are:

- Meter readings from HOFOR and Radius
- Power spot prices from Nordpool

For these data sources, validation of data is not necessary because the provider of the data does this validation before sharing the data with project. The data connection must be monitored and if it fails, an alarm must be raised. The data connection must be re-established and any lost or non-transmitted data must be re-transmitted and read in to the data warehouse.

2.2 Data from non-reproducible but validated sources

These data are characterized by being supplied from a source that it self performs validation but if the data connection is lost, data will be lost and will not be reproducible / reloaded. An example of such data in the project is:

- Data from the battery in the parking garage

For such data sources, validation of data is not necessary. The data connection must be monitored and when it fails, an alarm must be raised. The data connection must be re-established and it must be ensured that data is received.

2.3 Data from non-reproducible and not validated sources

This is data where the project itself has been responsible for the installation of the sensors. Therefore, there is no third party that also dependent on this data and therefore makes a validation before it is shared with project. Therefore, no validation of data and third party monitoring is performed. The data connection to the data source must be monitored and when it fails an alarm must be raised and the connection must be restored. Examples of such data are:

- Observations from the Smart Grid Unit
- KNX data from buildings

Received data must be validated and action taken if suspicious data is received.

2.4 Data validation

Currently there is no data validation implemented in the Data Management System (DMS) solution. Data is received and stored in the database as they are received. This is definitely not a wanted situation and data validation must be implemented in a not too far next release of the data warehouse. The remaining of this paragraph must therefore be read as input to the specification of the data validation function.

The data validation is far from simple to specify and implement as illustrated in Figure 1 and Figure 2

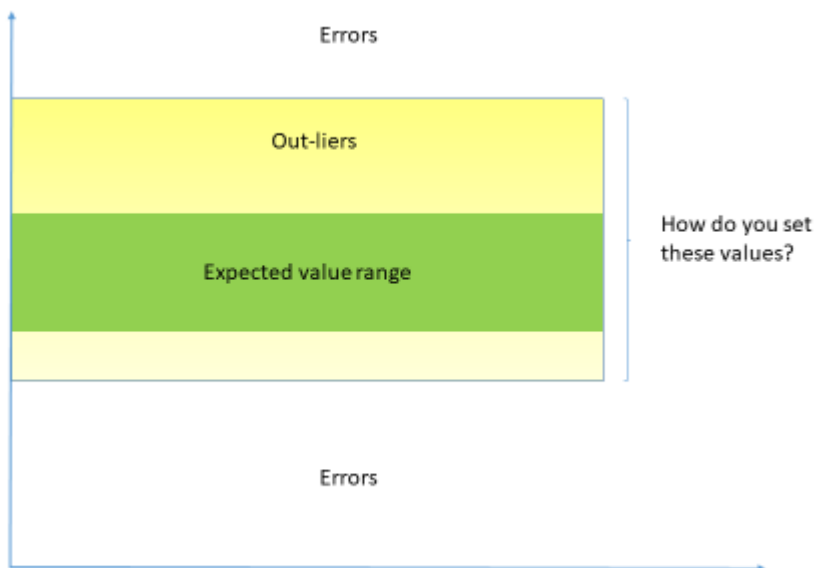


Figure 1 Normal or outlier

Figure 1 illustrate the challenge of validating a received measurement as either a value:

- Within expected value range
- As an outlier
- As a faulty value

The room temperature in an apartment can serve as an example. The expected value range could be from 5 – 40 C, 5 C during winter in case and no one's lives in the apartment and 40 C during a hot summer day. But what is the outlier range(s) and when becomes an outlier a faulty measurement indicating an error on the sensor?

To handle this type of validation the application receiving the measurements must have the above value range band implemented and the responsible for defining and setting up the experiment must define the above bands.

Further it shall be possible to define what should happen to values classified as outlier or faulty. Should they just be stored, should they be stored with a kind of note and when an alarm be raised in order to have someone looking into what could be an error on the measurement sensor?

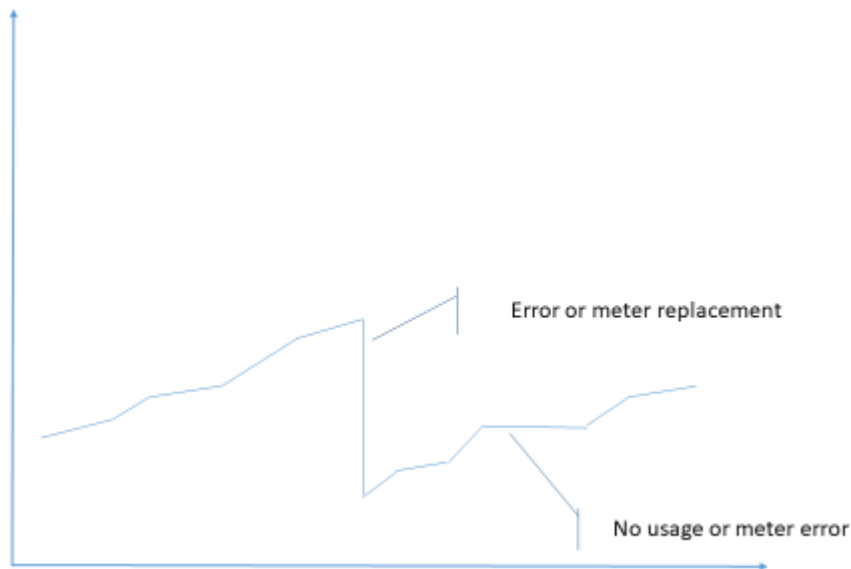


Figure 2 Error or expected values

Figure 2 illustrate another problem with data validation. The example could be a power meter in an apartment. The meter measures the total cumulative used power and is therefore expected to report increasing values over time.

But what if the reported values suddenly drop as shown in Figure 2. It could be caused by a simple regular meter replacement. To conclude this, more meter readings are needed – for one day? but when can you conclude and what should be done until a conclusion is taken? Of course the preferable solution is to be informed by the responsible for the meter installations, if no replacement has taken place, then it must be an error and action must be taken.

Next, what is causing the period with no use as illustrated by the horizontal line in Figure 2. No use or a fault on the device? For how long time can you accept the “no use” before you take action? Can other measurements either support the “no use” or error situation?

3. Supervision of data connections

The data connections can be monitored in two different systems:

- 1) The Meraki dashboard
- 2) The dashboard for the Energydata.dk

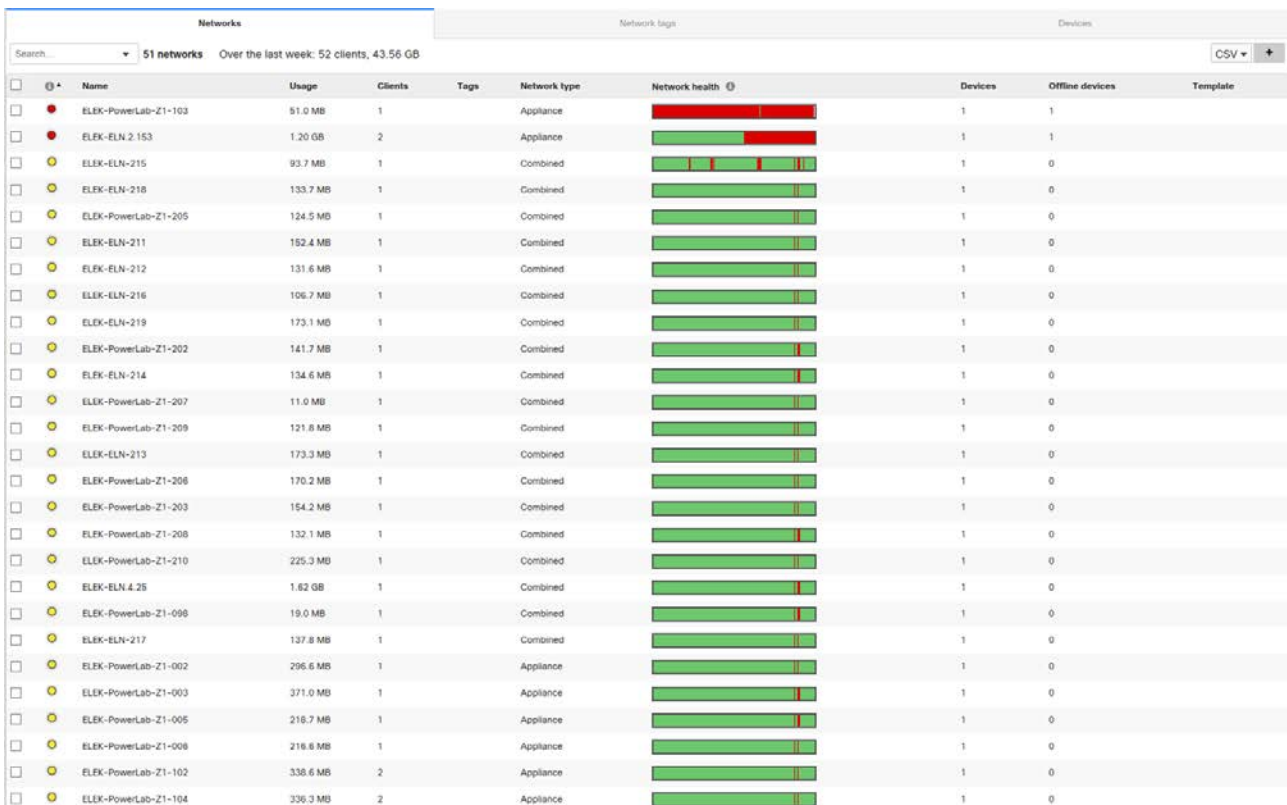
The Meraki system provides secure VPN connections between a central router located at Risø and connected to Energydata.dk and connected to the local deployed routers at premises where data is collected, e.g. in an apartment. This communication network can be

supervised by a dashboard application provided the Meraki system. This dashboard serves both as a monitoring tool and configuration tool.

The communication can alternatively be supervised from the Energydata.dk system. Here the communication is supervised indirectly by monitoring the incoming data. If no data is received this indicate a communication error.

3.1 The Meraki dashbord

The communication between the Data Management System (DMS) and the installations in Nordhavn is monitored by the Meraki Dashboard. In Figure 3 an overview is presented showing the availability the last week.



| Name | Usage | Clients | Tags | Network type | Network health | Devices | Offline devices |
|----------------------|----------|---------|------|--------------|----------------------------|---------|-----------------|
| ELEK-PowerLab-Z1-103 | 51.0 MB | 1 | | Appliance | Network health: 100% green | 1 | 1 |
| ELEK-ELN 2.153 | 1.20 GB | 2 | | Appliance | Network health: 100% green | 1 | 1 |
| ELEK-ELN-215 | 93.7 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-ELN-218 | 133.7 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-PowerLab-Z1-205 | 124.5 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-ELN-211 | 152.4 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-ELN-212 | 131.6 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-ELN-216 | 106.7 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-ELN-219 | 173.1 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-PowerLab-Z1-202 | 141.7 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-ELN-214 | 134.6 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-PowerLab-Z1-207 | 11.0 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-PowerLab-Z1-209 | 121.8 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-ELN-213 | 173.3 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-PowerLab-Z1-208 | 170.2 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-PowerLab-Z1-203 | 154.2 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-PowerLab-Z1-208 | 132.1 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-PowerLab-Z1-210 | 225.3 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-ELN 4.25 | 1.62 GB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-PowerLab-Z1-098 | 19.0 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-ELN-217 | 137.8 MB | 1 | | Combined | Network health: 100% green | 1 | 0 |
| ELEK-PowerLab-Z1-002 | 296.6 MB | 1 | | Appliance | Network health: 100% green | 1 | 0 |
| ELEK-PowerLab-Z1-003 | 371.0 MB | 1 | | Appliance | Network health: 100% green | 1 | 0 |
| ELEK-PowerLab-Z1-005 | 218.7 MB | 1 | | Appliance | Network health: 100% green | 1 | 0 |
| ELEK-PowerLab-Z1-006 | 218.8 MB | 1 | | Appliance | Network health: 100% green | 1 | 0 |
| ELEK-PowerLab-Z1-102 | 338.6 MB | 2 | | Appliance | Network health: 100% green | 1 | 0 |
| ELEK-PowerLab-Z1-104 | 336.3 MB | 2 | | Appliance | Network health: 100% green | 1 | 0 |

Figure 3 Meraki Dashboard, overview

As it can be seen in Figure 3 most networks are available but all have an incident at the same time, as seen as the tiny red line. This is due to an incident at the central router which can be seen Figure 4.



Figure 4 Status for central Meraki router

As shown in Figure 3 device ELN-215 has some “hick ups” which is seen as the red blocks. A more detailed look into ELN-215 is presented in Figure 5.



Figure 5 ELN-215 availability for 1 week

A more detailed analysis is needed if the root cause for the un-availability is required as it could be caused by many issues, e.g. the cellular network.

In Figure 6 the live traffic for ELN-215 is shown.

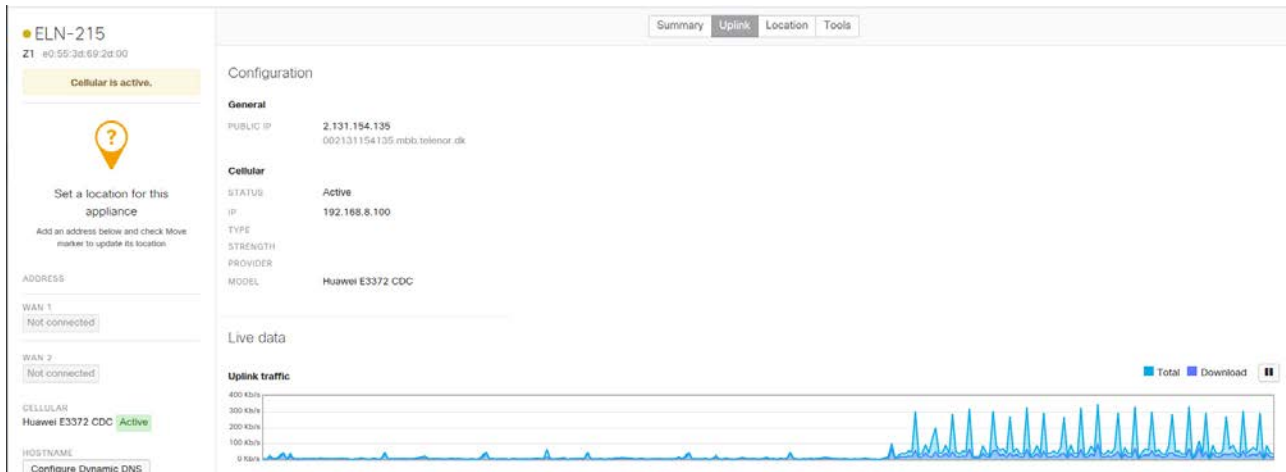


Figure 6 Live traffic from ELN-215

By using the Meraki Dashboard the network status can be supervised and further it can be used for trouble shooting.

The Meraki dashboard supports sending alarms if a data connection to the Meraki router breaks or if connected equipment (to the Meraki router) disconnect. Enabling this requires that an Operation and Maintenance team is appointed in order to receive the alarms and to act upon the received alarms.

3.2 Monitoring in Energydata.dk

The dataflow can also be monitored from the Prometheus monitoring and alerting system used in Energydata.dk. This solution is described in D2.7a: “MTBF or similar system stability analysis report”.

The defined alarms are listed in Table 1. The alarms are for internal use for the operation and maintenance of the Data Management System (DMS) and therefore the alarm names and slogans requires knowledge to the Data Management System (DMS) architecture.

| Alarm | Description |
|------------------|---|
| CisDown | Interface is down |
| CisError | Not all mqtt messages were published |
| CleanChargeDown | Interface is down |
| CleanChargeError | <ul style="list-style-type: none"> • FTP server might be down • The file does not exist • Cannot connect to MQTT |
| HeatboostDown | Interface is down |

| | |
|----------------------------------|--|
| HeatboostError | <ul style="list-style-type: none"> • Error in publishing MQTT messages • Exception thrown while processing file, topic does not exist or error in file format |
| HoforDown | Interface is down |
| HoforError | <ul style="list-style-type: none"> • Topic is wrong • Failed publishing some messages |
| KafkaMqttNoMessagesIn | Kafka is not receiving any messages on the "mqtt" topic |
| KafkaRawNoMessagesIn | Kafka is not receiving any messages on the "raw" topic |
| KennyXMissingDeviceTelegrams | Kennyx is not receiving any telegrams from the KNX gateway given by the label "knx_gateway" |
| MeteoblueError | <ul style="list-style-type: none"> • Unable to connect to Meteoblue • Error when publishing messages |
| MeteoblueNoData25Hours | No data has been published in over 25 hours |
| mSCADADischargeError | <ul style="list-style-type: none"> • Error when fetching discharge schedule (response was not 200, fetching timed out) • Error when fetching activation value (response was not 200, fetching timed out, response unparseable) |
| mSCADADischargeNoData | No data has been sent |
| NordpoolerError | Error when publishing messages |
| RadiusError | <ul style="list-style-type: none"> • Error when publishing messages • Error while processing file |
| RadiusMissingMetadata | Radius is missing metadata |
| SamlyCsvMessages | Samly has not written any telegrams to CSV files from the KNX gateway given by the label "knx_gateway" |
| SamlyMqttMessages | Samly has not sent any telegrams to the MQTT broker from the KNX gateway given by the label "knx_gateway" |
| SamlyRedisMessages | Samly is not receiving any messages from Redis which correspond to the KNX gateway given by the label "knx_gateway" |
| ThiimSGUInterface_x_Down5Minutes | Thiim SGU interface has been down for more than 5 minutes |
| TomorrowDown | Interface is down |

| | |
|---------------------------|--|
| TomorrowError | <ul style="list-style-type: none"> • Error in dataset (might not contain topics 'carbon-density' or 'power-break-down') • Error when processing response for dataset • Response returned status which was not 200 |
| VerneMQNoMessagesReceived | VerneMQ is not receiving any messages |
| VerneMQNoMessagesSent | VerneMQ is not sending any messages |

Table 1 Defined alarms in Prometheus

Having the Data Management System (DMS) architecture in mind it can be seen from Table 1 that all interfaces are supervised and alarms are raised if data is missing or if data cannot be parsed (interpreted) and inserted to the database.

3.3 Outstanding monitoring

The monitoring described in the previous supervises the connections and raise alarms if no data is received, but the granularity of supervision is not detailed enough to detect if certain data is missing. As an example, it is supervised that KNX telegrams are received from a gateway, but it cannot be detected if certain specific telegram is missing, e.g. if telegram from a presence sensor is not received.

Therefore the monitoring must be refined to handle that data becomes available differently. Either data is available according to a schema, e.g. data can be fetch from a FTP server according to a plan, e.g., once a day at a certain time, the data is available due to sampling at a certain frequency, e.g. every minute or other know regularity. Supervision of these types of data provisioning is relative simple. It is known when data shall be available and if it is outstanding, an alarm shall be set.

Other measurements (data) is only available when there is a change in the system being monitored, i.e. data is event based; e.g. when there is a change in presence in a room, a door is being opened, the power frequency change.

The event based data is challenging to monitor as it is unknown when there shall be an observation and therefore when to set an alarm. These type of measurement should be combined with a sampling at a low frequency so that data is becomes available at changes and at least e.g. once a day. The regular measurement in this set-up can be seen as a kind of heart beat from the device, so it is known that it is alive.

There is no supervision of pure event based data connection in Energydata.dk and therefore such set-up should be avoided if possible.

4. System stability and fault tolerance

The architecture of the core Data Management System (DMS) is presented in Figure 7. As it can be seen from Figure 7 the Data Management System (DMS) is deployed on 8

servers. All software applications are replicated on two or more servers except for the VerneMQ and other related MQTT applications. This means, that the Data Management System (DMS) is tolerant for a fault on any of the servers and applications expect for the server holding the VerneMQ and related MQTT applications. The reason for not making the MQTT applications redundant is to secure a consistent update of the database. If these applications were shared between two or more servers, then the database update could become inconsistent if one of thread was down.



Figure 7 Core servers of energydata.dk

The 8 servers housing the Data Management System (DMS) are all deployed at the same physical location at Risø. This makes the Data Management System (DMS) vulnerable to incident at this location, e.g. fire. However it shall be noted, that the servers are hosted in a server room providing:

- Redundant power supply from two independent power sources
- Battery backup
- Redundant networks connections
- Supervision of access to the server room
- Fire supervision and firefighting equipment

As it is planned, that the Data Management System (DMS) shall be used after the EnergyLab Nordhavn project is terminated, providing access to the gather data and expanded to hold more and other data, a redundant deployment solution is recommended.

5. Backup

The Data Management System (DMS) consists of two database systems working together and each holding different types of data.

Metadata is stored in a SQL database using a PostgreSQL application and measurements are stored in a time series database based on Cassandra. The deployment on the servers is shown in Figure 7.

5.1 Backup of metadata

A complete backup of the metadata in the PostgreSQL database is performed every day in the morning (04.10 AM). A data dump of the PostgreSQL is generated and this file is backup'ed by the AIT Tivoli backup system. Each backup file is a few MB and is kept in the AIT backup system for 1 year. The backup process is monitored by AIT.

5.2 Backup of measurements

Measurement is stored in the Cassandra database. The Cassandra database has a simple replication security where data is distributed on a number of servers, in the Energydata.dk Data Management System (DMS) implementation on three servers. In the Cassandra cluster, data is replicated among these servers so that one server can fail without data loss. However, when a single server fails, the performance of the system will be reduced. This implementation and architecture therefore imply a backup of measurement data as the system is protected against a single server failure. However the deployment on one geographically location makes the system vulnerable for incident on this site.

A backup solution for the measurements are therefore required. This could be provided by either:

- 1) A backup solution based on external backup to a AIT provided backup store
- 2) Expanding the Cassandra cluster to two different geographical location
- 3) Create a backup solution as part of a long term storage for the measurement data

The backup solution for Cassandra (measurements) must be decided and implemented.

6. Conclusion

The Data Management System (DMS) has been in operation for almost one year. During this year the supervision of the data connection has performed acceptable; alarms have been raised when measurements have not been received.

The back-up solution is acceptable but not perfect. Meta data is secured by a AIT provided backup system but the measurements are only protected by a replication at the same physical location. This imply that data is at risk in case of a disaster at the datacentre. This must be mitigated by a backup solution that secure data at a geographical separate location.

The data validation of received data is not implemented and this must be implemented for already existing data sources and be integrated into coming new data connections. This must be given high priority because the errors here are first detected when someone start to analyse the data and errors can therefore exist for some time before they are detected and corrected causing data for a long period being faulty.