# Final and complete LEUVENAIR dataset

The final Leuvenair dataset contains all Leuvenair measurements from 2018-01-30 till 2020-02-02 (included), 29 077 573 records in total.

| Timeframe | Download link | Size | Integrity (md5sum) |
|---|---|---|---|
| **Complete datadump 2018 (from 2018-01-30)** | LEUVENAIRfulldump2018.csv.gz | 208 Mb | 327888b614605dff16de2fcde0bc8c0a |
| **Complete datadump 2019** | LEUVENAIRfulldump2019.csv.gz | 198 Mb | 2746ddfd120f7d13eae04dfc038f6cc8 |
| **Complete datadump 2020 (till 2020-02-02)** | LEUVENAIRfulldump2020.csv.gz | 16 Mb | b2649bd235a181188cfc800b15b1f395 |
| **5-Minute median of all active sensors** | LEUVENAIRfulldumpMedian.csv.gz | 1.5 Mb | b5cf5635b99f086c30af6b1afc761b2b |

# Some guidelines for the analysis

(will be updated when new questions come in).

- The PM2.5 measurements of the SDS011 sensor are much more reliable than the PM10 measurements. For long-term analyses, stick to the PM2.5 measurements. PM10 measurement can be used for peak detection only.
- The sensor will overestimate the PM concentration in humid conditions (relative humidty RH > ~75%) and will understimate the PM concentrations in dry conditions (RH < ~50%).
- The measurements of the DHT sensor (temperature + humidty) are not so realiable. Try to use an external source for temperature and relative humidity.
- We saw some cases where the SDS011 sensor freaked out for several hours reporting very high values, and then gradually returned to "normal" values. We believe this effect could be caused by some dirt inside the sensor.
- There is some intrinsic sensor-to-sensor variability present in the sensors, which might even change over time. We have three sensors on exact the same location (8765, 13526 and 13528) that can be used to quantify this effect.
- Don't look for 1-to-1 relations with traffic: you won't find one! Traffic is not the main source of (primary) fine dust; woodburning is! Use the $NO_2$ data from Curieuzeneuzen (also in metadata table below) if you want to study traffic relations.
- For reference values, use the RIO-model of IRCEL. More information here and here.
- Take a look at the available literature on the SDS011 sensor in this open dir before you reinvent the wheel.
- Take a look at the available analysis scripts provided by Influencair on Github.
- All timestamps are in UTC.
- For a crash course on air pollution, take a look at the introductory presentation of Frans Fierens (VMM/IRCEL).
- For some background on the Leuvenair project, take a look at this powerpoint presentation.

# Metadata sensors

Link to a tabular overview of the LEUVENAIR sensors or downloadable as a JSON file.

- **EXPORT** EXPORT = 0 means that the sensor was of too low quality to be integrated in the final datadump files.
- **SENSOR POSITION** (qualitative index given by participant) 1 = on the garden side, very well shielded from all streets, 10 = the sensor is on a house wall directly on the street. With this value it is irrelevant how big the street is, it is only about where the sensor is attached to the house.

- **INDUSTRY** (qualitative index given by participant) 10 = a lot of polluting industry in the immediate neighbourhood of the sensor, 1 = very rural area without industrial PM sources.
- **WOODSTOVES** (qualitative index given by participant) 10 = a lot of woodburning in the immediate neighbourhood of the sensor, 1 = no woodburning in the area.
- **TRAFFIC** (qualitative index given by participant) 10 = sensor close to very busy roads, 1 = very quiet area in terms of traffic.
- **SVL** 1 = StratenVolLeuven sponsored sensor
- **NO2_CN** is the $NO_2$ concentration (yearly mean) at the same location as the PM sensor, taken from the Curieuzeneuzen experiment.

# Licence



Author: Maarten Reyniers. Comments, questions, remarks via the Leuvenair contact form.