From particle to screen

How bioaerosol data gets to the users

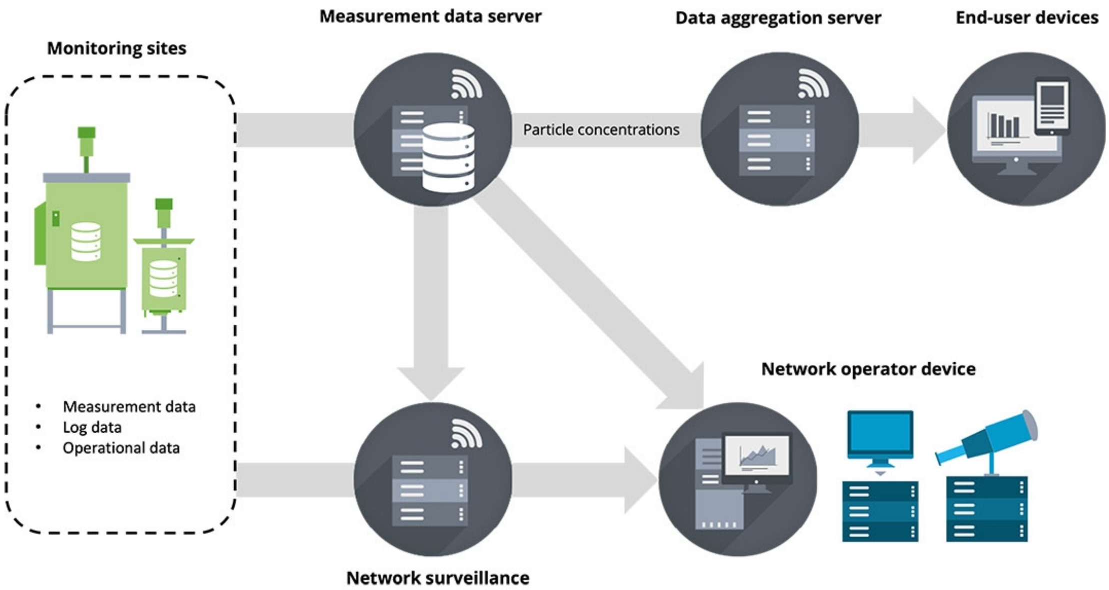


Table of Contents

Introduction

This document describes how we convert single bioaerosol particles to data which is available on our screen. In general we try to answer one main question: How does real-time bioaerosol data gets to the users? Be it as a network operator at his working desk or as an end-user on the smartphone. Starting with a short description on how SwisensPoleno converts aerosol particles into data, you will also find detailed descriptions on how the data can be processed to finally reach its destination. All within the components and services Swisens has to offer for real-time bioaerosol monitoring.

1. Requirements

An appropriate installation site for the SwisensPoleno must be available:

- A location that will provide suitable input to dispersion models of the measured particles/pollen.
- A reliable fixture to securely mount the system.
- AC power

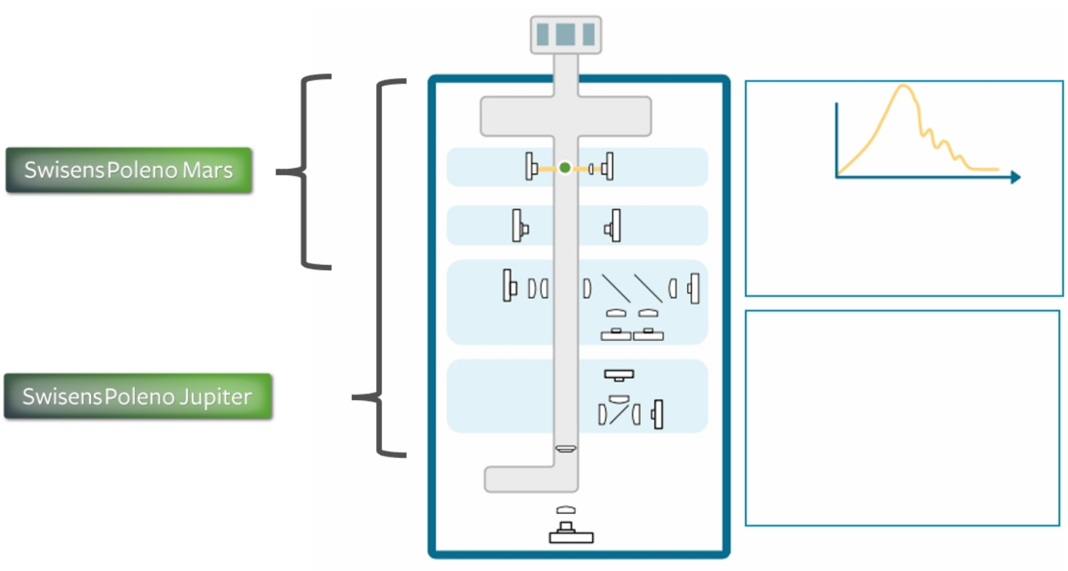
The instrument requires a network/internet connection:

- 10/100/1000MBit/s Ethernet, or
- a SIM card with an adequate data plan for the integrated LTE router (mobile network), assuming that there is adequate coverage

2. How data from bioaerosol particles is generated

2.1. Sampling

A particle is picked up by the inlet airflow of the SwisensPoleno and enters the measurement chamber. The particle passes the trigger light source, causing scattered light, which is then picked up by the photodetector. The detector waveform is acquired and stored in the memory of the control electronics. The controller detects one (SwisensPoleno Mars) or two (SwisensPoleno Jupiter) peaks in the signal, which qualifies as a valid particle. A new measurement event is created.

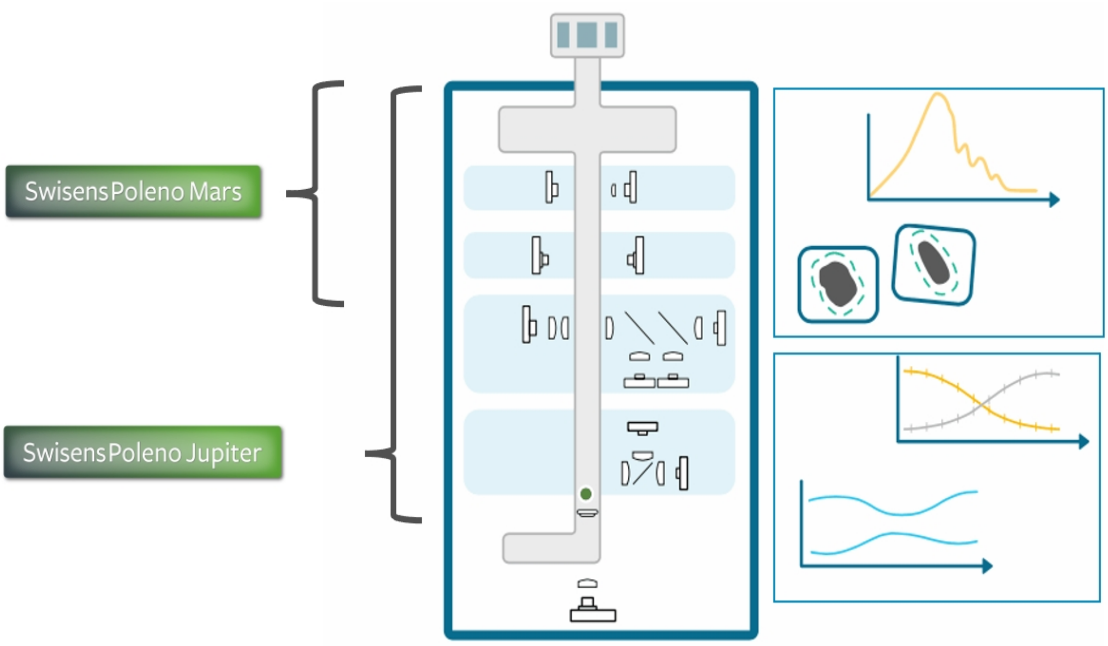


2.2. Measurement

The particle passes the measurement systems, causing more data to be acquired:

- Two holograms
- Spectrally resolved fluorescence (SwisensPoleno Jupiter only)
- Optical polarization (SwisensPoleno Jupiter only)

For each measurement event, approx. 10MB of raw data is acquired, which is dominated by the two holograms (images).



3. How the data is processed

The raw data is processed by the instrument software, thereby extracting the relevant information:

- Basic information like particle position and velocity is calculated.
- The holograms are normalized and reconstructed into shadow images and the relevant part is cut out of the full image.
- Fluorescence measurement vectors are turned into spectra and lifetime information.
- The machine learning model is used to classify the particle as one of the known and trained classes.

- Metadata like an unique identifier, the originating instrument and relevant operational data is added.

4. Variations on how data can be stored

4.1. Local file storage

If enabled, the processed measurement event is stored to the file system on the instrument's solid-state disk as a set of files:

- Most information as a JSON-formatted text file
- Images as PNG to compress the data without losing information
- Data vectors/tables as NumPy files to avoid overloading the JSON too much and providing efficient export and import functionality

What information to save is highly configurable and so is the amount of storage required per event. However, experience has shown that 250kB to 400kB per measurement event is a good estimate for most cases.

4.2. Manual Backups with harddisk

If desired and available, event files are regularly moved from the internal SSD to a large external harddisk for backup purposes.

4.3. File transfer by scp, rsync etc.

If desired, measured event files can be transferred from the instrument to another location using networking protocols like scp, rsync, etc.

4.4. Local data based on SwisensPoleno

If enabled, the processed data is written into the SwisensPoleno's internal MySQL database. The information is the same as when writing it out into files, but the format is optimized for database storage.

In addition to the data in the database's tables, a binary log (binlog) is also written, which is used to replicate the data onto an external data server.

Besides the events, the database also contains operational data, which provides information about the instrument's health and performance.

If desired, old and replicated events and binlog data can be regularly cleaned from the internal database to free up space on the SSD.

4.5. Data replication onto external data base

The data in the instrument's database (or, more exactly, in its binlog) is replicated to the database on an external server via the network connection. This external server can be Swisens infrastructure if the full service is desired or it can be integrated into the customer's infrastructure for on-premise hosting.

5. How the data can be accessed as a network operator

5.1. Access local data on SwisensPoleno

The data currently in the instrument's internal database can be accessed via the SwisensDataExplorer installed on the instrument. This software also provides access to the instrument's operational data and events using a public REST-API on the basis of simple HTTP requests.

5.2. Access replicated data from SwisensPoleno

The replicated data can be accessed via the SwisensDataExplorer installed on the external server. This provides extensive tools to check the acquired events, export time series, create data sets for analysis or machine-learning trainings, plot operational data, and much more. In addition to the easy to use web interface, the SwisensDataExplorer also implements a REST-API, which makes the data easily accessible for external tools.

For advanced users:
The SwisensDataAnalyzer can directly connect to the external data server and fetch events from there to perform extensive analyses, up to train new machine-learning models. It also allows existing data to be reclassified using new models for performance comparisons.

6. Surveillance tools for proper operation

Surveillance is an important tool to ensure that a SwisensPoleno is operating properly and to detect abnormal conditions before they become a problem. This surveillance can be done by Swisens even if the data is hosted on-premise, or it can be done by the customer. The recommended surveillance tool is Zabbix, but since only HTTP requests are required, other tools may be used as well. Ideally, the surveillance software can access the DataExplorer on both the SwisensPoleno, as well as on the external data server, to also check the data replication progress.

7. Final processing for end-user applications

From now on, the measured particle data can be aggregated with other environmental data, e.g. from weather models, to derive particle trajectories and concentrations in the vicinity of the system. From there, the acquired and aggregated data is post-processed and made available to end users, e.g. to provide pollen forecasts to allergic people via a web service or app.