Urban flows observatory, Sheffield
Data storage and sharing strategies

Author: Patricio F. Ortiz

Date: February, 2020

This document describes the way in which the Urban Flows Observatory in Sheffield stores and shares its data collection.

There are currently four broad categories of quantities which we are measuring/collecting data, and up until this point in time, all sensors have been deployed **outdoors**

- Meteorological data: Air temperature, Bulb-temperature, Relative Humidity, Air Pressure, Wind speed, Wind direction, amount of precipitation, type of precipitation
- Air Quality data:
  - Particulate material: PM1, PM2.5, PM4, PM10, Total-Suspended-Particles, Particle-Density
  - Gaseous pollutants: NO, NO2, NOX, O3, SO2, CO
- Environmental/physical: Noise, CO2, solar radiation level.
- Traffic: flow and occupancy

Internal storage

We designed the system considering three kinds of assets which need to be linked to the collected detector data

- **Sites**: are the physical places where we install instrumentation. A site is mainly characterised by its geolocation: longitude, latitude and altitude (height above sea level). We keep information about its street location and postal code (when available). In addition, we keep track of when a site has entered service and when it has been decommissioned. Each site has a unique ID
- **Sensors** or pods are the actual instrument packages containing the different detectors. Sensors can be deployed anywhere, and they can be redeployed as well, hence, they do not attach any geolocation to them. The information about the sensors include the quantities they measure, the units in which they are measured and a metadata identifier, unique to each quantity measured (we call it the UCD o unified content descriptor). The manufacturer, the sensor's serial number, and the date it was deployed (and decommissioned) are recorded.
- **Pairs** are entities created when a **sensor** (pod) is deployed at a **site**. Pairs get created when a sensor (pod) is deployed on a site, and disappear when the pod is taken down for maintenance or to be deployed elsewhere. Data requests examine this information to link sensor data with a geo-location.

All this information varies very slowly and it is designed to be changed by a person, as opposed to the sensor data acquisition which is part of an automated pipeline.

We group sensors by what we call families to represent instruments acquired from a single provider or which share similar characteristics.

Static sensor data is stored in tables for each family, and the only thing they store is a time-stamp (in seconds since 1970.0) and the detector data, in other words, a table representing time stamped observation. No geolocation is added to these tables to save space.

Only the instruments mounted on Mobius, our mobile platform will contain geolocation associated with each individual observation, as Mobius is not a static asset.

We chose to use a time-oriented database system in order to optimise access time and responsiveness to queries. In that sense, asking data for a few hours or days involves accessing a few storage buckets and not the whole dataset.

Three kinds of buckets are used: daily, monthly and yearly. Daily buckets are updated every time we harvest data, monthly and yearly buckets are updated once a day.

This storage strategy ensures scalability and quick response time.

Data export

This field is still open to the needs of our “clients”.

So far, we provide data on a “per pair” basis, that is, we provide tables of sensor data for each individual pair active during the requested period.

Each table includes geolocation data associated to the site where the sensor is deployed

Although only time in “seconds since 1970.0” is stored, queries and output can involve other quantities related to time: year, month, day, day of the week, hours+minutes, minutes since midnight, etc. These quantities are computed on the fly.

In terms of data format we offer a few options: CSV, NetCDF and geo-json for the data, and the metadata (location of sites, sensor description, pairs details) can be recovered separately.

The observatory is in the position to accommodate the development of other forms of data distribution suitable for partners like DAFNI.

Importing data from the observatory can be done automatically, as described in the API document.