

The logo for Air Quality Data Commons is displayed on a dark blue rectangular background. The text "air quality" is in a smaller, lowercase sans-serif font, and "data commons" is in a larger, bold, lowercase sans-serif font. To the right of the text is a stylized graphic of a map of the United States, composed of a grid of white dots of varying density, representing data points.

air quality data commons



Contents

Quick Start	2
Sign Up and Log In	3
Create a Chart	4
Create a Dashboard	5
SQL Lab	6
Tutorial	8
Appendix- Standardized Terms	10

Quick Start

The Air Quality Data Commons (AQDC) is a cloud-native software application designed to collect, standardize, transform and integrate air quality data recorded by the low- and medium-cost air quality sensors around the globe. The AQDC is an effort to democratize air quality data and increase transparency, by providing infrastructure: to collect, store and perform big data computation; to promote data products that allow an understanding of sensor responses; and to encourage the community to develop data-driven solutions for cleaner air. We encourage anyone seeking to analyze air quality data to use the AQDC application and contribute their work back to the platform so that its value increases over time. This guide explains some of the ways in which users can use the AQDC in order to grow as a community working to understand and improve the air quality surrounding us.

Click [here](#) for a list of standardized terms defined by the Air Sensor Workgroup, which can help you navigate the AQDC more easily.


1. From the main page, you can explore existing dashboards directly on the map or by expanding the map by clicking . When you find a dashboard that interests you, click on the orange box to find more information. Click on the “View Dataset’s Default Dashboard” button to find the data. If necessary, review Terms of Use and click 

2. You can also explore data by clicking on the “Explore” tab at the top of the page.



Here you can view published datasets and dashboards. You can filter the data by status, mobile or stationary monitoring, measurand, organization, sensor manufacturer, model and cost.

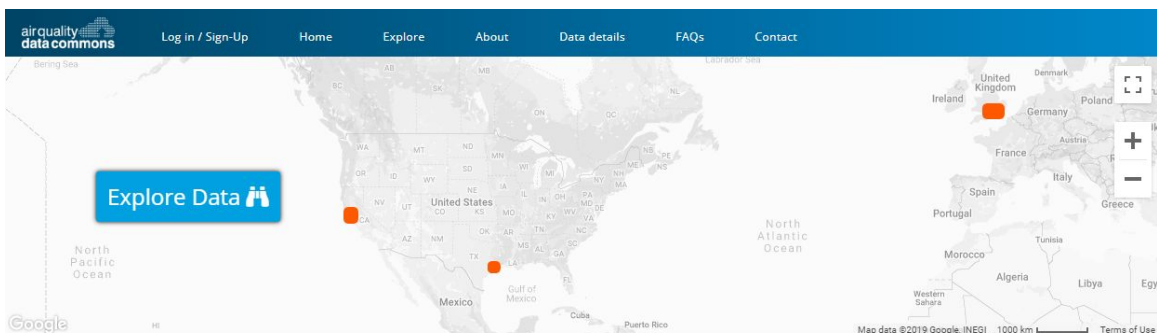
3. Users can download a CSV file of the data directly from a dashboard.

Note: This function allows you to obtain only a portion of the selected dataset. Click  in the top right of a given chart and select “Export CSV”.

Sign Up and Log In

While data may be viewed and downloaded without logging in, creating a free account unlocks more data analysis capabilities including the ability to create and save custom charts and maps and run queries in SQL Lab.

1. Access the AQDC website at aqdatacommons.org (not available using Internet Explorer).



Welcome

The AQDC seeks to accelerate solutions to air pollution by standardizing and sharing air quality data.

The AQDC is an open-access, open-source data platform that allows people to share and use FAIR (Findable, Accessible, Interoperable, Reusable) data from low- and medium-cost air quality sensors while maintaining necessary data privacy and security. In addition to centralized storage, the AQDC provides the infrastructure to analyze and visualize neighborhood-level air quality data.

2. Click “Log in/Sign-Up”.
3. For new users, click on “Sign Up (with Google Account)”.
 - a. You do not need to create a new user account for the AQDC; you can use your Google account credentials.
 - b. The AQDC does not store your password; we use Google’s authentication and authorization system.
 - c. We have simplified sign up and login with a single interface for both actions. The platform determines whether you are signing up for the first time or logging in subsequently, and responds accordingly.
 - d. Note that we currently only support Google account based authentication. By doing this, we can ensure a high level of security for the application, data and our users.
 - e. You can use a Gmail account/G Suite for Business account or you can create a new Google account linked to your enterprise email or other existing email. Follow this link to create a [Google account without Gmail](#).
4. After you authenticate successfully, you will arrive at “**Your account**” page where you can see your details and make edits as needed.
5. You are now ready to explore the AQDC!

Create a Chart

A chart is an individual report, created with AQDC data. If you click on charts under the Superset landing page, you will see a list of published charts that you can access.

1. To create a chart, click on Account > Superset, and then select the “Charts” tab.



2. Click the blue  in the top right corner to add a new record.

3. Next, select a data source. Data tables are named by the dataset, followed by meas (measurand), snsr (sensor), snsr_node (node), and snsr_sys (sensor system). Finish by clicking “Create new chart”.

- a. Most users will select “meas” (measurands) to see the testing values. Select from the other tables for metadata on a specific sensor or node.

Create a new chart

Choose a datasource

Choose a datasource

- aqdc.bay_area_mobile_analysis_meas
- aqdc.bay_area_mobile_analysis_snsr
- aqdc.bay_area_mobile_analysis_snsr_node
- aqdc.bay_area_mobile_analysis_snsr_sys

4. Your datasource is selected. Here you can select your visualization type.
5. You can also change the settings for time, most users will change to “No filter”.
 - a. For a scatterplot visualization, make sure to click on “Longitude and Latitude” and, using the dropdowns, match with longitude and latitude, respectively.
 - b. For the bar chart visualization, select your options for time, metrics and series.

6. Make any other selections and click 

7. Observe your query output and make any necessary edits using the “Point Size” and “Point Color” functions.

8. Edit your title and click  Save

Here you can:

- a. Save your chart
- b. Add your chart to an existing dashboard
- c. Add your chart to a new dashboard

Save A Chart

Overwrite chart Map Houston Test

Save as

Do not add to a dashboard

Add chart to existing dashboard

Select Dashboard



Add to new dashboard

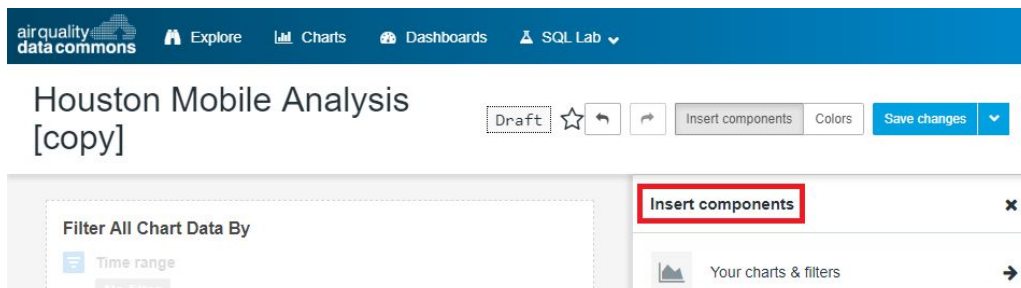
Save

Save & go to dashboard

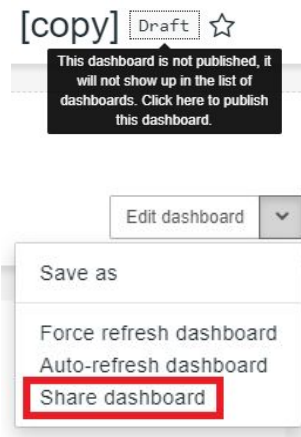
Create a Dashboard

A dashboard is a collection of charts. Users can create new dashboards, copy and edit existing dashboards and share dashboards with others. Dashboard charts refresh when underlying data changes.

1. To create a dashboard, click on Account > Superset, and select the “Dashboards” tab.
2. Click the blue  in the top right corner to add a new record.
3. Add a title and fill in any other information, if necessary, and click .
4. Your dashboard will appear in the list, but will be labeled “False” under the Published column. This means that only you will be able to see your dashboard.
5. Click on your dashboard, followed by “Edit dashboard” on the right. Here you are able to insert component charts. You can also return to a chart under the “Charts” tab, and add to your new Dashboard.



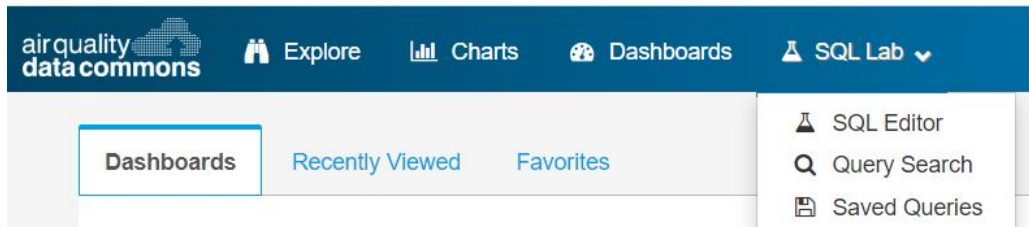
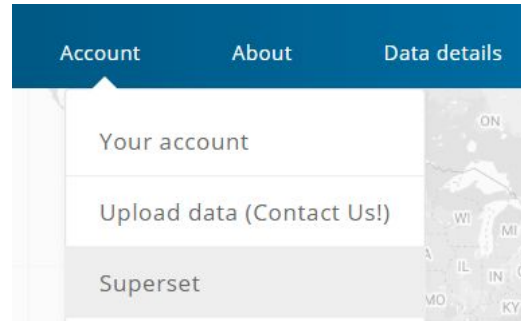
6. You can continue to add charts to your dashboard. Bounce between the “Charts” and “Dashboards” tabs when creating and adding information to your final dashboard. Dashboards can include charts with different data sources.
7. When finished creating the dashboard, you can publish it by clicking “Draft” in the title of your dashboard.
8. Use the dropdown at the top right of the page and click the “Share dashboard” button to get a shareable link and send it to anyone you would want to share this particular dashboard with. Once shared, the recipients will be able to see all changes and updates made to the dashboard and the reports it contains.



SQL Lab

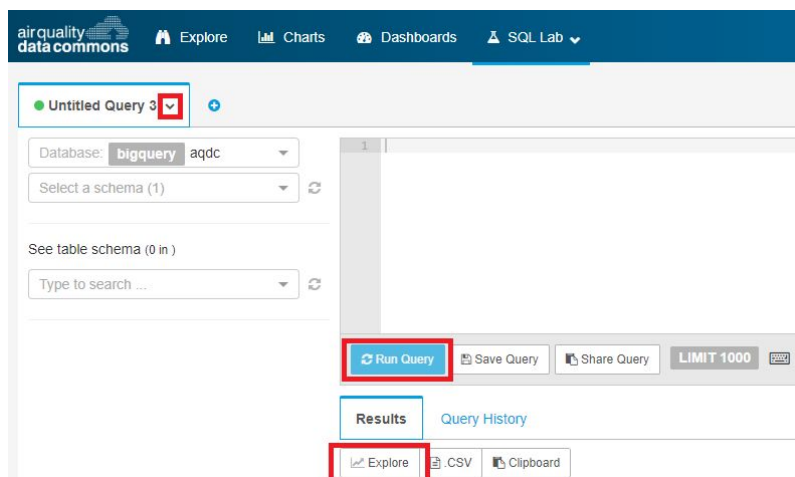
The AQDC uses structured query language (SQL) for data analysis and to generate the data behind visualizations. The SQL Lab includes a SQL Editor that allows users to edit or create their own SQL queries. This guide includes example queries for beginner SQL Lab users to practice and familiarize themselves with the more advanced analysis tools of the AQDC.

1. To start creating your own SQL query, click on Account > Superset.
2. On the SQL tab, you can create a new query by clicking "SQL Editor". You can browse other users' queries by clicking "Query Search" and review your past queries in "Saved Queries".

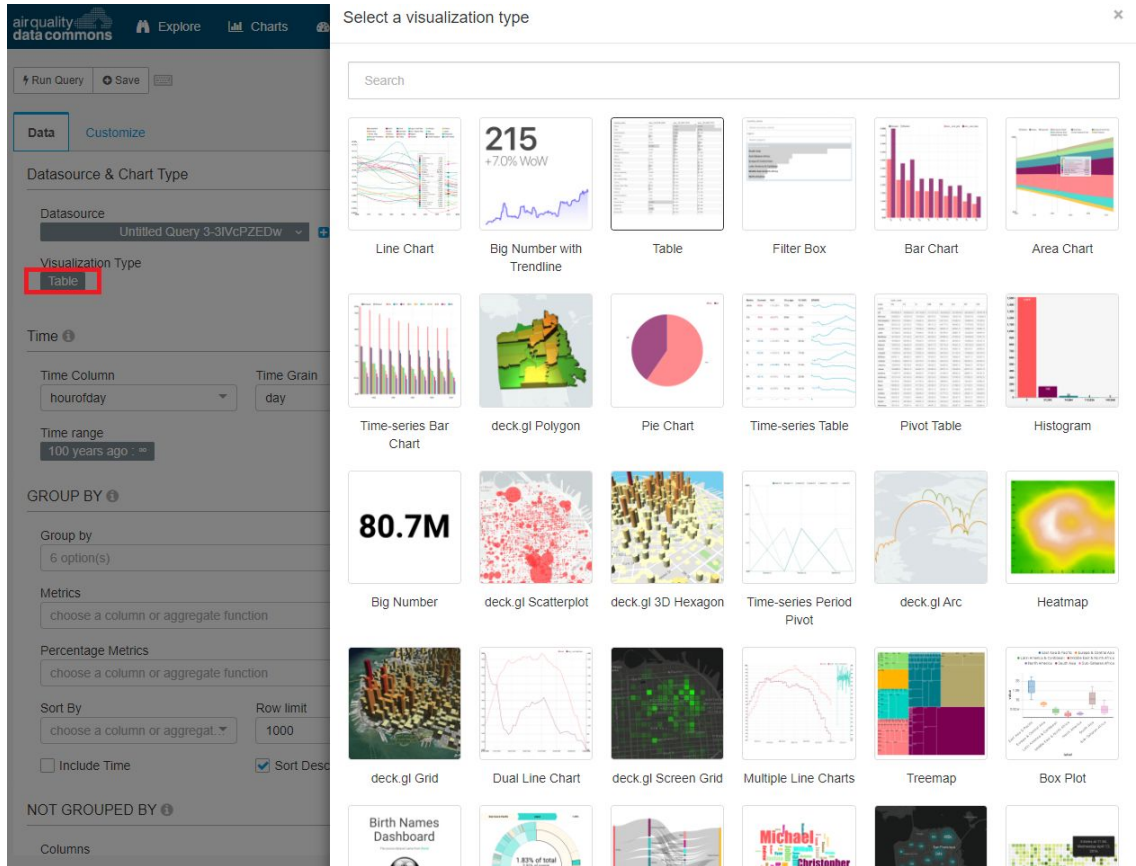


3. By clicking "SQL Editor", you will arrive at the main page and can start writing your own query.
 - a. *Note: Unless you save your query, these tabs will NOT persist if you clear your cookies or change browsers.*
 - b. By clicking the dropdown, you can edit the name of your query.

4. Enter your query and click "Run Query" to see your results.
5. To explore your data further and run a visualization, click on "Explore".



6. You will be taken to a new page and can visualize your data by adjusting time, metrics, filters, series, etc. By clicking a visualization type, the AQDC will prompt you to fill in the appropriate fields. Here you can see some of the visualization types available:



7. Once you have filled in the appropriate fields for your visualization type, click on “Run Query” in the top left corner.
 - a. If your query is not running, try refreshing your browser and clicking “Run” again.
8. You can edit your visualization by clicking on the “Customize” tab.
9. To save your new visualization, click **Save** in the top left corner.

Tutorial

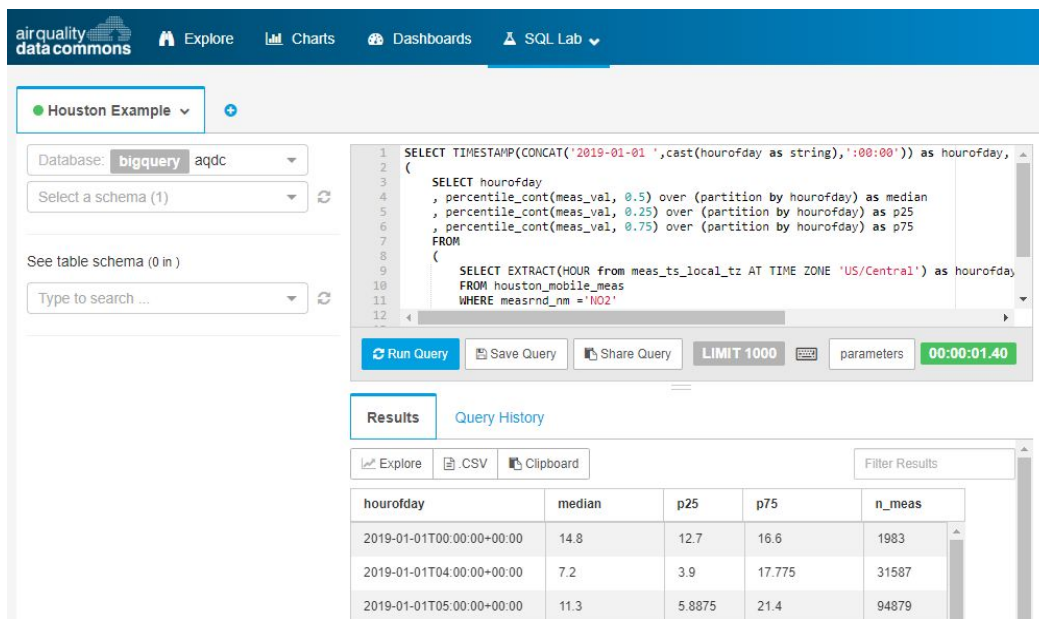
Practice running your own query by copying and pasting the following example query into the dialog box in the “SQL Lab” tab.

- **Create a line chart and a time-series bar chart.** This query accesses the “Houston Mobile Analysis” dataset and finds the hourly median, 25th and 75th percentiles of NO₂ measurements to visualize the diurnal cycle of pollution.

1. Copy and paste the following into the dialog box:

```
SELECT TIMESTAMP(CONCAT('2019-01-01 ',cast(hourofday as string),':00:00')) as hourofday,
median, p25, p75, count(hourofday) as n_meas FROM
(SELECT hourofday
, percentile_cont(meas_val, 0.5) over (partition by hourofday) as median
, percentile_cont(meas_val, 0.25) over (partition by hourofday) as p25
, percentile_cont(meas_val, 0.75) over (partition by hourofday) as p75
FROM
(SELECT EXTRACT(HOUR from meas_ts_local_tz AT TIME ZONE 'US/Central') as hourofday, meas_val
FROM houston_mobile_meas
WHERE measrnd_nm ='NO2'))
GROUP BY hourofday, median, p25, p75
ORDER BY hourofday
```

2. Click “Run Query”, and your results should look something like this:



The screenshot shows the Air Quality Data Commons SQL Lab interface. The top navigation bar includes 'airquality data commons', 'Explore', 'Charts', 'Dashboards', and 'SQL Lab'. The main area is titled 'Houston Example' and shows a query editor with the following SQL code:

```
1 SELECT TIMESTAMP(CONCAT('2019-01-01 ',cast(hourofday as string),':00:00')) as hourofday,
2 (
3   SELECT hourofday
4     , percentile_cont(meas_val, 0.5) over (partition by hourofday) as median
5     , percentile_cont(meas_val, 0.25) over (partition by hourofday) as p25
6     , percentile_cont(meas_val, 0.75) over (partition by hourofday) as p75
7   FROM
8   (
9     SELECT EXTRACT(HOUR from meas_ts_local_tz AT TIME ZONE 'US/Central') as hourofday, meas_val
10    FROM houston_mobile_meas
11    WHERE measrnd_nm ='NO2')
12 )
```

Below the query editor, there are buttons for 'Run Query', 'Save Query', and 'Share Query', along with a 'LIMIT 1000' dropdown and a 'parameters' section showing '00:00:01.40'. The 'Results' tab is active, displaying a table with the following data:

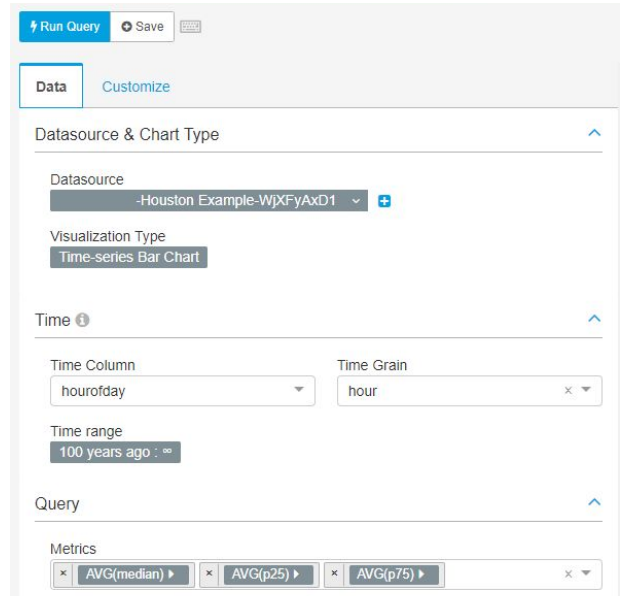
hourofday	median	p25	p75	n_meas
2019-01-01T00:00:00+00:00	14.8	12.7	16.6	1983
2019-01-01T04:00:00+00:00	7.2	3.9	17.775	31587
2019-01-01T05:00:00+00:00	11.3	5.8875	21.4	94879

3. Click on “Explore”. Change the visualization type to “Time-series Bar Chart”. You will be prompted that the “Metrics” field cannot be empty.

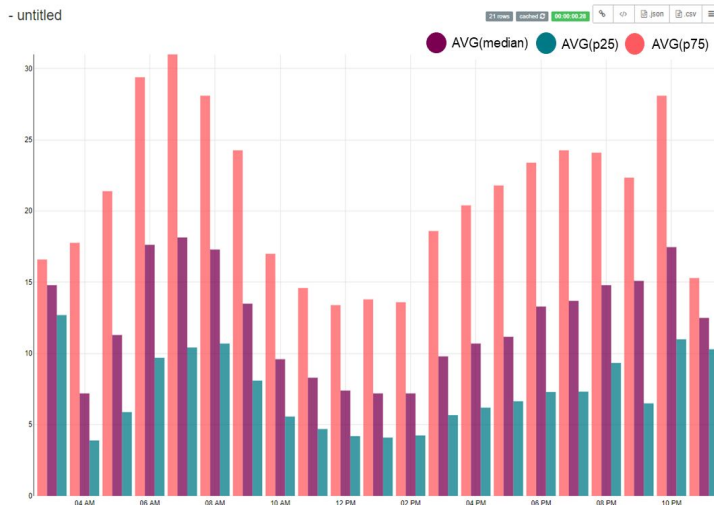
4. Change the “Time Grain” setting to “hour”.

- For the metrics, you can either click or type the following:
AVG(median), AVG(p25), AVG(p75).
Click “save” when prompted.

- Click “Run Query”.
 - If your query is not running, try refreshing your browser.

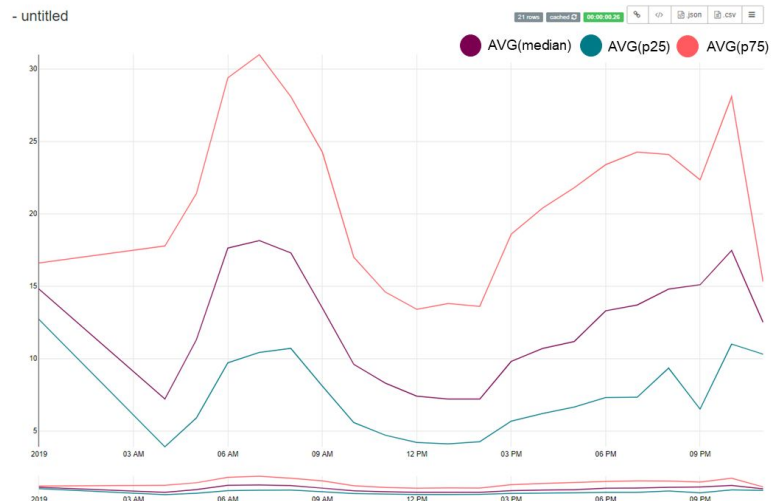


- Your results should look like this:



- Change your visualization to “Line Chart” and click “Run Query”.

Your results should look like this:



Appendix- Standardized Terms

Pollutant: Any substance, as certain chemicals or waste products, that renders the air, soil, water, or other natural resource harmful or unsuitable for a specific purpose. Specifically in our context, pollutant is the primary substance being measured by the sensors.

Meteorological Parameter: Non-pollutant parameters like temperature, humidity, atmospheric pressure, wind speed, wind direction, etc., that are typically measured alongside pollutant concentrations in the air sensors; they may or may not affect the pollutant concentration and may be used while analyzing pollutant measurement data.

Sensor: The most basic piece of hardware in a system. A chemical cell or physical device that produces an analytically useful signal by detecting or measuring the analyte (pollutant or meteorological parameter in our space).

Sensor System: An integrated set of hardware that uses one or more sensors to detect and/or measure the analyte, process signals, output parameters. May or may not include visual displays, battery for power, and ethernet or wireless internet.

Sensor Network: A deliberate arrangement of connected sensor nodes measuring pollutant(s) concentration.

(Sensor) Node: A point in a network of sensor devices that is capable of performing some processing, gathering sensory information and communicating with the gateway and/or other connected nodes in the network.

Reference Method: Federal reference method, Federal equivalent method or test method, typically used by Governmental (e.g., State) agencies to measure pollutants.

Reference Monitor: Instrument operated and maintained according to initial and ongoing Federal equivalent method (FEM)/Federal reference method (FRM) and Quality assurance (QA) protocols, often by a local, state, or federal government agency.

Data Model: Software design technique for databases intended to support transaction capture and end-user queries. A data model organizes data in tables, describes the relationships across data elements and tables, eliminates ambiguity and conflict, and provides efficient ways to access data according to user requirements.

Data Architecture: Defines the structure for data collection and storage along with data standards, metadata management, master data management (MDM), data quality rules and policies. Data architecture also includes data models and a high-level data flow of the system.

Data Management: Covers the end-to-end lifecycle of data, from acquisition/collection to archival and purging, and the flow of data through the various stages in the data system. Data Management includes data architecture, data governance, overall policies for handling, storing and sharing data. Data management minimizes the risks and costs of regulatory non-compliance, legal complications and security breaches.

Metadata: Metadata is data about data; it describes the data. Metadata provides context for data but metadata is not considered the principal data itself. A data system without metadata has very low chance of being highly useful. Metadata is the information and documentation which makes data understandable and shareable for users over time. Data remain usable, shareable, and understandable as long as the metadata remain accessible.

Data + Context (metadata) = Information