Transforming Open Urban Data into Infrastructure Supporting Air Quality Interventions

Kumar, Vijay; Jolly, Kevin; Gunner, Sam; Pregnolato, Maria; Tully, Patrick; Tryfonas, Theo

DOI
10.1109/HPCC67675.2025.00147

Publication date
2025

Document Version
Final published version

Published in
Proceedings - 2025 IEEE International Conference on High Performance Computing and Communications (HPCC)

2025 IEEE International Conference on High Performance Computing and Communications (HPCC) | 979-8-3315-6874-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/HPCC67675.2025.00147

Transforming Open Urban Data into Infrastructure Supporting Air Quality Interventions

Vijay Kumar
Wales Cyber Innovation Hub
Cardiff, UK
KumarV14@cardiff.ac.uk

Kévin Jolly
Accenture
Toulouse, France
kevin.jolly@accenture.com

Sam Gunner
Civil, Aerospace & Design Engineering
University of Bristol
Bristol, UK
0000-0002-2467-6385

Maria Pregnolato
Civil Engineering and Geosciences
Delft University of Technology
Delft, Netherlands
0000-0003-0796-9618

Patrick Tully
Industrial Liaison Office
University of Bristol
Bristol, UK
patrick.tully@bristol.ac.uk

Theo Tryfonas
Civil, Aerospace & Design Engineering
University of Bristol
Bristol, UK
0000-0003-4024-8003

***Abstract*— Urbanisation has led to urban population growth affecting the economy and the environment, including degrading air quality via pollution. Air pollution has been linked to a variety of conditions and health risks including heart disease, stroke, asthma, Alzheimer's and neurodevelopmental disorders. However, it is difficult for a citizen to find precise air pollution data at a particular location. Smart City strategies usually stipulate that city councils should focus on delivering platforms for active citizen participation using existing technology. Existing civic data hubs such as the London Datastore, Open Data Bristol etc., provide air pollution data but lack elaborate representations for user-defined locations. Existing air quality initiatives such as the Smart Citizen platform and Sensor.Community provide more advanced graphical representations. However, they restrict themselves to showing data coming from their respective devices. The paper presents the Open City Air Quality Platform (*OpenCAQP*), a development that merges a wide range of data sources and air pollution parameters into a single platform. The *OpenCAQP* allows citizens, environmentalists, data analysts, and developers to access and visualise data. The proposed solution contributes to two key objectives: i) analysis of the air pollution data sources available in a city; ii) a replicable scalable, modular open source capability aggregating and visualising air pollution data from multiple sources. Its effectiveness has been evaluated by measuring quality, usability and increased awareness of users through a feedback questionnaire.**

Keywords—open data, air quality, smart city infrastructure

I. Introduction

Air quality is essential for the environment and human well-being. Particulate matter is a threat to health by substantially increasing the risks of contracting respiratory or cardiovascular diseases [1] and is even estimated to reduce life expectancy [2]. It also causes fatigue and reduces human productivity [3], which harms the economy and increases health costs. Therefore, there are both health and economic reasons to tackle air pollution. Citizens and decision-makers alike need to be aware of the air pollution around them. However, finding accurate, precise and complete air quality data for a given location is challenging. In the era of big data, many cities aspire to use low-cost sensing technologies to enable citizens' involvement in collecting such data [4, 5]. However, ingesting, storing and analysing air pollution data from different data sources with varying types of inputs is challenging [6, 7, 8], and data analysts, especially citizens, lose the comprehensive view of the data. It is difficult to produce meaningful observations either because of lack of air pollution-related data [9], small sample sizes [7, 10], differing data types between sources. Platforms with advanced visual analytics for citizens and decision-makers in city council are usually bespoke and not widely available.

Our work aims at developing an open platform for air pollution data that helps by gathering and integrating information from multiple data sources. The platform is intended to support interventions of multiple stakeholders, including city councils wishing to intervene in areas of low air quality; citizens desiring to make informed choices (such as buying a house) and change their daily behaviour (such as asthmatic people avoiding walking through areas of low air quality [11]); or tech developers that could use the application programming interface (API) to develop applications. The implementation of the system contributes towards

- the analysis of different data sources of air pollution available in a city;
- replicable modular and scalable platforms that aggregate and visualise air pollution data from multiple sources;

Finally, we evaluate our open-source implementation of the presented work with regards to its quality, utility and awareness raising capacity using a feedback questionnaire.

The remainder of this paper is organized as follows: §II provides air quality background and explains how other platforms work. §III presents an analysis of existing data sources and data platforms. §IV provides the implementation details of our *Open City Air Quality Platform* (*OpenCAQP*) prototype. In §V, we provide an evaluation of our prototype and §VI concludes the paper, discussing future work.

II. Background And Motivation

Air pollution concerns the introduction of particulates, biological molecules, and harmful materials into the earth's atmosphere and is divided into primary and secondary pollutants [7, 12, 13, 14]. Degrading air quality is one of the major challenges faced by cities because of population growth, fossil fuel dependency, climate change impacts etc. (Fig. 1). Previous research has shown that air pollution and mortality are closely related [9, 15] - indicatively, household and ambient air pollution were responsible for an estimated 7 million deaths some 10 years ago. Globally, over 90% of the world's population lives in places exceeding WHO air quality guideline limits. In the UK, government initiatives such as

979-8-3315-6874-0/25/$31.00 ©2025 IEEE
DOI 10.1109/HPCC67675.2025.00147

Automatic Urban and Rural Network (AURN) consist of 150 stations recording air pollution measurements used for compliance reporting against ambient air quality directives. However, such stations provide data only at limited locations; for instance, there are only two monitoring locations in Bristol. Limited monitoring leads to missing data about the actual concentrations of the pollutants across locations [13] and may not offer an accurate representation of air quality in the city. In addition, pollution concentration fluctuates highly with terrain [9] and is vastly dependent on local emissions [16, 17]. Hence, it is valuable to extend data collection by building sensor grids such as WAPMS [18] in the Mauritius island to enrich the granularity of the data collected [8] and act as alternatives to traditional governmental monitoring stations [13, 19, 20]. Higher data granularity helps to better understand critical areas, microclimates, and the interactions between pollutants and traffic levels. To address this, interventions encouraging active citizen participation with the use of sensing technology have been developed [21, 22]. Typically, citizens are called to measure air pollution by collecting data using sensors such as Atmotube[1] or Smart Citizen Kit (SCK)[2].

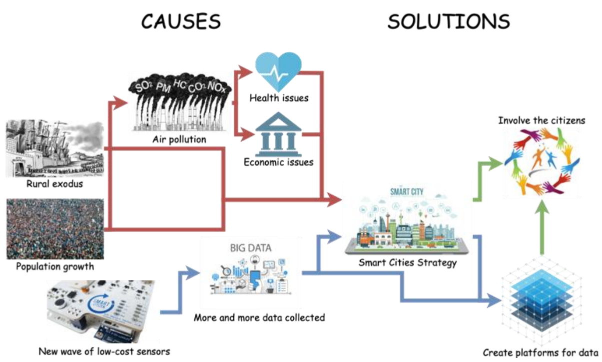


Fig. 1: Urban air pollution challenges and novel resolution approaches

Subsequently, data analysis reveals any correlations of air pollution measurements with third-party data such as traffic, weather and health [8]. E.g. previous research [23, 24] has shown correlations between air, noise pollution with transportation [13, 25]). Furthermore, a positive correlation between vehicle density and air pollution has been discovered [7]. Other studies relate fluctuations in air pollution data to seasonal patterns and temperature [7, 10]. Data analysis can predict the causality effect of weather conditions on pollution by inspecting air pollution measurements [18]. Weather can influence pollution in different ways, e.g. by transporting pollutants by wind [7, 26] or by preventing pollutants from dispersing because of the cold [17]. Finally, there are detectable relationships between household size, population, urbanisation, and pollution [2].

To meet the demand for local collection and the required granularity [27], innovative low-cost sensors [28] allow capillary and precise data to be streamed and reported in nearreal-time [13] enabling richer air quality insights. Studies suggest that data retrieved from conventional air monitoring high precision stations and data from low-cost lowprecision sensors are almost in the same range [10]. Further, measurements from high-precision governmental stations may act as a reference to help calibrate low-cost sensors [27] as performed in the Dutch Smart Emission project [4] in the city of Nijmegen. Other initiatives include the development of sensor grids or networks such as the Discovery Net project [8], Amsterdam Smart Citizens Lab Approach [19] and the FabCity project in Barcelona [29]. Most notably, the Sensor.Community[3] project (ex *Luftdaten*) built 2,349 stations across Germany to better monitor German cities over particulate matter pollution.

Data portals have long been set up to collect city data, e.g., Milton Keynes (MK) Data Hub [6, 30] makes 3,000 local and national data sets from different sectors such as energy or transport available to everyone, enabling developers to create data-intensive applications by using APIs to retrieve the data. Similarly, the Smart Emission Spatial Data Infrastructure (SESDI) [4] allows citizens to collect environmental quality data using low-cost sensors and aims at offering data for different purposes such as health determination, spatial planning, environmental monitoring, traffic management, climate adaptation etc. However, there has been no development after the project's closure in 2018; SESDI does not provide the ability to get pollution data from user location or collect data from multiple sources. Some projects such as SCK [21], and Sensor.Community display real-time data visualisations from a single type of data source (their own sensors) and do not integrate data from multiple sources. Other platforms including AirText[4], which logs air quality, UV, pollen and temperature forecasts for Greater London and the South East, provide limited such capability, i.e. map of overall air quality but do not provide detailed data.

Another capability, the UK Collaboratorium for Research in Infrastructure and Cities (UKCRIC) is a network of facilities comprising laboratories, testbeds, methods and tools for research, innovation and pedagogy. UKCRIC was established as a distributed research capability in response to the need for investment and regeneration of UK infrastructure, as set out in the Armitt Review [31]. It has evolved into four interconnected missions [5] focussing on problem-specific challenges that cannot be addressed individually, e.g. climate change. In addition to large-scale research laboratories, a network of facilities collecting urban data by leveraging emerging technologies and digitalisation was developed, aimed at increasing the understanding of how cities function and at supporting decision makers in managing city infrastructures. Each of these *Urban Observatories* (UO), provides a platform for gathering and analysing data for informing decision making at local, city, regional or national level. UO help stakeholders understand challenges and perform evidence-based interventions. The UKCRIC UO network provides an exemplar helping cities in establishing their own observatories by sharing learning and best practices. An example UO which collects urban data about noise, weather, traffic, air quality, footfall etc. is Newcastle's[6].

III. Air Quality Data Sources, Existing Platforms And Requirements For A New Approach

The discussion of §II highlighted the importance of considering multiple data sources and various parameters (e.g. pollutant concentrations, demographics, and weather) to gain

1 https://atmotube.de/?view=en
2 https://smartcitizen.me/kits/
3 https://sensor.community/en/
4 https://www.airtext.info/
5 https://www.ukcric.com/about-ukcric/missions/
6 https://newcastle.urbanobservatory.ac.uk/

accuracy, reliability and for citizens and policymakers to develop insights from air quality data.

A. Input data sources available

By investigating the different sources of available air quality data, two categories of data sources emerged. Firstly, there are traditional data sources (high-precision, high cost, sparse in numbers) and often deployed by city councils or national agencies such as Bristol City Council (BCC), Department for Environment Food & Rural Affairs (DEFRA), European Environment Agency (EEA). Secondly, citizens often deploy low-cost, affordable sensors such as Sensor.Community, the Smart Citizen, Atmotube etc. to monitor their homes, offices, or other spaces. Different data sources provide different kinds of measured parameters (such as temperature, PM2.5/10 etc.) and can be sorted into two categories. First, pollutant concentrations (typical pollution monitoring parameters) can be separated between primary pollutants, secondary pollutants, and particulate matter. Second, parameters that act as influencers and are linked to air pollution. For instance, demographics per ward, the number of households per ward, cars per household, and weather data such as temperature, relative humidity, wind direction and speed, noise, or light. Fig. 2 shows a summary of the data sources available, and the air quality parameters that are taken into account for the design of *OpenCAQP*.

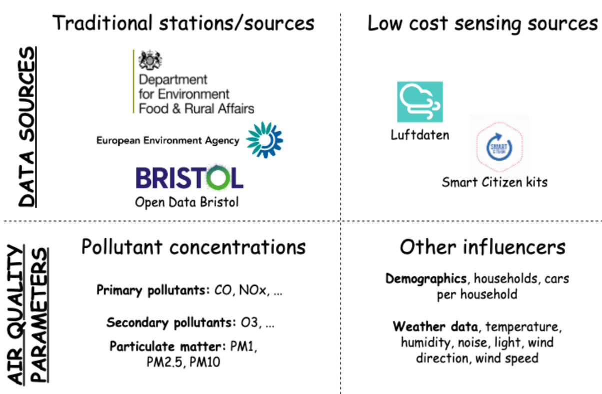


Fig. 2: OpenCAQP uses data from multiple sources

B. Traditional sources

OpenCAQP considers two traditional sources which are DEFRA AURN and Open Data Bristol (ODB). AURN performs the most substantial automatic monitoring of the UK used for compliance reporting against the ambient air quality directives and includes automatic air quality monitoring stations measuring multiple pollutant concentrations. It provides openair R library[7] to import data from monitoring networks. In contrast, ODB [8] is a data catalog website developed by Bristol City Council containing a variety of datasets. *OpenCAQP* uses multiple datasets from ODB such as air quality data (hourly updated data from different sites across the city), the nitrogen dioxide (NO_2) diffusion tube dataset (contains yearly levels of NO_2), the population estimates dataset (containing yearly estimation of the number of inhabitants per ward), the car availability dataset (contains yearly estimation of the number of cars per ward), wards dataset (contains the coordinates that delimit each ward). *OpenCAQP* uses the API provided by ODB returning JavaScript Object Notation (JSON) data. For e.g., the URL https://opendata.bristol.gov.uk/api/records/1.0/search/?dataset=air-quality-data-continuous&q=date_time%3D2019-08-15T14&rows=500&sort=date_time&timezone=UTC retrieves the data recorded on the 15^{th} of August 2019 at 2 pm from the air quality data continuous dataset.

C. Low-cost sensors

Our approach considers low-cost sensors as sources, such as Sensor.Community and SCK. E.g., the former initiative allows citizens to build sensors by themselves from parts available online and provides a data visualisation platform[9]. *OpenCAQP* uses the Sensor.Community API endpoints provided by setting the area coordinates URL to get the data related to Bristol (area=51.454762,-2.597043,25). The API endpoint allows retrieving the latest data from all sensors within a radius of 25 Km around a centre located in the middle of Bristol. Similarly, the latter Smart Citizen initiative allows citizens to measure air quality using SCK sensors capturing multiple parameters (temperature, humidity, lux, dust, pressure, PM2.5, PM10). It also provides a publicly available interface allowing anyone to develop applications and experiment on top of the Smart Citizen platform. *OpenCAQP* uses the parameter "near" which takes coordinates of a point, and the URL https://api.smartcitizen.me/v0/devices returns JSON data containing the devices that are the closest to the point. The retrieved data returns only 25 devices per request and requires a pagination system using another parameter. The Geohash parameter provides the location of the device. The data of a particular SCK device providing the average value over 1 hour is retrieved using the URL https://api.smartcitizen.me/v0/devices/<device_id>/readings?sensor_id=58&rollup=61m&from=<from_date_time>&to=<to_date_time>.

D. Technical challenges in ingesting different data sources

There are many challenges with gathering air pollution data from different data sources [13, 18]. Firstly, the data source can provide access to data in different ways. For instance, the SCK platform provides an API whereas DEFRA outputs a CSV file. Secondly, once data is read, even though different devices get data for the same parameters, the parameter names may be different from one source to another. For instance, particulate matter 10 (PM_{10}) is named **`P1`** by a Sensor.Community device, **`PM10`** by an SCK, or **`pm10`** by a DEFRA station and in ODB datasets. In addition, the units in which the parameter is recorded can be different for each data source; timestamps also bring a particular issue. All timestamps should refer to the same time-zone and have the same format. Also, the platform needs to integrate diverse data sets from pollutant concentrations to the wind, temperature, or traffic-related data. Thus, a crucial step is to sort the data by finding similarities that could be used between datasets.

To resolve these issues, *OpenCAQP* uses the existing methods available to fetch data from the different sources, defines a format with specific variable names, processes the data from various sources, and stores it into a standard unit and variable names. It further converts all timestamps to the Coordinated Universal Time (UTC) time zone and the location of a measurement to the GeoJSON (points or polygons). The *OpenCAQP*'s main feature is to provide air quality data at a given location provided by the user.

7 https://davidcarslaw.github.io/openair/reference/importAURN.html
8 https://opendata.bristol.gov.uk/
9 http://deutschland.maps.sensor.community/#6/51.165/10.455

OpenCAQP takes air pollution data from multiple datasets at the same time and divides the data based on two factors; first, the record's location (a point with given coordinates and a polygon with given demarcations), and second, the data retrieval interval of the dataset (hourly, daily and yearly).

E. *Investigation of existing platforms*

Analysis of existing platforms (§II) provided insight about key requirements such that an air quality platform must be easy to use, have simple and clear design, provide accurate and interactive visualisations, and be available open-source with open access for transparency. It must provide API for developers to build on top of it [27]. *OpenCAQP* has two main components: gathering air quality data from multiple sources and visualising of that data to inform the end user. For the data storage layer of *OpenCAQP* we evaluated multiple projects such as Dataverse, Socrata, OpenDataSoft or DKAN, CKAN. We selected the latter as suggested by [6]. Comprehensive Knowledge Archive Network (CKAN), a web-based opensource management system for the storage and distribution of open data, is widely used by institutions and provides robust documentation, datastore functionality extension (ad hoc database for storage of structured data) and has inbuilt visualisation tools for data stored in the form of a table, charts and maps. It also provides the functionality necessary to *OpenCAQP*, i.e., a python extension to insert data into the datastore and a JavaScript extension to extract data from the datastore in RESTful JSON API.

To inform the design of *OpenCAQP*'s visualisations we studied projects such as Breathe Heathrow, Smart Citizen and others. Breathe Heathrow [32] used Mapbox, an open source mapping platform for custom designed maps. SCK uses Open Street Map, a widely used free map with Leaflet, a JavaScript library for creating interactive maps, on top [21]. The BBC provides an open source visualisation R library BBPLOT[10] to create graphs, along with the JavaScript library D3.js.

IV. *OpenCAQP* Implementation

Fig. 3 presents some of the ways *OpenCAQP* displays data to end users via clear and easy-to-understand visualisations. It uses geospatial representations to show overall pollution over an area, charts and tables to display trends of how pollution changes over time, and an API provides a way to access the data programmatically. The visualizations can meet the requirements of different types of end users stakeholders: citizens, environmentalists, policy-makers and developers. Citizens can use the platform for checking daily pollution levels or historical pollution levels (e.g. in case of known area history when buying a house). Environmentalists and policymakers can check how the historical pollution levels correlate with weather and vehicle traffic parameters to understand how seasons and mobility patterns affect pollution. Developers can use the API provided to develop applications.

Architecturally *OpenCAQP* separates data storage from web server into two separate virtual machines (VMs). The first VM imports data from input sources and stores it in the data storage platform. The second VM contains the web server that extracts data from the database based on user requests.

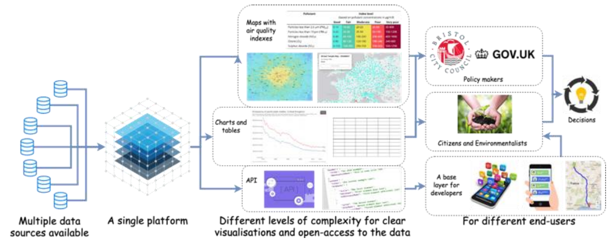


Fig. 3: Overall OpenCAQP architecture

A. *The data structure of our CKAN instance*

Fig. 4 summarises the structure of the CKAN instance used in *OpenCAQP*. Our platform sorts the datasets by the type of location of the record (point or polygon) and the data retrieval interval of the dataset (hour, day, year). The CKAN instance contains multiple datasets. First is a dataset for the latest records of the data sources and containing four major resources. Each resource corresponds to one category of data: hourly-point, daily-point, yearly-point, and yearly-polygon. Another resource named "necessary fields" contains the essential fields to be provided to the user for each type of location: point (pm1, pm2.5, temperature) or polygon (population estimates, number of cars). Secondly, other datasets correspond to each data source with all the data, including the latest records to track the previous data. Thirdly, another dataset contains units corresponding to a measurement (such as ug/m3 for pm1) and is useful for displaying the units next to the value. Separating the latest records from the rest of the data allows for the portal to quickly retrieve the relevant data to fulfill the objective of providing a user with the current air pollution data at a given location. *OpenCAQP* groups datasets into organisations that correspond to the origin of the data source. For instance, ODB organisation contains five packages corresponding to the five datasets imported from this source. Regarding the units and latest records packages, they are associated with the University of Bristol organisation as they aggregate data from multiple sources. *OpenCAQP* further aggregates the datasets in a group named Bristol to indicate the data location.

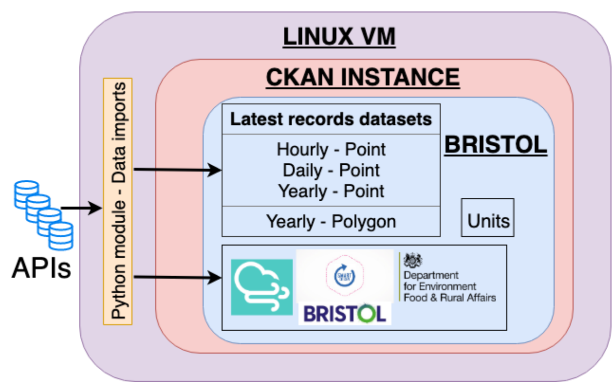


Fig 4: Structure of the data catalogue

B. *The website portal*

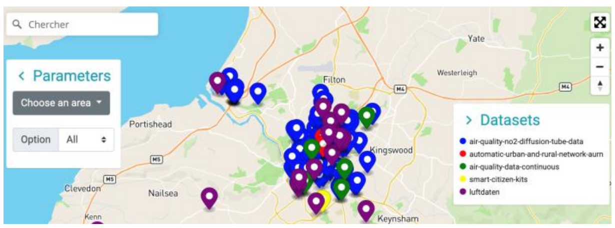
It is important to present data to the end-users in clear, easy-to-access visualisations. *OpenCAQP* contains a website to display its air pollution data. For developing efficient, fast, and scalable web applications for the website back-end Node.js[11], a JavaScript platform, along with Express.js[12], a

10 https://github.com/bbc/bbplot
11 https://nodejs.org/en/
12 https://expressjs.com/

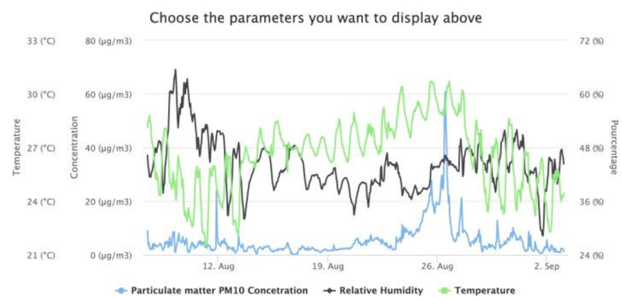
JavaScript library that greatly simplifies back-end serverside programming, especially for developing an efficient and correct routing component, were used. For rendering the view of the website, the Mustache.js[13] templating engine was used. It also uses other libraries such as Mapbox, an online map provider that uses OpenStreetMap[14], the free collaborative project, allowing a user to look up a location by name, returning its geographic coordinates. In addition, Highcharts[15], a JavaScript library for creating interactive well-designed charts, and Turf.js[16], another JavaScript library for spatial analysis. Turf.js is used for creating a grid of squares out of a polygon and check if a point is inside a polygon. For design and responsiveness purposes, Bootstrap and jQuery were also used.

device, such as its source, the device's location on a map; the latest record from this device; a tool to create charts with the data from that device. The end user can plot one or many parameters from a particular device over time to understand how air quality fluctuates with time. Fig. 5b presents a graph of multiple chosen parameters (PM_{10}, humidity, temperature) with different vertical axes. The chart is highly interactive, and the end user can perform zoom in/out by clicking and dragging or pinching on a touchscreen.

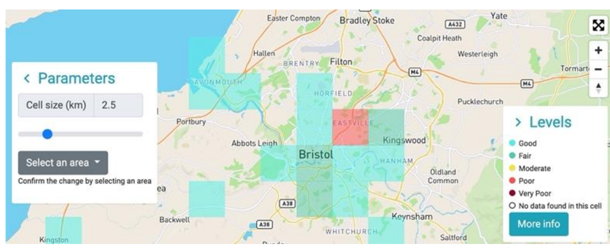
3) Air quality maps. This dashboard provides a coloured map indicating pollution levels using the European Air Quality Index[17]. It divides the polygon into square cells using the Turf library and then finds the average of the latest records for each air quality parameter inside each cell. Then,



(a) All available devices in Bristol



(b) SCK combined parameters visualisation



(c) Air quality map of Bristol

Fig. 5: OpenCAQP portal dashboards

C. Graphical elements and visualisations

The dashboarding portal provides four main features:

1) Identify all the available devices collecting data in an area. The feature allows a user to get the latest data from all available devices and get a better idea of the current coverage over the provided areas. Fig. 5a shows all the sensor devices available for Bristol. The end user can choose to show the devices that update their data every hour, day, or year. Like the previous feature, the end user can find additional data about the device, such as the previous data, by clicking on the more data link in the popup.

2) Retrieve data from a specific device over time. The end user can find historical data about a particular device. The device profile page provides basic information about the

it displays the colour of the worst level of any of the parameters. The cell size can be modified, and there is an option to change the area. Fig. 5c shows an air quality map of Bristol with a cell size of 2.5Km. The end user can find the values used to determine a cell's color by clicking on a cell.

4) Retrieve air pollution data from a location. The feature allows a user to get the closest air pollution readings available, according to the user's location. Fig. 6 shows the air quality and area data for the particular location of Bristol University's Queen's Building).

D. API and extensibility

The API provides developers with the ability to download data from *OpenCAQP*. It allows them to retrieve results of the different functionalities of the portal, such as air quality data at a location when provided with the location's coordinates. In addition, to get the latest record of devices in an area by giving

13 https://github.com/janl/mustache.js
14 https://www.openstreetmap.org/
15 https://www.highcharts.com/docs/
16 https://turfjs.org/
17 https://airindex.eea.europa.eu/

the area and an optional parameter (hourly, daily or yearly); or fetch the air quality map grid by providing the size of a grid cell and which region. It also provides functionalities to import the data directly from the datasets of the CKAN instance. The API returns data in JSON format and is provided by a URL. An option is provided to download a JSON file containing the result of the API request. Existing OpenDataSoft APIs, such as ODB's, was an inspiration for the development of our API.

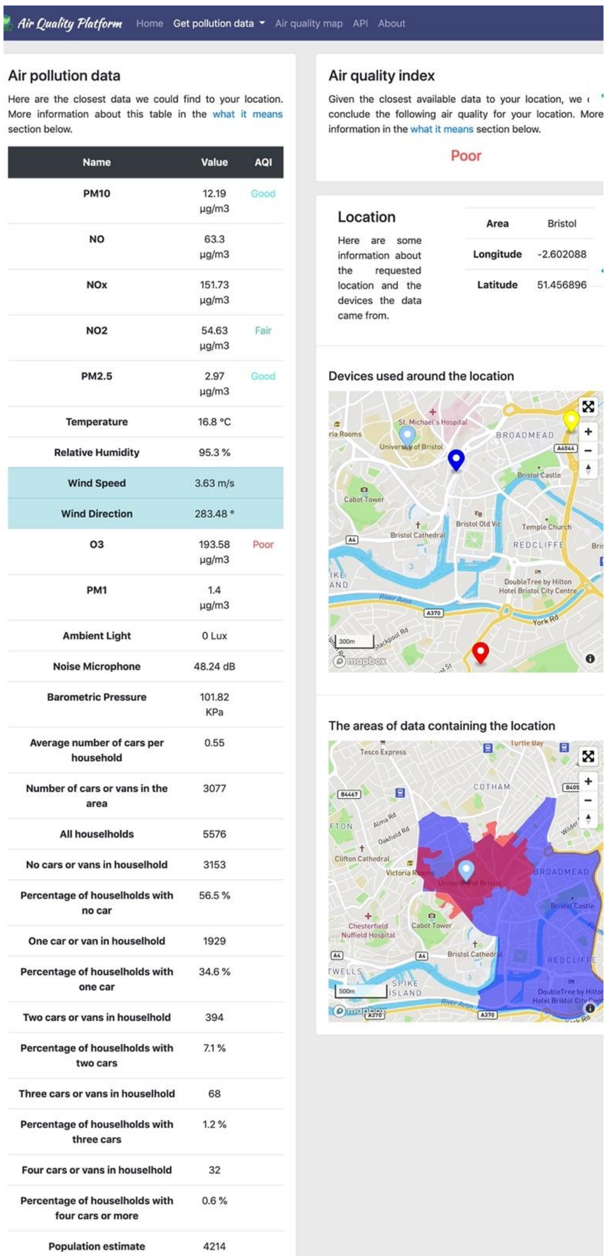


Fig. 6: Air Quality at a user requested location

The end-user can choose the fields they require to extract from the selected dataset. After selecting a dataset, the fields are provided to the user with other relevant information. It finally allows the addition of query parameters such as the field values to filter or sort the data on, as well as the fields or the maximum number of records to return.

V. Evaluation And Impact

OpenCAQP was implemented with scalability in mind. To add more data sources and the region (city), the administrator managing *OpenCAQP* can easily define the areas, and *OpenCAQP* would automatically add any new device in the database added in that area in the future. It is also straightforward to add a new data source and a new area (city) to the CKAN instance. The functionality was tested by adding the available data for Greater London from existing sources. The website can also scale to mobile and tablet screens efficiently and adjust to screen resizing due to libraries like Bootstrap and jQuery.

The authors sought feedback from potential users to evaluate design, usefulness, user friendliness, design of the platform, and how it compares with existing ones. First, we designed a questionnaire[18] for a **qualitative** evaluation. This was focussed on a small group of a dozen users[19] in Bristol comprising students, working-age adults and a retired person.

Citizen awareness of air quality platforms: The questionnaire included preliminary questions about their knowledge of air pollution and air quality platforms. All respondents believed that it is essential to care about air pollution. Only a quarter of the respondents thought they are well-informed about Bristol's air pollution. The majority (three quarters) were not aware of any existing platform for air pollution before taking this questionnaire. Respondents that were aware of some, mentioned London Air and Heathrow Airwatch as examples. Finally, respondents were asked what they would expect from such a platform. The answers mainly suggested that the air quality platform should be user-friendly, display clear information, be granular in both time and space; show data at a specific location; clearly define most and least polluted areas and provide a strong API. A recommendation suggested that the platform should identify root causes and propose action plans to reduce pollution.

Citizen experience of Data Hubs providing air quality data: The users were introduced to the data hubs providing air pollution data such as DEFRA's UK Air Information Resource[20], ODB[21] or the Milton Keynes Data Hub[22]), followed by questions after an attempt to get users to try and find air pollution data through these websites. All survey participants found it demanding to get air pollution data at a specific location and most of those questioned opined that the visualisations were not straightforward. Furthermore, over half of the sample found that the data's graphical representations were not user-friendly. Respondents explained that the navigation was designed poorly, and the visualisations look complicated to a lay person.

18 https://forms.gle/sf4u4AXkA8qZzQRP7
19 https://github.com/bitvijays/AirPollutionDataPlatform/blob/master/feedback_questionnaire_answers.xlsx
20 https://uk-air.defra.gov.uk/data/data_selector_service?show=auto&submit=Reset&f_limit_was=1
21 https://experience.arcgis.com/experience/bcf5a6312bc04ffeb43db67cd57f5439
22 https://datahub.mksmart.org/

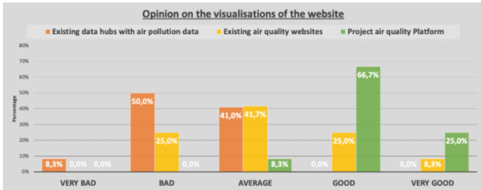
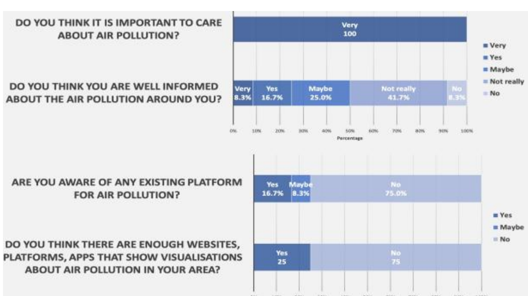


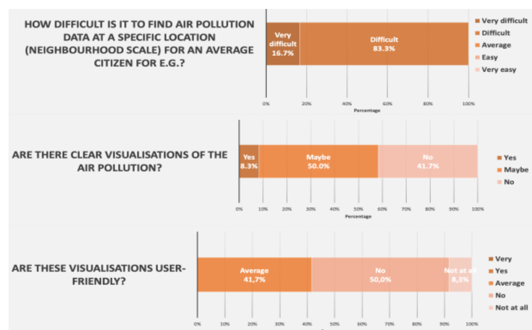
Fig. 7: OpenCAQP visualisations vs existing platforms

Citizen experience of existing air quality websites: We requested participants to analyse existing air quality websites such as AirText, Sensor.Community, SCK and the European Air Quality Index (AQI) to get users to query air pollution data and to evaluate their available visualisation tools. Participants' opinion on the user-friendliness of the websites was mixed as almost half of the participants find it average, and the rest divided themselves on the two ends of the spectrum. The same feedback applies to the overall rating of the design of these data portals. A slight majority of participants felt that the visualisation tools were simpler to use and more straightforward than those provided in the data hubs. However, regarding finding air pollution data for a specific location, the respondents are dispersed between the different options. Furthermore, a third of the respondents think that the data are not readily available for advanced users. The respondents mention that contrary to ODB, the examples of air quality websites provided in this section do not have an API or at least they are challenging to find if they exist.

Citizen experience of *OpenCAQP*: Finally, *OpenCAQP* was presented to the users to compare with existing platforms. All respondents reported a positive experience with the platform and its data presentation. They rated the site's design as useful to very good. The main reasons for these positive opinions on the design and the visualisations were that the website was easy to use, the tools provided were user friendly, and the graphical representations are simple and clear. Participants were then asked to compare their experience on this platform with the ones they had been using. Almost all found *OpenCAQP* experience better or way better than their experience on other portals. A hierarchy in preference emerges through the questions regarding user friendliness and quality of the visualisations on the compared websites, as shown in Fig. 7. The *OpenCAQP* visualisations were overall preferred to the ones of existing air quality websites. Those questioned explained that this preference was due to the new website had more simplicity, user friendliness regarding the visualisation tools, the intelligibility of the data, and accessibility for the lay user. Finally, respondents appreciated particularly the functionality to provide an address and get precise local data and the mobile portability. The final question solicited ideas for future improvements to our portal. Replies included that *OpenCAQP* can be further improved by adding new datasets, data sources to gain further granularity and coverage, provide recommendations for citizens to reduce their carbon footprint, and remembering the previous location of a user to get the data faster the next time they visit the website among others. Fig. 8 includes the results of our evaluation questionnaire.

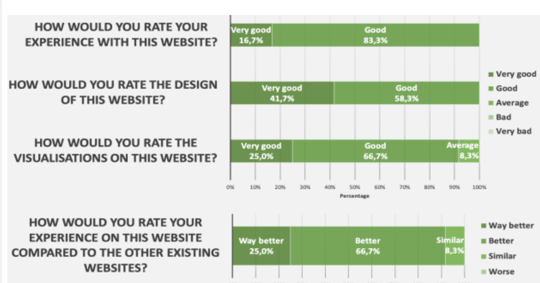


(a) Introduction



(b) Existing data hubs

(c) Existing air quality dashboards



(d) OpenCAQP

Fig. 8: Respondent answers to evaluation questions

VI. Conclusions And Future Work

This work aimed at developing an open data platform for air quality data by combining information from multiple sources and providing graphical analysis of that data to key stakeholders to increase awareness of air pollution. A platform such as *OpenCAQP* is relevant to a variety of stakeholders including citizens, environmentalists, policy officers and technology developers. The platform provides an API to retrieve the data easily and can be deployed at low cost both by city councils and communities of citizens, as demonstrated in Bristol, UK. Such capabilities improve the understanding of air pollution and provide clear, interactive visualisations to citizens. *OpenCAQP* offers a local platform to host air pollution data and engage citizens. Furthermore, this platform can eventually support communities and policy makers in their decision making, e.g. monitor air quality in schools and advise parents to drop off their kids in ways that protects them better from exposure to pollution. Ultimately, city councils can implement the OpenCAQP in other cities in the UK/world, providing visual analytics capability to citizens and increasing their awareness of air pollution.

Ways to improve the current version of the platform include getting access to non-public air quality data to enhance the platform's granularity and coverage. New graphical data representations could be added, in particular charts that allow comparing readings from multiple devices. Furthermore, addressing design improvements suggested by respondents in the feedback questionnaire. The platform opens the potential for developers to build useful third party applications. Some ideas of applications that could use the data from the platform include augmented reality showing real-time air pollution data in the streets; mobile apps providing areas to avoid while walking, driving; chatbot that provides the appropriate air pollution data upon a user's request etc. Finally, an interface can be provided to the end users to upload the data collected by personal devices such as Atmotube containing the location data and other environmental parameters.

Acknowledgments

This work was funded in part by EPSRC grant *'UKCRIC: Urban Observatories (Strand B)'* (EP/P016782/1) and Horizon Europe project *TwinAIR* (grant ID 101057779). BCC hosted a public instance of *OpenCAQP* and Knowle West Media Centre (KWMC) facilitated citizen participation during Horizon 2020 project *REPLICATE* (grant ID 691735).

References

[1] FJ Kelly & JC Fussell. Air pollution and public health: emerging hazards and improved understanding of risk. *Environmental Geochemistry and Health*, 37(4):631–649, 2015.

[2] M Cole & E Neumayer. Examining the impact of demographic factors on air pollution. *Population and Environment*, 26(1):5–21, 2004.

[3] JG Zivin & M Neidell. The impact of pollution on worker productivity. *American Economic Review*, 102(7):3652–73, 2012.

[4] M Grothe, J Vanden Broecke, LJ Carton, H Volten & R Kieboom. Smart emission - building a spatial data infrastructure for an environmental citizen sensor network. In *Proc. of Geospatial Sensor Webs Conf. GSW 2016 Münster, Germany*.

[5] N Castell, FR Dauge, P Schneider, M Vogt et al. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environment International*, 99:293 – 302, 2017.

[6] M d'Aquin, A Adamou, E Daga et al. Dealing with diversity in a smartcity datahub. In *Proc. of 5th Workshop on Semantics for Smarter Cities, CEUR Workshop Proceedings, CEUR-WS.org*, pp. 68–82, 2014.

[7] B Jatinder. Complete study of factors contributing to air pollution. Technical report, Bangalore, India, 10 2016.

[8] M Ghanem, Y Guo, J Hassard, M Osmond & M Richards. Sensor grids for air pollution monitoring. In *Proc. of 3rd UK e-Science All Hands Meeting, Nottingham, UK*, 2004.

[9] LB Lave & EP Seskin. An analysis of the association between u.s. mortality and air pollution. *Journal of the American Statistical Association*, 68(342):284–290, 1973.

[10] N Ilieva, Y Nikolova, I Predyov & N Kozarev. Sensors applicability for PM2.5 and PM10 air concentration measurements. *Journal of Chemical Technology and Metallurgy*, 53(3):496–503, 2018.

[11] F Bamashmoos, P Woznowski & T Tryfonas. A review of air quality sensing technologies and their potential interfaces with IoT for asthma management. In *Proc. of 11th PErvasive Technologies Related to Assistive Environments Conf., PETRA 2018, Corfu, GR, pp.* 470–475.

[12] DC Carslaw & G Rhys-Tyler. New insights from comprehensive on-road measurements of nox, no2 and nh3 from vehicle emission remote sensing in london, uk. *Atmospheric Environment*, 81:339–347, 2013.

[13] P Kumar, L Morawska, C Martani, G Biskos et al. The rise of low-cost sensing for managing air pollution in cities. *Environment international*, 75:199– 205, 2015.

[14] European Environment Agency (EEA). Air quality in Europe. 2018.

[15] N Fann, AD Lamson, SC Anenberg, K Wesson et al. Estimating the national public health burden associated with exposure to ambient PM2.5 and ozone. *Risk Analysis*, 32(1):81–95, 2012.

[16] RE Britter & SR Hanna. Flow and dispersion in urban areas. *Annual Review of Fluid Mechanics*, 35(1):469–496, 2003.

[17] H Ritchie & M Roser. Air pollution. *Our World in Data*, 2019.

[18] KK Khedo, R Perseedoss & A Mungur. A wireless sensor network air pollution monitoring system. *CoRR*, abs/1005.1737, 2010.

[19] Q Jiang, F Kresin, AK Bregt, L Kooistra et al. Citizen sensing for improved urban environmental monitoring. *Journal of Sensors*, 5656245, 2016.

[20] E Cogliani. Air pollution forecast in cities by an air pollution index highly correlated with meteorological variables. *Atmospheric Environment*, 35(16):2871–2877, 2001.

[21] T Diez & A Posada. The fab and the smart city: The use of machines and technology for the city production by its citizens. In *Proc of 7th Intl Conf. on Tangible, Embedded and Embodied Interaction*, TEI '13, pages 447–454, New York, NY, USA, 2013. ACM.

[22] K Austen. Environmental science: Pollution patrol. *Nature News*, 517(7533):136–138, 2015.

[23] M Agbali, C Trillo, T Fernando, IA Ibrahim & Y Arayici. Conceptual smart city KPI model: A system dynamics modelling approach. In *2018 Second World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, pages 163–171, 2018.

[24] M Chen, T Ho & C Jan. A system dynamics model of sustainable urban development: Assessing air purification policies at Taipei City. 2006.

[25] HW Davies, JJ Vlaanderen, SB Henderson & M Brauer. Correlation between co-exposures to noise and air pollution from traffic sources. *Occupational and Environmental Medicine*, 66(5):347–350, 2009.

[26] J Gabrys. Air Walk: Monitoring Pollution and Experimenting with Speculative Forms of Participation. In: Charlotte Bates and Alex Rhys-Taylor, eds. *Walking Through Social Research.* Routledge.

[27] H Volten, JLA Devilee, A Apituley, LJ Carton, M Grothe, C Keller, F Kresin, A Land-Zandstra, E Noordijk, E van Putten et al. Citizen science with small sensor networks. Collaboration between a Dutch EPA (RIVM) and local initiatives. 2016.

[28] M Balestrini, T Diez & P Marshall. Beyond boundaries: The home as city infrastructure for smart citizens. In *Proc. of ACM Intl Joint Conf. on Pervasive and Ubiquitous Computing*, UbiComp '14, pp 987–990.

[29] R Besson. La fabcity de barcelone ou la réinvention du droit à la ville. *Revue Urbanews*, Mar 2015.

[30] E Daga, M d'Aquin, A Adamou & E Motta. Addressing exploitability of smart city data. In *Proc. of 2016 IEEE Intl Smart Cities Conference (ISC2)*, pages 1–6.

[31] J Armitt. An independent review of infrastructure. 2013.

[32] J Landry, K Webster, B Wylie & P Robinson. How can we improve urban resilience with open data. *Open Data Institute*, 2016.