
openaq
Open Air Quality Data: The Global State of Play
Written by the OpenAQ Team
with contributions from the OpenAQ Community
July 2020

Contents

Important Notes before reading:

(1) This assessment represents the current state of the OpenAQ Community's shared knowledge, as gathered via public resources. If you are a government official or other person who would like to contribute new information or submit corrections, please email: info@openaq.org.

(2) For the purposes of this assessment "air quality open data" will be shorthand for ambient, ground-level and station-specific pollutant (PM2.5, PM10, SO2, NO2, CO, O3, and/or BC) data collected by governments.

(3) For the purposes of this assessment, the term "real-time data" is shorthand for data produced on daily or sub-daily levels and released on an approximate daily or sub-daily level.

About OpenAQ

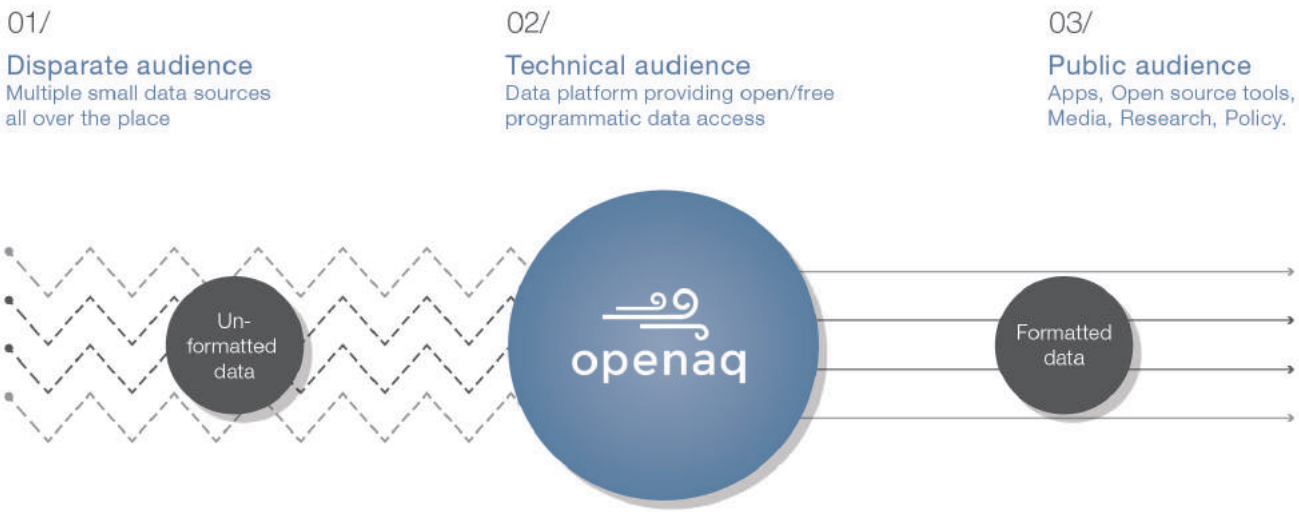


Figure 1: OpenAQ harmonizes air quality data from disparate sources into one single format in order to allow the public to more fully maximize the uses - and ultimate impact - of the data.

OpenAQ's mission is to aggregate and harmonize open air quality data across the globe and to build an ecosystem of people across sectors and geographies to use these open data to fight the unequal access to clean air to breathe – air inequality.

The OpenAQ Platform currently houses more than 500 million data points from 11,000+ stations in 93 countries and is the largest open air quality data platform of its kind, in terms of measurements and geography. It receives ~35 million data requests/month to the open API.

The underlying platform that fetches, stores, and serves data is completely open-source and has subsequently been built by a global community of scientists, software developers and others invested in creating a global air quality air quality open data-sharing infrastructure over the past 5 years. The data are used for a wide variety of applications, including everything from air quality forecasts produced by NASA scientists to platforms communicating air quality in India to data-driven media reports by the general public. By filling this basic data-access gap, OpenAQ empowers a wide variety of individuals, organizations, and sectors across the globe to do more with these existing data.

1. Goals of This Assessment

This assessment reviews the global air quality data landscape.

Air quality data underpins all actions on air pollution. It tells us how much pollution is in the air we breathe. It is used to predict how pollution will change in space and time and shapes our understanding of how pollution impacts our health and the economy. This information is foundational to design, implement, enforce and assess the effectiveness of pollution control policies.

This global air quality data assessment determines the set of countries for which a small, strategic investment in existing air quality data-sharing infrastructure could dramatically open up data access to the public. It also identifies the places where no known air quality data are currently generated by government programs, pinpointing the most strategic locations for launching new monitoring and data-sharing efforts.

- Develop key criteria for sharing openly-accessible air quality data provided by government entities.
- Use these key criteria to describe the current landscape of global air quality data accessibility to the best of our knowledge and to identify opportunities to fill global air quality data collection and data-sharing gaps.
- Share this information in such a way that a broader community can provide feedback, corrections, and updates that track progress of the open air quality data landscape over time.

OpenAQ intends to update this assessment on a regular basis, as a service to the air quality community and to gauge progress over time.

2. Motivation for Focusing on Government Air Quality Monitoring

Broad access to open air quality data, especially when data are accurate, timely, and sustained over multi-year periods, is critical for public, private, and civil society initiatives to improve the air we breathe.

When shared openly, air quality ground-level monitoring performed by governments, often provide unique value to broad sets of audiences in several ways compared to other monitoring efforts. Government air quality ground monitoring efforts offer:

- **Sustained, temporally-fine data:** Multiple pollutant data are often provided at sub-daily level levels (most frequently at hourly intervals) and measurements are sustained over the course of multiple years at a given location.

- **Timely availability of data:** Government air quality data often generated and provided to the public in real time in some form or, if data are collected manually, made available in a timely fashion (e.g. in a few days rather than a year).

- **High perception of legitimacy and credibility:** Government-produced data is of natural public interest and has a higher capacity to be held to public account.

- **More uniformity of measurements and methodology across countries:** While governments' air quality monitoring practices and their ability to assure quality of data can be variable, in terms of management strategies, implementation, and resources, most countries that conduct ambient air quality monitoring, do so with more similar frameworks, objectives, and pollutants of key interest than compared to other air quality monitoring activities (e.g. research, community-based, or individual efforts, few of which are often coordinated across projects or have shared, specific goals). Combined with the fact that government-generated measurements are typically sustained over a longer term, this allows a higher degree of comparability of air quality data collected simultaneously across countries than other ground-based efforts.

Yet in the most polluted countries, there are the least amounts of open air quality data available from governments. For example, using open data accessed from the OpenAQ platform to calculate country-level PM2.5 average values over the entire historical record for each country in

the platform and comparing that to the number of corresponding stations in each country, we see a clear trend:

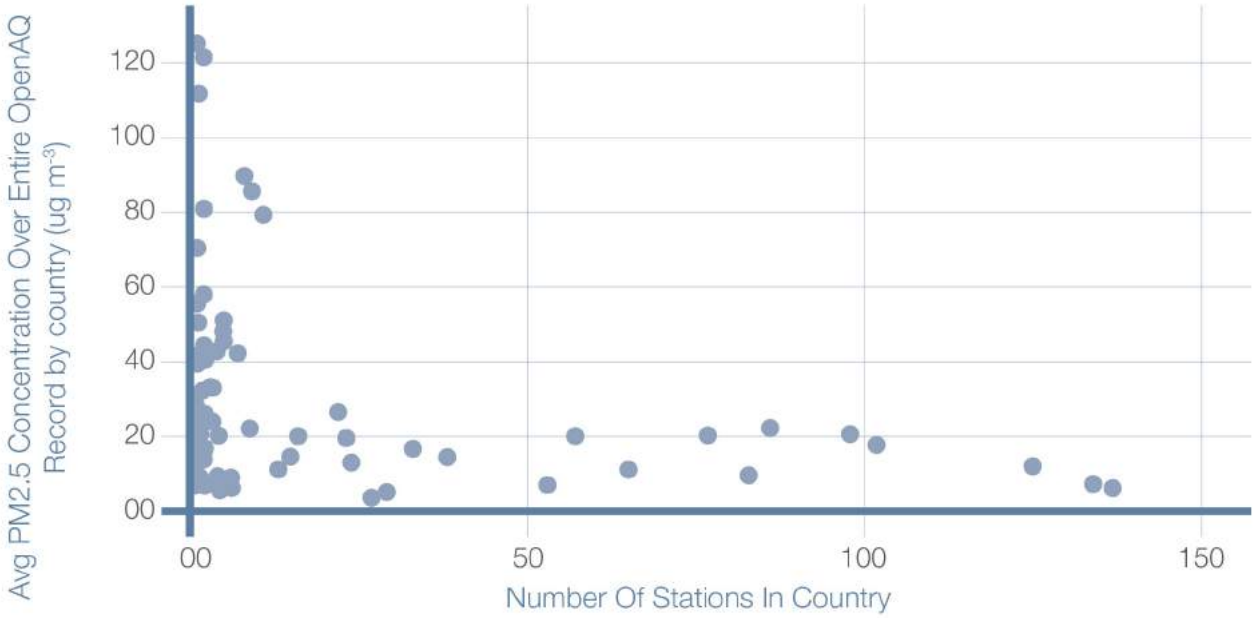


Figure 2: Where there is more severe air pollution, there tends to be less open data.

Notes on this figure: (1) Data accessed from openaq.org on 20 January 2020. (2) Six countries (China, France, Germany, India, Spain, and the United States) that had greater than 150 stations were excluded in the figure. (3) The time-average for each country varies; it depends on how long the country's data have been ingested in OpenAQ. (4) The number of stations reporting to OpenAQ is not the absolute total number of government stations that exist. Rather, it should be viewed as a proxy for the number of stations reporting availably, easily accessible data. (5) When average PM2.5 concentrations are presented versus "millions of people per monitor in a country currently reporting to the OpenAQ Platform," the trend disappears. This suggests that although countries reporting high pollution levels of PM2.5 appear to have fewer stations, there is not a correlation between air pollution levels and number of stations reporting to OpenAQ relative to country population per country.

We aim to define the global air quality open data landscape created by governmental sources, and pinpoint specific ways in which it can be improved and therefore more fully accessed and used for impact by the public.

3. Four Key Criteria for Fully Open AQ Data

While air quality is monitored and publicly reported in at least 103 countries, there is considerable variability in how these data are shared and their consequent ability to be maximally and most impactfully used by the public.

Below, we describe the simple system used in this assessment for categorizing ambient air quality open data access provided by governments from the perspective of making data available in formats that: (a) are easily 'harmonizable' with other data sources into a single format, and (b) allow the most use-cases to be built on top of the underlying data.

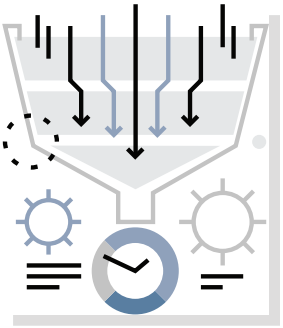
More specifically, a given governmental source is deemed to provide fully accessible open air quality data if the following **Four Key Open Data Criteria** are met:

(A) **Physical Data:** Data are shared in physical units, as opposed to a country- (or organization-) specific Air Quality Index (AQI), Air Pollution Index, or AQI-like quantities.

Example: PM2.5 data are shared in units of micrograms per cubic meter, as opposed to a single country-defined and unitless Air Quality Index system.

(B) **Station-Specific & Coordinates:** Data are provided at the most transparent geographic scale at which they are collected - station-scale - and with location metadata in the form of readily-available geographic coordinates, ideally to five places past the decimal.

Example: Data that are provided originate from a single station, as opposed to an aggregation or average of several stations over a city. Additionally, the geographic coordinate single station from which the data originate are also shared (e.g. The station is located at: 101.89322 N, 30.29571 E).



(C) **Timely Fine-Scale Temporal Information:** Data are provided at daily or sub-daily levels in near-real time or in a timely manner with time-of-collection stamps and averaging periods.

Example: Data are shared as hourly averages within a few hours after they are produced, as opposed to yearly averages of pollution level shared a couple of years later.

(D) **Programmatic Access:** Data and metadata as defined in the preceding criteria are publicly accessible in a programmatic or machine-readable form.

Example: Data are shared via an Application Programming Interface (API), ftp server, or other method that allows machine-to-machine interaction), as opposed to data shared as a pdf or on a website as a graphic, table, or even in a spreadsheet form that requires a user to click a 'download' button.

A special note on programmatic access: This is a particularly key but misunderstood and overlooked criterion. While sharing air quality data in a manner readily-understandable by the public on a website is a valuable first step, it is one particular use-case for the data. If the data are also provided in a machine-readable, analysis-ready form, many more use-cases - and ultimately impact - can be derived from the data. The key is to enable full data use to those who are not the data producers. An analogy can be made with art: Imagine a world in which only those who produced art supplies (data producers) were able to be artists (data users). How greatly limited the creative scope of art in the world would be! The same goes for data: If full access is limited to those who produce the data, the uses of the data will be much more limited in scope than if the data were programmatically accessible and available for the public at large to apply their expertise.

We base these Key Criteria on the qualities of a government data source that make it easily digestible into the OpenAQ Platform and converted to OpenAQ's harmonized data format. The harmonized data format, in turn, is designed to capture the minimum amount of data and metadata that allow broad usage downstream of the OpenAQ platform and are also commonly shared by government entities.

We have applied these Four Key Criteria to the publicly-available status of government-produced air quality data in 212 countries. A variety of sources were used in assessing the status of such data, including: government agency websites or reports, the World Health Organization's 2018 Ambient Air Pollution Database, AQICN.org, openaq.org, and OpenAQ's repository on GitHub, where the existence of government data are reported by community members. While we may have informally heard of other sources, we have not included any 'word of mouth' reports in this assessment.

4. OpenAQ Global Assessment of Government Monitoring

The above assessment and below findings were made by applying the Four Key Criteria for Fully Open AQ Data in §3 to 212 governments across the world.

5. Four Key Findings

FINDING #1: Only half of the world's governments produce air quality data.

We find evidence for 49% of countries' governments producing air quality data in some capacity. The remaining 51% of countries represent a total population of **1.4 billion people**.

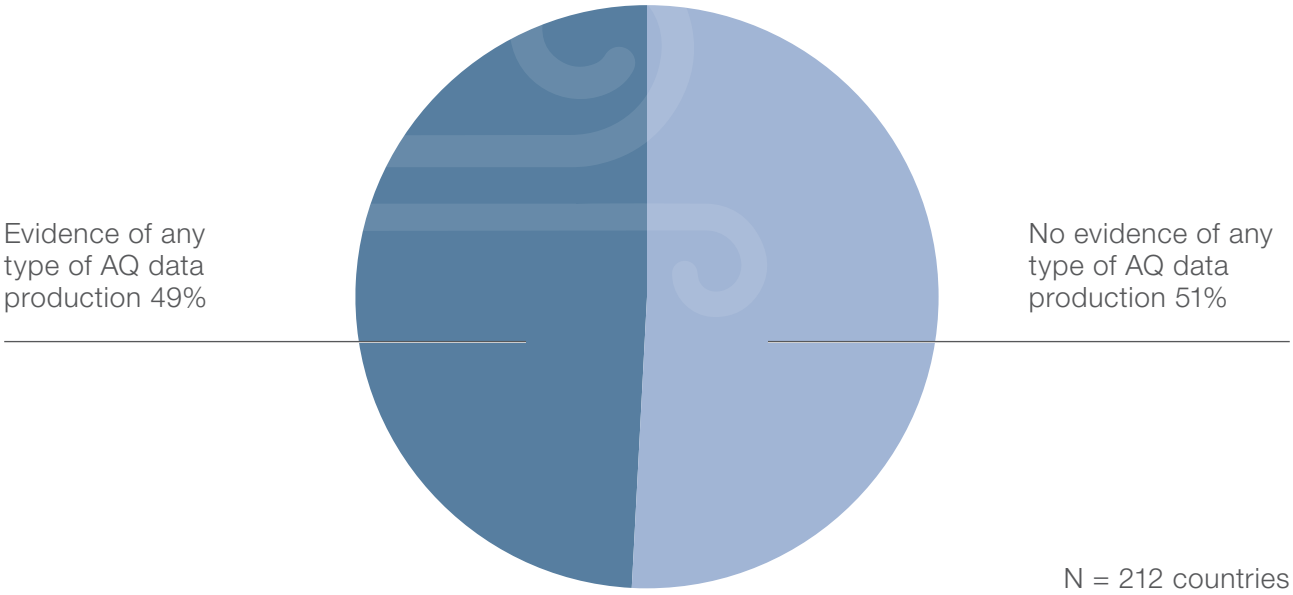


Figure 3: Percentage of countries with evidence of home government producing ambient air monitoring quality data of any pollutant type according to the four outlined criteria.

FINDING #2: Four out of ten governments share real-time air quality data.

We find evidence that 38% of governments share real-time air quality data in some capacity, even if not in a fully open form. The remaining 62% of countries represent a total population of **2.1 billion people**.

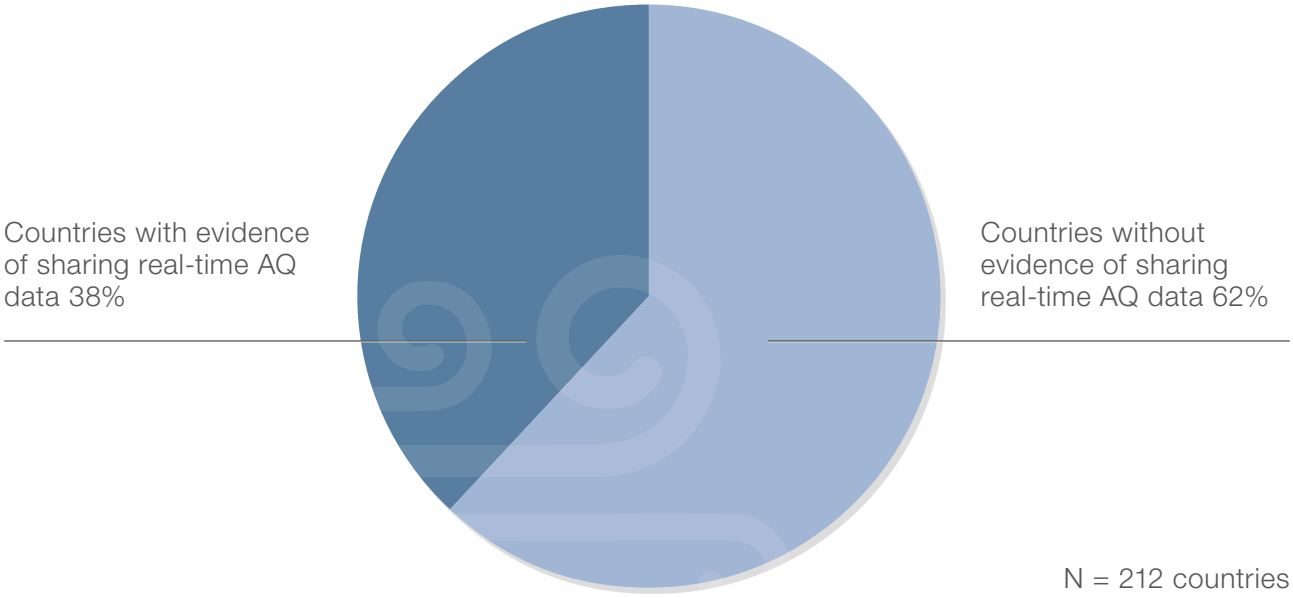


Figure 4: Percentage of countries with evidence of home government sharing real-time air quality data.

FINDING #3: Investing in government air quality monitoring programs in just 13 countries could affect 1 billion people.

Air pollution and its impact on a country's population is not immediately solved by standing up air quality monitoring; however, a baseline of measurements - and the commensurate capacity of governments to make such measurements and to share this information publicly - are key steps in this process.

The 13 most populous countries, equating to a total population of 1 billion people, in which there appears to be no evidence of a public national-level* government program for long-term ground-level ambient air quality monitoring:

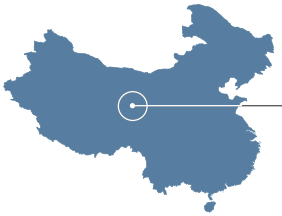
Country	**Population (in millions)****	**2017 ranking of 'air pollution' as an in-country risk factor for death and disability, as determined by the Global Burden of Disease** Note: This includes ambient PM2.5, ozone and household air pollution.
Pakistan	221	#5
Nigeria	206	#3
Ethiopia	115	#4
Congo, Dem. Rep.	90	#7
Tanzania	60	#3
Kenya	54	#5
Uganda	46	#4
Algeria	44	#8
Sudan	44	#6
Iraq	40	#7
Afghanistan	39	#2
Uzbekistan	33	#8
Angola	33	#4

* "Public National-level" is meant to connotate a nationally-coordinated effort, as opposed to a single city-led or sub-national led efforts, where information on the program is available on a publicly-accessible website.

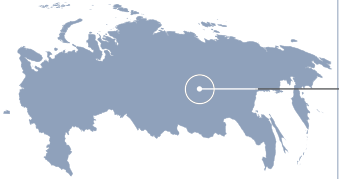
** 2020 Estimated Country Population from World Prospects by the UN

FINDING #4: At least 30 governments generate real-time data but don't yet share them in a fully open manner. Making these existing data more fully open* would affect 4.4 billion people.

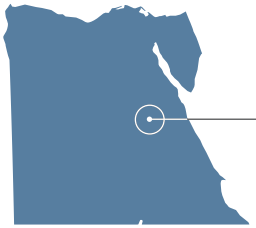
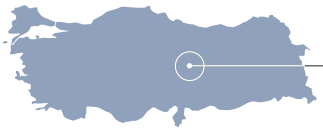
Countries with populations over 50 million in which real-time air quality data are produced in some format, but not in a fully open manner on a national-level by the home government*:

Country	Population (in millions)**	Key Criteria of Fully Open Data that are Not Met
 China	1439	• Programmatic Access • Station-Level & Coordinates
 India	1380	• Programmatic Access

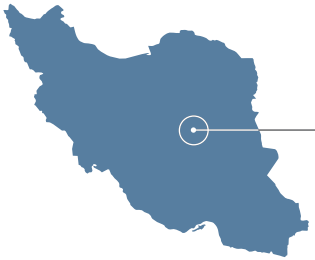
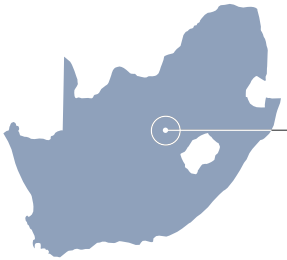
The maps provided in this report are provided for informational purposes only. OpenAQ is not responsible for any misrepresentation or delimitation of borders.

	Indonesia	**274**	• Physical Data • Station-Level & Coordinates • Programmatic Access
	Brazil	**213**	• Station-Level & Coordinates • Programmatic Access
	Russian Federation	**146**	• Physical Data • Station-Level & Coordinates • Programmatic Access
	Japan	**126**	• Station-Level & Coordinates • Programmatic Access

The maps provided in this report are provided for informational purposes only. OpenAQ is not responsible for any misrepresentation or delimitation of borders.

	Philippines	**110**	• Physical Data • Station-Level & Coordinates • Timely Fine-Scale Temporal Information • Programmatic Access
	Egypt	**102**	• Physical Data • Station-Level & Coordinates • Timely Fine-Scale Temporal Information • Programmatic Access
	Vietnam	**97**	• Physical Data • Station-Level & Coordinates • Programmatic Access
	Turkey	**84**	• Programmatic Access

The maps provided in this report are provided for informational purposes only. OpenAQ is not responsible for any misrepresentation or delimitation of borders.

	Iran	84	• Timely Fine-Scale Temporal Information • Physical Data • Programmatic Access
	Thailand	70	• Programmatic Access
	South Africa	59	• Programmatic Access

*"Fully-open data" is defined as a data-source meeting the four key criteria articulated in §3.

**2020 Estimated Country Population from World Prospects by the UN

Data is fundamental to enabling society to improve its air. This report shows that the current air quality data landscape is inadequate: we need more data that are made more fully open and accessible to more people.

The maps provided in this report are provided for informational purposes only. OpenAQ is not responsible for any misrepresentation or delimitation of borders.

In conclusion, we recommend:

Organizations and governments that support air quality programs must ensure that their investment promotes data transparency and openness. Doing so - through the criteria outlined in this report – will unlock the full potential applications from the data and lead to improved air quality as a result.

6. Acknowledgements

This assessment has been put together by the OpenAQ Team and has greatly benefited from several comments from community contributors, listed below. Please note: Any errors, omissions, or opinions belong to OpenAQ and in no way should be attributed to the individuals below.

Michael Brauer
Joe Flasher
Sarath Guttikunda
Andrew Grieshop
Emily Hasenkopf
Darby Jack
Pallavi Pant
Heidi Yoon
Dan Westervelt

This document also benefited from contributors who wish to remain anonymous.

OpenAQ received support for this work from *Bloomberg Philanthropies, Clean Air Fund,* and *ClimateWorks Foundation.*


openaq