Open Air Quality Data Platforms for Environmental Health Research and Action

Rosales, Colleen Marciel F.,[1,2*] Bratburd, Jennifer R.,[3] Diez, Sebastian,[4] Duncan, Sara,[5] Malings, Carl,[6] and Pant, Pallavi[7]

[1] OpenAQ, Washington, D.C., USA (ORCID: 0000-0002-8925-8352)
[2] Air Quality Research Center, University of California, Davis, California, USA
[3] University of Wisconsin-Madison, WI, USA (ORCID: 0000-0001-5787-1470)
[4] Centro de Investigacion en Tecnologias para la Sociedad, Universidad del Desarollo, Chile (ORCID: 0000-0001-9659-0356)
[5] School of Health Sciences, Western Carolina University, Cullowhee, NC, USA (ORCID: 0000-0002-9200-9680)
[6] Morgan State University, Baltimore, MD, USA & NASA Goddard Space Flight Center, Greenbelt, MD, USA (ORCID: 0000-0002-2242-4328)
[7] Health Effects Institute, Boston, MA, USA (ORCID: 0000-0003-0251-6580)

*corresponding author: colleen@openaq.org; cfrosales@ucdavis.edu

ABSTRACT

Purpose of Review. Air pollution remains the greatest environmental threat to public health, and significantly impacts many aspects of human life. This review surveys platforms that provide data openly and make air quality data available for use and analysis by environmental health researchers to inform air quality action.

Recent Findings. Successful programs that rely on different types of open air quality data have been observed and exist as models for regions that have yet to improve their air quality. However, disparities remain in the availability and granularity of generated air quality data in resource-rich and resource-poor regions. Even if air quality data are made available and open, their usability and actionability still remain challenging. Determinants of usability may include the user's technical know-how and ability to deal with disparate data; compute capabilities; barriers in data sharing and data generation; and misuse and misclassification of data and levels and sources of exposure. However, the synthesis of different sources of data and advancements in modeling may help fill in spatiotemporal gaps in many areas of the world. In addition, the democratization of air quality data is facilitated by advancements in air quality data generation (*e.g.,* data collection by open-source air sensors) and the ability to contribute to global, open-access, publicly available repositories that openly share data.

Summary. Many open data platforms exist for sharing and accessing ground-based/in-situ data, satellite remote sensing data, and modeled or otherwise derived outputs. Online platforms that allow visualization and analysis without the need to download software are also now widely available. Despite wide availability, disparities in accessibility exist—expansion and support through interdisciplinary collaboration and continuous funding support for air quality data is essential for the continuous global benefit and community-specific action to address air quality as an environmental health issue.

KEYWORDS

open data, air quality, air pollution, air quality databases, air quality repositories, satellite remote sensing data, ground-level monitoring, air sensors

MAIN TEXT

Introduction

Air pollution remains the greatest environmental threat to public health, and significantly impacts well-being, economy, trade, food security, and many other aspects of human life [1]. While the world has experienced an average slight decline in air pollution in recent years, the burden due to non-communicable diseases resulting from air pollution is high and growing in many regions worldwide [2]. In particular, populations from low- and middle-income countries (LMICs) are exposed to 1.3–4 times higher levels of ambient particulate matter ($PM_{2.5}$) than those in high-income countries [2]. LMICs may lack air quality monitoring programs, national air quality standards, and other critical analysis tools to assess pollution emissions and impacts on health [3–5]. For example, a report by the United Nations Environment Programme (UNEP) in 2021 stated that 34% of countries do not have protections against ambient air quality and 86% of these countries do not have air quality standards at all [6]. In 2025, the World Health Organization (WHO) released an Air Quality Standards database (https://www.who.int/tools/air-quality-standards), where it was reported that there are still many countries that still do not implement standards for pollutants [7]. In addition, while regional ambient pollution has improved drastically in high-income countries, concentrations can vary drastically at smaller scales [8]. Confounding these issues are threats to federal or national-level monitoring and enforcement programs and possible laxes in regulation that may lead to increased exposures [9].

Inequalities in air pollution exposure correspond with disparities in the technical computational capacities required to collect, access, and use air quality data effectively. Globally accessible open data platforms that make air quality information freely available and accessible can address some of these disparities. However, many users in resource-limited regions may lack the infrastructure, skills, or even connectivity to work with large, machine-readable datasets, which can hinder their ability to leverage these tools for local decision-making. Addressing the uneven access and ability to utilize large-scale datasets is critical for ensuring that open data platforms achieve their full potential in promoting equitable access to air quality information and action. This review will focus on open data platforms for air quality, particularly those making data universally and widely available for air pollutants identified in the World Health Organization (WHO)'s air quality guidelines, *i.e.*, fine particulate matter ($PM_{2.5}$), coarse particulate matter (PM_{10}) nitrogen dioxide (NO_2), ozone (O_3), sulfur dioxide (SO_2) and carbon monoxide (CO) [10].

According to the Open Data Toolkit, data are "considered 'open' if anyone can freely access, use, re-use, and redistribute them, for any purpose, without restrictions" [11]. Furthermore, true openness requires making the data available in open formats that software can read (computable and machine-readable), allowing users to readily query and re-use the data; for example, publishing a portable document format (PDF) on a government website is not considered "open." [12, 13] In the case of air quality data, an example "use case" would be downloading ground-level air quality measurements and feeding them into a machine-learning model that combines these measurements with satellite measurements to create air quality forecasts [14]; such a process would be considered "open" if both ground level and satellite datasets were freely and publicly accessible, straightforward to use, and searchable. Additionally, machine learning models used must be open-source as well as the resulting forecast and its metadata, with the code and data products shared in a public repository.

Case-studies

Accessible measurements of air pollution are critical for decision-support applications: for example, having air quality forecasts to help people plan their day to reduce exposure to ambient air pollutants; helping communities, researchers and governments measure exposures to populations to understand and study impacts on health; and evaluating emission sources and effective strategies to reduce emissions. Open data should be used to inform the public, while ideally reducing exposure through changed policies or other actions. This can occur at the continental, country, state, or smaller municipality level. Often, locally applied data can help to increase understanding of and involvement in improving air pollution in communities [15].

A variety of open data networks have been used to better understand local air pollution, particularly at the hyper-local level (10 meters or less). A well known example is in the United States (US), where the Clean Air Act has successfully led to the improvement of air quality, with extensive air quality monitoring and the requirement to make the data publicly available. Setting up regulatory networks, as well as informational networks helps continuously inform regulations. For example, the update to the US National Ambient Air Quality Standard for $PM_{2.5}$ from 12.0 to 9.0 μg/m^3 was informed by, alongside other data, long-term air quality data trends as seen from air quality monitoring networks [16]. Likewise, the WHO updated its recommended annual $PM_{2.5}$ guideline in 2021 from 10 μg/m^3 to 5 μg/m^3, and for NO_2 from 40 μg/m^3 to 10 μg/m^3. Another non-regulatory example is the *Aires Nuevos* project in which low-cost Air Visual Sensors were placed throughout cities in Central and South America. The data from these monitors are readily available on the IQAir website (see https://www.iqair.com/us/profile/airesnuevos as of Feb 21, 2025) and were used for local reporting such as that described by Silva *et al* [17].

For air quality forecasts, accessible public data helps improve tools. The US EPA operates the Fire and Smoke Map (https://fire.airnow.gov/), which aims to provide the public with timely, interpretable air quality data as it relates to fires and smoke, such as those from wildfires. The US Department of State's app ZephAir uses data from embassies and consulates around the world, plus satellites, forecast models, and other ground monitoring stations, combined with machine learning to provide air quality forecasts for U.S. personnel, as well as publicly available for anyone who wants to access the information [18]. Similarly, the Copernicus Atmosphere Monitoring System (CAMS) provides five-day forecasts for air pollutant concentrations (https://atmosphere.copernicus.eu/charts/packages/cams/) for the entire globe. The data have been leveraged for a variety of applications, including an app service for residents and visitors in Greece (https://atmosphere.copernicus.eu/discovair).

Data types

Data related to air quality can be either quantitative or qualitative. Numerical air quality measurements in physical units (reported in concentration units such as μg/m^3 for $PM_{2.5}$) as collected by an air sensor or a reference monitor, and associated readings like timestamps (date and time) and geographical coordinates are quantitative measurements. On the other hand, qualitative data may include descriptive or subjective information (*e.g.,* observations of visibility ranked on a Likert scale), reports and anecdotes of symptoms from individuals, narrative descriptions of air pollution sources, and stories of lived experiences of communities. This paper describes quantitative air quality data openly available for use in environmental health research. The use and availability of qualitative data for environmental health research in air quality is outside the scope of this review.

Ground-based/in-situ data

Ground-based, or in-situ, measurements are key for accurate quantitative characterization of air pollutants. Importantly, these direct measurements also serve as: (a) "ground truth" for the calibration and validation of other air quality data (*e.g.,* model outputs, satellite products, sensors, derived data products, etc.) and (b) if "ingested" during the data generation process, they can help enhance the accuracy of models, satellite data, sensors, derived products, etc. These observations capture fine-scale local and/or short-term variations that are often missed by satellite data or models, making them indispensable for air quality monitoring and research. Table 1 provides some examples of currently existing air monitoring networks.

Although the spatial coverage of ground-based networks is more limited compared to satellites, they remain crucial for understanding air quality in both urban and rural areas. These networks typically support regulatory monitoring and can also be used for exposure assessments and epidemiological studies, particularly in densely populated urban settings. Real-time data also provide actionable insights during pollution episodes, such as wildfires or industrial accidents, aiding emergency responses. In rural areas, ground-based data are particularly valuable for monitoring agricultural pollutants (*e.g.*, ammonia, pesticides, dust), as well as the impacts of transported urban emissions (*e.g.*, ozone) and emissions from fires. In LMICs, ground monitoring in rural areas can also be helpful in estimating the contribution of household air pollution due to the use of solid fuels for cooking and other residential activities to ambient air pollution. These measurements also help assess their effects on sensitive ecosystems, including crops and forests, and provide baseline data to evaluate future changes related to land-use changes or industrial development.

However, ground-based monitoring systems require regular maintenance and calibration to ensure data accuracy and comparability, which can be particularly challenging in resource-limited settings. High costs and limited technical capacity—both in terms of skilled personnel and infrastructure—may pose significant barriers in less advantaged regions, often hindering the sustainability and effectiveness of these networks [13, 20–22]. Even in resource-rich countries with more dense monitoring networks, gaps in coverage can have policy implications that limit the effectiveness of air quality policies in unmonitored regions, though there is potential to supplement this with other data sources [23, 24]. Strengthening local capacity through training programs is essential for enabling teams to install, maintain, and calibrate these systems, thereby ensuring their long-term viability [25, 26]. In addition, procurement of spare parts, especially for reference monitors, is a significant barrier in resource-constrained environments, and frequent power outages can result in loss of data; for example, between 27-33% of data were estimated to have been lost due to load shedding in South Africa [27, 28]. Additionally, the lack of standardized protocols and interoperability among networks managed by different entities often limits their integration into open data platforms and complicates comparisons across datasets. Addressing these challenges is crucial to maximizing the utility of in-situ networks and fostering equitable access to high-quality air pollution data worldwide.

A significant opportunity to complement traditional ground-based networks in regions with limited monitoring infrastructure is the use of emerging "low-cost sensors". While their accuracy does not match that of reference-grade monitors, their affordability and ease of deployment make them a practical choice especially for community-led monitoring to address localized environmental challenges [30].

Table 1. Some examples of existing air monitoring networks.

			Examples	**Geographical Coverage**

In-situ air quality networks			Acid Deposition Monitoring Network in East Asia (EANET) [31] Chemical Speciation Network (CSN) Interagency Monitoring of PROtected Visual Environments (IMPROVE) ACTRIS [32] ECA-INEA AURN SPARTAN [33] ASCENT	Asia USA USA Europe Mexico United Kingdom Global USA
Remote-sensing ground-based networks			PANDONIA AERONET EARLINET LALINET [34] TCCON MPLNET SHADOZ GAW	Global Global Europe Latin America Global Global Southern Hemisphere Global
"Low-cost" sensing networks			Breathe Cities Sistema de Alerta Temprana del Valle de Aburrá (SIATA) sensor.community opensensemap.org verpm	14 cities globally Colombia Global Global Mexico

Satellite remote sensing data

Satellite remote sensing is another tool for understanding air quality at regional to global scales. Such sensors routinely collect information on the presence and concentrations of aerosols and trace gases in the atmosphere. The spatial and temporal resolution of these data are continuously improving, allowing better urban-scale analysis and near-real-time information. For example, the European Space Agency (ESA) TROPOMI instrument (launched 2017) provides daily global coverage of up to 5.5 km by 3.5 km resolution for trace gasses like NO_2 and SO_2, while geostationary missions like Geostationary Environment Monitoring Spectrometer (GEMS) over East Asia (launched 2020) and TEMPO over North America (launched 2023), and the planned Sentinel-4 mission over Europe and North Africa provide similar information on an hourly basis for their respective regions. Air quality and its health impacts are also becoming a higher priority; the Multi-Angle Imager for Aerosols (MAIA) mission planned for launch in 2026 is the first explicitly public health-focused mission of the US National Aeronautics and Space Administration (NASA), and includes both ground-based and space-based components as well as planned epidemiological studies to better connect aerosol concentrations and properties with their health impacts in several cities across the globe.

Satellite data complements ground-based measurements by capturing spatial patterns and trends not observable using spatially limited in-situ measurements [45]. Satellite data are also suitable for the detection of emission sources (*e.g.,* wildfires and point sources such as oil and gas fields) and the development or validation of emissions inventories using a top-down approach [46]. An inability to directly measure near-surface pollutant concentrations is a notable limitation of satellite data in air quality applications; other information is needed to connect remote sensing data to the near-surface conditions most relevant for human and environmental health. The inability to make certain measurements at night or under dense cloud or smoke conditions can also bias a purely satellite-based analysis toward clear-sky daytime conditions [47]. Thus, while not sufficient on their own, satellite-retrieved data can be used in conjunction with other data to gain a comprehensive picture of air pollution.

Publicly funded agencies that operate satellite missions, such as NASA and ESA, have mandates to make their data open. While commercial satellite data providers are becoming increasingly relevant for optical and radar imagery, they are as yet not prevalent in the provision of data in spectral bands relevant for air quality (although this may change in the future), and the highest quality satellite datasets for atmospheric pollutants remain available cost-free. However, there are still barriers to accessibility and usability which limit the openness of satellite data. Large file sizes and unfamiliar, sophisticated data formats can make these data inaccessible to those without appropriate computational resources and subject matter expertise. The myriad of missions and datasets to choose from can also be overwhelming to new users. These problems, however, have been recognized and there are institutional efforts to address them, *e.g.*, via NASA's Open Science initiative, NASA Applied Remote Sensing Training Program (ARSET) capacity building, and NASA Health and Air Quality Applied Sciences Team (HAQAST) partnerships with data users.

Modeling

Modeling via numerical simulation of air quality is important for better understanding the processes affecting air quality and providing context for observations. It can also be useful for forecasting and prediction that enables protection and informs action. Atmospheric composition models, such as zero-dimensional box models, transport models, or global three-dimensional (3D) models are particularly useful for forecasting future air quality and for running counterfactual simulations to investigate policies aimed at improving air quality. Creating a realistic model simulation, however, requires information on historical and/or projected emissions of various pollutants. Observational data are also required to validate model performance. Finally, observational data can be used to update and adjust models to bring them more in line with real-world conditions via data assimilation. In all cases, the open availability of these input data would facilitate their use. Open provision of modeling results is also an important consideration to support open science (*e.g.*, even those without the resources or expertise to operate the models can still analyze the outputs). Open model architectures (*e.g.*, those based on open-source software and allowing for community involvement and updating) and open modeling platforms (*e.g.*, accessible computational resources such as cloud computing) are further dimensions to consider regarding the openness of modeling.

Derived datasets

Derived datasets synthesize other air quality information, such as in-situ measurements and satellite remote sensing data, using numerical and/or statistical modeling, often with the goal of producing spatially and temporally complete datasets. Creating derived datasets can also reduce biases and uncertainties versus a single information source, especially in the case of relating satellite and/or atmospheric model outputs to surface-level pollutant concentrations and air quality. A common example is land use regression models, which use spatial characteristics of land use and cover to extrapolate from sparse in-situ measurements to spatially complete exposure maps. In many cases, derived datasets are customized and adapted to a specific study domain, time interval, and objective to maximize their local applicability and relevance in a specific use case, at the expense of generalizability. There are also more explicitly general datasets, such as global atmospheric composition reanalyses (*e.g.*, MERRA-2, CAMS EACS4) and derived datasets for $PM_{2.5}$ (*e.g.*, https://sites.wustl.edu/acag/datasets/surface-pm2-5/) [48], which aim for global accuracy at the potential expense of applicability to local and hyperlocal conditions (*e.g.*, [49]).

There are innumerable approaches to creating such derived datasets, which have different benefits and drawbacks and rely on different assumptions, leading to potentially different outputs for what are ostensibly the same quantities. The variety of such derived air quality datasets can be a strength or a weakness. On the one hand, it can be difficult to know which dataset is suitable for a particular region or application, versus those that might have biases or make assumptions that are unsuitable. On the other hand, if these datasets are openly available and interoperable, it may be possible to quickly test multiple datasets to find the one best suited to a particular use or to use several datasets via an ensemble approach (*e.g.*, [50]). It is also important to have access to other open data, especially in-situ measurements, both for developing new and validating existing derived datasets; the lack of such open data will bias derived datasets towards a better representation of regions with more open data.

Data Repositories

Data repositories where users can contribute data and metadata (*i.e.*, descriptive narratives or references accompanying the air quality data) are a go-to resource for many air quality researchers. Table 2 shows some examples of platforms that aggregate and share air quality data. Many of them allow user contributions; others perform the search of air quality data sources themselves.

On the other hand, many air quality researchers use generic open data platforms like Zenodo, osf.io, or other online data commons that host any type of data set (not just air quality) to provide their datasets publicly [51–54]. For example, a project called SensEURcity ("a multi-city air quality dataset collected for 2020/2021 using open low-cost sensor systems"), which involves open-source sensors (developed by the involved research groups), provides their data and metadata publicly and freely through Zenodo, an online platform funded by CERN, OpenAIRE and the EU [55]. Similarly, the QUANT project [52] has made its multi-year, high-resolution measurements (both reference-grade and sensor data) and metadata available through CEDA for 14 commercially available sensor brands [56].

Table 2. Some examples of third-party platforms that aggregate and share air quality data and metadata.

Organization or entity			Website URL	Description
OpenAQ			https://explore.openaq.org/	Aggregates and harmonizes open air quality data from across the globe onto an open-source, open-access data platform
Berkeley Earth			https://berkeleyearth.org/data/	Data accessible at the global, national/regional, and local levels. Source and intermediate data are available as well.
sensor.community			https://sensor.community/	Community-driven, open environmental data where users can submit air quality data from DIY sensors
AQICN			https://aqicn.org/	Real-time air quality information (air quality indices) available globally
canair.io			https://canair.io/	Citizen science project using mobile and static sensors to measure air quality with cell phones and low-cost technology

Some data repositories more specifically serve data related to air quality. For example, the World Meteorological Organization's (WMO) OSCAR is a "resource developed by WMO in support of Earth Observation applications, studies and global coordination." From here, quantitative, user-defined requirements for observation of physical variables such as those related to weather, water and climate can be accessed. Specific information on all earth observation satellites and instruments, and expert analyses of space-based capabilities may also be accessed from OSCAR. The Tropospheric Ozone Assessment Report (TOAR), provides access to long-term ozone measurements from over 10,000 stations around the world as well as satellite, aircraft, and model data [57]. All data are freely accessible for research and have been used across air quality and climate change assessments at national and global scales [58, 59].

Visualization and analysis

Finally, there are openly available online platforms for visualizing air quality information. These are often created by the data providers or aggregators, with the goal of facilitating accessibility of their data, especially for those without the expertise or infrastructure needed to work directly with the raw data. For example, open platforms for satellite data visualization include Worldview, Giovanni, VEDA/JPL data mapper, FIRMS, AerosolWatch, JSTAR Mapper, RAMMB/CIRA, EUMETSAT viewer and CAMS; these resources provide visualizations of data from various missions relevant to different aspects of air quality, along with the ability to perform basic analyses or comparisons. The US EPA provides AirNow, a popular tool for air quality information, along with the Fire and Smoke map. Many companies that provide dashboards for their air quality instruments also provide data visualization capabilities, often relying on publicly available data.

A key advantage of these platforms is increasing the accessibility and usability of data by non-expert users, as well as allowing experts to easily browse databases in search of events of interest for further analysis. A risk of such platforms, however, is that their simplified representations of data may obfuscate nuances, leading casual users to incorrect conclusions. The balance between accessibility and clarity of data representations is an area of ongoing development and dialogue between the data experts and their user communities.

Challenges

Themes that emerge as challenges include data accessibility. Considering what is “accessible” may depend on the user's perspective. For example, accessing online air quality platforms may require stable internet connectivity, and can be an initial barrier to access. Other times, it is the technical know-how in interpreting and analyzing the data that proves to be a barrier to access, especially in underserved communities. Some of the key challenges are outlined below, along with potential mitigation strategies.

Data generation. A potential barrier to collecting air quality data is the maintenance needs of reference-grade monitors. Specifically, maintaining such monitors is hindered by limitations in access to necessary resources, supplies, and spare parts, coupled with the requirement for specialized technical skills. These barriers may be overcome by leveraging international expertise and partnerships. Intercountry and local government air monitoring efforts, as well as air monitoring projects by non-governmental organizations (such as academic institutions, community-based organizations, civic groups and citizen scientists), are also increasing worldwide [13]. For example, some countries participate in a regional intergovernmental effort that includes monitoring, such as the International Centre for Integrated Mountain Development (ICIMOD) which serves the Hindu Kush Himalaya region and Air Quality Central Asia which supports monitoring in Kazakhstan, Kyrgyzstan, Uzbekistan and Tajikistan. Breathe Cities is a combined initiative of a growing number of global cities to combine data from stationary and mobile air sensors with grassroots campaigns to build awareness of air pollution and support municipal action. One of the cities is Accra, and their data is made available through https://breatheaccra.org/.

Data sharing. There are reasons why sharing data might not be seen as beneficial. For example, Ivey et al. mention that traditional academic practices become problematic when researchers pursue "helicopter" or "parachute science," where data from communities are published without representation or validation of accuracy from the community being studied [25] or the government authority operating the monitors. These practices can lead to data being preemptively closed and unavailable to the public, but can also lead to hampering action that benefits environmental and public health. Another reason is the threat to data privacy and security: as data platforms increasingly incorporate real-time hyperlocal data (*e.g.*, from sensors or monitors in communities), issues around data privacy and security become significant, requiring actions to protect sensitive information while maintaining openness.

When data sharing is warranted, however, upstream mandates requiring open data with specific stipulations could be influential and necessary for open data access. For example, in the United States, projects receiving federal funding are mandated to provide accessible data. One example of a platform that results from such mandates is the Climate and Health Outcomes Research Data Systems (CHORDS) platform of the National Institute of Environmental Health Sciences: (https://chordshealth.org/). The European Union (EU) also requires its member states to provide up-to-date ambient concentrations of various pollutants [60]. One example of a portal from the EU is https://aqportal.discomap.eea.europa.eu/. In addition, there are also philanthropic organizations that specify and outline definitions of "openness" [61, 62]. However, funding limitations and platform sustainability could be put in jeopardy when grant funding or governmental support, which may not be permanent, is not sustained. For some countries, monetizing the data is a way to raise funds. This poses a risk to the long-term storage and availability of air quality data, especially in regions where funding is limited. Long-term data storage can be both a logistical issue and dependent on the current state of government affairs. For example, in early 2025, U.S. EPA EJScreen, a tool used in operation since 2015 to explore disparate pollution impacts, was retired without ample public notice. Likewise, the NASA SERVIR program, a partnership between NASA and the United States Agency for International Development (USAID) that provided support for international dashboards, was scrubbed and removed from NASA websites after abrupt funding freezes and firings at USAID.

Disparate data types, formats, and applicability. Overlapping data from different sources can make it difficult to tell what dataset to use in what situations. In these cases, flowcharts like what NASA HAQAST provides can be very useful (see https://haqast.org/data-and-tools/). Large datasets can be difficult to download and manipulate and can be offered in formats unfamiliar to health practitioners and researchers. How remotely sensed data relates to surface air quality may be unclear in some instances, *e.g.* in what conditions can aerosol optical depth (AOD) be used as a proxy for ground $PM_{2.5}$? Models can have bias errors, especially when compared to hyperlocal air quality. For air quality monitoring required by regulation, air quality data collected may still be widely variable in terms of data accuracy, spatial distribution, and temporal frequency. Harmonization may or may not fall into the responsibilities of the complying body. Some governments may develop and impose data formats and data standards to reduce barriers to access. In addition, “data intermediaries” *i.e.*, third-party aggregators and harmonizers, can benefit the community by performing harmonization [66].

Computational power. Users might face challenges in the implementation of complex air quality models. Effective models might be constrained by the need for robust infrastructure and sufficient computational power, which might not be accessible to users unaffiliated with research institutions. For example, many users who only rely on personal and laptop computers may lack the computing resources needed to work with large datasets. Easy-to-use platforms (*e.g.,* NASA Worldview and Giovanni) can allow users to work with datasets without having to download and manipulate large files.

Exposure misclassification In-situ, satellite-derived, and modeled data are almost exclusively measurements or estimates of outdoor air concentrations and often at large spatial scales. However, air pollutant concentrations can vary widely across distances and outdoor concentrations are not always representative of those experienced indoors. In general, air pollutant concentrations are lower indoors compared to outdoors, but can be elevated, often dependent on occupant behavior [67]. For example, Chambliss *et al.* found concentrations of ultrafine particles measured by mobile monitoring sites to be more than twice as high as concentrations provided from land-use regression models [49]. In LMICs in particular, indoor exposures to air pollution can be quite high in areas where solid fuels are used for household cooking and heating and contribute substantially to respiratory disease burden risk [68]. When using open data containing concentrations of outdoor air pollution, researchers, policymakers, and the public should be aware that these concentrations are not always reflective of actual exposures.

Misuse and determining appropriate levels of use. Assessing what level of rigor and accuracy is necessary for different purposes and objectives is helpful such that potential negative impacts from the misuse of data (*e.g.*, non-validated data) can be minimized. The requirement of sensor assessment and data validation will depend on the intended use of the data, for example, data that are used for increased public understanding (such as community engagement and education) will not be necessary to undergo as much quality assurance as data that are used for legal and policy actions, such as regulatory standard setting and enforcement [69]. Post-processing algorithms may or may not be necessary depending on the purpose**.** Open data availability impacts conversations and policy action on air quality.

Perspectives and Future Recommendations

When thinking about "openness", data providers should be guided by a framework that allows an open data platform to increase its impact in data-scarce regions, such as LMICs in the Global South. Third-party aggregators and data intermediaries can also help ensure that platforms that serve data-scarce regions equitably can exist. Addressing gaps in technical training can make open data impactful.

Ultimately, in-situ "ground truth" data will always be needed. Efforts should be made to support existing monitor networks with infrastructure, resources, and local technical capacity building, as well as to expand and supplement these networks to close global data gaps, and finally, to encourage and support openness for the collected data. Support for such efforts should be international in scope, considering the global public good provided by these data. However, efforts should be locally directed to ensure that community needs are being met and local conditions are taken into consideration. To further enhance the utility and accuracy of satellite data, there is a critical need to expand ground-based measurement networks *(e.g.,* AERONET and PANDONIA), particularly in LMICs, to support robust validation of satellite-derived products. Additionally, the current predominance of geo-stationary satellite missions in the Northern Hemisphere highlights a geographical disparity that limits their applicability to other regions [9]. International multilateral agencies should consider funding projects aimed at expanding this coverage, which would benefit the entire world, especially considering that entangled issues like air pollution and climate change require a global perspective. Existing global networks can be leveraged to increase capacity (*e.g.,* enabling local managers and citizen science groups to effectively use its data).

Simple and accessible tools are needed to bring together multiple sources of information and easily make appropriate comparisons between them. Co-development of these tools between subject matter experts and data users is essential to ensure they provide trustworthy, actionable information. The tools themselves, as well as their development, should follow open science principles.

KEY REFERENCES

World Meteorological Organization (2024) https://library.wmo.int/idurl/4/68924

World Health Organization (2023) Overview of methods to assess population exposure to ambient air pollution https://www.who.int/publications/i/item/9789240073494

Van Beveren R (2024) Environmental decision makers have a data plumbing problem. Data intermediaries can help. https://www.policyinnovation.org/blog/dataintermediaries?mc_cid=54f8893a22.

Duncan BN, Malings CA, Knowland KE, Anderson DC, Prados AI, Keller CA, Cromar KR, Pawson S, Ensz H (2021) Augmenting the Standard Operating Procedures of Health and Air Quality Stakeholders With NASA Resources. GeoHealth 5:e2021GH000451

United States Government Accountability Office (2024) Air Quality Sensors: Policy Options to Help Address Implementation Challenges.

STATEMENTS AND DECLARATIONS

Funding. CMFR's contribution to this work is supported by OpenAQ, a nonprofit organization working on open air quality data. OpenAQ's primary source of funding is philanthropic organizations, supplemented by individual donations, corporate sponsorships, and subcontracts funded by public and private partners. Open access support is provided by the University of California Davis Libraries.

JRB was supported by the NASA Grant 80NSSC21K0427 for the NASA Health and Air Quality Applied Sciences Team (HAQAST).

Sebastian Diez is supported by the Agencia Nacional de Investigación y Desarrollo de Chile (ANID) through FONDECYT (Grant 1221951).

Sara Duncan is employed by Western Carolina University and is not supported by grant funding directly related to this project.

CM is supported by the NASA Health and Air Quality Program (Grant 80NSSC22K1473).

PP is employed by the Health Effects Institute which receives funding from the United States Environmental Protection Agency and the worldwide motor vehicle industry as well as philanthropic organizations. The views expressed here are those of the authors and do not necessarily reflect the views of the Health Effects Institute, or its sponsors. PP also serves as the Chair of the Governing Board of OpenAQ.

Human and Animal Rights and Informed Consent. This article does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

1. World Health Organization (2024) Ambient (outdoor) air pollution. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.

2. Health Effects Institute (2024) State of Global Air 2024. Boston, MA.

3. Li C, Donkelaar A, Hammer MS, et al (2023) Reversal of trends in global fine particulate matter air pollution. Nat Commun 14:1–12

4. Greenstone M, Ganguly T, Hasenkopf C, Sharma N, Gautam H (2024) Air Quality Life Index 2024 Annual Update.

5. United Nations Environment Programme (2023) Guide on Ambient Air Quality Legislation - Air Pollution Series.

6. United Nations Environment Programme (2021) Regulating Air Quality: The First Global Assessment of Air Pollution Legislation - Air Pollution Series.

7. World Health Organization Air Quality Standards database. https://www.who.int/tools/air-quality-standards.

8. Kerr GH, Goldberg DL, Harris MH, Henderson BH, Hystad P, Roy A, Anenberg SC (2023) Ethnoracial Disparities in Nitrogen Dioxide Pollution in the United States: Comparing Data Sets from Satellites, Models, and Monitors. Environ Sci Technol 57:19532–19544

9. Buonocore JJ, Fisher J, Prull D, Willis MD, Arunachalam S, Perera F, Kinney P, Sousa B, Levy JI (2025) Federal Policy Platforms and Public Health: Reinforcing the Benefits of Air Pollution Control Devices at Power Plants in the United States. Am J Public Health 115:30–33

10. World Health Organization (2021) WHO global air quality guidelines. Particulate matter (PM2.5 and PM 10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide.

11. World Bank Group (2025) Open Data Toolkit: Open Data in 60 Seconds. https://opendatatoolkit.worldbank.org/en/data/opendatatoolkit/open-data-in-60-seconds.

12. Open Knowledge Foundation Open Definition 2.1. https://opendefinition.org/od/2.1/en/.

13. Rosales CMF, Hagerbaumer C, Sawant V, Marte J, Biggs R (2024) Open Air Quality Data: The Global Landscape. Washington, D.C

14. Sawant V, Hagerbaumer C, Rosales CMF, Isied M, Biggs R (2022) 2022 Open Air Quality Data: The Global Landscape.

15. Guttikunda SK (2024) Designating Airsheds in India for Urban and Regional Air Quality Management. Air 2:247–257

16. (2024) United States Environmental Protection Agency. https://www.federalregister.gov/documents/2024/03/06/2024-02637/reconsideration-of-the-national-ambient-air-quality-standards-for-particulate-matter

17. Silva CM da, Moreira Junior DP, Rodrigues JRB de A, Siciliano B, Arbilla G (2023) Clean air for a good start: Children are the future of the planet. Ambiente Soc 26:00041

18. Tran M (2024) AI Air Quality Forecasting: A State-NASA Partnership. Foreign Serv. J.

19. Chowdhury S, Pillarisetti A, Oberholzer A, Jetter J, Mitchell J, Cappuccilli E, Aamaas B, Aunan K, Pozzer A, Alexander D (2023) A global review of the state of the evidence of

household air pollution's contribution to ambient fine particulate matter and their related health impacts. Environ Int 173:107835

20. Clean Air Fund (2024) The State of Global Air Quality Funding 2024.

21. Our Common Air Commission (2024) Accelerating Country-led Air Quality Reporting to Achieve Clean Air.

22. Seddon J, Verhulst SG, Maron A From Knowing to Doing: Operationalizing the 100 Questions for Air Quality Initiative.

23. American Lung Association (2024) Something in the Air: Bridging the Air Quality Gap with Satellite Technology - A State of the Air Supplemental Report.

24. Wang Y, Marshall JD, Apte JS (2024) U.S. Ambient Air Monitoring Network Has Inadequate Coverage under New PM2.5 Standard. Environ Sci Technol Lett 11:1220–1226

25. Ivey CE, Amegah AK, Hodoli CG, et al (2022) Academic Incentives May Hamper Public Good. Env Sci Technol. https://doi.org/10.1021/acs.est.2c05721.

26. Gani S (2022) Environ. Res Lett. https://doi.org/10.1088/1748-9326/ac4a9e

27. Wright CY, Benyon M, Mahlangeni N, Kapwata T, Laban T, Garland RM (2023) Data gaps will leave scientists "in the dark": How load shedding is obscuring our understanding of air quality. South Afr J Sci 119:1–5

28. Saeed T, Khaliq MM, Bergin MH, et al (2024) Sustaining Low-Cost PM2.5 Monitoring Networks in South Asia. https://doi.org/10.5194/egusphere-2024-1932

29. Oguge NO (2025) Personal communication.

30. World Meteorological Organization (2024) Integrating Low-cost Sensor Systems and Networks to Enhance Air Quality Applications. https://library.wmo.int/idurl/4/68924

31. EANET (2024) Progress Report of the EANET Core Activities Since the Twenty-Fifth Session of the Intergovernmental Meeting (IG25) on the EANET. Kuala Lumpur, Malaysia

32. Laj P, Lund Myhre C, Riffault V, et al (2024) Aerosol, Clouds and Trace Gases Research Infrastructure (ACTRIS): The European Research Infrastructure Supporting Atmospheric Science. Bull Am Meteorol Soc 105:E1098–E1136

33. Liu X, Turner JR, Oxford CR, et al (2024) Elemental Characterization of Ambient Particulate Matter for a Globally Distributed Monitoring Network: Methodology and Implications. ACS EST Air 1:283–293

34. Antuña-Marrero JC, Landulfo E, Estevan R, et al (2017) LALINET: The First Latin American–Born Regional Atmospheric Observational Network. https://doi.org/10.1175/BAMS-D-15-00228.1

35. Samet JM, Dominici F, Curriero FC, Coursac I, Zeger SL (2000) Fine Particulate Air Pollution and Mortality in 20 U.S. Cities, 1987–1994. N Engl J Med 343:1742–1749

36. Pope III CA, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston GD (2002) Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution. JAMA 287:1132–1141

37. Kan H, London SJ, Chen G, Zhang Y, Song G, Zhao N, Jiang L, Chen B (2007) Differentiating the effects of fine and coarse particles on daily mortality in Shanghai, China. Environ Int 33:376–384

38. Liu C, Chen R, Sera F, et al (2019) Ambient Particulate Air Pollution and Daily Mortality in 652 Cities. N Engl J Med 381:705–715

39. Hayes L, Mejia-Arangure JM, Errington A, et al (2024) Relationship between air quality and asthma-related emergency hospital admissions in Mexico City 2017–2019. Thorax 79:43–49

40. Luong LTM, Dang TN, Thanh Huong NT, Phung D, Tran LK, Van Dung D, Thai PK (2020) Particulate air pollution in Ho Chi Minh city and risk of hospital admission for acute lower respiratory infection (ALRI) among young children. Environ Pollut 257:113424

41. Dong J, Wang Y, Wang J, Bao H (2021) Association between atmospheric PM2.5 and daily outpatient visits for children's respiratory diseases in Lanzhou. Int J Biometeorol 65:989–999

42. Wu T, Yang X, Chu A, Xie X, Bai M, Peng Y, Zhang Z (2022) Acute effects of fine particulate matter (PM2.5) on hospital admissions for cardiovascular diseases in Lanzhou, China: a time-series study. Environ Sci Eur 34:55

43. Checkley W (2023) Tuberculosis, Wildfires, and Case-crossover Studies: An Epidemiological Trifecta? Am J Respir Crit Care Med 207:242–243

44. Ciciretti R, Barraza F, De la Barrera F, Urquieta L, Cortes S (2022) Relationship between Wildfire Smoke and Children's Respiratory Health in the Metropolitan Cities of Central-Chile. Atmosphere 13:58

45. Duncan BN, Malings CA, Knowland KE, Anderson DC, Prados AI, Keller CA, Cromar KR, Pawson S, Ensz H (2021) Augmenting the Standard Operating Procedures of Health and Air Quality Stakeholders With NASA Resources. GeoHealth 5:e2021GH000451

46. Liu F, Beirle S, Joiner J, et al (2024) High-resolution mapping of nitrogen oxide emissions in large US cities from TROPOMI retrievals of tropospheric nitrogen dioxide columns. Atmospheric Chem Phys 24:3717–3728

47. Geddes JA, Murphy JG, O'Brien JM, Celarier EA (2012) Biases in long-term NO2 averages inferred from satellite observations due to cloud selection criteria. Remote Sens Environ 124:210–216

48. Shen S, Li C, van Donkelaar A, Jacobs N, Wang C, Martin RV (2024) Enhancing Global Estimation of Fine Particulate Matter Concentrations by Including Geophysical a Priori Information in Deep Learning. ACS EST Air 1:332–345

49. Chambliss SE, Campmier MJ, Audirac M, Apte JS, Zigler CM (2024) Local exposure misclassification in national models: Relationships with urban infrastructure and

demographics. J Expo Sci Environ Epidemiol 34:761–769

50. Makkaroon P, Tong DQ, Li Y, et al (2023) Development and Evaluation of a North America Ensemble Wildfire Air Quality Forecast: Initial Application to the 2020 Western United States "Gigafire." J Geophys Res Atmospheres 128:e2022JD037298

51. McDuffie EE, Martin RV, Spadaro JV, et al (2021) Source sector and fuel contributions to ambient PM2.5 and attributable mortality across multiple spatial scales. Nat Commun 12:1–12

52. Diez S, Lacy S, Pete E, Read K, Josefina U (2024) QUANT: A Three-Year. Multi-City Air Qual Dataset Commer Air Sens Ref Data Perform Eval Data Set Zenodo. https://doi.org/10.5281/zenodo.10775692

53. Kawano A, Kelp M, Qiu M, Singh K, Chaturvedi E, Azevedo I, Burke M (2024) High-quality daily PM2.5 datasets at a 10 km resolution for India. https://doi.org/10.5281/zenodo.10807119

54. Xie Y (2024) Data for "Recent improvements in PM2.5 air quality in India benefited from meteorological variation." https://doi.org/10.34770/xtje-mj26

55. Van Poppel M, Schneider P, Peters J, et al (2023) SensEURCity: A multi-city air quality dataset collected for 2020/2021 using open low-cost sensor systems. Sci Data 10:322

56. Lacy S, Diez S, Edwards P (2023) Quantification of Utility of Atmospheric Network Technologies: (QUANT): Low-cost air quality measurements from 52 commerical devices at three UK urban monitoring sites. CEDA

57. Schröder S, Schultz MG, Selke N, et al (2021) TOAR Data Infrastructure. FZ-Juelich. https://doi.org/10.34730/4D9A287DEC0B42F1AA6D244DE8F19EB3

58. Mills G, Pleijel H, Malley CS, et al (2018) Tropospheric Ozone Assessment Report: Present-day tropospheric ozone distribution and trends relevant to vegetation. Elem Sci Anthr. https://doi.org/10.1525/elementa.302

59. Tarasick D, Galbally IE, Cooper OR, et al (2019) Tropospheric Ozone Assessment Report: Tropospheric ozone from 1877 to 2016, observed levels, trends and uncertainties. Elem Sci Anthr. https://doi.org/10.1525/elementa.376

60. Data and reporting - European Commission. https://environment.ec.europa.eu/topics/air/air-quality/data-and-reporting_en. Accessed 28 Feb 2025

61. The University of Chicago Energy Policy Institute EPIC Air Quality Fund: Applications. In: EPIC Air Qual. Fund Appl. https://epic.uchicago.edu/epic-air-quality-fund-applications/. Accessed 28 Feb 2025

62. Hasenkopf C (2025) "Open-ready" Air Quality Policies (LinkedIn post). In: LinkedIn. https://www.linkedin.com/posts/chrishasenkopf_at-the-close-of-the-world-health-organizations-activity-7315329878738796544-u_s3/. Accessed 29 May 2025

63. Yang Z (2025) US State Department Kills Global Air Monitoring Program Researchers Say Paid for Itself. Wired

64. Jha A, Nauze AL (2022) US Embassy air-quality tweets led to global health benefits. Proc Natl Acad Sci 119:e2201092119

65. CREA (2025) U.S. embassy AQ data halt: Ramped up government monitoring could save hundreds of millions of dollars per city. Cent. Res. Energy Clean Air

66. Van Beveren R (2024) Environmental decision makers have a data plumbing problem. Data intermediaries can help. https://www.policyinnovation.org/blog/dataintermediaries?mc_cid=54f8893a22.

67. Stamp S, Burman E, Chatzidiakou L, Cooper E, Wang Y, Mumovic D (2022) A critical evaluation of the dynamic nature of indoor-outdoor air quality ratios. Atmos Environ 273:118955

68. Pillarisetti A, Ye W, Chowdhury S (2022) Indoor Air Pollution and Health: Bridging Perspectives from Developing and Developed Countries. Annu Rev Environ Resour 47(Volume 47:197–229

69. United States Government Accountability Office (2024) Technology Assessment: Air Quality Sensors. Policy Options to Help Address Implementation Challenges.