Commentary
To combat air inequality, governments and researchers must open their data

Christa A. Hasenkopf*[1], David C. Adukpo[2], Michael Brauer[3], H. Langley Dewitt[4], Sarath Guttikunda[5], Alaa I. Ibrahim[6], Delgerzul Lodoisamba[7], Nyambura Mutanyi[8], Gustavo Olivares[9], Pallavi Pant[10], Maëlle Salmon[11], and Lodoysamba Sereeter[12]

[1]OpenAQ, Washington, DC, USA
[2]Department of Physics, School of Physical Sciences, University of Cape Coast, Cape Coast, Ghana
[3]The University of British Columbia, Vancouver, BC, Canada
[4]MIT Center for Global Change Science and Rwanda Climate Observatory, Kigali, Rwanda
[5]Urban Emissions, India
[6]Zewail City of Science and Technology; and American University in Cairo, Cairo, Egypt
[7]Department of Environmental Health, School of Public Health, Mongolian National University of Medical, Ulaanbaatar, Mongolia
[8]Kenya Open Data Initiative, Nairobi, Kenya,
[9]National Institute of Water and Atmospheric Research, Auckland, New Zealand
[10]University of Massachusetts, Amherst, Amherst, Massachusetts, USA
[11]ISGlobal, Centre for Research in Environmental Epidemiology,
Universitat Pompeu Fabra, CIBER Epidemiología y Salud Pública, Barcelona, Spain
[12]German-Mongolian Institute for Resources and Technology, Nalaikh, Mongolia
*Corresponding Author: christa@openaq.org

This piece was drafted and endorsed by members of the OpenAQ Community. In addition to the authors listed above, substantive contributions were also made by those who are not able to publicly be affiliated with this statement due to their organizational constraints. After the first author, publicly-listed authors are shown alphabetically. The views expressed in this piece are the authors' and OpenAQ's but do not necessarily express the views of other organizations.

http://dx.doi.org/10.17159/2410-972X/2016/v26n2a5

Why open air quality data matters

Air pollution data measured by governments across the world are a public good that can lead to transformational advances in public health when made openly available. Such advances are needed because, according to the WHO, one out of every eight deaths in the world is due to air pollution (WHO 2014). These deaths disproportionately occur in high population density, lower income countries (WHO 2016), which, where data are available, often correspond to regions with higher long-term levels of ambient pollution (Figure 1).

Our community of scientists, software developers, journalists and open environmental data enthusiasts, OpenAQ (openaq.org), terms this unequal access to clean air to breathe across the world as 'air inequality.' While the health statistics are dire, the tools to fight air inequality abound: decades-worth of policy solutions achieved in cities and regions around the world serve as powerful examples. Yet, often there is a gap between the problem and the implementation of effective policy solutions. Government-collected data, when shared openly and in a timely manner, can bridge this gap by galvanizing communities, guiding effective data-driven mitigation policies and external, independent analyses, and helping answer remaining public health and economic policy-relevant science questions. It can also open up entirely new economic potentials, similar to how access to weather data spurred public and private innovations a few decades ago.

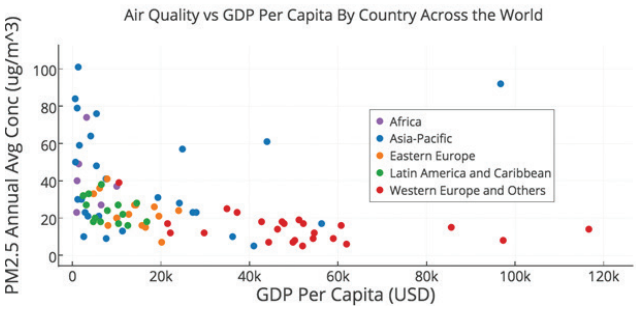

Figure 1: *$PM_{2.5}$ annual concentrations on a country-level basis versus GDP per capita of each country, grouped by region. The three outliers from the Asia-Pacific are, from left to right, Bahrain, UAE and Qatar. $PM_{2.5}$ data is from the 2014 WHO Outdoor Air Pollution Database (WHO Database 2014). The country GDP per capita data are from 2014 World Bank data (World Bank 2015). The regional categories are based on UN National Regional Groups Definitions. View the interactive graphic here: http://tinyurl.com/GDP-PM25-country*

For an example on an existing policy-relevant science question that could be answered with the help of open data, take the typical dose-response curve for relative risk of mortality and ambient $PM_{2.5}$ concentration exposure (Figure 2, adapted from Burnett, et al. 2014). The scientific community's understanding of the impact of $PM_{2.5}$ exposure on health is well-defined at low-level exposures. This understanding is based on a robust research inventory of large-cohort epidemiological studies conducted in the U.S. and E.U. over several decades. Open air

quality data supplied by governments played a pivotal role in these research studies; 7 out of 10 of the most highly-cited air pollution-related epidemiological health studies used government open data (OpenAQ 2016). These studies were also not just science for science's sake; they drove national level air quality standards policymaking in the U.S., as well influenced international guidelines.

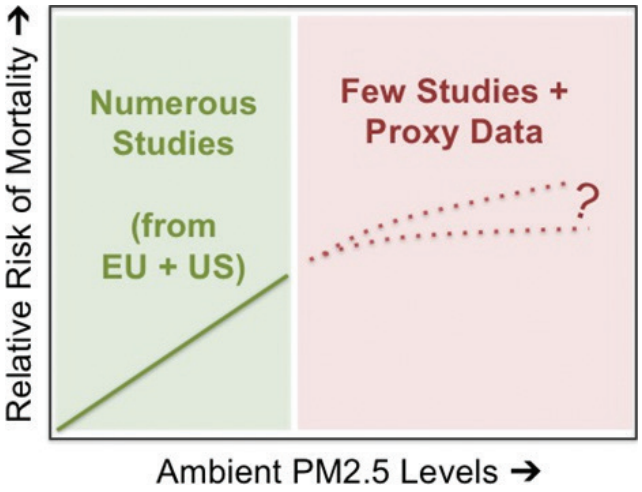

Figure 2: *A simplistic and generalized $PM_{2.5}$ dose response curve, adapted from figures in Burnett et al. (2014). The schematic depicts the gap in scientific understanding of the relationship between relative risk of mortality and high $PM_{2.5}$ levels, relative to the U.S. and E.U.*

Meanwhile, there are far fewer studies in places where ambient $PM_{2.5}$ levels are significantly higher, such as throughout much of Asia and Africa, and proxy studies on second-hand smoke or indoor air exposure are used instead to estimate health impact. Yet, the exact shape of that dose-response curve at high levels of $PM_{2.5}$ has serious policy implications affecting the health and economies of countries across the world and billions of people. With existing available data, UNICEF recently estimated that 520 million children in Africa breathe outdoor air that does not meet minimum air quality guidelines advised by the World Health Organization (UNICEF 2016). The World Bank estimates that annual labor income losses will cost Sub-Saharan African countries the equivalent of 0.61% of GDP (World Bank 2016).

On aqicn.org, which shows global real-time maps of ground-based, government-measured air quality data, there are only a handful of stations reporting data in Africa - a couple in Addis Ababa, Ethiopia and a couple outside Durban, South Africa. While more government-run and research-grade stations do exist across Africa and more sources emerging each year, this data is neither online nor accessible. A key problem is that there is no default means for the producers of these data to share. This warrants a universal open air quality data platform, in order to maximize the societal, economic, and health benefits of these data.

In the case of Africa, the dearth of air quality data makes existing data anywhere on the continent all the more valuable and the need of an open, universal data-sharing platform all the more urgent.

How to truly open up air quality data

National governments are in singular positions to build and maintain continuous air quality networks, as well as to make such valuable datasets open to the public. While approximately 70 countries across the world share real-time air quality data in some form, many do not yet do this in the ways that maximize the utility of the data to the public. To do this, governments must allow access to real-time physical data (e.g. an open Application Program Interface or API - a way that allows one computer system to easily communicate with another - or even files sitting on an FTP server) and related documentation. Currently, it is too often the case that only air quality index-type data are available on a screen as a temporarily-existing graph or table. Some nations report air quality data at several sub-national levels on different sites, which can leave the true national air quality landscape inscrutable to the public. Researchers also generate scores of data that could be used for other applications besides a single study's objective.

An even more powerful step beyond national-level or research-grade open air quality data availability is a global platform that transparently aggregates and freely shares air quality data in a universal format and in a programmatic manner. Providing data in such a way allows various stakeholders - from scientists to journalists to policymakers - to build universal tools, and enables open international collaborations. An example from a different field is OpenStreetMap, a collaboratively edited map powered by an open-source ecosystem, which has become a foundational data layer for a diverse set of humanitarian, government, and other public and private sector efforts. A universal air quality platform of this type could also provide a repository and structure for quickly emerging low-cost sensor technologies that promise to soon generate unprecedented amounts of air quality data.

While we saw several private-sector, academic, and informal public sector entities aggregating real-time government-level air quality data for various purposes, no group was sharing their methodology openly, nor, most importantly, freely sharing the physical data in real-time and historically. To fill this gap, over the past year, our global and grassroots community of software developers, scientists, journalists and open environmental data enthusiasts began building the world's first global platform that automatically aggregates and freely shares real-time and historical air quality data. To date, we have aggregated more than 27 million air quality data points from 25 countries, much of which would be impossible or difficult to access otherwise. Members of our community have been using these data to create real-time air quality apps, open source packages in the statistical language R (github.com/ropenscilabs/ropenaq) and python (github.com/dhhagan/py-openaq), data-driven articles for the media, and more. The existence of this system can also facilitate the development and implementation of low-cost data handling standards by providing an outlet for air quality data.
We welcome contributions of new data sources from any government-level or research-grade entities to our platform, especially those available through an automated, centralized system.

Data sources from Africa are especially welcome, since we currently only have two sources: one real-time source from Addis Ababa, Ethiopia and another non-real-time source from Ilorin, Nigeria.

However, knowledge of or access to a new real-time data source is only half the battle for our community; we rely on volunteers with the software development skills necessary to build the pieces of code that connect a given data resource to our platform. We are always seeking more individuals with these skills to contribute to our open-source platform. To ease the burden of individuals to contribute historical data, in December 2016, we anticipate launching a new portal on openaq.org that will allow researchers and governments to upload historical data to share on our platform.

We invite all other forms of collaboration and input that help shape our open-source project (github.com/openaq) to best meet the air pollution and health community's needs. Lastly, we especially welcome collaborations with intergovernmental entities, international non-governmental organizations, and international aid agencies with missions that overlap with ours to improve public health and the environment.

The OpenAQ Community firmly believes that a universal open air quality platform enables powerful, collaborative ways for local and national governmental bodies, intergovernmental organizations, international non-governmental organizations, academic groups, and the general public to fight air inequality.

References

Burnett, R.T., C. A. Pope III, M. Ezzati, C. Olives, S. S. Lim, S. Mehta, H. H. Shin, G. Singh, B. Hubbell, M. Brauer, H. R. Anderson, K. R. Smith, J. R. Balmes, N. G. Bruce, H. Kan, F. Laden, A. Prüss-Ustün, M. C. Turner, S. M. Gapstur, W. R. Diver, and A. Cohen. (2014). An Integrated Risk Function for Estimating the Global Burden of Disease Attributable to Ambient Fine Particulate Matter Exposure. *Environmental Health Perspectives*, 122(4), pp. 397-403.

OpenAQ (2016). Open air quality data packs a powerful punch. *Blog*, Available at: https://medium.com/@openaq/open-air-quality-data-packs-a-powerful-punch-38ca9462910b#.6n36axj8b [Accessed: 7 Nov. 2016]

WHO (2014). 7 million premature deaths annually linked to air pollution. *News Release*, Available at: http://www.who.int/mediacentre/news/releases/2014/air-pollution/en/ [Accessed: 7 Nov. 2016]

WHO Database (2014). Ambient (outdoor) air pollution in cities database 2014.*Database*, Available at: http://www.who.int/phe/health_topics/outdoorair/databases/cities-2014/en/ [Accessed: 7 Nov. 2016]

WHO (2016). Ambient (outdoor) air quality and health. *Fact Sheet*, Available at: http://www.who.int/mediacentre/factsheets/fs313/en/ [Accessed: 7 Nov. 2016]

World Bank (2014). Available at: http://databank.worldbank.org/data/reports.aspx?source=2&series=NY.GDP.PCAP.PP.CD&country= [Accessed: 7 Nov. 2016]

World Bank (2016). Air Pollution Deaths Cost Global Economy US$225 Billion. *Press Release*, Available at: http://www.worldbank.org/en/news/press-release/2016/09/08/air-pollution-deaths-cost-global-economy-225-billion [Accessed: 7 Nov. 2016]

UN National Regional Groups Definitions. Available at: http://www.un.org/depts/DGACM/RegionalGroups.shtml [Accessed: 7 Nov. 2016]

UNICEF (2016). 300 million children breathing toxic air - UNICEF report. *Press Centre*, http://www.unicef.org/media/media_92979.html [Accessed: 7 Nov. 2016]