Visual Unearthing Project

RESEARCH REPORT

Michael P. Canares, Step Up Consulting

WITH CONTRIBUTIONS FROM

Glenn Maail, Ben Hur Pintor, Peck Sangiambut, Ngo Hai Ly, Mary Therese Pepito, and **Jean Celeste Paredes**

This report on the “Visual Unearthing” project is commissioned
by the Goethe-Institut Indonesien with conceptual support
from Nana Karlstetter.

Table of Contents

List of Tables

List of Figures

1. Introduction

1a. Background

The regional project "visible unearthing", implemented by Goethe Institut aimed to use open data to analyze the interactions of air-water quality and other indicators (groundwater level, etc.) that are important, especially in climate change in very specific environments (cities, regions, ecosystems). As an important part of the process, a data inventory was conducted to identify the datasets that can capture a condition of interest and visualize it in ways that could generate meaningful discussions.

With COVID-19 impacting the Southeast Asian region and globally, the initial plan was to look at environmental data with a certain level of relationship with COVID 19. Given that restrictions in movement have significantly impacted

mobility during lockdowns, and with transportation as one of the identified contributors of air quality (EPA 2019), the main focus of the assessment was the availability of open air quality data.

The data inventory, and corresponding visualization afterwards, was guided by the following key parameters:

a. Geographic – the data collection was confined to the following cities – Jakarta (Indonesia), Hanoi/Ho Chi Minh (Vietnam), Manila (Philippines), and Bangkok (Thailand)
b. Data Focus – the focus was environmental data, primarily air quality, and other related datasets such as water quality, pollution, solid and liquid waste, among others.
c. Time frame – as a comparison of context or situation is desired, at least two data points will be gathered (2019 and a period after) for the data to be visualized.
d. Open data – The engagement would use open data available. The openness of the data (its availability, freedom to use, reuse, and redistribute) would determine the data sets subjected to visualization.

1b. Data Collection Process

The data collection process consisted of two steps:

Step 1: Inventory of Relevant Data Assets

For each country, a country researcher was appointed.

The country researcher conducted a data inventory of environment-related data, as follows:

	Data Set	Definition used in this research
1	Urban settlements area	Percentage of area with urban settlements
2	Urban population	Number of people living in an urban area, spatially identified, when possible, with computations of population density
3	Population growth	Change in the size of the population from year-to-year
4	Water consumption data	Average water consumed by a person per period (day/month/year)
5	Total volume of untreated sewage	Volume of sewage disposed of in a given period (day/month/year)
6	Water quality data	Percentage of freshwater with concentrations of faecal material bacteria above the levels recommended by WHO (% of samples analyzed with results of faecal material; number of days when WHO standards are exceeded)
7	Modes of transportation data	Number of private cars, trains and trolleybuses, buses and microbuses and non-motorized vehicles
8	Motorization index	Number of cars per inhabitant
9	Air quality data	Air quality data is dependent on the standards set by a country
10	Volume of solid waste generated	Total solid waste produced over a particular period (day/month/year)
11	Volume of liquid waste generated	Total liquid waste produced over a particular period (day/month/year)

Table 1. Data Sets Assessed in Each City

At least five of these datasets pertain directly to the quality of urban environment (volume of untreated sewage, water quality data, air quality data, volume of solid waste generated, volume of liquid waste generated), while the rest can be considered as contextual urban indicators data (e.g. population, water consumption).

The inventory was not just limited to assessing which data was available and which are not. Each data was assessed using the following questions.

a. Does the data exist? (5 pts)

b. Is it available online from the government in any form? (10 pts)
c. Is the dataset provided in machine-readable formats? (15 pts)
d. Is the machine-readable data available in bulk? (15 pts)
e. Is the dataset available free of charge? (15 pts)
f. Is the data openly licensed? (15 pts)
g. Is the dataset up to date? (10 pts)
h. Is the publication of the dataset sustainable? (5 pts)
i. Was it easy to find information about this dataset? (5 pts)
j. Are (linked) data URIs provided for key elements of the data? (5 pts)

A perfect score of 100 meant that the data passes the criteria of availability, machine-readability, licensing, recency, ease of use, and sustainability of publication.

Step 2: Collection of Air Quality Data

As earlier stated, there are two data assets that we hope to collect in this phase:

a. PRIORITY: air quality data
b. Another set of data peculiar to each city based on the inventory results or suggested by the researcher. This was, however, foregone at a later stage, in agreement with Goethe Institut.

Country researchers collected the data on air quality and transformed the collected data into CSV. At the minimum, three data time periods will be required for collection:

a. a date and time in 2019
b. a date and time in 2020
c. a date and time in 2021

Ideally, the three time periods should be similar. For example, if April 1, 2019, is the first time period, the succeeding should be April 1, 2020, and April 1, 2021. However, if data is available daily, then data should be collected from March 2019 to the last day of fieldwork.

For each data collected, a metadata and methodological note are necessary.

a. The following metadata were necessary for each file collected:
    a. filename/title
    b. subject
    c. type
    d. date created
    e. source
    f. url source
    g. publisher/author/provider
    h. tabnames and explanation
    i. column names and explanation
b. The methodological note included:
    a. Sources of raw data used – if data is not raw, indicate the source of the raw data used.
    b. Manner of collection – how researchers collected the data.
    c. Computational elements – if data is a result of a formulaic computation, discuss how the values were arrived at.
    d. Basis of computation – discuss the basis used for arriving at the formula (e.g. UNEP standards, WHO standards).
    e. Author – identify who generated the computational elements of the data.

Detailed methodological notes are contained in the annexes.

1c. Data Visualization

Using the data gathered from the immediately preceding steps, sample data visualizations were generated for each city, focusing on air quality data. Different processes were undertaken to answer the question – "What is the impact of COVID-19 restrictions on air quality in the four cities covered by the study".

1d. Research Team

The research team is composed of the following:

- Michael P. Canares - Research Lead
- Glenn Maail - Researcher, Jakarta
- Ben Hur Pintor - Researcher, Manila
- Peck Sangiambut - Researcher, Bangkok
- Ngo Hai Ly - Researcher, Hanoi
- Mary Therese Pepito - Research Assistant
- Jean Celeste Paredes - Research Assistant

2. Individual City Results from the Data Inventory

2a. Manila (Philippines)

2a1. Status of Data and Data Sources

There aren't any open, machine-readable air quality data available from the government of the Philippines or the corresponding subnational units. Air quality data—e.g. TSP, PM2.5, PM10—is available as annual average values (2012-2019) in a PowerBI dashboard (https://emb.gov.ph/air-quality-data/) and not downloadable as raw values. A request via the eFOI portal or a partnership letter addressed to the government agency mandated to collect the data is needed to access the raw data.

However, air quality data is collected from monitoring stations by home enthusiasts and air quality professionals and curated by websites/projects such as OpenAQ and PurpleAir. The data from these monitoring stations are readily

available online and can be downloaded in machine-readable formats. The data obtained from these sensors is limited by the length of time that the sensors have been operational or sharing their data with the network.

Population, demographic, and settlement data are available from government and non-government sources. The Philippine Statistics Authority releases demographic and census data. Another good source of data is the Humanitarian Data Exchange.

As to the secondary data, the following are readily available online:

1. Water quality
    a. from the Environmental Management Bureau
    b. https://emb.gov.ph/water-quality-management-data/
2. Modes of transportation
    a. from the Land Transportation Office
    b. https://lto.gov.ph/transparency-seal/annual-reports.html
3. Volume of solid waste generated
    a. from the Metropolitan Manila Development Agency, 2018
    b. https://data.gov.ph/?q=dataset/solid-waste-management-metro-manila-2018

The following are not readily available online:

1. Total volume of untreated sewage
2. Volume of liquid waste generated

Meanwhile, the following data, although not readily available, can be computed from existing data:

1. Water consumption data
    a. No direct water consumption data is available, but it can be computed from secondary datasets found in some MWSS reports such as:
        i. Population served by water concessionaires

ii. Total production of water by water concessionaires

2. Motorization index
   a. The data does not exist as defined but can be computed from the registered MV (modes of transportation) data and the population data

2a2. Results of Data Scoring

Data	Source	a	b	c	d	e	f	g	h	i	j	Score
Urban settlements area	Philippine Statistics Authority	y	y	y	n	y	n	y	y	y	n	65
	High-Resolution Settlement Layer	y	n	y	y	y	y	y	y	y	n	70
Urban population (spatial)	Philippine Statistics Authority	y	y	y	n	y	n	y	y	y	n	65
	High-Resolution Settlement Layer	y	n	y	y	y	y	y	y	y	n	70
Population growth	Philippine Statistics Authority	y	y	y	n	y	n	y	y	y	n	65
Water consumption data		n	n	n	n	n	n	n	n	n	n	0
Total volume of untreated sewage		n	n	n	n	n	n	n	n	n	n	0
Water quality data	Environmental Management Bureau	y	y	n	n	y	n	y	n	n	n	15
Modes of transportation	Land Transportation Office	y	y	n	n	y	n	y	y	n	n	15
Motorization index		n	n	n	n	n	n	n	n	n	n	0

Data	Source	a	b	c	d	e	f	g	h	i	j	Score
Air quality data	Environmental Management Bureau	y	y	n	n	y	n	y	y	n	n	15
	IQAir	y	n	n	n	y	n	y	y	y	n	5
	AQICN	y	n	n	n	y	n	y	y	y	n	5
	OpenAQ	y	n	y	y	y	y	y	y	y	n	70
	PurpleAir	y	n	y	y	y	y	y	y	y	n	70
Volume of solid waste generated	Metropolitan Manila Development Agency	y	y	y	n	y	y	n	n	n	n	55
	Environmental Management Bureau	y	y	n	n	y	n	y	n	n	n	15
Volume of liquid waste generated		n	n	n	n	n	n	n	n	n	n	0

Table 2. Data Openness Score - Manila

The data scores would show that most of the contextual urban data indicators are available and accessible. In contrast, only one of the data directly related to urban environmental conditions is available and accessible – air quality data- largely from agencies outside of government. It is interesting to note that critical data on waste generation is not available publicly.

2b. Jakarta (Indonesia)

2b1. Status of Data and Data Sources

The key environmental datasets are generally available, although the level of granularity could be further improved. Nine out of the eleven datasets assessed are available online. The datasets are mostly made available on the website of the national statistical office (BPS Indonesia). Others can be accessed from the Ministry of Environment and Forestry of the Republic of Indonesia's website and the Jakarta Open Data portal.

However, there are some limitations with the published data. With the exemption of the air quality data, the datasets are available only every year. The level of granularity for daily or monthly data as required by the assessment guideline is not yet available. For example, the water consumption data, total volume of untreated sewage, volume of solid waste generated, and volume of liquid waste generated are only available annually.

All datasets are available free of charge, although licensing rules for some datasets are often unclear. Since the datasets are all available from government institutions, they are available free of charge. The terms and conditions available on the BPS' website stated that the information available on the website could be used freely (including for commercial purposes), copied, translated, or modified on terms and provisions based on the terms and conditions referencing rules. The remaining datasets are available on the Ministry of Environment's website. However, there are no details of charging regimes or licensing rules available on the website.

The available data is kept up-to-date, and there is also an indication that the publication is sustainable. For the annual data, the data is available at least until 2019. For daily data such as the air quality data, the data is available as recent as May 2021. The publication of data on the website appears to be well maintained since the publication of the environmental data is also a part of an annual report of Environment Statistics of Indonesia co-produced between BPS Indonesia and the Ministry of Environment and Forestry. Finally, most data can be found using a simple keyword search or by visiting the BPS or the Ministry of Environment and Forestry's website.

2b2. Results of Data Scoring

Data	Source	a	b	c	d	e	f	g	h	i	j	Score
Urban settlements area		n	n	n	n	n	n	n	n	n	n	0

Data	Source	a	b	c	d	e	f	g	h	i	j	Score
Urban population (spatial)	National Statistics Office	y	y	y	n	y	y	n	n	y	n	65
Population growth	National Statistics Office	y	y	y	y	y	y	y	y	y	n	95
Water consumption data	National Statistics Office	y	y	y	y	y	y	y	y	y	n	95
Total volume of untreated sewage	National Statistics Office	y	y	n	n	y	y	y	y	n	n	60
Water quality data	National Statistics Office	y	y	n	n	y	y	y	y	y	n	65
Modes of transportation	National Statistics Office	y	y	n	n	y	n	y	y	n	n	45
Motorization index		n	n	n	n	n	n	n	n	n	n	0
Air quality data	Ministry Environment and Forestry	y	y	y	y	y	y	y	y	y	n	95
Volume of solid waste generated	Ministry Environment and Forestry	y	y	n	n	y	n	y	y	n	n	45
Volume of liquid waste generated	Ministry Environment and Forestry	y	y	n	n	y	n	y	y	y	n	50

Table 3. Data Openness Score - Jakarta

Table 3 above shows that all environmental data and half of the contextual urban data are available in Jakarta, indicating a high degree of data availability. Still, some lack access and usability, especially on the volume of liquid and solid waste generated. While the different ministries collect and store the data, the National Statistics Office regularly publishes them on their website.

2c. Hanoi (Vietnam)

2c1. Status of Data and Data Sources

Air quality monitoring data is available in Vietnam, but they are not open and readily accessible from government websites in machine-readable quality. Historical air quality data is not easy to access, but this is visualized (without access to underlying data) on a web page and in real-time. For example, in the Hanoi Air Quality Monitoring Network case, while data is visualized and available for viewing every month, data cannot be downloaded to use for analysis.

Air quality index data (e.g., PM2.5, PM10, O3) - is available as annual average values (2014-present) in the aqicn.org website (https://aqicn.org) and daily average values are downloadable as well.

Nevertheless, historical air quality data in Vietnam is collected from 2 monitoring stations by the government of Vietnam and the US Embassy in Hanoi. The processed data belongs to 2 different AQI calculation systems: the Vietnamese government (VN_AQI standards) and US Embassy (US EPA standards). The data from these monitoring stations are readily available online and can be downloaded in machine-readable formats. The data obtained from these sensors is limited by the length of time that the sensors have been operational.

Population, demographic, and settlement data are available from the government. There are two sources: the General Statistics Office and the Hanoi Statistics Office page. The General Statistics Office releases demographic and census data on a provincial and national scale, while Hanoi Statistics Office releases the statistics on a municipal scale.

Water quality data is too technical and may only be collected and reported by research Institutions, either from a government ministry or a university. But in

most cases, these institutions do not disclose data the to public. These institutions include the National Center for Water Resources Planning and Investigation (NAWAPI), the School of Environmental Science and Technology, Under Hanoi University of Science and Technology, and the Institute of Environmental Technology.

For transportation data, there are two sources: the ASEAN statistics portal and the Vietnam Register Department. However, both sources are at the national level and do not contain granular data at the subnational level. ASEAN portal is in the beta version, but we can export data into a machine-readable format. The Vietnam Register Department, on the other hand, only publishes the data as a bar chart.

There are two sources for the volume of solid waste generated: the General Statistics Office of Vietnam (GSO) and the Vietnam Environment Administration under the Ministry of Resource and Environment (MONRE). The data from the General Statistics Office is machine-readable but is collected only at the province level and updated annually. On the other hand, the data from the Vietnam Environment Administration are contained in data tables presented as reports. The data table includes annual data for four years: 2010, 2015, 2018 and 2019.

For the volume of liquid waste generated, data comes from a published report from the MONRE, which shows waste generated from industrial water waste and hospital water waste. Domestic water waste data is not available, and this is a huge limitation because this accounts for more than 30% of water waste, especially in big cities like Hanoi. Unlike solid waste, liquid waste statistics are limited.

Water consumption data is available, but only at the country level, so we could not extract relevant data for cities. On the other hand, it was not easy to find statistics about untreated sewage in Vietnam, much more at the subnational level.

2c2. Results of Data Scoring

Data	Source	a	b	c	d	e	f	g	h	i	j	Score
Urban settlements area	Hanoi Statistics Office	y	y	n	n	y	y	y	y	y	n	65
Urban population	Hanoi Statistics Office	y	y	n	n	y	n	y	y	y	n	65
Population growth	Hanoi Statistics Office	y	y	y	n	y	n	y	y	y	n	65
	General Statistics Office	y	y	y	y	y	y	y	y	y	n	95
Water consumption data		n	n	n	n	n	n	n	n	n	n	0
Total volume of untreated sewage		n	n	n	n	n	n	n	n	n	n	0
Water quality data		n	n	n	n	n	n	n	n	n	n	0
Modes of transportation		n	n	n	n	n	n	n	n	n	n	0
Motorization index		n	n	n	n	n	n	n	n	n	n	0
Air quality data	AQICN	y	y	y	y	y	y	y	y	y	n	95
	CEM	y	y	n	n	y	n	y	y	y	n	50
	PamAir	y	y	n	n	y	n	y	y	y	n	50
	Hanoi Committee	y	y	n	n	y	n	y	y	y	n	50
	IQAir	y	y	n	n	y	n	y	y	y	n	50
Volume of solid waste generated	General Statistics Office	y	y	y	y	y	n	y	y	y	n	80
	Ministry of Natural Resources and Environment	y	y	n	n	y	n	y	y	y	n	50
Volume of liquid waste generated	Department of Water Resources	y	y	n	y	n	n	n	n	n	n	30

Table 4. Data Openness Score - Hanoi

Except for air quality data and volume of solid waste generated, most

environmental datasets assessed in this research are not available in Hanoi. While other urban contextual datasets are available, these are not published in machine-readable formats, affecting ease of use. Some data are also available but not reusable because of licensing limitations, such as the volume of waste generated and even population and population growth.

2d. Bangkok (Thailand)

2d1. Status of Data and Data Sources

Amongst all cities studied, air quality data in Bangkok is the most advanced in terms of the regularity of collection, the accessibility of the datasets online, and the number of data collection points across the metropolis. The data is published in at least two sources, the Air Quality and Noise Management Division, Bangkok Metropolitan Administration and the Pollution Control Department, Ministry of Natural Resources and Environment.

The one published at the Air Quality and Noise Management Division is the friendliest, especially to non-Thai speaking researchers. The data is also readily downloadable in .csv format, which makes it easier for analysis and visualization. It monitors air quality in 78 data collection stations across Bangkok and publishes data in real-time on an hourly basis. The website also features the option to view data as a graph. While there is no direct download link, the data can be downloaded and converted into a machine-readable format.

Among the urban contextual data assessed, urban population data, while available, are not accessible in open formats. The data is published on the Ministry of the Interior website as a pdf file with no access to underlying data. The data set is also not openly licensed, and further use is restricted. Other demographic data also is in the same state. This confirms previous assessments in the Open Data Barometer, which notes that while the country is data-rich, its accessibility and openness to the public is a significant challenge.

Among the key environmental data assessed, data on untreated sewage is not available, though water quality data, along with volume of liquid waste generated, are available as excel files and published regularly. Like other cities in this study, motorization index, which is data derived from computation, is not readily available, although relevant agencies collect base data for the computation.

The Bangkok Metropolitan Administration and its departments and attached agencies are the main sources of environmental data. However, there is no integrated portal that shows the different data sets. There is a need to visit individual websites to be able to view or download data.

2d2. Results of Data Scoring

Data	Source	a	b	c	d	e	f	g	h	i	j	Score
Urban settlements area	Land Development Department	y	y	n	n	y	n	n	y	n	n	40
Urban population	Ministry of the Interior	y	y	n	n	y	n	y	y	y	n	50
Population growth	National Statistics Office	y	y	y	n	y	n	y	y	y	n	65
Water consumption data	Metropolitan Waterworks Authority	y	y	n	n	y	n	y	y	n	n	45
Total volume of untreated sewage		n	n	n	n	n	n	n	n	n	n	0
Water quality data	Metropolitan Waterworks Authority	y	y	y	y	y	n	y	y	y	n	80
Modes of transportation	Department of Land Transport	y	y	y	y	y	n	y	y	y	n	80
Motorization index		n	n	n	n	n	n	n	n	n	n	0
Air quality data	Bangkok Metropolitan Administration (BMA)	y	y	y	y	y	y	y	y	y	n	95

Data	Source	a	b	c	d	e	f	g	h	i	j	Score
Volume of solid waste generated	Bangkok Metropolitan Administration	y	y	y	y	y	n	y	y	y	n	95
Volume of liquid waste generated	Drainage and Sewerage Department, BMA	y	y	y	n	y	n	y	y	y	n	65

Table 5. Data Openness Score - Bangkok

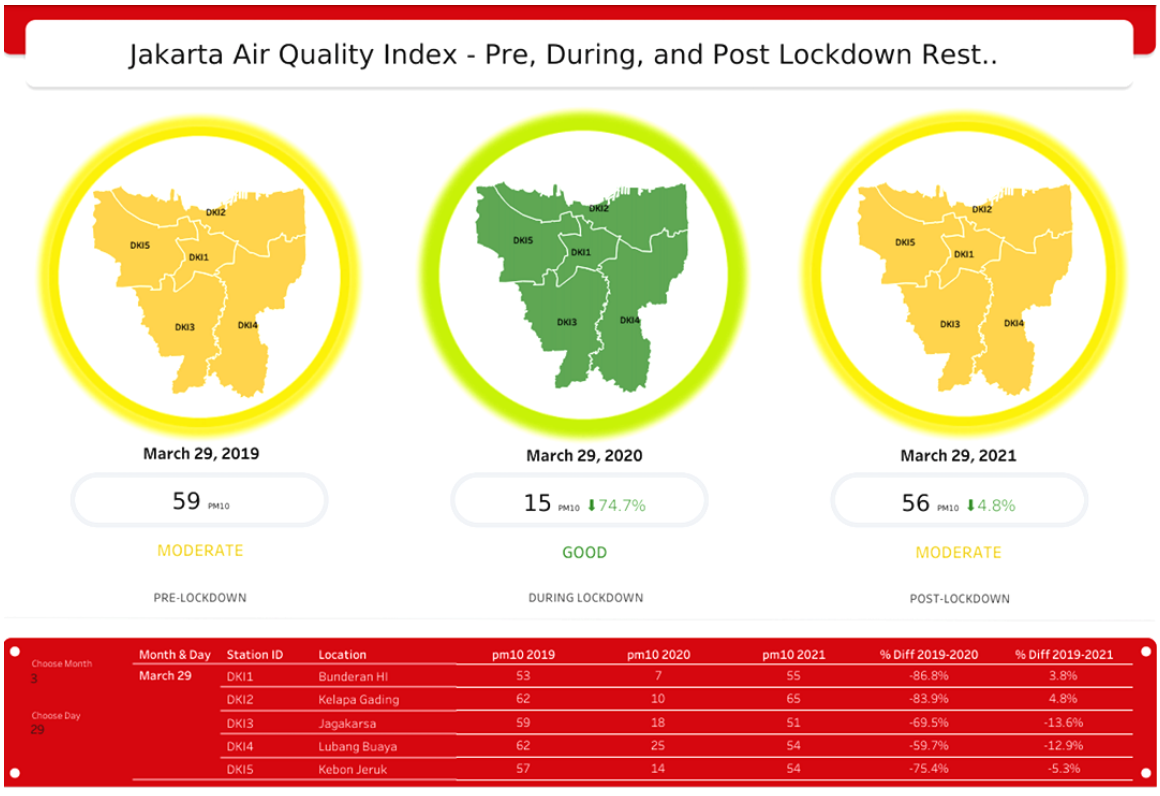
As shown in Table 5 above, contextual urban data is difficult to access in open formats, even though they are available. The only data where there is no evident digital source is untreated sewage and motorization index. The latter is a derived data and may be computed from raw data sources.

3. Options for Data Visualization

In this section, we present the different options for visualization based on the data available. The focus, in this case, is air quality data. The primary question that the visualization would like to answer is whether there is a change in air quality during the COVID-19 pandemic due to restrictions in movement imposed by the countries covered by this study.

3a. Comparison Across Specific Dates

The visualization below compares the air quality index in 5 data collection points across three years. A specific date was chosen using the lockdown scenario as the primary determinant. Jakarta, in this case, implemented its first hard lockdown in the second half of March 2021, imposing work from home arrangements and restricting religious worship. The choice of the specific date (March 29) is conditioned by data availability within the three-year



Month & Day	Station ID	Location	pm10 2019	pm10 2020	pm10 2021	% Diff 2019-2020	% Diff 2019-2021
March 29	DKI1	Bunderan HI	53	7	55	-86.8%	3.8%
	DKI2	Kelapa Gading	62	10	65	-83.9%	4.8%
	DKI3	Jagakarsa	59	18	51	-69.5%	-13.6%
	DKI4	Lubang Buaya	62	25	54	-59.7%	-12.9%
	DKI5	Kebon Jeruk	57	14	54	-75.4%	-5.3%

Figure 1. Air Quality Index in Jakarta, Comparing Specific Time Periods

period from 2019 across the different data collection points where we gathered the average.

Figure 1 indicates a significant improvement in air quality index when we compare 2019 with 2020 data when successive lockdowns were imposed in the city by the government. Towards the end of March 2021, lesser restrictions were imposed by the city government.

3b. Comparing Current data with Average Baseline Figures

What can be done when historical data at specific dates (like the example in Figure 1 above) is unavailable? This was our dilemma when we looked at the Manila data for 2019 to 2020.

The per city PM2.5 and PM10 readings in Manila are incomplete in 2019. This makes the comparison of the data for Paranaque and Mandaluyong, the two cities where sensor data is available for 2020 and 2021, very challenging. Ideally, the comparison would have been for the three periods using the same areas where sensor data is available. But because of the limitations, we utilized the average air quality index for 2019 for the whole National Capital Region, where Manila is located, and compared it with the 2020-2021 data from the sensors for the two areas, Paranaque and Mandaluyong.

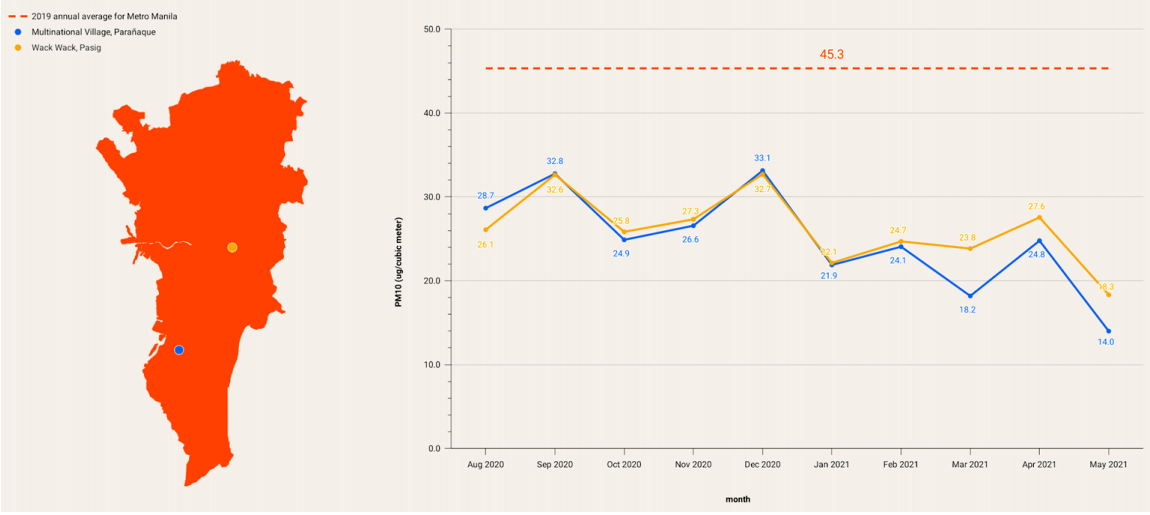


Figure 2. Changes in PM 10 values comparing 2020 and 2021 with baseline figures - Manila

Figure 2 shows that during the pandemic, PM 10 (particularate matter 10), which refers to inhalable particles, with diameters that are generally 10 micrometers and smaller, is lower than the baseline figures. PM 10 contains microscopic solids or liquid droplets that are so small that they can be inhaled and cause serious health problems to individuals and are thus considered one of the primary air quality indicators.

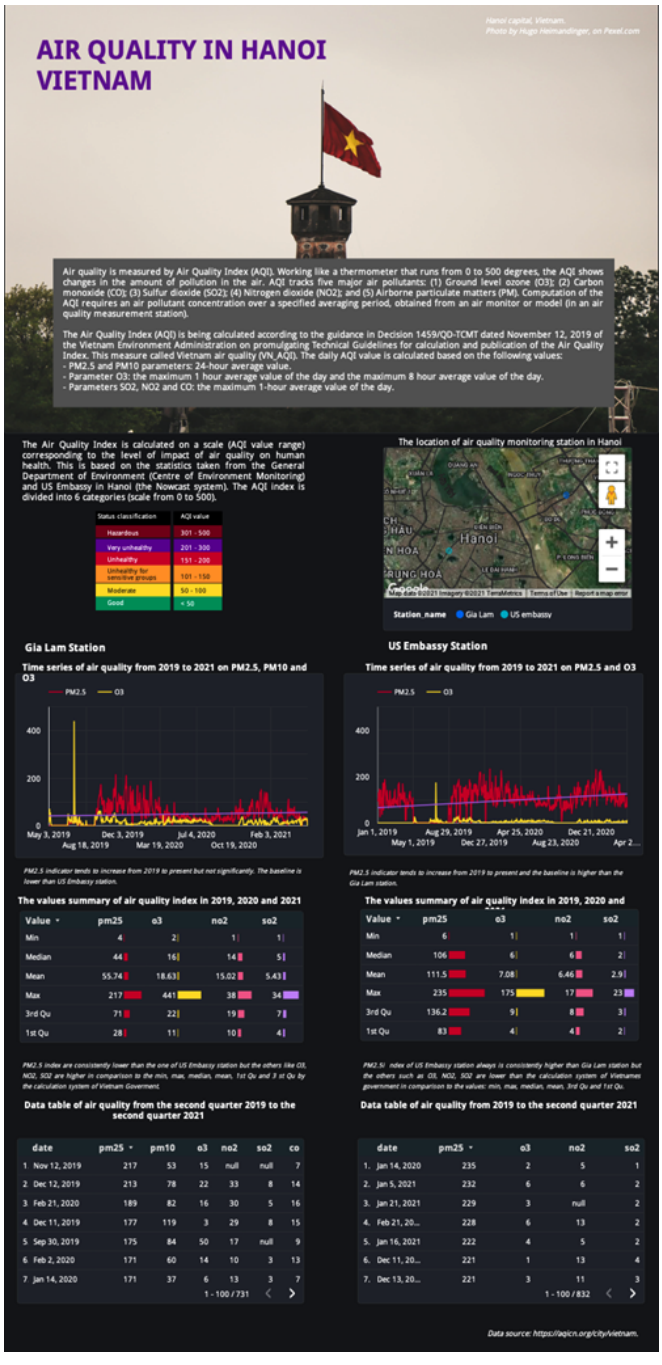


Figure 3. Sample Data Visualization - Hanoi

3c. Comparing Time Series Data in Two Data Collection Sites

While air quality data is available in some contexts, these are not provided as machine-readable data to allow visualization. This happens in Hanoi where air quality data and its corresponding visualization are published online but not in machine-readable formats. Also, air quality data in Hanoi published by the government do not show historical figures and only present time series data on a monthly basis, precluding the ability to present time series data on a more extended period covering pre-, during, and post-lockdown scenarios.

Figure 3 shows a sample visualization of the Hanoi data, covering only two data collection points where historical data is available and downloadable. The data collection stations – Gia Lam and US Embassy, are located opposite each other (southwest and northeast)

and may be representative of the air quality conditions of neighboring areas.

Air quality in both stations (using PM 2.5) also indicate an improving scenario during the hard lockdown periods (April 2020). Air quality started to deteriorate when mobility restrictions were relaxed in the third quarter of 2020.

But the data in Hanoi is very limited in this case unless we use government data collected and published by the Hanoi Air Quality Monitoring Network, which only allows a 30-day cycle publication in closed formats. This will not enable us to use a time serious comparison across 2019 to 2021.

This data situation highlights the need for open data, which undoubtedly is a limitation in the Hanoi case.

3d. Combining Contextual Variables in the Visualization

Several factors impact air quality data. As earlier indicated, mobility of people using different forms of transport is one of the causes. To test whether lockdown policies and mobility restrictions impact air quality, the contextual variables can be visualized along with air quality data to provide some initial insights.

Figure 4 below combines data on lockdown procedures (taken from qualitative assessment reviews of governmental regulations), classified as medium or hard lockdown, mobility data (taken from Google mobility data), and air quality data (from the Ministry of Environment and Forestry).

At least three patterns are emerging from the visualization. First, during hard lockdown periods, mobility within Bangkok significantly decreased when compared to baseline figures. Second, air quality data is generally below the baseline figures during hard lockdown periods, except for December to January.

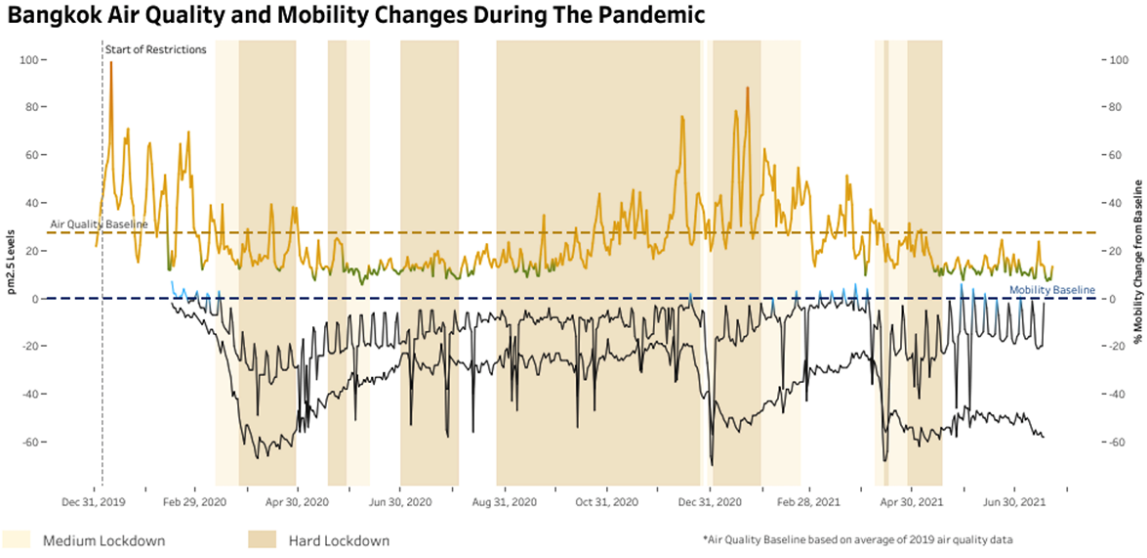


Figure 4. Air Quality, Lockdowns, and Mobility in Bangkok

Finally, lockdown impacts mobility significantly within the period immediately following its imposition and gradually increases towards baseline over time.

3e. Limitations of Data and Visualization

Looking at these options, it is quite evident that limitations abound. We discuss these limitations below:

a. Data unavailability precludes a more granular analysis of the air quality and demographic and urban profile. For example, it has been argued in several papers that mobile sources of pollution (cars, motorcycles, trucks, and busses) are responsible for the deteriorating air quality in Manila (UNIDO 2019), Jakarta (UNEP 2019), Hanoi (Tang et al, 2020) and Bangkok (Narita et al., 2019). Thus, there is sufficient evidence to show that lockdowns, which delimit movement, has the consequent effect of improving air quality, as some studies would show (e.g. Wetchayont et al 2021, Nguyen et al 2021). But failure to gather motorization index, concentration of urban population, traffic routes and severity, among

others, was a significant challenge for the researchers to undertake a more granular analysis.

b. Data openness is a significant factor in our ability to analyze and visualize data. There are a lot of data out there, but our ability to access and use them is limited because of format (e.g. some are in PDFs), license (e.g. publishers indicate an "all rights reserve" regime), and bulk download features (e.g. limitations in what can be viewed and accessed). A classic example, in this case, is Vietnam, where data is available but not in machine-readable formats, has a proprietary license, and features limited access.

c. Sustainability in data collection impacts the ability to generate time-series interpretation and insights. This is very stark in the Philippines, where data for air quality is not collected regularly. Only an annual average of PM2.5 and PM10 for NCR is available from DENR-EMB for 2012-2019. Not all cities/municipalities have complete readings from 2012-2019, unlike Bangkok, where data is available almost daily across all stations. Time-series comparisons for Manila, in this case, is virtually impossible.

d. Geolocating contextual data is not a standard practice across countries. For example, population data is published at the municipal or district levels, not at the commune or village level. The same is true with vehicle ownership, waste generation, water consumption, mode of transportation, among others. So one will know how much total waste is generated in Bangkok, but you will not know which districts generate more domestic waste than others. In this case, for example, heatmaps can not be generated, and if they can be, in some cases, the unit of analysis is relatively large (e.g. district level), that analysis becomes significantly limited.

4. Conclusion and Next Steps

4a. Summary of Findings and Conclusion

This research yielded the following key findings:

a. In most cities, data is available for at least 70% of the data inventoried, although the level of granularity remains low.

b. The availability of machine-readable datasets needs to be improved. Among those datasets that are available online, slightly more than half (are available in machine-readable formats. These datasets can be downloaded in .csv format. The remaining datasets are only available in pdf or .html format, making them difficult to be processed and analyzed further by users.

c. The majority of the datasets that are available online are free of charge. Users can access and download or scrape them without additional costs. There are a few cities where it requires adequate skills to locate the data,

download them, and convert them to machine-readable format to allow analysis and visualization.

d. Most of the datasets are not openly licensed. For some of the datasets, an explicit restriction is indicated on the dataset (e.g. "All rights reserved.) while the majority do not have any license information indicated. For purposes of the research, we consider the datasets that do not have information on license as proprietary data.
e. It appears that most datasets are sustainably collected, though the regularity of collection differs across cities. For some, like the air quality data in Bangkok, these are updated real-time, with historical data available, while in others, data is collected annually.

At least three issues are impacting data quality, and these are discussed briefly below, along with recommendations:

- Machine-readability. For data to have the greatest value, having the datasets in machine-readable formats must become a key goal for data providers. Government agencies or other institutions that release data should move towards providing key datasets in machine-readable formats at source.
- Regularity of collection and publication. Extracting data trends and generating insights from patterns and outliers will only become meaningful if data is collected and published regularly. City governments should prioritize regular collection and publication of those data that are needed for decision-making processes.
- License and proprietary rights. Licensing rules impact how data is used to benefit society. When public data is open-licensed (e.g. creative commons), its use can be widely maximized by triggering research and innovation.

The absence of government-wide data guidelines to ensure uniformity and standardization could partially explain differences in data quality between the various cities included in the study. City governments in these countries need to radically change their approach to data publication and disclosure and focus

their efforts on implementing adequate data governance standards to cover those concerns mentioned above that impacts data quality and its subsequent use and impact.

4b. Next Steps

While not required under the terms of reference governing this research, the following are the papers that the researchers will produce in the next two months:

a. Openness of environmental data and its implications on data governance. The paper will utilize the findings of the inventory conducted by the researchers and its implications on measuring environmental health, as well as on monitoring the Sustainable Development Goals
b. A review of alternative data sources to capture air quality data, with particular reference to the use of satellite data that can potentially reveal anomalies in the relationship between lockdown, mobility, and air quality. This is particularly true in Hanoi, where there seems to be only a slight improvement despite mobility restrictions. It has been argued that pollutants for the city are outside the city itself, particularly those coming from the powerplants and the industrial clusters.
c. A deeper investigation of lockdown, mobility, and air quality, using the results of this study (e.g. see Figure 4) and other analyses conducted by other researchers in the last six months.

The timeline for producing these results will be from August to September 2021.

References

Environmental Protection Agency. Undated. Available at https://www.epa.gov/transportation-air-pollution-and-climate-change/learn-about-air-pollution-transportation. Accessed on 12 April 2021.

Narita, D., Oanh, N.T.K., Sato, K., Huo, M., Permadi, D.A., Chi, N.N.H., Ratanajaratroj, T., Pawarmart, I. (2019). Pollution characteristics and policy actions on fine particulate matter in a growing Asian economy: The case of Bangkok metropolitan region. Atmosphere 10, 227. Available at https://doi.org/10.3390/atmos10050227. Accessed on 9 May 2021.

Nguyen, T.P.M, Bui, T.H., Nguyen, M.K., Nguyen, T.H., Vu, V.T., Pham, H.L. (2021). Impact of Covid-19 partial lockdown on PM2.5, SO 2, NO 2, O 3, and trace elements in PM2.5 in Hanoi, Vietnam. *Environmental Science Pollution Research*. Available at https://doi.org/10.1007/s11356-021-13792-y. Accessed on 12 June 2021.

Tang, V. T., Oanh, N. T., Rene, E. R., & Binh, T. N. (2020). Analysis of roadside air pollutant concentrations and potential health risk of exposure in Hanoi, Vietnam. *Journal of Environmental Science and Health, Part A, 55(8), 975-988*. Available at https://doi.org/10.1080/10934529.2020.1763091. Accessed on 22 May 2021.

United Nations Industrial Development Organization. (2019). *Health & Pollution Action Plan: in support of the National Environmental Health Action Plan 2017-2022*. Inter-Agency Committee on Environmental Health of the Republic of the Philippines. Available at https://www.unido.org/sites/default/files/files/2019-10/Philippines%20HPAP.English.pdf. Accessed on 18 May 2021.

United Nations Environment Programme. (2019). *Fiscal policies to address air pollution from road transport in cities and improve health: Insights from country*

experiences and lessons for Indonesia - Inclusive green economy. Available at https://wedocs.unep.org/handle/20.500.11822/33583. Accessed on 19 July 2021.

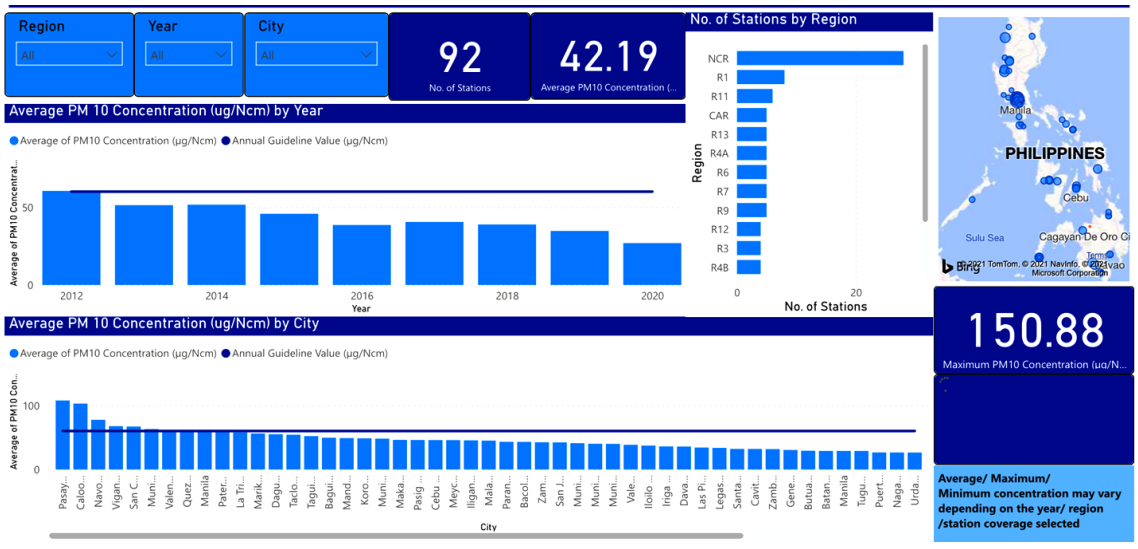
Wetchayont, P., Hayasaka, T., & Khatri, P. (2021). Air quality improvement during COVID-19 lockdown in Bangkok metropolitan, Thailand: Effect of the long-range transport of air pollutants. *Aerosol and Air Quality Research, 21, 200662*. Available at https://doi.org/10.4209/aaqr.200662. Accessed on 12 June 2021.

Annexes

Annex 1. Methodological Note: Manila

Datasets used

- Annual average PM10 data for the National Capital Region / Metro Manila
  - from the Department of Environment and Natural Resources - Environmental Management Bureau (DENR-EMB) covering 2012-2019



https://app.powerbi.com/view?r=eyJrIjoiM2ZhN2QzNTEtMmUxMi00M2M1LTlhZmItMWUzNDFiNWY0MDE0IiwidCI6ImY2ZjRhNjkyLTQzYjMtNDMzYi05MmIyLTY1YzRlNmNjZDkyMCIsImMiOjEwfQ%3D%3D

  - PM10 sensor readings from two PurpleAir stations - https://www.purpleair.com/map?opt=1/c/mPM25/a1440/cC0#12.27/14.54757/121.03803
    - Multinational Village, Paranaque; Wack-Wack, Pasig
    - August 2020 - May 2021
    - Has per-minute, hourly, and daily average readings

Constraints and limitations

1. For pre-pandemic data (before 2020)
   a. We only have annual average air quality data for the entire NCR.

   b. No monthly or daily averages are available.
   c. The annual averages for individual NCR administrative units are incomplete.
      i. Only Makati, Manila, Marikina, Muntinlupa, Pateros, and Quezon City have data for 2019
      ii. Only Caloocan, Makati, Manila, Marikina, Muntinlupa, Parañaque, Pateros, Quezon City, San Juan have data for 2018
2. For data during the pandemic (2020-present)
   a. This data is not currently readily available from the government
   b. The data is available from PurpleAir but only from two sensors
      i. Multinational Village, Paranaque
      ii. Wack-Wack, Pasig
   c. The sensors have data from Aug 2020 - present

What we can't do

1. Compare the air quality data on a per-administrative unit level since the 2020 data we have (Paranaque and Pasig) have no annual average values for 2019.
2. Compare the air quality data on an annual basis since we only have data for Aug 2020-Dec 2020 and Jan 2021-May 2021.

What we can do

1. We can compare the annual average NCR data to the monthly average data obtained from the two PurpleAir sensors. This will allow us to show the air quality trend during the pandemic and compare it with the annual average value for 2019.

Data Visualization

The colors used in the map and the chart correspond to one another. The color for the annual average value for Metro Manila in 2019 is used as the fill color for the Metro Manila administrative boundary. The colors of the lines in the chart and the colors of the markers for the location of the sensors on the map

correspond to each other. We simply mapped the locations of the sensors using the coordinates provided in the data.

The chart shows the monthly average PM10 values from the two PurpleAir sensors (blue and yellow) from Aug 2020-May 2021, together with the annual average PM10 for the entire Metro Manila in 2019 (orange)

Annex 2. Methodological Note: Jakarta

Jakarta's Air Quality Data 2019 – 2021 are available on the Open Data Portal[1]. It is published and maintained by Jakarta's Environmental Agency. The AQI data were collected from five air quality monitoring stations located in each of the five municipalities in the greater Jakarta area (Table 1). These stations are DKI1-Bundaran HI in Central Jakarta, DKI2-Kelapa Gading in North Jakarta, DKI3-Jagakarsa at South Jakarta, DKI4-Lubang Buaya at East Jakarta), and DKI5-Kebon Jeruk at West Jakarta.

Table 1. Air Quality Monitoring Stations

Station ID	Station Name	Municipality
DKI1	Bundaran HI	Central Jakarta
DKI2	Kelapa Gading	North Jakarta
DKI3	Jagakarsa	South Jakarta
DKI4	Lubang Buaya	East Jakarta
DKI5	Kebon Jeruk	West Jakarta

The dataset consists of several columns. The first two columns are the date of collection and the name of the measurement station. The following five columns listed the value for the several air quality indicators as listed below.

• Date	Date of collection
• Station	Name of measurement station
• pm10:so2	sulfur dioxide
• co	Carbon monoxide
• o3	ozone
• no2	Nitrogen Dioxide
• max	the maximum AQIx value of all indicators
• category	air pollution lever

As seen in Table 2, the indicator for PM10 is recorded as the daily average value.

1 https://data.jakarta.go.id/

Indicator O3 indicates the highest average value over one-hour observation and eight-hour observation of the day. All the recorded values of the three air pollutants (SO2, NO2, and CO) indicate the highest one-hour average values of the day. Finally, the max column indicates the maximum AQIx value of all indicators, then listed under the "critical" column. Finally, the last column (i.e. category) describes the level of air pollution level. The pollution level is divided into good (AQI= 0-50), moderate (AQI=51-100), unhealthy (AQI=101-200), very unhealthy (AQI=201+).

Table 2. Methodological Note on Air Quality Dataset

Key Aspect		Description
a.	Sources of raw data used – if data is not raw, where is the source of the raw data used	Air quality monitoring station
b.	Manner of collection – how did they collect the data	Data were collected through the air quality monitoring station in five locations in Jakarta
c.	Computational elements – if data is a result of a formulaic computation, discuss how the values are arrived at	• Indicator for PM10: 24-hour average value. • Indicator O3: highest one-hour average value of the day and the highest eight-hour average value of the day. • Indicators SO2, NO2 and CO: highest one-hour average values of the day. • Combined daily AQI: the maximum AQIx values of all indicators.
d.	Basis of computation – discuss what is the basis used for arriving at the formula used (e.g. UNEP standards, WHO standards)	WHO standards
e.	Author – who generated the computational elements of the data	Jakarta's Environmental Agency

Data visualization

The covid-19 pandemic has a significant effect on environmental quality. Due to restrictions of people movement, which then slowed down much of economic and social activities, air quality has improved, and water pollution decreased in many cities in different parts of the world[2].

2 https://www.sciencedirect.com/science/article/pii/S2405844020318089

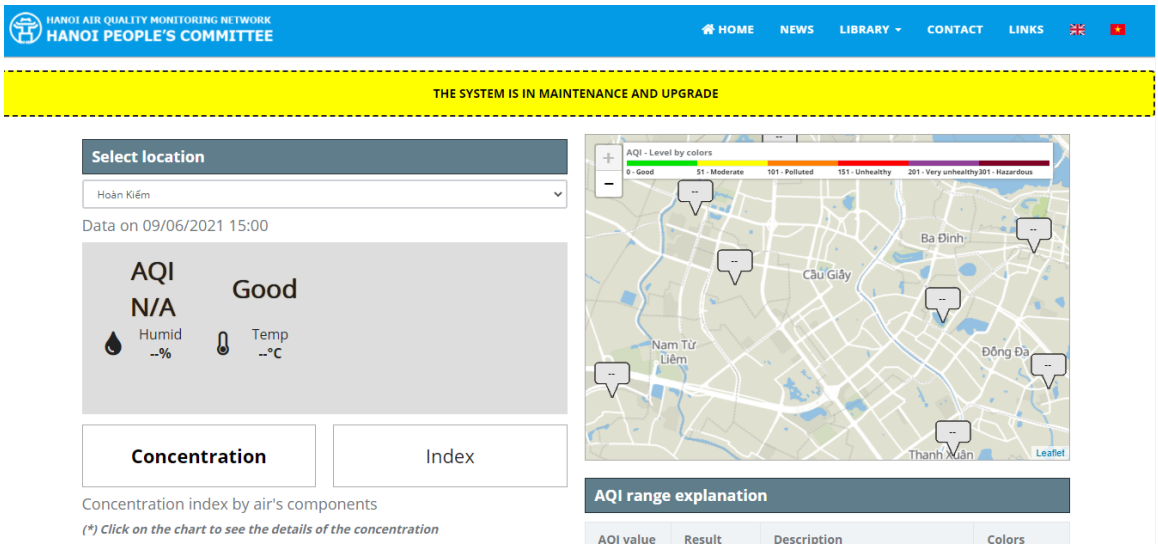
In Indonesia, the first case of Covid-19 was found on March 2, 2020. As the pandemic worsened, the national government, including the provincial and municipal governments, reacted with a state of emergency declaration and instituted lockdown and travel restrictions. In Jakarta, once the first case was detected, the governor halted the issuance of permits for large gatherings and closed several tourist destinations. On March 14, all school activities were also suspended, and a week later (March 20), the government of Jakarta declared a state of emergency[3].

3 https://www.thejakartapost.com/news/2020/03/20/breaking-jakarta-declares-covid-19-emergency-urges-offices-to-suspend-operation-for-14-days.html

Annex 3. Methodological Note: Hanoi

There are several websites on air quality monitoring in Vietnam, but they are not published in machine-readable format. No historical data is also provided, and the data is visualized on the web page in real-time.

Hanoi Air Quality Monitoring Network. In 2016, Hanoi City invested, built, and established ten automatic air quality monitoring stations, including two fixed stations and eight sensor stations. The Environmental Protection Sub-department's fixed station is located on the 6th floor of the Environmental Protection Sub-Department building, located in a residential area, far from the main road, to assess the daily air quality of the surrounding residential area. Minh

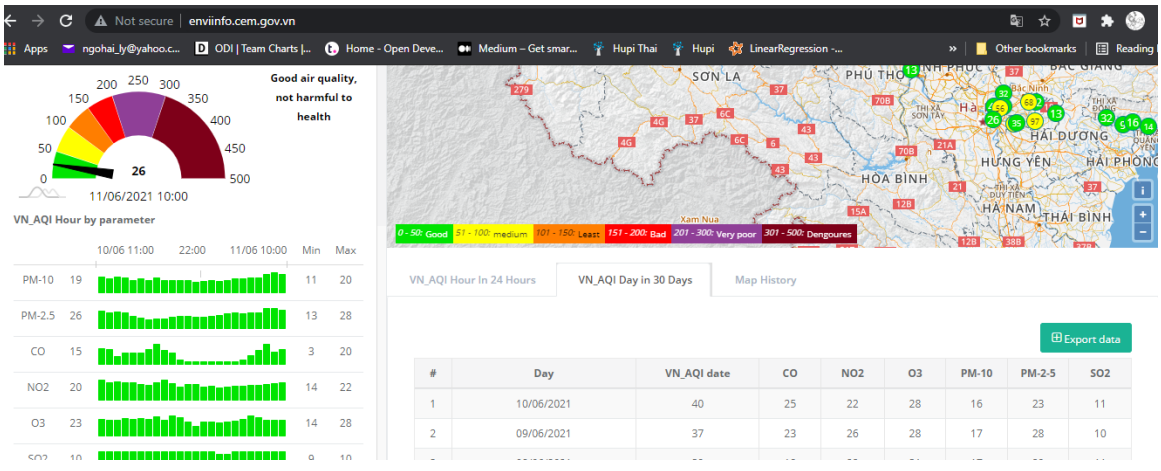


Khai fixed station is located at the People's Committee of Minh Khai Ward, Bac Tu Liem District, used to monitor the air quality of ring traffic. Eight sensor stations are installed in Hang Dau, Hoan Kiem, Kim Lien, My Dinh, Pham Van Dong, Thanh Cong, Tan Mai, Tay Mo. In July 2019, the French Embassy installed an air monitoring station, PM2.5 dust target at the embassy campus and integrated it into the monitoring network of Hanoi, bringing the total number of monitoring

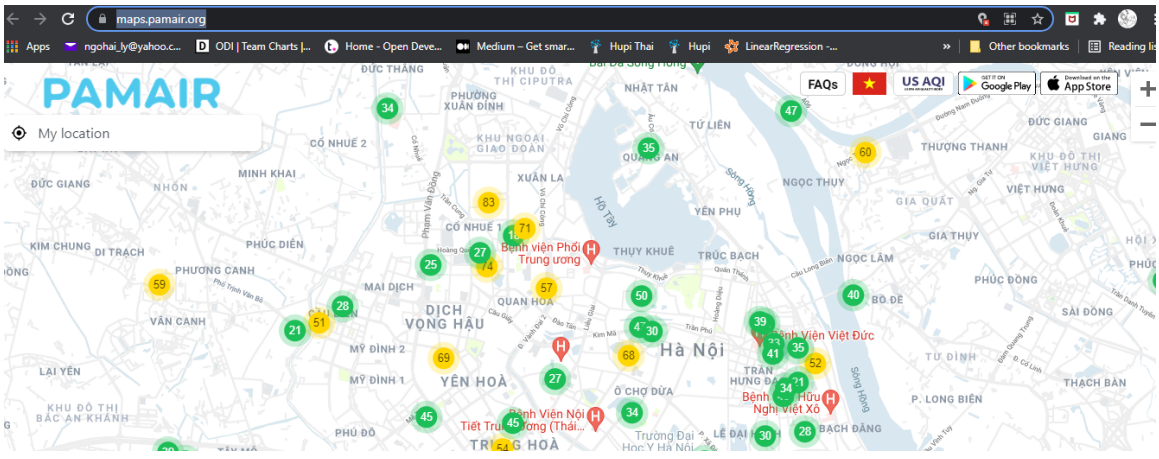
stations to 11 stations. By May 2020, Hanoi received 24 more automatic air monitoring sensor stations sponsored by THT (Korea) Co., Ltd. This increased the total number of air monitoring stations in the city to 35.

The Hanoi Air Quality Monitoring Network provides data in real-time and visualizes the data on its web page. However, this system is on maintenance and upgrading during the conduct of the research, so we could not extract data from this source.

Center for Environmental Monitoring (CEM). The CEM is the focal point of the national environmental monitoring network system. The center has a nationwide monitoring system. However, they have a fixed monitoring station in Hanoi, Gia Lam station near the suburban Long Bien district. CEM only allows the extraction of historical air quality monitoring data for the previous 30 days.



PamAir. Pam Air is an organization founded in 2017, the first and only private air quality monitoring network in Vietnam, operating over 400 monitoring points across 63 cities and provinces. They published the data in real-time on their web page and visualized the data. But they do not provide the raw data publicly.



IQAIR. IQAIR only provides data in visualization format on their web page in real-time. They do not provide the data in machine-readable format. There is no possibility of accessing the data in bulk. No historical data is also available.

AQICN. Air quality index data - e.g., PM2.5, PM10, O3, among others, are available as annual average values (2014-present) in the aqicn.org website (https://aqicn.org). Daily average values are, however, downloadable. A request via the aqicn portal or an email registered to the agency mandated to collect the data is needed to access these data.

However, Vietnam's historical air quality data is collected only from 2 monitoring stations by the government of Vietnam and the US Embassy in Hanoi. The processed data belongs to 2 different calculation AQI systems: the Vietnamese government (VN_AQI standards) and US Embassy (US EPA standards). The data from these monitoring stations are readily available online and can be downloaded in machine-readable formats. The data obtained from these sensors is limited by the length of time that the sensors have been operational.

For this report, we used the AQICN data because it's the only one we can access. Air quality is measured by Air Quality Index (AQI). Working like a thermometer that runs from 0 to 500 degrees, the AQI shows changes in air pollution. AQI tracks five major air pollutants: (1) Ground-level ozone (O3); (2) Carbon monoxide (CO); (3) Sulfur dioxide (SO2); (4) Nitrogen dioxide (NO2); and (5) Airborne particulate matters (PM). Computation of the AQI requires an air pollutant concentration over a specified averaging period, obtained from an air monitor or model (in an air quality measurement station).

The PM2.5 baseline of the Vietnamese government is lower and flatter than the one of the US Embassy in Hanoi (linear calculation by time) The PM2.5 index maximum value of Gia Lam station is 217 on 12 November 2019. In contrast, it's 235 on 14 January 2020 of US Embassy station.

Annex 4. Methodological Note: Bangkok

Air quality data for Bangkok was derived from the Air Quality and Noise Management Division of Bangkok's Metropolitan Authority. The data is accessible on this website - https://bangkokairquality.com/bma/aqi?lang=en. The data is downloadable from the site. To generate the visualization, downloaded data was cleaned and tested and uploaded to Tableau for visualization. Data collection points prioritized were from 5 data collection points in Bangkok's city centre, where most of the work offices are located, based on a work concentration index. Mobility data was derived from Google's COVID-19 mobility reports, focusing largely on work-related mobility comparing changes from baseline. For Google's methodology, please access this link.

Lockdown descriptions and severity is sourced from government pronouncements and media reports. The severity of the lockdown is dependent on the pronouncements issued by the Centre for COVID-19 Situation Awareness in Thailand. A hard lockdown is used to describe situations where the CCSA declared a state of emergency, which carries more stringent mobility requirements.

All data is plotted on Tableau, using time series comparison, including the qualitative data on lockdown descriptions.

STEP UP CONSULTING SERVICES