

The need for open data sharing in the era of global pandemics

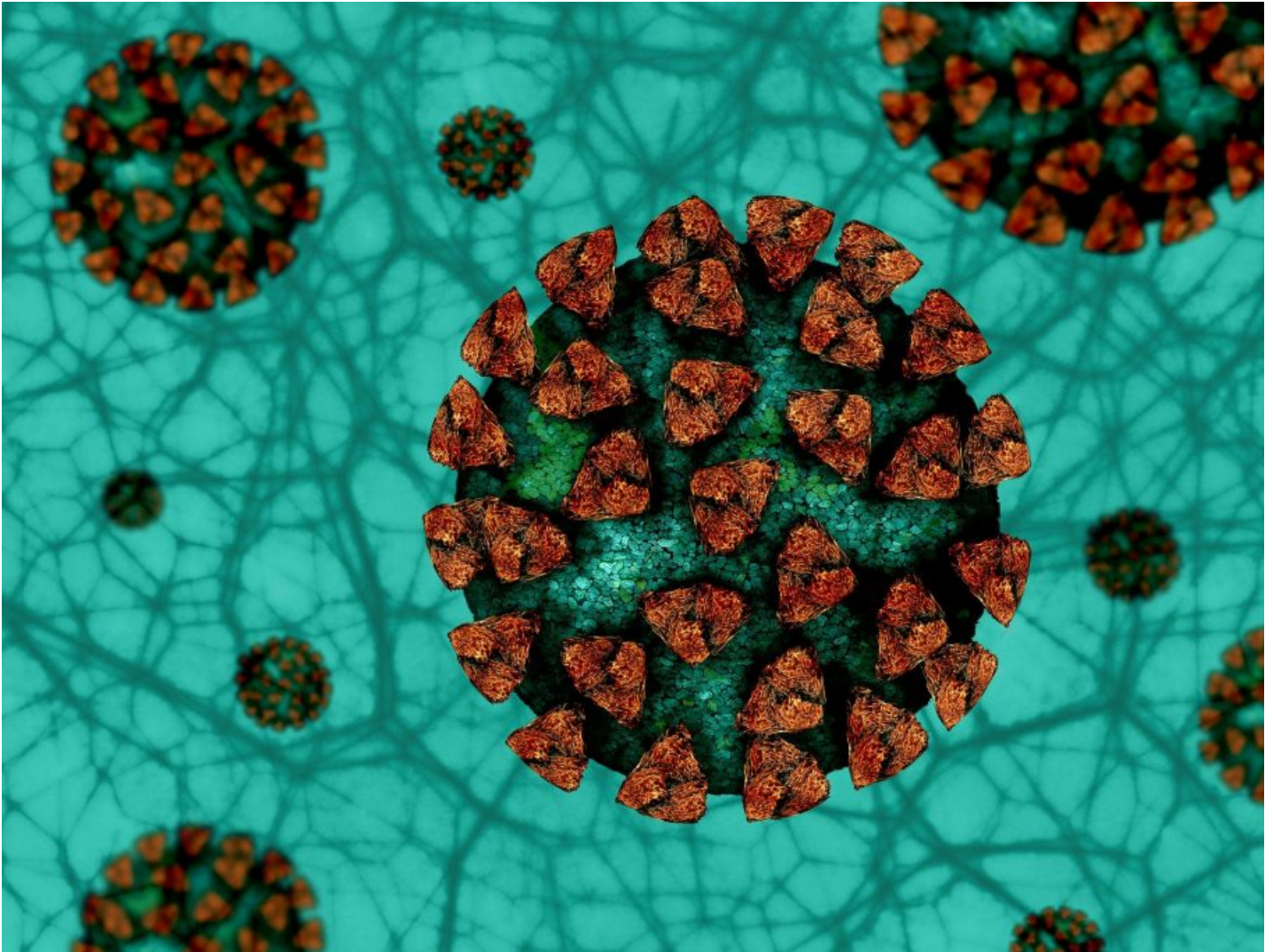
*Since the start of the pandemic, data on different countries' case counts has been readily available. However, not all data is equally useful. In this post, **Bernardo Gutierrez** and **Sabrina Li**, from the Open COVID-19 Data Working Group and the Oxford Martin School, outline the need for much more detailed, open and accessible sharing to inform science and policymaking. The authors argue for open, accessible data sharing and highlight the need for data collection processes to be increasingly homogenised.*

Many people closely followed the COVID-19 pandemic between January and May, or at least that was the case for us. This meant refreshing Twitter and looking at online dashboards to scout for the latest news on case counts and finding out where cases were popping up. Interactive maps that display case counts and their locations in near real time, such as the well-known [Johns Hopkins University COVID-19 dashboard](#) (JHU), have become ubiquitous tools for presenting accessible information on the pandemic's progression to the general public. The JHU dashboard reports the total number of cases for a country by aggregating data reported by various health agencies and institutions around the globe. Some countries may report cases at a high geographic resolution (e.g. for counties, municipalities, etc.), while others only present the total number of cases at the national level.

How to make data “good” data

- **Detail**

Varying levels of detail certainly affect how a map illustrates where cases are being reported, but it can go much further. Aggregated case counts are easier to collect and report, but they lack additional details about the characteristics of the individual who became infected. What is their age and gender? Have they recently travelled to other (affected) countries? When did they start showing symptoms and were they mild or severe? While this information may have limited use for the public, researchers and policymakers make use of these extra bits of information to tell us about how a pathogen spreads, what populations are the most vulnerable, and the best ways to mitigate its effects.



- **Open Access**

For this reason, there is immense value in making this information *open access*, or freely available to any and all interested parties through various sharing mechanisms. The open sharing of data during a pandemic is more important than ever, and can be used as a method to ensure the transparency and reliability of the data. It fosters interdisciplinary collaboration that further advances our understanding of the infectious agent and its behaviour, and facilitates coordinated and timely responses that transcend national borders.

- **Accessibility**

Good data is the fuel of good science, but it is challenging to acquire. Effort taken to acquire data is squandered if the data never reaches the people that analyse it to draw insights or produce appropriate policies and interventions. The issue of data accessibility is not new, but has become increasingly noteworthy as epidemics become more prominent. The 2014-2016 Ebola virus epidemic in Western Africa and the 2015 Zika virus epidemic in the Americas highlighted that information about new cases is not always easily accessible, including infectious disease researchers. Furthermore, transparent data collection processes allow for a better understanding of the potential issues of how information is obtained to improve its reliability.

The scientific community has increasingly argued for the value of sharing data openly, as it allows for cross-validation of results which makes new discoveries more robust and reliable

- **Multiple data sources**

An ideal scenario of data collecting would rely predominantly on primary data sources, usually falling back to national health agencies who collect this information as part of their surveillance. Initiatives like [HealthMap](#) or [Promed](#) have also taken the task of collating information from less traditional sources: the plethora of news outlets and publications shared via social media available in the contemporary landscape of the internet. This diversity of data sources further pushes the reliability and transparency of the process, allowing cross-validation of individual data sets to be made possible.

- **Scientific collaboration**

The notion of multiple data sources is bound to the independent work of different teams and institutions; therefore, it follows that the analysis procedure also benefits from the collaboration of the wider scientific community. Having different researchers investigate the data and explore similar research questions adds to the robustness of new evidence and the conclusions drawn from it. The scientific community has increasingly argued for the value of sharing data openly, as it allows for cross-validation of results which makes new discoveries more robust and reliable. The same concept applies to any evidence-based endeavour: robustness and reliability can be seen as throwing wrenches at a machine to test its breaking points so it can be improved. Facilitating access to raw data that drives the decision-making process tests how decisions hold up, and can improve their efficacy and practicality in the long run.

- **Speed: Prompt reporting, prompt political responses**

Accessible data gains a new level of utility when it is promptly reported. A pandemic such as the one we're experiencing requires different sectors to coordinate a consistent response. While academic researchers respond by providing key information about the pathogen and its transmission through the application of science, these insights become more useful when they're translated into expedited action and policy. The interplay between academics and government officials responsible for enacting mitigation policies has been developed over recent times, but the global dimensions of modern pandemics require an increasingly coordinated response from multiple actors. Research from different groups has highlighted the important [role of human mobility on the spread of SARS-CoV-2](#), showing the ease of travel in the modern world. This means that the speed of response and action from countries can have big implications for the broader global community.

It is imperative to highlight the importance of homogenising data collection approaches, especially between countries.

Open Data: The challenges and opportunities

As the pandemic pushes on, some of the challenges regarding data availability have been addressed by multiple groups focusing on specific shortcomings. One of these key issues, the trade-off between scale and precision (i.e. despite having the most up-to-date data sets, the JHU dashboard may ignore detail on individual-level case information), has seen progress through alternative approaches proposed by the likes of the [Open COVID-19 Data Working Group](#), the [MIDAS Network](#) and other [local collaborative efforts](#), who collect detailed data through a crowd-sourcing approach. However, the complexity of this data comes with limitations such as human error and case double-counting, inherent biases from data reporting systems (due to countries adopting different data collection methodologies), delays between the occurrence of new cases and their reporting, and incomplete case counting.

Considerable improvements can be achieved, but it is imperative to highlight the importance of homogenising data collection approaches, especially between countries. Herculean efforts are put into data collection by workers on the ground, and vast amounts of resources are poured into securing this vital information. In our hyper-connected modern world, these efforts are best put into good use when data is available to all actors. Pushing for the common goal of rapidly advancing new discoveries and coordinating responses can maximise the benefits of data collection and enable us to understand the progression of this pandemic and potentially future ones.

Note: This review gives the views of the author, and not the position of the LSE Impact blog, or of the London School of Economics.

Image credit: [Gerd Altmann](#) via [Pixabay](#)
