On the responsible use of digital data to tackle the COVID-19 pandemic

Large-scale collection of data could help curb the COVID-19 pandemic, but it should not neglect privacy and public trust. Best practices should be identified to maintain responsible data-collection and data-processing standards at a global scale.

Marcello Ienca and Effy Vayena

On 30 January 2020, the World Health Organization (WHO) director-general declared the coronavirus disease 2019 (COVID-19) outbreak a public-health emergency of international concern (PHEIC). Six weeks later, the outbreak was categorized as a pandemic. COVID-19 has already caused 24 times more cases (as of 18 March 2020) than the previous coronavirus-induced PHEIC—the 2002–2003 severe acute respiratory syndrome (SARS) outbreak—and the COVID-19 numbers are expected to grow. Compared with the 2002–2003 outbreak, however, the COVID-19 emergency is occurring in a much more digitized and connected world. The amount of data produced from the dawn of humankind through 2003 is generated today within a few minutes. Furthermore, advanced computational models, such as those based on machine learning, have shown great potential in tracing the source or predicting the future spread of infectious diseases[1,2]. It is therefore imperative to leverage big data and intelligent analytics and put them to good use for public health.

Relying on digital data sources, such as data from mobile phones and other digital devices, is of particular value in outbreaks caused by newly discovered pathogens, for which official data and reliable forecasts are still scarce. A recent study has shown the possibility of forecasting the spread of the COVID-19 outbreak by combining data from the Official Aviation Guide with data on human mobility from the WeChat app and other digital services owned by Chinese tech giant Tencent[3]. Mobile-phone data already showed potential in predicting the spatial spread of cholera during the 2010 Haiti cholera epidemic[4], while leveraging big-data analytics showed effectiveness during the 2014–2016 Western African Ebola crisis[5].

However, during those recent epidemics, the large-scale collection of mobile data from millions of users—especially call-data records and social-media reports—also raised privacy and data-protection concerns. In 2014, privacy concerns urged the GSM Association (an industry organization that represents the interests of mobile-network operators worldwide) to issue guidelines on the protection of privacy in the use of mobile-phone data for responding to the Ebola outbreak[6].

In the data-intensive world of 2020, ubiquitous data points and digital surveillance tools can easily exacerbate those concerns. China, the country most affected by COVID-19, is reportedly using ubiquitous sensor data and health-check apps to curb the disease spread[7]. According to a *New York Times* report[8], there is little transparency in how these data are cross-checked and reused for surveillance purposes. For example, the report said that Alipay Health Code, an Alibaba-backed government-run app that supports decisions about who should be quarantined for COVID-19, also seems to share information with the police[8]. In Italy, the European country recording the largest number of COVID-19 cases, the local data-protection authority was urged, on 2 March 2 2020, to issue a statement to clarify the conditions of lawful data use for mitigation and containment purposes. In its statement, the authority warned against the privacy-infringing collection and processing of data by non-institutional actors (e.g., private employers).

Two weeks later, the European Data Protection Board issued a statement on the importance of protecting personal data when used in the fight against COVID-19 and flagged specific articles of the General Data Protection Regulation that provide the legal grounds for processing personal data in the context of epidemics[9]. For example, Article 9 allows the processing of personal data "for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health," provided such processing is proportionate to the aim pursued, respects the essence of the right to data protection and safeguards the rights and freedoms of the data subject.

As big data will be critical for managing the COVID-19 pandemic in today's digital world, the conditions for responsible data collection and processing at a global scale must be clear. We argue that the use of digitally available data and algorithms for prediction and surveillance—e.g., identifying people who have traveled to areas where the disease has spread or tracing and isolating the contacts of infected people—is of paramount importance in the fight against the COVID-19 pandemic. It is equally important, however, to use these data and algorithms in a responsible manner, in compliance with data-protection regulations and with due respect for privacy and confidentiality. Failing to do so will undermine public trust, which will make people less likely to follow public-health advice or recommendations and more likely to have poorer health outcomes[10].

Careful data-management practices should govern both data collection and data processing. In the collection of data from affected people, the principle of proportionality should apply, which means that the data collection must (i) be proportional to the seriousness of the public-health threat, (ii) be limited to what is necessary to achieve a specific public-health objective, and (iii) be scientifically justified. Gaining access to data from personal devices for contact tracing purposes, for example, can be justified if it occurs within specific bounds, has a clear purpose—e.g., warning and isolating people who may have been exposed to the virus—and no less-invasive alternative—e.g., using anonymized mobile positioning data—is suitable for that purpose. Furthermore, 'do it yourself' health surveillance, as it was labeled by the Italian data-protection authority, should be avoided.

At the data-processing level, data quality and security controls are needed. Data-integrity weaknesses, which are

common when data from personal digital devices are used, can introduce small errors in one or multiple factors, which in turn can have an outsized effect on large-scale predictive models. Furthermore, data breaches, insufficient or ineffective de-identification and biases in datasets can become major causes of distrust in public-health services.

Data privacy challenges not only are of a technical nature but also depend on political and judicial decisions. Requesting or warranting access to personal devices can, for purposes such as contact tracing, be more effective than simply leveraging anonymized mobile positioning data. However, compelling providers to allow access to or even assist in decrypting cryptographically protected data (similar to what occurred during the 2016 US Federal Bureau of Investigation–Apple encryption dispute) may be counterproductive, especially if the agreements between (inter) national authorities and service providers lack transparency or proportion. Similar trade-offs apply to health apps that require users to register with their names or national identification numbers.

National authorities should be mindful that precisely because personal data may contain valuable information about the social interactions and recent movements of infected people, they should be handled responsibly. Overriding consent and privacy rights in the name of disease surveillance may fuel distrust and ultimately turn out to be disadvantageous. There have been reports that China's digital epidemic control might have exacerbated stigmatization and public mistrust. This risk of mistrust is even greater in countries in which citizens place a much lower level of trust in their government, such as Italy, France and the USA[11]. Therefore, whenever access to these data sources is required and is deemed proportional, the public should be adequately informed. Secrecy about data access and use should be avoided. Transparent public communication about data processing for the common good should be pursued. Data-processing agreements, for example, should disclose which data are transmitted to third parties and for which purpose.

Reports from Taiwan show a promising way to leverage big-data analytics to respond to the COVID-19 crisis without fuelling public mistrust. Taiwanese authorities integrated their national health insurance database with travel-history data from customs databases to aid in case identification. Other technologies, such as QR code scanning and online reporting, were also used for containment purposes. These measures were combined with public communication strategies involving frequent health checks and encouragement for those under quarantine[12].

As more countries are gearing up to use digital technologies in the fight against the ongoing COVID-19 pandemic, data and algorithms are among the best arrows in our quiver—if they are used properly. ❐

Marcello Ienca ✉ and Effy Vayena
Department of Health Sciences & Technology, Swiss Federal Institute of Technology in Zurich, Zurich, Switzerland.
✉*e-mail: marcello.ienca@hest.ethz.ch*

Published online: 27 March 2020
https://doi.org/10.1038/s41591-020-0832-5

References

1. Scarpino, S. V. & Petri, G. *Nat. Commun.* **10**, 898 (2019).
2. Wheeler, N. E. *Nat. Rev. Microbiol.* **17**, 269 (2019).
3. Wu, J. T., Leung, K. & Leung, G. M. *Lancet* **395**, 689–697 (2020).
4. Bengtsson, L. et al. *Sci. Rep.* **5**, 8923 (2015).
5. Bates, M. *IEEE Pulse* **8**, 18–22 (2017).
6. GSMA. https://www.gsma.com/mobilefordevelopment/resources/gsma-guidelines-on-the-protection-of-privacy-in-the-use-of-mobile-phone-data-for-responding-to-the-ebola-outbreak/ (2014).
7. *The Economist*. https://www.economist.com/china/2020/02/29/to-curb-covid-19-china-is-using-its-high-tech-surveillance-tools (2020).
8. Mozur, P., Zhong, R. & Krolik, A. *The New York Times* https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html (2020).
9. European Data Protection Board. https://edpb.europa.eu/news/news/2020/statement-edpb-chair-processing-personal-data-context-covid-19-outbreak_en (16 March 2020).
10. Ward, P. R. *Front. Public Health* **5**, 22–22 (2017).
11. OECD. https://www.oecd.org/gov/government-at-a-glance-22214399.htm (2019).
12. Wang, C. J., Ng, C. Y. & Brook, R. H. *JAMA* https://doi.org/10.1001/jama.2020.3151 (2020)

Competing interests

The authors declare no competing interests.