

Effective public health data sharing in the context of the COVID-19 pandemic¹

Anjula Gurtoo² and Inder S. Gopal

The unprecedented crisis caused by COVID-19 has triggered several discussions on long term global public health uncertainty. One of the critical resources leading the solutions to these issues is Data. Experts from around the world are highlighting the criticality of publicly and privately owned health information that can be freely used, without unnecessary national or international legal restrictions on access or usage. The COVID-19 crisis has brought forth the practical benefits of sharing accurate and trusted public health and pandemic related data.

This article attempts to ignite the discussion on sustainable data access and security at the end of the beginning of the coronavirus crisis. We, here, try to shed some light on harmonizing health data management in general, and during a pandemic crisis. We explore the role of data friction in disseminating health information and its transformation to knowledge through alignment between public and private involvement. As the situation evolves rapidly, the predictions about economic, environmental, and social impact become highly dynamic. Public data, however, will play a critical role in mitigating the crisis remains undisputed and clear. The article forms a part of the series to understand the various data access options and data security issues in the post COVID-19 world.

How can data help in Public Health?

In the COVID-19 world, data has become central to the public discourse in a manner never seen before. Every newscast leads with data, often compiled on a state-by-state basis. How many new cases? What is the current doubling rate? How many tests? What is the positive test rate? Sophisticated statistical concepts such as geometric growth, or the mysterious-sounding “ r_0 factor”, have become commonplace and everyone cheers when r_0 drops by a few hundredth of a point.

However, the data being shared is mostly official data from Governmental sources. If Governmental data can be complemented with data from non-governmental and private sources, it can enable a vast new set of Public Health applications. These new applications will correlate data, apply sophisticated analytics and inferencing, and will be able to proactively address Public Health problems and deliver new value to citizens.

¹We will like to gratefully acknowledge the contributions of the Ministry of Housing and Urban Affairs, Government of India, and Robert Bosch Centre for Cyber Physical Systems, IISc Bangalore, in these discussions.

²Professor, Indian Institute of Science, INDIA

A prime example of such a new application is *effective contact tracing*. The recently released Aarogya Setu app from the Ministry of Health enables a form of contact tracing by using Bluetooth proximity. If location and proximity information from mobile phones could be shared without compromising privacy, a more effective application could be created. Airtel, Jio and handset software providers such as Google and Apple routinely collect data on location and location history. But privacy and security considerations severely limit the extent of their sharing, even for efforts such as contact tracing which are clearly for public good. Can a combination of technology and policy solutions help open up this flow?

Other examples include proactive *crowd management and dissipation*. This relies on crowd density information which can be inferred from aggregated location information and analysis of other sources such as video camera feeds. While privacy considerations limit sharing, crowd density data here can be more easily anonymized than location information and so solutions may be easier to create. *Quarantine enforcement* is another important application. This will require data from public health, police, or immigration officials that identify areas of lockdown or individuals that need to maintain isolation. Such information is currently disaggregated and unavailable in any consistent form.

Testing can benefit greatly from broader access to Public Health data as well. For example, a valuable service would be helping citizens find and interact with testing facilities that screen for disease or antibodies, or facilities that provide vaccine or other treatment services. In the US, Verily, a company affiliated with Google, announced about providing such a citizen app for COVID-19. However, the rollout was limited, partly by the lack of standardized data from public & private clinics and hospitals on availability and types of COVID-19 testing. Also, Public Health researchers will greatly benefit from anonymized information on testing results, categorized by location and other demographics. Such data remains hard to get in a form that lends itself to any effective analysis.

Some of these new data-enabled applications will be for Governmental or administrative use, while some will be available to consumers. While many applications are created by Governmental or public agencies for public good, some may be created by private companies for commercial benefit. All these applications will work towards better management of pandemics and improved chances or avoiding one. The one thing they all have in common is the need for accurate, secure and understandable Public Health data, that does not violate any individual's privacy.

What inhibits effective data sharing?

The public health data made available for sharing remains restricted to official data from the Ministry of Health or other governmental sources. This is only a small fraction of the total amount of public health related data. Plethora of data from public and private sources never reaches the desks of decision makers and certainly not to the nightly newscast. The generic term, *data friction*, describes the various inhibitors that create barriers in sharing data. Eliminating data friction becomes a necessary (but not sufficient) step in taking full advantage of data.

Some of these inhibitors are described below.

- *Poor quality or non-existent data* –Several critical health metrics are simply not being measured, or the measurements are sporadic and often inaccurate. Some of the reasons are cost of measurement, lack of appreciation of the value of data, and lack of access due to large numbers and their constant mobility. Poor and inefficient communication channels are also a frequent cause. The COVID crisis has spawned many creative ways to communicate data- ranging from WhatsApp texts, scanned scraps of paper, Memes, short videos to telephone calls. These methods however need to scale as well be standardized.
- *Privacy and Security concerns*-Personal health data about an individual should never to be shared without the consent of the individual. However, these concerns are often painted with too broad a brush and used as an excuse for a blanket ban on sharing data. Personal data can sometimes be shared in aggregate or with appropriate anonymization without violating privacy. And, while non-personal data is usually not subject to privacy considerations but the privacy considerations are often invoked because of lack of clarity in interpretation.
- *Finding and understanding pertinent data*- If the data universe expands to a multiplicity of sources, how does a prospective data consumer locate pertinent data? Currently no catalog or directory of data exists for use to identify and describe data. Such a searchable catalog becomes essential to create a useful data ecosystem. Even if the data source gets identified, the same type of data is often represented differently from different sources. This may be as simple as different units of measurement (Celsius vs Fahrenheit for temperature), or something more sophisticated as the object model for a data object.
- *Other sources of data friction*- Policy, cultural, legal and economic barriers to data sharing faced abound. These include erroneous secondary analyses of data, unwarranted litigation, desire to protect confidential commercial information, desire to protect intellectual autonomy, etc.

What can facilitate data sharing?

Lowering data friction facilitates data sharing. Reducing data friction in a multi-faceted task that requires a combination of technical and non-technical approaches. Some of the necessary aspects include:

- *Data sharing platforms* - Data Sharing Platforms are a basic prerequisite for receiving data from many sources, manage privacy and security in a consistent fashion, and normalize all the data through a common set of API's. These platforms also ensure control by data providers on who gets access to their data and enables them to obtain monetary compensation if appropriate. They are often called "data exchanges", and an example being the India Urban Data Exchange (IUDX)³ as an excellent base to build upon. The platforms, however, by themselves are insufficient unless accompanied by a coherent data policy for both personal data as well as non-personal data.
- *Economic models that create incentives for sharing* –Creating a broad-based data ecosystem is difficult if the impetus for sharing data is entirely top-down driven. Public and Private parties will share data if needed in response to an emergency or government mandate. But routine sharing in normal times becomes far more likely if there are economic benefits for all parties involved. Public private partnerships and data management ecosystems where government and private share for monitoring, supporting and managing for public good.
- *A culture of data sharing* - Building a culture of data sharing becomes imperative. Diverse data types, different subject domains, multiple locations, and host institutions, highlight the broad range of existing agencies and capacities that require to come together for effective use of public health data. The default behaviour of parties that have public health data should be to share, unless there are privacy reasons not to do so. Research⁴ demonstrates change in overall culture (versus agency behaviour) has long term positive implications for policy design and implementation. Culture change would include an environment of trust, common data sharing values and norms, commonality in rules, and institutionalization of interest of all parties.
- *Development of an Application Ecosystem* – The iPhone did not become a success until there were a critical mass of independent application developers building innovative applications and services for the platform. A similar application ecosystem will be required to make a data platform successful. The application developers will require nurturing and encouragement and will need clear business

³iudx.org.in

⁴<https://doi.org/10.3897/rio.3.e21773>

models for managing interactions and dependencies, for establishing and operating infrastructure systems, for monetization of data and associated services, and for developing long term expertise. Other significant factors for a cohesive, comprehensive ecosystem would include tools and support, and setting up the basic values cum philosophy of the ecosystem.

Conclusions and Cautions

Maximum value gets created through the ability to combine data from different sources at the same time. For example, by combining crowd density data with incidence and location of new cases, one can gain analytic insights into the efficacy of social distancing. Other examples of new applications possible by combining data from different sources include superior contact tracing, proactive crowd dispersal, and improved quarantine tracing and enforcement.

However, some cautions remain. Surveillance for public health benefits is necessary but at what stage does the surveillance become an unethical and violation of human rights? Recently, information on COVID-19 patients and those placed in quarantine has found its way to the public domain. District administrations of Mohali and Karnataka published a list of people quarantined. Such disclosures may seem right from a common man's point of view, but they have a devastating impact on individual liberty. Several doctors and nurses treating coronavirus patients were forcibly evicted by their landlords fearing that would make them susceptible to COVID-19. Stigma and harassment await those who are undergoing quarantine as well.

The role of government here, therefore, becomes critical and needs discussion. While, the COVID-19 experience in China with an exclusively government-controlled information flow, has shown that such information monopoly can have a deadly impact, a disaggregated and distributed government system as exists in India or the US, such information monopolies can arise at different levels of government. Thus, this becomes a matter of national security to propagate the tools and systems that allow non-governmental players to participate in the ecosystem of public health data, both as providers and consumers.