Inter-American Development Bank

Is data privacy the price we must pay to survive a pandemic?

Social Sector

Marcelo Cabrol
Ricardo Baeza-Yates
Natalia González Alarcón
Cristina Pombo

DISCUSSION
PAPER Nº
IDB-DP-00763

April 2020

Is data privacy the price we must pay to survive a pandemic?

Marcelo Cabrol
Ricardo Baeza-Yates
Natalia González Alarcón
Cristina Pombo

April 2020

IDB
Inter-American
Development Bank

Is data privacy the price we must pay to survive a pandemic?[1]

Marcelo Cabrol (IADB), Ricardo Baeza-Yates (Northeastern University), Natalia González Alarcón (IADB) y Cristina Pombo (IADB)

The Inter-American Development Bank's (IADB) fAIr LAC initiative[2] promotes the ethical[3] and responsible use of data and artificial intelligence -based systems, in order to center them on the individual, according to principles of equity. To this end, all actors in the ecosystem must respect, among other things, the democratic values of citizens, such as privacy and data protection. Even in times of maximum risk to public health, the individual choice between personal privacy and social choice of our well-being must be made compatible. This is an ethical issue of first order that is especially relevant during the current pandemic. Therefore, we must discuss it thoroughly.

In the fight against COVID19, billions of personal geo-location data points are being used by countries around the world to "flatten the curve" of contagion, in order to re-establish the circulation of people, and to better manage physical distancing among individuals. We are referring to the surveillance systems that several governments have begun using **to track or trace individuals and their physical contacts** (contact-tracing). Through these digital tools, governments seek to track the movement of infected individuals, identify those who might have been exposed to coronavirus, trace their physical contacts to alert and warn of future risks, monitor physical distancing orders, publish maps that identify risky areas and thus successfully implement and enforce measures to contain the contagion. This use of technology can be controversial given the implications it has regarding privacy risks and the decisions about it that some countries are taking. If we understand data as a public good (non-rival and non-exclusive) necessary to improve and accelerate response in the midst of a pandemic, is it possible, then, to justify the relaxation of privacy standards? Does the goal of tracking and controlling contagion as a measure to ensure everyone's health and re-establish some social normalcy justify the possibility of intrusive government surveillance? And if so, are there conditions that can mitigate the privacy risks to which citizens are exposed?

1 The authors thank Marcelo D'Agostino, Senior Advisor on Information Systems and Digital Health at the Pan American Health Organization / WHO, Alejandro Pardo of the IDB Lab, Jennifer Nelson and Luis Tejerina of the IDB Health and Social Protection division, Pablo Picón and Roberto Sánchez of the IDB Social Sector for their comments

2 As a result of the region's growing interest in using artificial intelligence to solve social problems, and the challenges that this implies, 2019 fAIr LAC was born in with the aim of promoting the ethical and responsible use of AI to improve the provision of social services.

3 Everything that falls within the ethical principles defined by the OECD.

Tracking People and Their Physical Contacts

According to the WHO, **contact-tracing systems** work in three steps, (following the traditional method of contact-tracing through specialized, trained public health personnel and personal interviews):[4]

1. In confirmed cases, identify individuals who encountered the infected patient,
2. Record potential physical contacts of infected patients and contact them, and
3. Follow up with the list of contacts either to test them or to warn them that they have been in contact with someone infected.

These contact-tracing tools have become part of the pandemic-management strategy of several Asian countries that have suffered the first wave of the pandemic. The strategy is gaining strength among several groups of experts[5] who have reached a significant level of consensus on the effectiveness of contact-tracing to halt contagion without having to do full lock downs. This is because the application of these contact-tracing tools would allow users to know of their potential exposure to an infected person, to trace their history of in-person contacts, or to identify risk zones, and thus take appropriate actions, such as physical distancing. The effect would be to reduce contact-rate faster than potential cases (those exposed to the virus), given that these would be eliminated from the chain of contact. One weakness of these applications, however, is the large percentage of asymptomatic cases, estimated to be between 18% and 42%, that go undetected and may well be the main source of infection.

These kinds of public health interventions can be backed up with different technologies. The state of Massachusetts has decided to adopt the traditional, manual form of contact-tracing, hiring over a thousand people to contact diagnosed individuals and thereby conduct tracing. However, the process takes too long to make manual contact-tracing efficient.

In South Korea they track people via GPS to have control on COVID-19 cases

South Korea's case is being studied attentively due to the success that has had in controlling the pandemic. The government has used different technologies to prevent and control contagion, but the base has been an elevated sanitary capacity capable of implementing mass COVID19 testing and making it available to the population followed by a high digital management capacity that has allowed rigorous monitoring via tracing contacts of infected and quarantined people, to identify any individual with whom carriers of the virus may have been in contact. Up to March 25th, South Korea has carried out more than 350,000 tests, the highest number of tests per capita worldwide. Meanwhile, other Asian countries are turning to the use of high-tech tracing bracelets to monitor positive cases. More than 20,000 travelers who arrived in Hong Kong received tracking bracelets that the government acquired to monitor the movements of people in quarantine.

Singapore, meanwhile, launched TraceTogether on March 20th; within five days, it had been downloaded by 735,000 people, approximately 13% of the population.

4 The WHO has its own app, Go Data (without AI) that does this: http://socialdigital.iadb.org/en/solutions/go-data-covid-19

5 Such as the group of experts from the Massachusetts Institute of Technology that is developing Safe Paths, as well as the group of experts of various specializations from Oxford University that analyzes the implications and requirements of the use of contact-tracing.

This app, unlike aforementioned apps, works as a peer-to-peer distributed computer network, using short-distance Bluetooth connections. This creates a decentralized network, without a control center that receives all the data, and by dint of that, protects individual privacy. Taiwan uses a different approach, tracking the location of quarantined people's phones by using cellphone tower data. If the system detects people straying outside their boundaries, it sends them a text message and alerts the authorities.

Now, the chief challenge of the deployment of these applications are the limits of privacy and the management of informed consent.[6] In a privacy-friendly system, the individual owner of personal data must understand and accept the way their data will be used or treated, and with the use of contact tracing, these risks exist as much for the individual as for the general public.

Since the application must analyze the travel paths of **carrier patients** for the past 15 days in order to identify other people potentially at risk, these carrier patients run the greatest risk of having their privacy being exposed. Even if their personal information is not explicitly made public, they may be identified when limited number of location data points is made public, as was the case with Patient 31 in South Korea. This problem is not new and already occurs with purchasable data from gaming applications, allowing specific individuals to be identified and security agents from heads of state to be tracked.

Application users are also at risk, since their location data is used to establish whether they have encounter or crossed paths with a diagnosed patient. The problem here is when a third party, commonly the government, has access to location data. For example, Alipay App, the Chinese application used in more than 200 cities to help citizens identify their symptoms and risk of contagion, is based on a QR code system where each user is assigned one of three colors: green, yellow, or red, according to their location, basic health information, and travel history. Several problems arise. In the first place, there is a dearth of information:[7] neither the company nor the government have explained in detail how the system sorts people into each color, which causes fear among people who receive a self-isolation order without being told why. In the second place, there is a lack of transparency regarding how their personal data is being used and where it is being stored. The application requests contact details, passport information, recent trips, and a medical certification. According to a New York Times analysis, everything seems to indicate that the application shares information with the police, establishing a new method of automated social control that might continue in use even after the epidemic. Meanwhile, Apple and Google have joined efforts, to ensure that all the tracking applications recommended by the governments automatically download themselves when a cellphone updates its operating system.

In China QR codes identify the risk of contagion

Finally, the risks for **the general public** spring from the fact that users and non-users are linked by social relationships and spatial proximity. If a family or friend is outed as a carrier of the virus, close non-users may suffer stigmatization and social repercussions. Consider for a moment that you are in quarantine for the illness: do your neighbors need to know? There have already been cases of health profes-

6 Access to individual location data depends on active regulation in each country or state. See the USA's case here.

7 For more information, see page 25 of fAIr LAC's technical note here.

sionals who have been asked to relocate temporarily, due to neighbors' fears of contracting the illness.

What should be prioritized in the use of data to monitor the pandemic?

When facing the dilemma between protecting privacy and protecting health, how does the population of Latin America and the Caribbean perceive these measures during the pandemic? We do not yet know. A study conducted in 2006 showed a not very favorable public perception towards the use of various measures during a possible outbreak, amongst the four regions included in the survey. It is important to have conversations with the public and to include them so they can participate in the decision of whether to use these measures.

Some of these applications have designed privacy safeguards applying the principle of "privacy by design," clearly reconciling the right to privacy with the needs of pandemic control. Singapore's TraceTogether affirms that it does not collect or use any kind of location data, and that it does not access the user's contact list or address book. It only uses Bluetooth data to establish a contact and it does not store information about where the contact occurred in a centralized place. Additionally, the government is in the process of opening the source-code so that it can be audited.

According to Ramesh Raskar, MIT Media Lab researcher and the lead developer of Private Kit: Safe Paths, a contact-tracing application that seeks to protect user privacy, infected individuals using Safe Paths can erase sensitive or potentially identity-revealing location data points. Moreover, the location data of all non-infected users is stored only on the users' devices, and not on the server. In this manner, users are the only ones who will know whether or not they crossed paths with an infected patient. Another similar initiative, also developed in MIT, uses Bluetooth to alert people without revealing their identity.

Bluetooth technology is capable of alerting to possible nearby infections without leaving a geolocator trace

Meanwhile, Europe, the region with the strictest privacy protection laws in the world, has begun to make decisions on the matter at hand. According to the European Data Protection Board, location data can only be used by the operator when anonymized or with user consent. However, GDPR allows competent authorities to process personal data during an epidemic in accordance with national legislation, since in exceptional circumstances prior consent is not necessary.[8] German Chancellor Angela Merkel announced the development of the Pan-European Privacy Preserving Proximity Tracing (PEPP-PT), a people-tracking application that **protects data and does not store user location.** Like TraceTogether, it will use Bluetooth to record the proximity of a user to other users with great precision, but with less range than GPS, so that people receive a message if they have been in contact with a virus carrier. Both, Belgium and Austria, are using databases from telecommunication companies, combined with health data, under the supervision of the corresponding authorities, to ensure the use of **aggregated and anonymized data** to assess how the virus spreads and which areas are at high risk. In Germany, Deut-

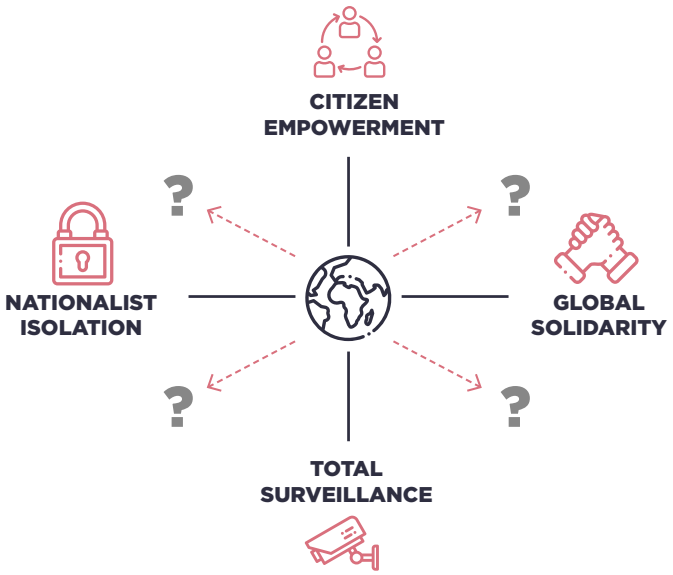
8 Moreover, these measures must be strictly limited to the duration of the emergency. In this case, the pandemic.

sche Telekom has provided data to the Robert Koch Institute, the government's public health agency, on an aggregate basis. Similarly, the British government is in talks with cell phone operators about accessing similar data. A recent study found that 51% of people surveyed in the UK were concerned about the privacy of their data is being increasingly used, which is even more relevant in the healthcare sector, because information which is shared through people-tracking apps is highly sensitive.

Conclusions

An essay published by Israeli historian Yuval Harari describes the management models of the privacy-health dilemma on two axes: The first axis spans between "totalitarian surveillance" and "citizen empowerment." The second, between "nationalist isolation" and "global solidarity." The following figure places Harari's world after coronavirus on these axes to illustrate the questions that this debate suggests, at least for the duration of the control and management of the pandemic.

Figure 1: graphic by Mario Tascón based on Yval Harari's article in the Financial Times



The different alternatives that governments take to track and geo-locate their citizens and the technologies they use have different implications regarding people's privacy. Some of these applications exceed the privacy standards that governments have under normal conditions in the use of citizens' data. The conundrum is to determine under which conditions, if any, a government should have access to this data, under which circumstances, and for how long. Is increased location monitoring necessary to protect public health amidst a pandemic? Or is it an intrusive government surveillance act? Can tools be developed that minimize privacy risks? Does exceeding privacy limits erode trust in authorities? What are the best practices and policies to apply during and after the pandemic? Is it possible for citizens to surrender their data only for the duration of the pandemic? When do exceptional powers end, and what is done with the collected data? It should be noted that the answers to these questions are different from one culture to another and even

from one country to another, since they depend on the level of democracy and the degree of trust people have in their governments.

To answer these questions and resolve this debate, we must bear in mind that people-tracking applications must have both social requirements, due to ethical aspects, and technological requirements, due to the objectives they want to achieve. For these requirements we must consider the following social and technological factors, among others:

- **Maintaining the balance among individual privacy, freedom of movement, and the collective interest of health is an ethical matter.** However, each environment will require the most appropriate technology for its particular circumstances. Solutions like TraceTogether have been successful because Singaporean society is committed to using it, and thanks to its rapid adoption, it has served its purpose. Several European countries have chosen to use mass information from their telecommunications operators to cross-check with health systems, while others have opted for more privacy-friendly alternatives. The same debate has only just begun in the United States.
- **We must consider the consequences of asking people to undergo voluntary testing and to make their situation public**, especially minorities or people targeted for their race, religion, or sexual orientation. The guise of public health must not be used to obtain information that could be used to violate human rights.
- **There are countries in which parts of the population are still without adequate mobile technology.** These areas will require specific solutions, as well as consideration of other socioeconomic factors, such as informal economies that can complicate the adoption of more adequate technologies.
- Another key element is **the effectiveness of the technological solution**. Once it is decided which technological model is the most appropriate for each country's circumstances, there might be already available alternatives developed and tested by other countries. Taking advantage of already-established solutions will reduce deployment risks and expedite their adoption.
- **The use of encryption protocols and/or differential privacy needs to be explored in greater depth** to improve the protection of people's privacy while managing the data needed to control an epidemic. The pandemic has not given us time to do this properly.

Lack of privacy also has another consequence: it **erodes the citizenry's trust** in the state. Furthermore, **lack of government transparency** regarding the use of personal data, combined with a **lack of information** on how these applications make decisions (as in the case of Alipay in China,) diminishes public trust. Although citizens are becoming accustomed to coexisting with the era of big data, and we hear of various success stories, there are also cases of lack of transparency and misuse of data, which reduce trust.

The above evidences the need to prioritize personal privacy. Therefore, governments ought to tend towards the use of aggregated and anonymized data, so that they can also make the data open, as has recently been proposed in Chile. On the other hand, most people surrender a part of their privacy whenever they install an

application, without even knowing which part of their privacy they are surrendering, since they do not read the terms and conditions that they accept. If this is the case, it's possible that losing something that has already long been lost will not be relevant for many people, particularly younger people Can it then be feasible to temporarily relax data privacy during a state of alarm?

Finally, it is crucial to evaluate **the quality of the data** that is being used to contain or report on the effects of COVID-19. In the absence of good and sufficient data, results will be weak and erroneous. Considering that we are dealing with health information, weak and erroneous results can be catastrophic (garbage in, garbage out). In fact, if the data is not high-quality or standardized (comparable indicators), it is often preferable not to have it. This also negatively affects credibility and trust towards technology and its promoters, which in this case are governments. An additional issue is how this data and how long can or should be archived for the future.

Returning to the title of this essay we see that the question we were asking poses a **false dichotomy,** since, as we have seen, it is possible to use technology well without needing to surrender either all our privacy or our fight against the pandemic. In the end, as in other aspects of life, and especially in health, the solution is not in extremes, but in consensus, networking and the search for collective solutions. The pandemic has not given us time to do this as it should, and there are already voices demanding that privacy be the norm.