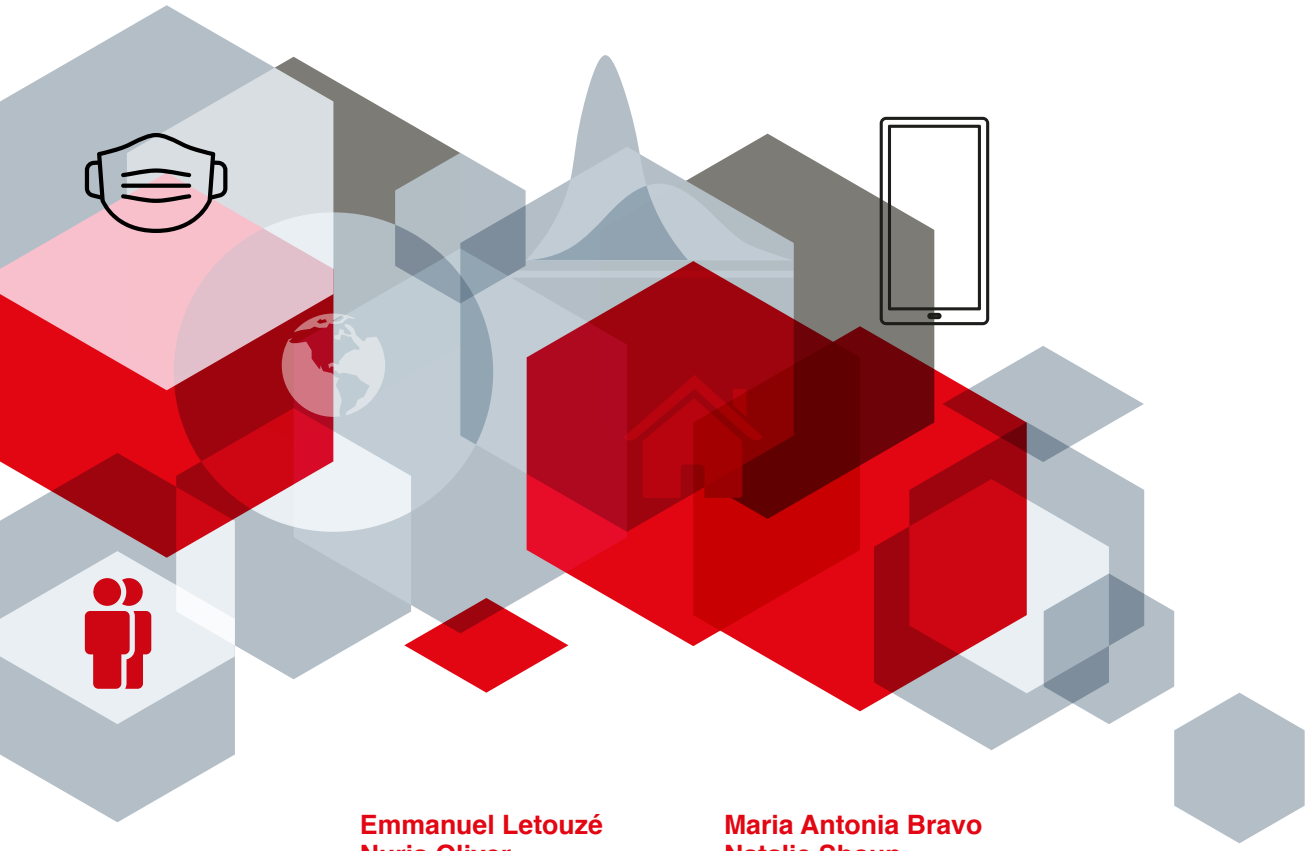
DATA-POP ALLIANCE

Vodafone Institute for Society and Communications

digitising europe

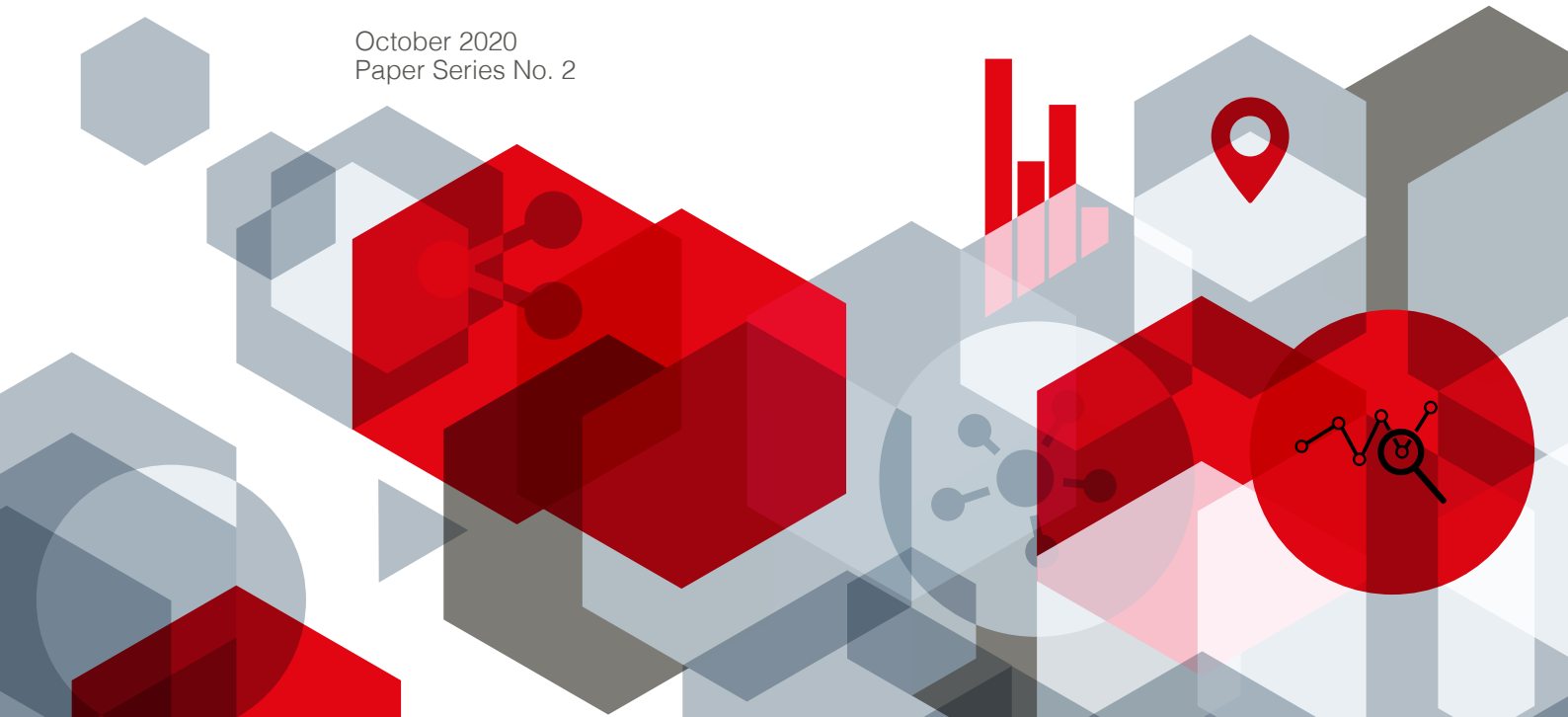


Emmanuel Letouzé
Nuria Oliver
Maria Antonia Bravo
Natalie Shoup

Using Data to Fight COVID-19

And Build Back Better

October 2020
Paper Series No. 2



Content

About this paper

This paper is the second in a Paper Series published jointly by Data-Pop Alliance and the Vodafone Institute for Society and Communications, following *"Sharing Is Caring: Four Requirements for Four Key Requirements for Sustainable Private Data Sharing and Use for Public Good*, published in November 2019.[11]

It was written by Emmanuel Letouzé, Nuria Oliver, María Antonia Bravo and Natalie Shoup. It benefited from comments and inputs from Inger Paus (Managing Director, Vodafone Institute for Society and Communications), Pedro Rente Lourenco (Lead Researcher and Data Scientist, Vodafone Group Big Data & AI), Matthew Allison (Senior Public Policy Manager, Data, Platforms & AI, Vodafone Group), Julia Ebert (Senior Research Manager, Vodafone Institute for Society and Communications) and Juan Camilo Mejía (Program and Research Manager, Data-Pop Alliance).

The report was copy-edited by Carola Miras and Juan Camilo Mejía.

1 https://www.vodafone-institut.de/wp-content/uploads/2019/11/DPA_VFI-Sharing-1.pdf

Executive summary

Since the start of the COVID-19 pandemic in the first quarter of 2020, numerous governments and public institutions around the globe have developed or promoted initiatives leveraging digital data and technology in support of response efforts. Some have sought to identify and predict hotspots, others to evaluate the effectiveness of containment policies or to detect and trace the close contacts of infected individuals. As many countries are still struggling with the first wave of the disease and several are fearing or already grappling with a second—often in tumultuous socio-political contexts—it is essential to take stock of the key features and expected benefits of major initiatives and summarize the main debates and questions they have raised—about their usefulness, implications, limitations, risks and requirements in the fight against the pandemic, and beyond.

The pandemic has also laid bare long-standing and deeply rooted structural fault lines in our world. Far from hurting everyone indiscriminately, the virus and its socioeconomic effects have affected disproportionately poor people, people of color, women, people with disabilities, migrants, and peoples governed by populists.

Some, such as the outgoing UN Special Rapporteur on Extreme Poverty and Human Rights, even consider that COVID-19 has "revealed a pandemic of poverty that benefits the rich".[2] Others have argued that the pandemic is more of a *syndemic*, which refers to a health issue that clusters along social lines.[3] With COVID-19, it is as if the veil of feigned ignorance about the features, drivers and effects of injustices, such as the indecent growth in income concentration and inequality, the differential impacts of environmental degradation and pollution, systemic racism and sexism, and even the risks posed to democracy around the globe, including those fueled by digital data and technology, had been shredded in a few months.

In this context, digital data and technology serve as lenses on the world and as levers of change, for good or bad. A decade into the "data revolution" and with a decade left to make progress towards the Sustainable Development Goals (SDGs), the current crisis provides a unique opportunity to ask how digital data and technologies can truly and structurally improve our world by both fighting the pandemic and "building back better". It is evident that reliable and timely data are of paramount importance to fight the pandemic. Yet, they are of no use if they are concealed or manipulated for and by officials interested in scoring political points, drowned in an ocean of dubious claims and rumors, or not effectively communicated and understood.

2 https://amp-theguardian-com.cdn.ampproject.org/c/s/amp.theguardian.com/global-development/2020/jul/11/covid-19-has-revealed-a-pre-existing-pandemic-of-poverty-that-benefits-the-rich

3 https://doi.org/10.1016/S0140-6736(20)32000-6

Similarly, it is clear that more advanced initiatives leveraging digital data and technology that are at the core of this paper—such as contact tracing applications or hotspot detection algorithms—can and must play a role in fighting the pandemic. But these digital "solutions" are not, as the saying goes, "silver bullets" that will solve our human-made problems by themselves. We are once again experiencing the very real risk of jumping to "technological solutionism" without understanding and addressing the key implications—technological and scientific, political, economic, ethical—of new data and technology.

Fundamentally, this crisis ought to be a moment in our lifetimes when we reassess our ways of life, our incentives, our priorities, and push for real change with some of the most powerful tools available: data and technology. We should use this crisis as a testbed and catalizer for how data and technology could help us set and achieve humanistic societal objectives, as underpinned by the SDGs and other frameworks—and not just serve the interests of surveillance agencies and large corporations. **This paper therefore explores how data can help fight COVID-19 and how COVID-19 also provides an opportunity to better use data to build back better.**

To realise this vision, four elements appear to be key:

One is context: we need to have a thorough understanding of the goals, implications and the impact on citizens and society of decisions in the longer term (from a science/technology, economic/commercial, social, political, legal and ethical point of view). It is also important to understand the different technologies being designed and used for real response, as well as the parameters and risks, benefits, limitations and impact of each. Furthermore, it is crucial to be mindful of the fact that not all responses can or must be digital, and that not all people will be able to access digital solutions. This means that solutions have to be thought in a holistic way so that everyone is included.

Another is education: citizens should be provided with clear, precise, understandable information. Huge amounts of dis- and mis- information are being produced about and around the pandemic, which makes it difficult for the non-expert to discern the differences between facts, hoaxes and everything in between, which feed on and fuel political polarisation. Social media companies and social platforms have a duty to the public to provide safeguards from theories that weaken trust in their governments and in science. Beyond citizens, the COVID-19 pandemic has brought to light the evident lack of data and digital literacy among many public officials and decision makers, with potentially devastating consequences. Education and the long-term collaboration of a diverse set of experts in relevant areas—such as data science, epidemiology, anthropology, computer science, immunology, public health, economy and sociology—with public administrators must be ensured to assist in more evidence and knowledge-driven decision making. These collaboration of a diverse set of experts need to analyze the incentives and constraints of participants and work together to accomplish beneficial outcomes for all parties.

A third one is evidently high-quality data: to fuel better human systems to both fight the pandemic and build back better, data are one of the most powerful tools at our disposal. Data must be allowed to be shared and analyzed in privacy-preserving, interoperable manners. Decision makers and citizens should be both informed and involved in what data are being collected and how; what they represent; how and why they are stored and potentially shared in raw or transformed forms. Data regulators and controllers have a key role to play in ensuring appropriate safeguards with regards to privacy, consent and inclusion of data subjects, and to help navigate the trade-offs between emergency situations and long-term conditions.

A fourth one is communication and trust: a privacy-sensitive society requires transparency and confidence in the use of the data collected. Honesty and transparency are key to building trust, in addition to competence (i.e. efficiently carrying out the task at hand) and reliability (i.e. competence sustained over time). The current situation has been enlightening for different stakeholders, showing that even though data could be the solution to some realities, there are many different groups that are inevitably less connected and therefore not accounted for. This reality means that data and technology may have contributed to spreading—just as much as to curbing—the pandemic, and this fact must be acknowledged, communicated and addressed. What can and cannot be achieved by these technologies must be communicated transparently so that citizens and societies can effectively use and accept them when they are deployed.

Social media companies and social platforms have a duty to the public to provide safeguards from theories that weaken trust in their governments and in science.

Once these various digital technologies are fully understood, it is important to critically interrogate the wider implications beyond immediate pandemic response. Such implications should lead to a set of guiding principles impacting how each of them is designed, developed and deployed.

With this in mind, we put forth six main recommendations:

1 Think and act boldly and decisively—now. This may be a once-in-our-lifetime opportunity for deep, ambitious and long-term thinking, especially to **fight deep-rooted inequalities** and excesses fueled by complacency and greed that have been exposed and exacerbated by the pandemic. Now is the time to design, deploy, test and scale digital data and technology approaches to enable long-term positive social transformation.

2 Only deploy data and technology that are fit for purpose. Despite its promise, technology is no silver bullet. Its strengths and limitations should be acknowledged. How to balance digital and non-digital technology solutions is of paramount importance. Furthermore, technological solutions should be thought of as **enablers**, **integrated** with existing structures when they perform well, such as public health systems. They should also have clearly stated rationale and purpose and be systematically evaluated. Given the already staggering digital divide, omnipresent structural inequities and biases, we need inclusive solutions so that large segments of the population are not left by the wayside.

3 Place people at the center and "in the loop" at all times. Privacy and human rights should be core considerations. Simulations of unintended consequences from ethical and human rights perspectives should be performed and potential risks minimized before implementing and deploying any technology. Social and behavioral responses to digital technology interventions need to be anticipated and embedded in the design of tools and apps. This requires large-scale **public consultations**, digital public spaces such as "online parks"[4], critical governance and accountability mechanisms, on line portals and local forums to ensure that citizens are informed and can actively participate in outcomes.

4 Develop "data literate" human and data systems. A major challenge and objective over the coming years will be to actively strengthen **"data literacy"** among both governmental agencies and citizens—defined as "the desire and ability to constructively engage in society through and about data".[5] This will mean building data skills and culture through capacity building support in order to base discussions and decisions on facts. Building a data culture and systems of interoperability is also key yet it is missing: it should work across distributed networks and systems thereby ensuring usability between different apps within or across different countries.

5 Test and scale sustainable business models. Now is also a good time to think broadly and boldly about sustainable business models for private-public data sharing and use. Today's data boom and raised visibility of digital solutions are great incentives for the private sector to allocate more resources into data sharing for the public interest, to formalize **public-private-people partnerships** (PPPP) and develop and test "freemium models" that would ensure financial sustainability. At a European level, research funding should be devoted to foster PPPP Data4Good research consortia within the next EU Horizon 2027 program.

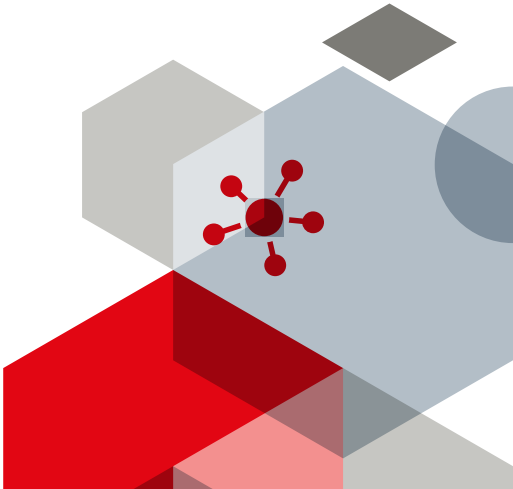
6 Consider and use regulation as an enabler. Regulations must support **enabling principles** such as (1) encouraging data sharing through voluntary, market-driven mechanisms; (2) sharing only under legally compliant, ethical and socially acceptable scenarios, in line with the principles of trustworthiness and privacy-by-design; (3) data for good initiatives should be subject to fair remuneration, thereby creating the conditions for products and services; (4) technologies should be fit for purpose and with a human(ity)-centric perspective. Let us not forget that that technological breakthroughs throughout history are often precipitated by a crisis, and then adapted and reused elsewhere, both for good and bad. Good harmonization of regulation is key to ensure that initiatives can be scaled up quickly, as appropriate, and sustained over time.

The COVID-19 pandemic—or syndemic—presents a historic opportunity for all parts of societies—the private and public sectors in collaboration—to organise themselves and collectively build back better following a human-centric approach to, and use of, digital data and technology. Let us not miss it.

Questions and comments about this paper can be sent to eletouze@datapopalliance.org.

4 https://www.wired.com/story/to-mend-a-broken-internet-create-online-parks/

5 Data-Pop Alliance Data Literacy White Paper, 2015



Introduction

Over the past few months, many governments and public institutions around the globe have developed or deployed initiatives leveraging digital technologies and privately held data in support of COVID-19 response efforts. Some resources aim to identify potential hotspots or demonstrate the effectiveness of containment policies, while others seek to trace infected individuals' close contacts, amongst others. The usefulness and implications of these initiatives—notably but not only contact tracing applications—have been widely debated. Meanwhile, many countries are still struggling with the first wave and several are in the midst of a second—often in tumultuous socio-political contexts.

Concomitantly, structural fault lines around the world have been laid bare in all available data: far from hurting everyone indiscriminately, the COVID-19 crisis and its effects have disproportionately affected people governed by populists, the poor, people of color, women, persons with disabilities, migrants and children. Some consider that COVID-19 has also "revealed a pandemic of poverty that benefits the rich".[6]

This historical context provides a unique, perhaps once-in-a lifetime, opportunity to reconsider our life styles and to optimise this significant potential for change. A key question posed is how digital data and technologies can truly and structurally improve our world by both fighting the pandemic and "building back better", i.e. not satisfying ourselves with returning to business as usual, but rather capitalizing on this dramatic event and allowing novel, ambitious projects and ideas not only to emerge but also to garner public support—pending the development of a vaccine or other effective treatment.

What makes infectious diseases unique is that they thrive on human interaction. In doing so they serve as a litmus test, revealing how societies function, rendering visible the world's inner workings and flaws. Thus, while data and technology are seen as increasingly relevant for pandemic response strategies, crises offer an opportunity to step back, examine and hopefully improve our current systems and societies. While there is no doubt that using data has significant potential for fighting COVID-19, challenges and questions about the requirements and long-term applicability of digital technologies must be identified and addressed.

Assessing the effectiveness, security, privacy, ethical and trust implications of these digital responses to the crisis is indispensable to combat the epidemic and overcome it rapidly. However, it is equally essential to ensure that the longer-term impacts of the models, protocols and applications created in the midst of this crisis safeguard fundamental rights and promote a renewed human-centric vision rather than a techno-solutionist approach that may enhance the very conditions that contributed to the magnitude of the pandemic's impact, such as structural inequalities. As we unpack these questions in a dire and urgent context, it is essential not to lose sight of the trade-offs and risks that putting our trust in technologies may entail, and how these tools could be leveraged to improve tomorrow's world.

With these points in mind, this paper is structured as follows:

Section I describes initiatives that use digital technologies and privately held data as part of pandemic response strategies, unpacking how these initiatives work and providing examples from several regions. Section II summarizes key questions and concerns these initiatives have raised across four main domains: technological and scientific, commercial and economic, ethical and legal, and political spheres.[7] Section III discusses recommendations to meet the challenges of today and tomorrow by leveraging data and tech in the fight against COVID-19 and potentially other pandemics, as well as the scourge of global poverty and inequality.

6 https://amp-theguardian-com.cdn.ampproject.org/c/s/amp.theguardian.com/global-development/2020/jul/11/covid-19-has-revealed-a-pre-existing-pandemic-of-poverty-that-benefits-the-rich

7 Based on the taxonomy proposed in the publication Sharing is Caring: Four Requirements for Sustainable Private Data Sharing and Use for Public Good co-developed and published by Data-Pop Alliance and the Vodafone Institute in November 2019. See https://www.vodafone-institut.de/studies/four-key-requirements-for-sustainable-private-data-sharing/

1. COVID-19 digital initiatives

The way that data and technologies are leveraged and positioned in the COVID-19 response presents a real opportunity for greater visibility, collaboration and evidence of impact for digital solutions. However, it also harbors significant risks given the speed with which governments and companies are obliged to react and make decisions about data use, privacy, oversight and accountability in developing and implementing these solutions. The scope of data considered in this paper centers mainly on that generated and/or enabled by interactions with, or between, mobile phones (both feature phones and smartphones): a highly sensitive issue with almost all publics.

Digital technologies based on the analysis of large-scale human behavioral data are being touted for their prospective usefulness to combat the pandemic. Given the ubiquity of cell phones, mobile phone network data has been one of the first sources of privately-held data that many countries—both developed and developing—have turned to in COVID-19 response efforts. Moreover, data capturing the interactions with, and between, smartphones has also emerged not only as a promising and rich source providing the public with information about the virus, but also as a critical source of information for decision makers and authorities.

The COVID-19 context has also opened discussions on using data collected by additional technologies such as smartphone apps (e.g. Facebook, Google Maps), search engines (e.g. Google searches) or social media platforms (e.g. Twitter feeds), facial recognition systems, satellite and surveillance devices, bank and credit card transactions, public transportation systems, electronic health records and funeral homes to aid governments in their responses to contain the spread of the virus. In this context, expectations as to what technologies and applications can really do to enable better responses and policies remain high. However, given that these applications frequently rely on collecting, sharing, storing and analyzing personal, and often quite sensitive data, it is critical to assess the possible unintended consequences that may arise from sharing and using such data.

A. Types and taxonomies of privately held data sources

The functions and promise of many of the mobile phone applications evoked above for combating the spread of COVID-19 are grounded in their ability to make use of mobile data in order to map hotspots of infection, determine changes in mobility patterns, or track contacts between at-risk or infected individuals. The ubiquitous nature of our mobile devices and the fact that human mobility is one of the key factors in the spread of an infectious disease make these devices a formidable tool to understand and measure our movements.

The most widely used types of location and proximity data collected by cell phones in the context of the pandemic are summarized in Box 1. Each of them has its strengths and weaknesses, with varying degrees of privacy implications.

Given that location and proximity data can be key sources of information for understanding the spread of a pandemic, analyzing how digital technologies and applications can be used to safely collect and harness data, and studying the ways they are being used—or proposed—is key to gage the opportunities and risks of these solutions for COVID-19 response efforts.

B. Technologies, applications and uses

In this paper we consider technological tools developed from the application of scientific knowledge to raw materials for practical purposes, i.e. digital technologies. Several applications have been identified below within the scope of pandemic containment using combinations of technologies and the different data sources detailed above.

Box 1. Location and proximity data: how mobile devices can be used to infer your position[8]

GPS: Mobile devices can determine their location using the global positioning system (GPS) through the device's GPS chip which receives signals from satellites orbiting the earth. Accuracy of GPS signals is variable and tends to be less so in urban areas or indoors. GPS signals are detected primarily through the device's operating system or through mobile applications where the user is asked to opt-in to sharing their location. They can also be detected by wearable devices or navigation systems to provide location data. When analyzed individually, GPS location data is subject to privacy regulations, given its sensitive nature.

Base transmitter stations (BTS): BTS—or cell towers—facilitate signal reception of cell phones and other wireless devices. Thus, carriers are able to know where devices are, based on which tower they connect to for services as well as the signal strength of the connection. Given that each tower has a unique ID, from the tower ID and the signal strength one can infer a device's location. BTS location information is useful for inferring aggregate mobility patterns but not highly accurate in location tracking of individuals, as their spatial granularity depends on the density of cell towers in a region. For instance, two devices connected to the same rural BTS could in fact be kilometers apart.

Wi-Fi: Wi-Fi signals tend to provide more accurate indoor location data and can often generate more granular data. Mobile devices can scan for nearby Wi-Fi networks and crowdsource location. Nearby networks or "access points" can include any Wi-Fi signals in the vicinity, such as that in cafes and shops or neighbors' homes.

Bluetooth: Bluetooth technology is common in portable devices and can be thought of as a beacon that broadcasts one-way signals which other devices can pick up (think of connecting your phone to wireless headphones) when enabled. This occurs through bursts of information packets dispersed into the electromagnetic spectrum, which other Bluetooth-enabled devices then detect. No direct connection has to be established, as devices exchange identifiers. Bluetooth can be used to infer location or proximity. In the case of location, a registered device in a known location can infer the locations of other devices that are visible to it via Bluetooth with a certain signal strength. In the case of proximity, Bluetooth-based signals can be sent to other devices within a certain range to collect proximity data rather than absolute location. Bluetooth-based proximity information is generally more privacy preserving than absolute (e.g. GPS, Wi-Fi) location data.

Many devices use a combination of GPS with other forms of location signals such as Wi-Fi and Bluetooth to improve the precision of the devices' location.

8 https://fpf.org/2020/03/25/a-closer-look-at-location-data-privacy-and-pandemics/; https://theintercept.com/2020/05/05/coronavirus-bluetooth-contact-tracing/. https://gimbal.com/location-data-guide/

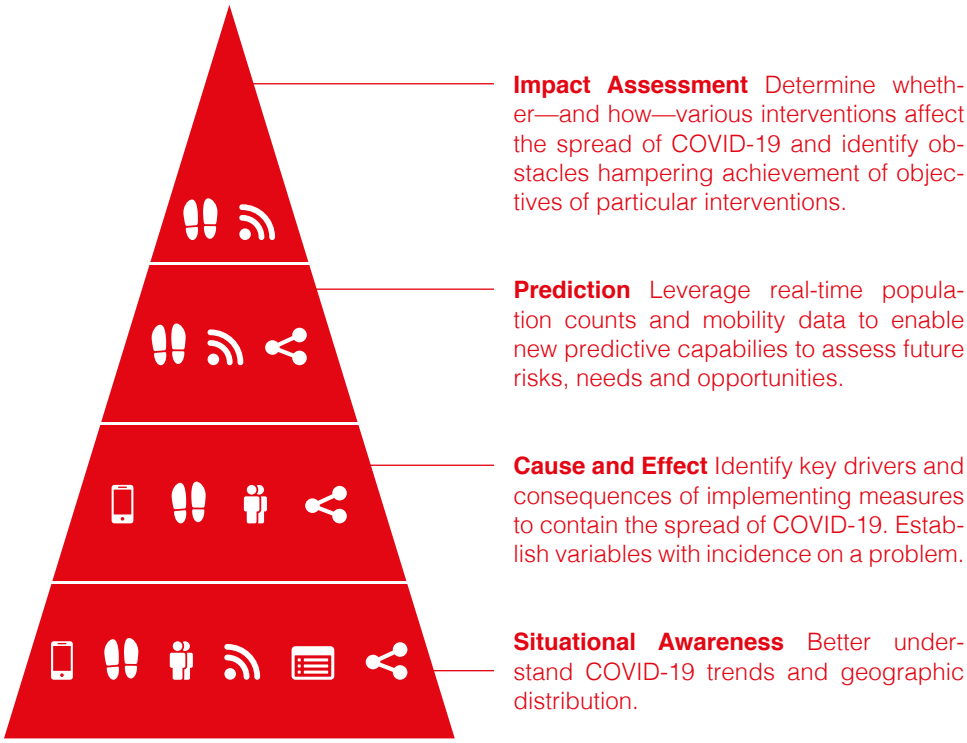
These include:

1 Self-assessment / symptom tracking

Self-assessment and symptom tracking websites or apps allow users to report their symptoms and get instant feedback on their assessed risk through interactive forms and surveys. In many markets with low smartphone penetration, symptom tracking messages can be displayed via a USSD menu (the user opens a menu and picks from a range of options) which enable two-way flash messages. These types of applications can help governments to better handle citizen demand for trustworthy feedback and information when faced with symptoms and to monitor disease outbreaks. This in turn enables a better use of resources and medical services: to an extent, these apps and websites relieve some of the strain put on hospitals and facilities as self-screening can help rule out infection and reduce the need for patients to seek a formal diagnosis. It can also suggest containment and control measures to individuals at risk. Given the large amounts of misinformation surrounding COVID-19, these apps and websites can also be a tool for citizens to access reliable, trustworthy information regarding symptoms and next steps to take when experiencing them.

While these applications do not necessarily require personal data in order to fulfill their promise, they do often collect information from users, including home address, phone number and location. These apps have been deployed widely by local and national governments in, for example, Afghanistan, Colombia, Kenya, Singapore and Turkey. A recent study by Zoe Global, Massachusetts General Hospital and King's College, which tracked self-assessment applications in Sweden, the UK and the US found that these apps could be "remarkably effective in predicting coronavirus infections".[9] Nevertheless, while self-reporting apps can be very useful there are caveats to consider, such as a high variance in self-reporting or misreporting due to a misperception of users' own realities. Awareness of the human factor in these types of applications is important.

9 https://www.nytimes.com/2020/05/11/health/coronavirus-symptoms-app.html?auth=login-email&login=email



Impact Assessment Determine whether—and how—various interventions affect the spread of COVID-19 and identify obstacles hampering achievement of objectives of particular interventions.

Prediction Leverage real-time population counts and mobility data to enable new predictive capabilities to assess future risks, needs and opportunities.

Cause and Effect Identify key drivers and consequences of implementing measures to contain the spread of COVID-19. Establish variables with incidence on a problem.

Situational Awareness Better understand COVID-19 trends and geographic distribution.

Centralized / Decentralized Contact Tracing

Social Media Analysis / Citizen Surveys

Flows Modelling / Mobility Mapping

Epidemiological

Surveillance and Enforcement

Self Assessmnt / Symptom Tracking

Figure 1. Purpose and applications of digital technologies for COVID-19 response and recovery efforts

Source: Data-Pop Alliance, 2020

2 Contact tracing applications

Contact tracing is a central technique often applied in epidemiology that has gained widespread attention amid COVID-19 response efforts. The objective here is to quickly identify potentially at-risk individuals who have been in close contact with a recently diagnosed positive case of an infectious disease requiring compulsory reporting, as is the case for SARS-CoV-2. Once these people have been identified, the main goal is to quickly test and isolate those with a confirmed coronavirus infection so as to break the chain of transmission.

Traditional contact tracing involves carrying out epidemiological interviews (typically performed over the phone) to collect relevant data about the symptoms, mobility and social behavior of patients. Personal information is commonly collected in these interviews, including the phone numbers of all the people with whom the patient has been in close contact within the past N days (for COVID-19 the latest recommendation is 48 hours). Traditional contact tracing has four intrinsic limitations where digital tools might help. First, it relies on the patient's memory. Second, all the close contacts need to be known to the patient such that (s)he can share their contact information with the contact tracer. Third, it does not work well across borders. Fourth, it is expensive, human-resource intensive and time consuming. Despite these limitations, it is an effective tool to help contain the spread of an infectious disease, assuming the contact tracing teams are properly scaled to the incidence of the disease and the information they collect is in digital form, ideally using state-of-the-art tools, so that it is readily available for analysis and decision making.

Since the outbreak of COVID-19 has exceeded the capacity of most manual contact tracing teams worldwide, public officials in many countries are turning to smartphones as a key tool to complement these existing initiatives. Thus, we are witnessing the emergence of dozens of smartphone-based contact tracing apps

Box 2. CoronaMadrid[10] and COVID-19 CDMX[11]

In early March, the autonomous community of Madrid in Spain released CoronaMadrid, a self-assessment application available for Android and Apple phones, as well as in web form. One of the main objectives with this application was to reduce traffic and demand on mobile hotlines, while allowing officials to continue providing instructions and recommendations to citizens. Individuals are only prompted to use this application if they experience symptoms. Users can opt-in to share their locations to provide public health organizations with more precise information to inform their responses. Users must share their phone numbers.

Conversely, the tool created by the government of Mexico City in March is a uniquely web-based form. It requires users to share personal information such as their cell phone number and complete address. After assessing symptoms, the tool will classify individuals according to their risk factor and recommend a series of actions to take. The privacy notice for this tool establishes that the data collected may be used by judicial and administrative, federal and local authorities.

10 https://coronavirus.comunidad.madrid/

11 https://test.covid19.cdmx.gob.mx/

globally. If indeed smartphone apps were able to passively record close contacts between individuals, they could automatically generate the necessary contact traces, such that at-risk individuals who had been in close contact with an infected person could be notified, tested, and isolated if positive. This process would enable the transmission chain to be broken and prevent community transmission of the disease.

Contact tracing applications rely on proximity technology and/or location traces to identify potential contacts between individuals. First efforts on this front—such as those deployed in China or South Korea—leverage the GPS location alone or in combination with other data, such as credit card transactions or visual surveillance camera footage. These applications infer close contacts if individuals have been within a radius of 1.5-2 meters of each other and for at least 15 minutes. However, limitations associated with GPS—including imprecisions in indoor (e.g. buildings) and transport (e.g. subway, planes) environments, as well privacy concerns, have led technologists and governments to turn towards Bluetooth as the main sensor to detect close contacts between individuals via smartphone apps.

Bluetooth-based contact tracing apps enable devices to share "digital handshakes" by sharing encrypted, unique identifiers (referred to in the literature as tokens, beacons, pseudonyms, temporary exposure keys (TEKs) or temporary contact numbers (TCNs) to record contacts that last "longer than a few minutes"[12] and located within a 1.5-2m radius.

The success of these apps depends on many factors, including high adoption rates and tight integration with public health systems, such that both doctors and infected individuals can report positive cases and at-risk individuals can be duly notified. Unlike manual contact tracing, these applications record contacts that a person may not remember or know they have come in close proximity with. However, these applications are not exempt from their own limitations and challenges, including difficulties in reliably detecting close contacts via Bluetooth, battery consumption, trolling and hacking scenarios,[13] human-centric challenges,[14] privacy and security risks and low adoption rates, particularly within the most vulnerable groups (e.g. the elderly, low socioeconomic levels and skeptics).

12 https://venturebeat.com/2020/04/13/what-privacy-preserving-coronavirus-tracing-apps-need-to-succeed/

13 In the UK, for example, hackers successfully launched phishing attacks with the National Health Service's app. A phishing message redirected victims to a fake website where they were asked to type in their personal details; www.itwire.com/guest-articles/guest-opinion/how-hackers-can-abuse-contact-tracing-apps-91032.html

14 The app's design needs to be based on how people can, need and want to perform tasks, rather than expecting users to adjust and accommodate their behaviors to the product.

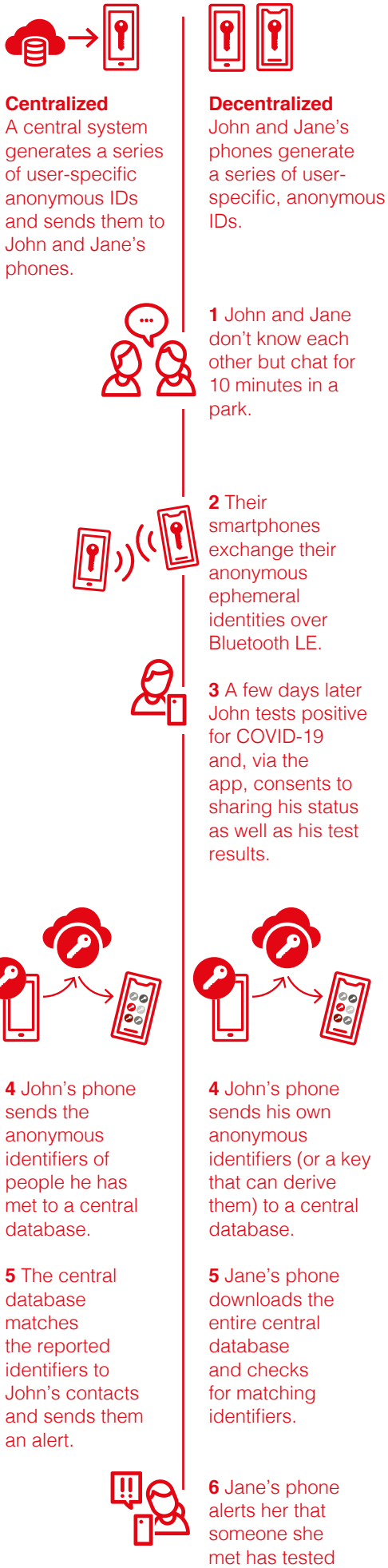


Figure 2. Centralized vs.decentralized contact tracing app-based models

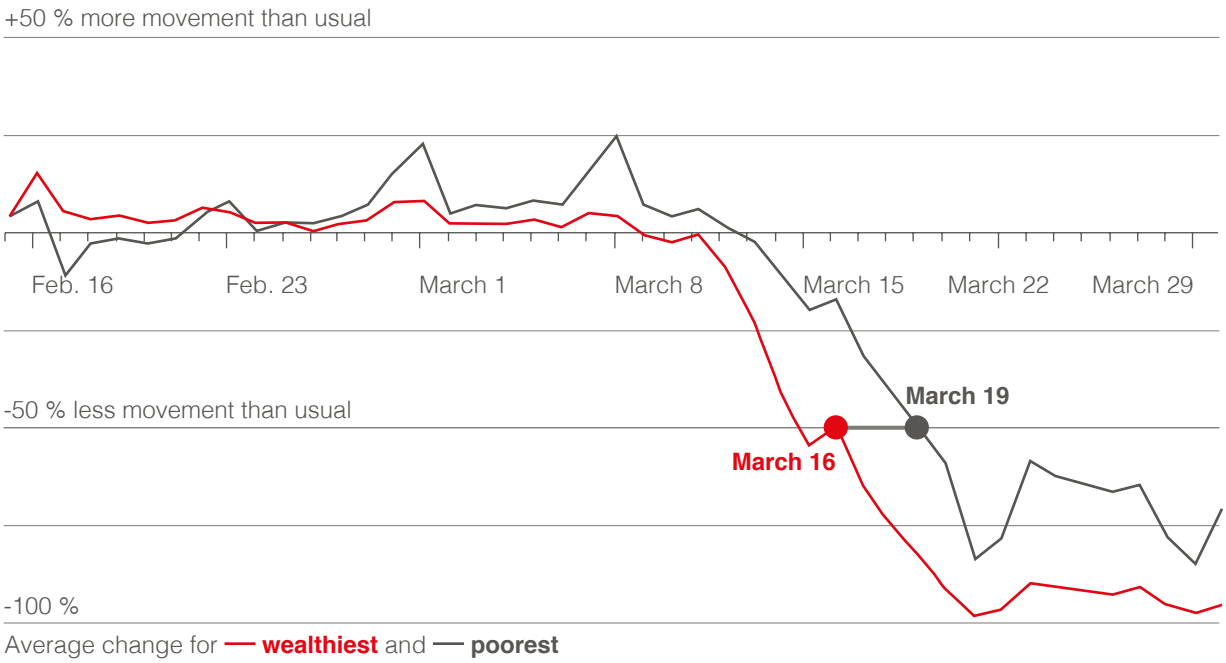


Figure 3. "Location Data Says It All: Staying at Home During Coronavirus Is a Luxury"

Source: NY Times, 3 April 2020, https://www.nytimes.com/interactive/2020/04/03/us/coronavirus-stay-home-rich-poor.html

3 Modelling and mapping population flows

Mapping flows, or mobility of people over time and space, has been one of the more common applications of private data for social good initiatives in the recent past. Mobile phones can often act as indicators of human mobility and give insights into behavior. Aggregated and anonymized location data can be sourced from various technologies such as GPS, mobile cell towers, Wi-Fi networks, Bluetooth connections, surveillance video, credit card records and wearables, as well as many other devices and apps. In the case of COVID-19 responses, the analysis of these data enhances findings by identifying risk and potential hotspots, assessing public responses and the effectiveness of social contact and mobility contention policies, and detecting where more resources may need to be channeled. Moreover, human mobility is a valuable input to computational epidemiological models. While the aim of flow modelling is descriptive, its use can cross the line and be used as a tool for control by authoritarian governments—and, even more surprisingly, by others perceived to be considerably less so—by applying stringent enforcement policies.

These analyses can also simply shed light on the effects and effectiveness of containment measures, especially across different demographics, potentially pointing to enabling and constraining factors. For example, a US study revealed that, when on March 16th people were asked to stay at home, those living in richer areas had already reduced their mobility by nearly half whereas people in poorer areas did not substantially reduce theirs until three days later, suggesting structural impediments for the latter to staying at home and limiting their exposure to the virus.

Box 3. Bluetooth-based contract tracing app models

The debate over app-based contact tracing models centers foremost on what data is captured so as to ensure that only strictly relevant information is collected. Second, there are debates over where this data should be stored. Concerns also exist around the aggregation of data, underlying privacy configurations and who should have access to this data—including public authorities. For the apps that exchange tokens via Bluetooth, two main architectures have been proposed: centralized and decentralized. In both cases, the token exchange takes place **locally** in the phones. The main difference stems from (1) who provides the phones with the initial seed used to generate such tokens, and (2) what information the phones send to a centralized server when their user is tested positive for coronavirus.

In the centralized model, the initial seed to generate the tokens is given by a trusted, centralized server typically controlled by administrators or public health authorities. Moreover, when an individual tests positive and upon recording this event in the application, their phone sends all tokens of the devices it has had close contact with (e.g. over the preceding 14 days) to a centralized server. The central server matches the tokens and alerts users to a potential contact. Resulting aggregated, anonymized data can help experts fine tune the risk calculation to determine whom to send a notification to and also allows administrations to detect infection patterns in society, which is crucial input when designing policies and measures aimed to curb the spread of a disease.

In the decentralized model, the initial seed to generate the tokens is given by the operating system (in the case of the Apple/Google API) or by the app itself. When an individual tests positive, upon recording this event in their phone, their app only sends to the central server their list of tokens. All the phones running the app periodically poll the central server for the list of tokens of positively diagnosed individuals. Given that the phones have the list of contact traces, they locally check if there is a match between their contact traces and the list of tokens associated with recently diagnosed individuals. If a match is found, the app triggers a notification with indications of what to do next. In this case, no central authority has visibility on how many users have been notified for each registered positive case.

Descriptive tools such as flow mapping are used to look at people's movement patterns locally to gage risks or potential hotspots, as well as to assess how people are responding to the virus and response measures to inform public response. In the Valencian region of Spain, a team of experts has been working closely with the president of the region on a variety of data-driven tasks related to COVID-19, including quantifying and modeling human mobility captured by the mobile network infrastructure. In a pioneering collaboration between the Spanish National Office of Statistics and the three largest telecommunication companies in Spain, experts have been able to assess the success of containment measures and their impact on the spread of the pandemic, estimating that over 40,000 lives were saved in the process.[15]

In many countries, pre-existing tools designed to look at flows of people for applications in the public transport or tourism sectors have also been adapted to the current COVID-19 context. In Austria, for example, Invenium, an existing collaboration between A1 Telekom Austria Group and a local university, developed a motion analysis application that was used to model human mobility flows for applications in traffic congestion or tourism sites to assess the effectiveness of response measures to reduce movement and social contact.[16] The COVID-19 Community Mobility Maps generated by Google are based on users' aggregated location data and reflect community-level behavior such as travel, for example, to grocery stores, parks and public transport centers.[17] Many of these mappings of concentration and movements of people use aggregated and anonymized data, further calibrating policy response and containment measures such as social distancing and contact tracing.

15 https://www.gva.es/es/inicio/area_de_prensa/not_detalle_area_prensa?id=858477

16 https://www.reuters.com/article/us-health-coronavirus-europe-telecoms/european-mobile-operators-share-data-for-corona-virus-fight-idUSKBN2152C2

17 https://www.google.com/covid19/mobility/

Box 4. Debate over contact tracing apps: moving from centralized to decentralized approaches

Early on, the general public in several European countries showed support for centralized models using pseudonymized proximity data. For example, the Pan-European Privacy Preserving Proximity Tracing (PEPP-PT) initiative developed an open protocol, defining standards for tracing apps built on it and uses a blend of centralized and decentralized methods.[18] The UK and France have also developed their own centralized apps, with France being the first to launch its voluntary app StopCovid, using a protocol known as Robert to complement existing manual contact tracing. While the data protection authority, CNIL, has not raised any major flags, concerns are being voiced over the use of pseudonymized data which necessitates a certain level of trust that the government is indeed respecting the limitations around data collection it has detailed.[19] Supporters of these models ensure that fully privacy-preserving techniques are in place, along with ready-to-use, well-tested, properly assessed mechanisms, and support for interoperability.

Other countries, like the US, quickly turned to decentralized models, using e.g. the Apple/Google API that combines Bluetooth, cryptography and location tracking. The debate between centralized and decentralized contact tracing models has continued within and across countries and, as response efforts develop, more countries, including Germany, are choosing to pursue decentralized models, mainly due to fears of function creep.[20]

Researchers at MIT have created Private Kit: Safe Paths,[21] a free and open-source application that uses both Bluetooth and GPS tracking based on the decentralized model. One key feature is its interoperable standards to ensure usability between different apps within or across different countries based on an open source Temporary Contact Number protocol to ensure interoperability. Decentralized Privacy-Preserving Proximity Tracing (DP-3T), developed by researchers in France, Germany, the Netherlands and Switzerland, creates a virus contraction risk score generated from an algorithm running on the user's data locally on their device. Decentralized apps are the subject of criticism particularly with respect to making it more difficult for health authorities to have the necessary data regarding how many close contacts receive a notification for each positive case; agility and practicality rely on cryptography which is complex and requires challenging and frequent updates of parameters, especially at the scale that would be needed to be effective in this epidemiological response.[22]

18 https://www.pepp-pt.org/

19 https://www.france24.com/en/20200602-france-rolls-out-covid-19-tracing-app-amid-privacy-debate

20 https://techcrunch.com/2020/04/27/germany-ditches-centralized-approach-to-app-for-covid-19-contacts-tracing/

21 https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-covid-tracing-tracker/

22 https://venturebeat.com/2020/04/13/what-privacy-preserving-coronavirus-tracing-apps-need-to-succeed/

Figure 4.
Predicted baseline mobility patterns for 28 January to 18 February 2020. Individual probability of moving between the top 20 European countries with the greatest outward mobility.

Source: https://science.sciencemag.org/content/sci/369/6510/1465.full.pdf

4 Surveillance and enforcement

In general, tools for surveillance and quarantine enforcement analyze sensitive personal data. More granular, often pseudonymized (but in the case of COVID-19 also increasingly sensitive) data sources have emerged to monitor people's movements with a view to containing the spread of the disease. This exceptional circumstance has led many governments to consider loosening or sacrificing individual privacy during the response period for the sake of curbing contagion and saving lives. Facial recognition systems, mobile tracking apps, wearables, geolocation, credit card and financial transactions and transport data are being used for real-time monitoring of compliance with response policies.

Examples of crisis response involving both new technologies and sensitive personal data along with emergency public health policies and law enforcement measures can be seen in several countries globally. Apps deployed in China, such as Alipay Health Code, are grounded in technology for symptom tracking, assigning a color-coded QR code to the user indicating their risk level.[24] This tool goes beyond self-assessment as all citizens are required to use it and personal data is sent to law enforcement bodies to enforce quarantine measures based on an individual's risk level. Amongst others, this app has been criticized for its lack of auditability as the rules behind the assignment of risk are not widely known. Though apparently not compulsory, access to many services and activities in China is dependent on receiving a green code.[25]

Taiwan was one of the first countries to lead the development of technologies for quarantine enforcement using mobile

24 https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html

25 https://www.afr.com/world/asia/how-china-s-health-code-app-is-used-to-fight-infection-20200424-p54mzk

Box 5. Applying research on malaria-related mobility flows to COVID-19: the case of Mozambique[23]

Novel mobility analyses by mobile networks have proven themselves useful for mapping the spread of many diseases. A partnership was established between Vodafone, the University of Southampton, the Clinton Health Access Initiative and the National Malaria Control Program; it was backed by the Bill and Melinda Gates Foundation. Together they analyzed mobility flows in Mozambique, a country where malaria poses a great burden on the economy and the general well-being of the population. By examining aggregated and anonymized population flows and malaria incidence in the country, the analysis allowed a better prioritization of resources and geographically stratified actions by providing malaria "sinks" and "sources" – thus showing how the disease moves across the country with population flows. The lessons learned from this analysis were then quickly applied to COVID-19 and, by leveraging the global reach of Vodafone's mobile networks, mobility insights were extracted. They were used not only for tracking how populations were responding to government measures, but were also fed into an epidemiological model on the effects of travel restrictions and lockdown behaviors during the spread of the disease.

23 https://www.vodafone.com/perspectives/blog/world-malaria-day-2020-vodafone-fighting-malaria.

data, implementing a "digital fence".[26] This integrates location data from cell phones to trigger an alert system if anyone moves too far from their home and issues a fine for breaking quarantine restrictions.[27] Hong Kong has introduced wearables in order to enforce the 14-day quarantine for anyone arriving at the airport: an electronic tracker wristband, paired with a mobile app used to calibrate the wristband, using geofencing technology.[28] In the case of the tracker wristband, the technology is said to preserve privacy as it does not track individuals' exact location, but simply signals whether an individual is inside or outside of their home.

There has been mixed public perception of the development and use of these applications. China, which has seen breaches of data collected for the COVID-19 response entailing negative consequences such as discrimination or stigma, has strengthened public debate around privacy in the country.[29] Similar concerns and debates have been raised about privacy and digital rights globally, such as the examples detailed in **Box 6**.

Transparency in the design and use of many of these apps across several countries has been called into question as there is often a lack of clear data privacy policies, communication with the public, or limitations on who has access, for what purpose and for how long. Complicating this is the ambiguity in legal and regulatory frameworks on data protection and privacy that has given some governments ground to implement measures that may infringe on digital and human rights in cases of emergency, with unclear limitations on these provisions. It is not impossible to imagine that certain abusive policies may linger long after any justification for them has disappeared.

26 This recent AI&I exchange with Audrey Tang, Taiwan's Digital Minister can be accessed here: www.youtube.com/watch?v=sfNESpLr0pk

27 https://qz.com/1825997/taiwan-phone-tracking-system-monitors-55000-under-coronavirus-quarantine/

28 https://qz.com/1822215/hong-kong-uses-tracking-wristbands-for-coronavirus-quarantine/

29 Tracing.Testing.Tweaking. Approaches to data-driven Covid-19 management in China (Meric paper)

Box 6. Trade-offs in quick containment and digital rights

In Israel, the response efforts went beyond introducing new technologies using sensitive personal data; the authorities also implemented an emergency law passed to specifically track infected individuals and their contacts in order to enforce individual quarantine measures. Importantly, there were time limitations involved in the implementation of this technology set out in the law, initially to thirty days. As of September, the security service program used for contact tracing was still in place.[30] However, other key questions such as who has access to this data for analysis, what other types of analysis may be performed and when the data will be deleted have not been specified. Digital rights advocates have pushed back on these measures, warning of the risks not only of mass surveillance but also of targeted law enforcement action, as there are fears of a slippery slope as these methods unfold.

South Korea has gone even further. It has not only used sensitive personal data from mobile phone tracking, credit card transactions, as well as face-to-face interview data with patients, but used this information to publish a publicly available map to allow citizens to verify their potential contacts and the patterns of those infected as well. While the data does not include personal identifiers, there is a high potential for re-identification of individuals due to the granularity of location data, mobility patterns and even personal descriptions of those infected. Though the transparency and openness of the government has allegedly increased trust in its containment efforts, the fear of social stigmatisation is high, given the amount of information usually released about confirmed patients.

India has become the only democratic nation in the world to require a majority of its citizens to download and use its tracking app, Aarogya Setu, with threats of fines, losing jobs, or jail if non-compliant. While official policy maintains that the application is voluntary, all government employees, many large private companies, landlords and even city governments are mandating its use. The technology underpinning the application differs from many others as it allows for enforcement as well, in that it goes beyond exposure notifications from proximity data to assigning color-coded risk badges, similar to China's Alipay Health Code app. Other concerns have been raised about the lack of legal frameworks around data privacy and lack of transparency around data access or use from the app as the developers' profiles are not fully disclosed to the public and include many private companies.[31]

30 https://hamodia.com/2020/09/08/contact-tracing-app-prevent-infection-spread-ineffective-kosher-phones/

31 https://www.technologyreview.com/2020/05/07/1001360/india-aarogya-setu-covid-app-mandatory/

5 Epidemiological modelling

Both metapopulation and individual computational epidemiological models benefit from high quality, real-time data about the number of people infected, hospitalized or in intensive care. Moreover, human mobility (computed from, for example, the mobile network infrastructure) and quarantine information enable building more accurate models and predictions. Having an underlying model enables running it under different scenarios—such as different social contention, mobility and contact tracing situations—to assess the impact that different non-pharmaceutical interventions (NPIs) might have on the incidence and spread of the disease.

Other technologies such as smart thermometers and AI-based diagnostic tools have been providing different ways to map and predict the evolution and spread of the virus. Data from smart thermometers were used in the US to create HealthWeather, a map which visualizes seasonal illness linked to fever, based on aggregated, anonymized data from the thermometers and mobile applications.[32] In the fight against COVID-19, the benefits of real-time data drew attention to similar sources of data such as wearable fitness and health devices, encouraging users to synchronize their existing devices to specific apps such as MyDataHelps to pool data. However, concerns were raised around the accuracy of these efforts as they are based on information about the behavior of flu-like illnesses, as well as the representativeness of these initiatives as their data collection appears to be biased towards people who have access to wearable devices.

Real-time modelling, and therefore real-time data access, are critical to enable timely response policies particularly in the case of outbreaks. Data access is also key to accelerate scientific research in order to better diagnose, treat and develop vaccines.

32 https://www.washingtonpost.com/national/health-science/start-ups-health-weather-map-may-help-forecast-spread-of-diseases-like-covid-19/2020/03/26/36c069b8-6ef0-11ea-a3ec-70d7479d83f0_story.html

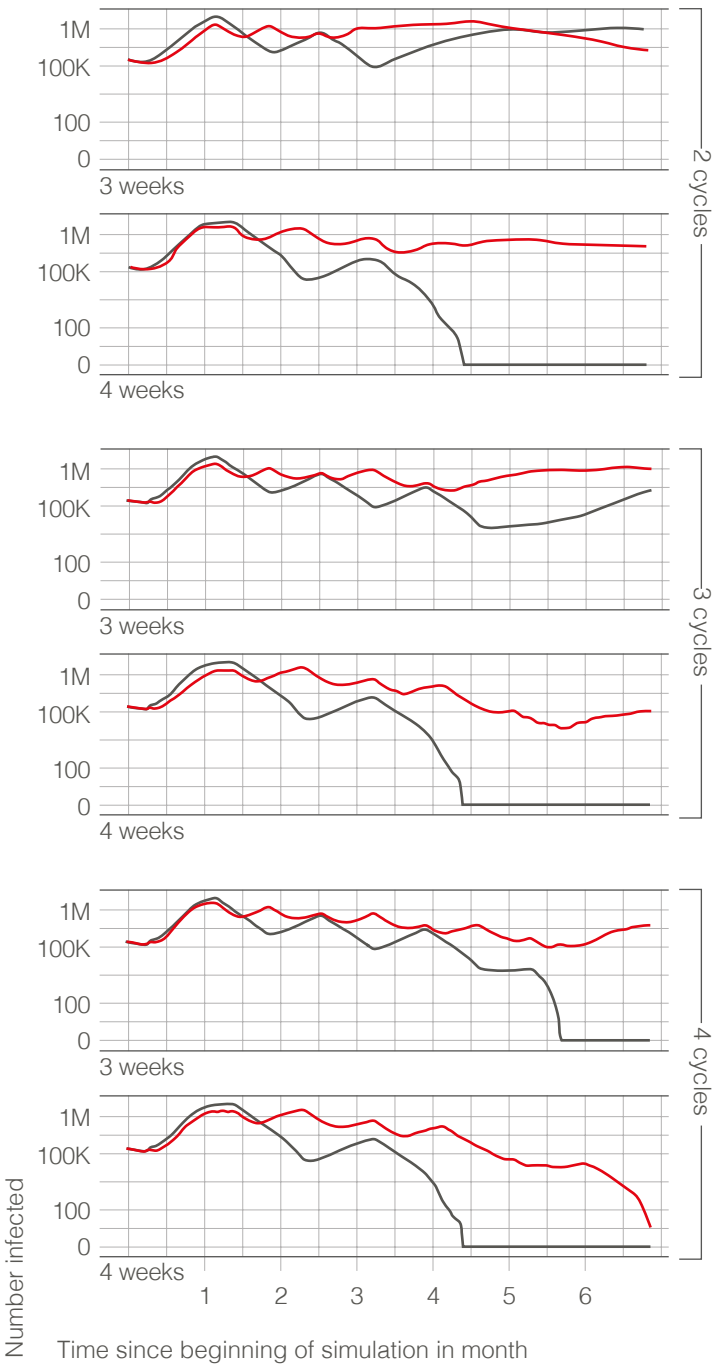


Figure 5. Cases over time, when lockdowns are synchronized or unsynchronized across all European countries.

Source: (N. W. Ruktanonchai, 2020) https://science.sciencemag.org/content/369/6510/1465

6 Social media analysis and citizen surveys

In order to further inform COVID-19 response efforts, social media analysis and citizen surveys are being employed to gather people's perceptions, experiences and behaviors with regards to the outbreak as well as the measures put in place to halt the pandemic. Issues around access to information, misinformation and information overload have been central to people's behaviors and views about the crisis and response efforts. COVID-19 has raised the visibility of public-private data sharing initiatives and the potential, as well as risks, of these technologies. As a result, citizens have been voicing their opinions and becoming more actively involved in related debates. Researchers and public institutions have launched large-scale surveys and begun digging into social media with the objective of shedding light on people's situations and perceptions regarding the pandemic and the confinement measures that they have endured for several months in many countries worldwide.

Misinformation (which is intentionally misleading) and information overload have been rampant with regard to the basics of data collection: from which source, what can it be used for and by whom. Disinformation (which is false) has also become a real problem. It has led to theories about a link between 5G and coronavirus resulting in attacks to 5G masts in the UK,[33] dangerous health advice by the US president resulting in poisoning from consuming products not medically recommended or safe,[34] or the President of Brazil insisting that the virus is a fantasy created by the media,[35] amongst others.

Box 7. Complementing epidemiological forecasting with mobility data

In epidemiological studies, mobility is a key factor to understand how diseases spread and what actions should be taken to minimize their impact on society.

The flow model aspect of the joint project conducted by the Vodafone Group and the University of Southampton was used to build a pan-European model of disease spread, simulating different scenarios of lockdown restrictions and synchronous vs asynchronous lifting of these, using data from multiple countries, as described in Box 5. These analyses provided key insights for policy makers as lockdown restrictions were implemented at different times and countries adopted different approaches. As the measures are eased, an evaluation of the effects of different countries' actions is critical for a successful return to normality.

A study of over 1200 simulations found that synchronised lockdown restrictions (non-pharmaceutical interventions [NPIs]; Figure 1) imposed over a 6-month period were more likely to halt the incidence of COVID than unsynchronized measures. One exception to this (Figure 5, top) is when there were 2 cycles of 3-week unsynchronized lockdowns were enough people were infected and herd immunity reduced transmission. Moreover, two synchronized 4-week lockdown cycles were able to stop community transmission, whereas when implementing unsynchronized lockdowns, four 4-week cycles were needed (see Figure 5).

Furthermore, in-country epidemiological modelling was conducted for several countries following Vodafone's example, evaluating the effects of travel restrictions and different social distancing measures and their implementation. This type of analysis allows for an in-country evaluation of policy and its effects, the aim being to striking the right balance between economic impact and societal health and well-being, by means of a comparative analysis of different scenarios.

33 https://www.bbc.com/news/technology-52281315

34 https://www.forbes.com/sites/robertglatter/2020/04/25/calls-to-poison-centers-spike--after-the-presidents-comments-about-using-disinfectants-to-treat-coronavirus/#569c4b541157

35 https://www.hrw.org/news/2020/04/10/brazil-bolsonaro-sabotages-anti-covid-19-efforts#

Trust is essential to COVID-19 response efforts, not only in the uptake and use of applications to make them significant, but also in relation to what types of applications people are willing to use, as well as whom they trust to collect and use it: government, private companies or researchers. The trust element depends on context: country, demographics, regulatory environments and the severity of the virus outbreak. It continues to change as the pandemic and response efforts evolve.

The COVID-19 crisis has landed the world with a headache it could have done without. But governments' sometimes blotched responses to contain the virus have led observers to realize that the current situation is a pivotal and potentially game-changing moment in the way we respond to public health crises. Certain countries (China, India, New Zealand, UAE and Vietnam) were perceived by their citizens to be good performers[36] and their governments have seen their popularity reinforced, but they are few and far between. Some governments dragged their heels to implement corrective steps at the outbreak of the pandemic, while practically all of them need to rethink initial solutions.

They need to move away from labor-intensive data collection measures to real-time and transparent tools that are perceived to be legitimate by the general public. Trust is all-essential.

Once endorsed, data collection can lead to improved responsiveness: in times of crisis, policy makers can be empowered to move swiftly and implement appropriate measures, thereby saving precious time and improving future outbreak preparedness. Reactivity can impact favorably on government approval in both developed and developing economies. Independent bodies and the press should be encouraged to monitor and report on sanitary measures; democratic governance, which has been put to the test in countries with authoritarian leadership, can but be reinforced as a result.

36 Toluna-Blackbox Index of Global Crisis Perceptions; https://blackbox.com.sg/everyone/2020/05/06/most-countries-covid-19-responses-rated-poorly-by-own-citizens-in-first-of-its-kind-global-survey

Box 8. Case Study: Brazil

Availability of reliable data for modeling the spread of the virus and public (lack of) transparency

Brazil is in the unenviable position of having the second highest daily incidence of new COVID-19 cases but faces a huge problem in predicting spread and therefore formulating effective policy responses. This is due to a significant problem with access to data or reliable, detailed enough data to be useful for understanding spread and responding with public health action. In response to this challenge, Open Knowledge Brazil launched a tool to assess transparency in the COVID response quality and availability of data. Each state was considered so as to create more accountability for public institutions in the response, provide tools for policy makers and government officials in their decisions and facilitate collaboration between other sectors.[37] However, Brazil's president further thwarted efforts at modeling the spread and impact of the virus by prohibiting the publication of figures on new cases and deaths, thereby raising great concern for how the situation will evolve in the country with the increased obscuring of public information.

37 https://transparenciacovid19.ok.org.br.

Do you believe that the measures the government has taken are enough to contain the spread of coronavirus?

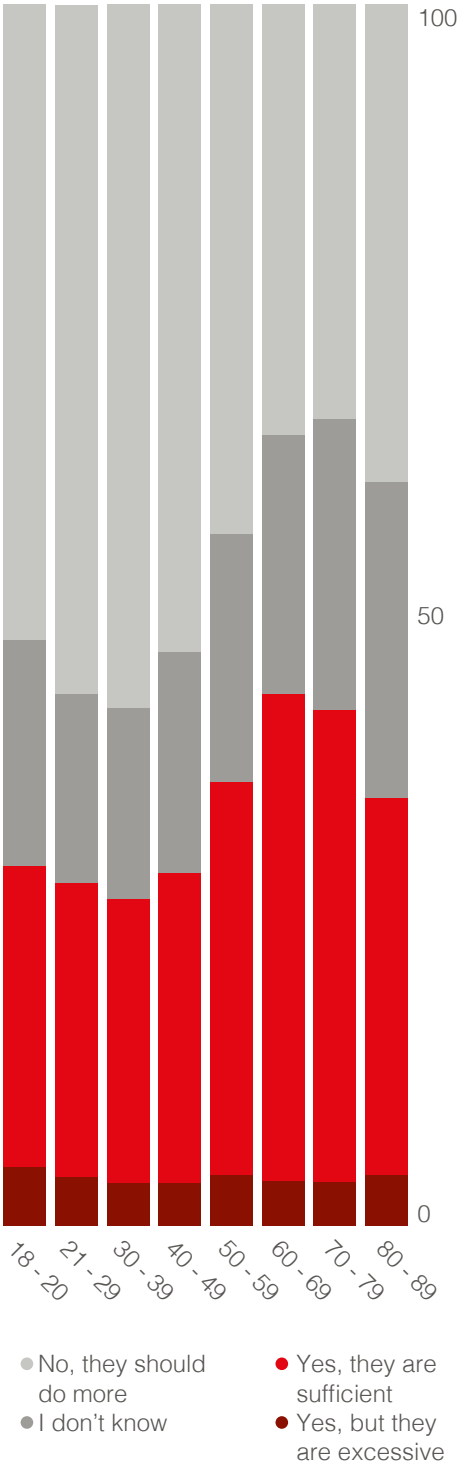


Figure 6. Covid19 Impact survey Dashboard results for Spain

Box 9.
Public perspectives on COVID-19

Imperial College London's Patient Experience Research Centre (PERC) launched a research project on perceptions and behavioral responses to COVID-19, establishing mechanisms for community engagement. It revealed that respondents had significant concerns and fears. The development of a vaccine was considered the most urgent priority;[38] they were also confused about the great, sometimes misleading, amount of information around the pandemic and response efforts.

Harvard, Cambridge, the IESE Business School and others formed a joint international initiative and launched a global survey, provided in many languages, to measure attitudes and beliefs around COVID-19.[39] According to results last updated on September 15, 2020, the survey found that respondents overall believe that citizen and government response was insufficient. The organizers of the survey have also made available the anonymized data on individual behaviors during the pandemic as well as perceptions about governments' policies captured via the survey.

A related effort has been launched in Spain in collaboration with the Valencian government via the Covid19impactsurvey.[40] With over 280,000 respondents in Spain and almost 50,000 from other countries, it is one of the largest citizen surveys in Spain to date. The researchers have published[41] their first results and visualizations[42] of all the data. Insights derived from the survey have regularly been used by the Valencian Government in their efforts to combat COVID-19.

Researchers in many countries are using social media data to track health concerns of people isolated at home. Social distancing and quarantine measures have proven to be a huge psychological disruption, worsened further in many cases by having loved ones fall ill and the economic impact of unemployment and economic precarity. Twitter in particular is being used as a valuable resource, as researchers can search for comments using tags such "telehealth". Tools can also be created, as illustrated by an initiative undertaken by researchers at Stony Brook and Stanford: they have used Twitter-based surveillance architecture and AI-based language assessment and statistical techniques to create measures of the impact of different response interventions and policies on mental health and well-being.[43] Citibeats, an AI company, and NTT Data in Japan created a dashboard to collect tweets and enable better capture and understanding of civic needs, dignity, trust, security and visibility around COVID with the aim of idem findings into local and regional action.[44]

38 https://spiral.imperial.ac.uk:8443/handle/10044/1/77842
39 https://covid19-survey.org
40 https://covid19impactsurvey.org
41 https://www.jmir.org/2020/9/e21319/
42 https://covid19impactsurvey.org/results
43 https://news.stonybrook.edu/homespotlight/team-using-twitter-to-track-covid-19-symptoms-and-mental-health/
44 https://covid19-japan.citibeats.com

2. Four sets of considerations and concerns raised by the use of digital data and technologies for COVID-19 response efforts

This section discusses the critical factors of the technological solutions and initiatives that have been deployed since the outbreak of COVID-19, following the taxonomy proposed in the publication *Sharing is Caring: Four Requirements for Sustainable Private Data Sharing and Use* co-developed and published by Data-Pop Alliance and the Vodafone Institute for Society and Communication in November 2019:[45] (a) technological and scientific; (b) commercial and economic; (c) ethical and legal; and (d) social and political. In doing so, it also points to generalizable considerations about how digital data and technologies can be harnessed to promote public good objectives.

1. Technological and scientific

Debates around permitting or promoting the use of initiatives created to learn from and solve the pandemic must take into account their technological and scientific soundness. Contact tracing applications must demonstrate that they can function accurately and respect guarantees; Self-assessment applications must provide the right advice and provide assurances that initiatives to map the spread of the disease or population distribution provide an accurate, non-biased picture. By way of an example, this section will examine contact tracing technologies where many of these questions are being drawn out and discussed.

Use of technological functionalities for standards not originally suitable

One of the main debates regarding contact tracing applications centers on whether close contacts can be accurately detected via Bluetooth. Researchers have noted that false positives—in short, when a red light lights up but shouldn't have—may occur when individuals share close spaces, such as in an apartment complex, through shared walls, or when they are in large outdoor settings.[46] Additionally, the accuracy with which proximity is determined is sensitive to the specific hardware available on phones and might be unreliable if the user has several Bluetooth-enabled accessories connected at the same time, such as wireless headphones, wearable devices or in-car systems, leading to potential false negatives—when a user should be alerted but is not.

Taking decisions regarding the risk of infection for individuals based on a potentially intrusive sensor could put people at risk of further surveillance or policing actions, when in fact their exposure levels were fairly limited. The European Union's Toolbox on Contact Tracing Applications notes that "Member States should consider specifications which allow contact detection to an accuracy of one meter, in order to minimize false positives".[47] Yet, Bluetooth inventors recently noted that how well one can assess the distance of one device from another depends on the "radio signal's path loss" which can vary greatly in open spaces or in spaces with obstacles between the two devices.[48] Given that many protocols use the strength of the signal to define how close contacts are to each other, this could lead to inaccuracy. Additional concerns with the precision of contact tracing applications are whether contacts can be recorded when phones are locked, whether systems are able to detect individuals by intentionally trolling the systems and whether these applications function when devices are running on low battery.

The challenges inherent to adoption rates

The success or even usefulness of contact tracing applications is a function of how many individuals install and use the application. If this technology is used in isolation as a response, one might need as high as a 60% adoption rate for the tools to be useful, which would be the case for COVID-19 due to its high reproduction number.[49] If not used in isolation but as a complement to manual contact tracing, researchers have estimated that 15% of adoption could already bring some value.[50] High adoption rates are problematic for several reasons. First, socioeconomic and demographic gaps are inevitable as only those with the latest models of handsets and operating systems can install many of the current contact tracing apps being proposed by governments. Second, having such a large portion of mobile devices running the same software poses a significant security risk. Third, data stemming from thousands of individuals could be traced and infections reported, while effectiveness may not be ensured. Access to data, however, is a precondition not only for contact tracing but also for self-assessment applications. Many of the tools made available rely on smartphones—which are not as ubiquitous in developing countries—or are only being launched in one operating system, leaving portions of the population without access.

If used in isolation, contract tracing apps need adoption rates of 60%

45 https://www.vodafone-institut.de/studies/four-key-requirements-for-sustainable-private-data-sharing/

46 https://arxiv.org/ftp/arxiv/papers/2004/2004.07463.pdf

47 https://theintercept.com/2020/05/05/coronavirus-bluetooth-contact-tracing/

48 https://theintercept.com/2020/05/05/coronavirus-bluetooth-contact-tracing/

49 https://www.research.ox.ac.uk/Article/2020-04-16-digital-contact-tracing-can-slow-or-even-stop-coronavirus-transmission-and-ease-us-out-of-lockdown

50 https://www.medrxiv.org/content/10.1101/2020.08.29.2018413 5v1

In Colombia, for example, the government launched an official self-assessment app on Android and only added it to other operating systems weeks later, leaving a significant portion of the population without access.[51] In France, 2 weeks after its launch, the app had been activated less than 2 million times and had sent only 14 notifications.[52]

Interoperability: a key requirement

Another important consideration for the success of these applications is the interoperability standards of devices, operating systems and applications, with existing systems in different institutions (ministry of health, hospitals, etc.) both at the regional/national and international levels. Moreover, it is important to assess whether authorities are able to cross the threshold needed for accuracy by aggregating data from multiple sources (protocols and back-ends vary across operating systems, for example). With regards to interoperability, in countries with porous borders, ensuring that applications are able to communicate with one another across borders is necessary in order to be able to trace contacts, a consideration especially relevant for Europe.

Fitting into the pandemic response puzzle

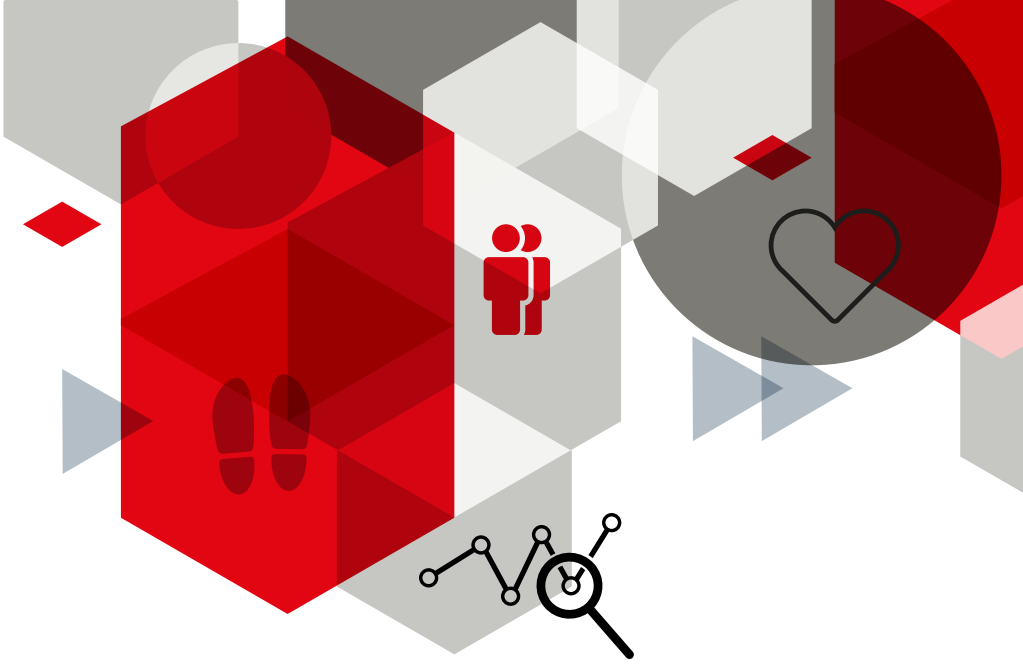
For some of these technological applications to succeed in COVID-19 response efforts, they must be embedded and enabled by their respective national public health systems. Lack of integration leads to very limited value, as seen in Spain in September with the RadarCOVID app: despite high adoption rates, the app was not integrated in many of the healthcare systems of the 17 Autonomous Communities in Spain, which might explain why the number of activated positive cases via the app was only 0.5% of all positive cases[53] vs 8% in Germany over the same time period.[54] These solutions are only a part of the complex puzzle of clinical and non-clinical processes which contribute to response measures. If other parts of the puzzle, notably non-technological considerations, including testing, ICU capacity, quarantine infrastructure, mask and protective gear availability or overall public health capacity do not function properly, then it is difficult for these solutions to be helpful. For example, in countries with deficient public health systems, low testing capacities and a population with limited options for self-isolation, a contact tracing application may not be useful at all given that these applications rely on testing certified by authorities as an indicator of infection.

51 https://threatpost.com/official-government-covid-19-apps-threats/154512/

52 https://techcrunch.com/2020/06/23/french-contact-tracing-app-stopcovid-has-been-activated-1-8-million-times-but-only-sent-14-notifications

53 https://github.com/pvieito/RadarCOVID-STATS#last-results

54 https://github.com/micb25/dka/blob/master/README.en.md



Box 10. Non-digital technologies and low-tech solutions for COVID-19

ACDC-tracing: towards anonymous citizen-driven contact tracing[55]

Though discussions seem to be centered on the use of digital technologies, it is worth considering whether and how low tech or non-digital technologies could also provide plausible and functional solutions for COVID-19 response efforts. Computer scientists Nuria Oliver (co-author of this paper) and Kristof Roomp (Microsoft, Principal Software Architect) offer a “simpler, anonymous, voucher-based” form of contact tracing, which issues vouchers to people who have tested positive. These individuals can then share the (physical or digital) voucher with individuals they have come across which in turn would allow them to be tested. This would continue to further track the path of infection and get adequate care or isolation measures in place. Thus, whilst this approach relies mostly on an individual‘s awareness of whom they have come in contact with, as well as the ability to contact them via a SMS, digital messaging or in person, it could produce interesting results even if adopted by a modest percentage of the population. As the authors of this paper have confidence in its scalability, they make a call for solutions that are privacy preserving and do not require a critical mass. Furthermore, these low-tech solutions could also complement existing initiatives such as embedding the single-use voucher functionality in mobile contact tracing apps.

55 https://arxiv.org/ftp/arxiv/papers/2004/2004.07463.pdf

Non-digital technologies and low-tech solutions for fostering access

When assessing the scientific and technological soundness of these applications it is also worth considering whether non-digital or low-tech solutions could be a feasible—and potentially overlooked—alternative to assist governments in their pandemic responses. More inclusive solutions, where a critical mass of the population does not have to engage with digital technologies, could also be relevant when access is a precondition to the effectiveness of a digital solution.

2. Commercial and economic

User incentives and technological uptake

Given that the success of most of these applications is determined by the percentage of the population that uses them or the share of the population they reflect, incentivizing users to engage with the application or to agree with the use of their data becomes a consideration with potentially great economic and commercial repercussions. For technologies that require informed and active interactions between individuals and devices or tools, ensuring that the right incentives are in place to encourage the adoption of these tools is critical for their usefulness. The main questions, thus, revolve around whether, why and how to incentivize individuals to install and use the apps and/or share their data.

Though this would not necessarily be the case for technologies using passively captured data that could be produced regardless of the pandemic, the use of these data for public health purposes, plus potentially additional uses, does warrant a close, careful look at whether the underlying permissions and consent given by the user can be legitimately and ethically justified for the declared uses. This will be further explored below: see point 3 on ethical and legal considerations. However, introducing these debates is helpful to address the key question of who benefits commercially or economically from the creation and use of these applications.

Examining underlying data economies and commercial incentives

Potentially billions of people have already or will in the near future engage with one of these applications and either proactively share their information, or their information will be shared by third parties. Under this scenario, it merits assessing who benefits economically from the use or production of these data, whether data and information monopolies are further cemented as an unintended consequence of these initiatives, whether the actions taken during the pandemic will profoundly impact the financial models of the data economy, or if surveillance and targeting of minorities is possible.

Although the future is harder to predict than ever,[56] at the moment it seems like big tech companies—along possibly with the pharmaceutical industry—may in the short term, and possibly in the long term, be some of the few to benefit from the crisis. This new situation begs one relevant question: whether the heightened power of large tech companies may further tip the economics of data for good to their benefit, for example, by fostering a gatekeeper position over data crucial for public good, for fighting pandemics or inequalities. Is that the new normality that we, as a society, collectively would want?

Long-term sustainability of current initiatives

Most, if not all, technology companies are currently providing their services to help fight COVID-19 "pro bono". At the same time, they are aware of mid- and long-term commercial considerations. Another key question is whether and how these initiatives are sustainable in the long term for the users, for the mobile operators (usually those who legally control the data), for the technology companies and for governments. The European Commission's high-level Expert Group on Business-to-Government Data Sharing published a report in February 2020 that gave their views on the technical, legal, ethical and economic barriers that prevent the use of privately held data for public interest.[57] The COVID-19 crisis may be an opportunity to realize that companies that want to contribute data for good causes incur costs and, accordingly, need to make investments that must be financed.

A key question is whether and how "data for good" initiatives are sustainable in the long term

As explored in the *Sharing is Caring* paper, most initiatives are currently based on data philanthropy models, self-funded by companies. A few telecommunication companies, such as Telefonica with SmartSteps, Orange with FluxVision or Vodafone with Vodafone Analytics, have developed commercial systems and products that can turn these activities into commercial offerings. At the same time, these solutions are proprietary and closed, such that it is difficult to ground policies on their

56 NYT:https://www.nytimes.com/2020/02/26/opinion/coronavirus-panic.html;https://www.nytimes.com/2020/05/22/opinion/sunday/coronavirus-prediction-future.html

57 https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=64954

basis. In the case of COVID-19, as for all other initiatives, indicators were provided free of charge for specific usages. As a consequence, "data for good" teams and initiatives are typically underfunded and fulfill the companies' corporate responsibility goals as they are, thereby not providing much incentive to go further.

Despite many initiatives and discussions, especially over the past five years with experiments such as the Open Algorithms project,[58] there are still no clear, sustainable "business" or financial models for this type of private-public data sharing work. Some ideas that have been explored include freemium models—whereby pricing for indicators is tiered, based on the granularity of results or use case—or the development of common data infrastructures and broadened supply/demand data sharing ecosystems. New models may have to include providing fiscal or other incentives for companies to share aggregated indicators, such as access to additional data sets through data pools, marketplaces, or other types of "data4good" collaborations. This would of course require that appropriate technological standards and safeguards be put in place as a result of extensive public discussions and debates leading to new laws, rules and regulations. Several models and modalities are possible, but it seems that the pandemic context and the resulting push for data sharing and enhanced visibility may be the right time to look seriously at the workings of the data economy. The COVID-19 crisis should be regarded as an opportunity to develop sustainable financial and economic models for data access.

3. Ethical and legal

The fact that many COVID-19 response efforts rely on personal data has naturally raised privacy flags. Such concerns need to be addressed with a combination of robust technical safeguards and privacy-preserving techniques. They also call for human oversight in the form of governance boards, ethical standards and committees and privacy principles.

Data regulation frameworks: abetting or hindering?

In many ways, the COVID-19 crisis can be seen as an "acid test" for the European General Data Protection Regulation (GDPR). Although it has been criticized as potentially hindering innovation that may save lives—which, if true, would raise severe ethical questions of its own—one main observation of these past few months is that the GDPR has not hindered digital solutions to contain the virus. Many initiatives leveraging data can be developed while respecting the GDPR framework. In that sense, the GDPR seems to have passed its first major test of value. Data regulation for COVID-19 response could usefully be enhanced by control and sanction mechanisms and evaluation procedures to assess whether they are effective in monitoring digital solutions. A good place to start is considering safeguards as well as privacy and impacts in the short and long term. It is also worth recalling that GDPR is not the only European privacy regulation that affects data sharing for scenarios like a pandemic. Indeed, as per Clause 20 of the ePrivacy Directive, "… the location of individuals by accessing the device's GPS capabilities (…) requires enhanced privacy protection".[59] This suggests both the need for strong privacy protection and greater harmonization across industries (including and beyond mobile network Operators - MNOs).

While rapid identification, response and containment are critical for curbing the spread of the pandemic, immediate implications for people's data security and digital rights as well as longer-term socio-political implications of this tradeoff are significant and should be discussed from the design stage through to the deployment of these technologies. It is evident that clear privacy regulations need to be implemented, across Europe in particular, to ensure that they are better aligned across member states. Privacy is a basic right; nevertheless, substantial insights can be gained from privacy-preserving data analysis, simultaneously protecting individuals and benefitting society. It is urgent to harmonize regulation so that initiatives can be scaled up faster. There are a number of drives underway to share best practices in "data for social good" fields (e.g. from the Global System for Mobile Communications—GSMA) and these could greatly benefit from harmonization with a view to fast deployment. Regulators will need to start thinking about how to preserve privacy for users while at the same time acknowledging the need

58 https://www.opalproject.org

59 https://edps.europa.eu/data-protection/our-work/subjects/eprivacy-directive_en;https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017PC0010&from=EN

to give new ideas and technologies the opportunity to fail fast (using tools such as "COVID19-data-sandboxes") to innovate faster for the benefit of consumers.

Appropriate safeguards

The main ethical question is whether there are fair and effective mechanisms in place (and how they are implemented) to ensure discussions on the right mix of different perspectives and considerations, oversight, auditability and, if needed, sanctions in cases of violations. Having strict regulations rooted in the principle that data subjects should have a say on how their data are used, and that their privacy be a (human) paramount right is key, as is ensuring that these regulations and principles are effectively enforced.

The lack of enforcement systems is especially salient outside of the EU. "[T]he U.S. lacks centralized data regulation but eschews government activities that would put privacy at risk. American companies are freer to gather and act upon potentially sensitive data, including location data, but seem to be getting less willing to act upon them."[60] In the Global South, for example, "[w]hile 32 African countries have data protection laws on the books, and five more have them underway, many fewer have the authorities and structures in place to implement and enforce these laws."[61]

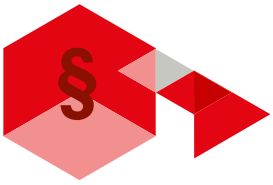
Another key principle is that of data minimization. In Mexico City's self-assessment application, the government collected personal data that, according to the privacy notice, "may be transferred to a vast array of judicial and administrative federal and local authorities".[62] In China, personal data has been misused by companies to collect data for their own commercial interest.[63] Other applications ask for multiple permissions, including geolocation and contacts' data, which are not essential for their functions. For example, the Google-Apple Exposure Notification (GAEN) interface requires enabling location services on Android phones—which also regularly send data to Google servers—for contact tracing apps to function properly.[64]

Short-term trade-offs for a long-term impact

A clear ethical risk is that the current sense of urgency to make applications and technologies available to save lives may silence concerns over potential harm and the larger political economy of data.[65] In making arguments for "more data and tech solutions", their advocates may tend to overestimate benefits while overlooking simple scientific issues, such as self-selection bias. This is clear in the case of contact tracing apps. Data suggesting that that groups that rely more on these technologies end up being less infected does not establish a causality; rather the relationship may be endogenous.

A key point is that serious ethical and legal decisions on the use of data and technologies must be informed by solid discussions and facts about the benefits versus risks, and how these trade-offs should be handled over time. This is all the more important given that there is a real risk that exceptional circumstances may make some solutions and practices palatable in the short term, only for them become the norm; part of the "new reality". In China, for emergency response purposes, vast amounts of personal data (people's location, facial recognition, infrared scans) were made available to the government and private companies which used them to quickly develop and improve other applications within and outside the scope of the pandemic response.[66] Globally, concerns have been raised around the absence of clear guidelines as to when data sets should be deleted as well as on permissible use of data outside of strict pandemic response.

Strict regulations rooted in the principle that data subjects should have a say on how their data are used are key, as is ensuring that these are effectively enforced



The European Commission requested all telecommunication operators to share data so as to create a pool of large amounts of aggregated, anonymized mobile data by region.[67] Though the announcements by the European Commissioner for the Internal Market, Financial Services, Financial Integration, Custom Rights and Taxes were made in early April, to date is are no news on the use of these data. Aggregate mobility estimations computed by the National Institute of Statistics in Spain from data shared by the three largest telecom operators are publicly available for download.[68] Other countries, such as Colombia, have followed suit, allowing telecom operators to share data more easily than in "normal" times. To date, there is no clear end to, nor exit strategy from, the new normal. This is especially worrying in countries that lack legal or regulatory protocols for oversight.

60 https://www.datanami.com/2020/04/07/contact-tracing-smartphone-apps-raise-privacy-concerns/

61 https://hewlett.org/data-in-the-time-of-covid-19-six-ways-to-reconcile-data-use-and-data-rights/

62 https://privacyinternational.org/long-read/3675/theres-app-coronavirus-apps

63 Tracing. Testing. Tweaking. Approaches to data-driven Covid-19 management in China (MERICs Paper)

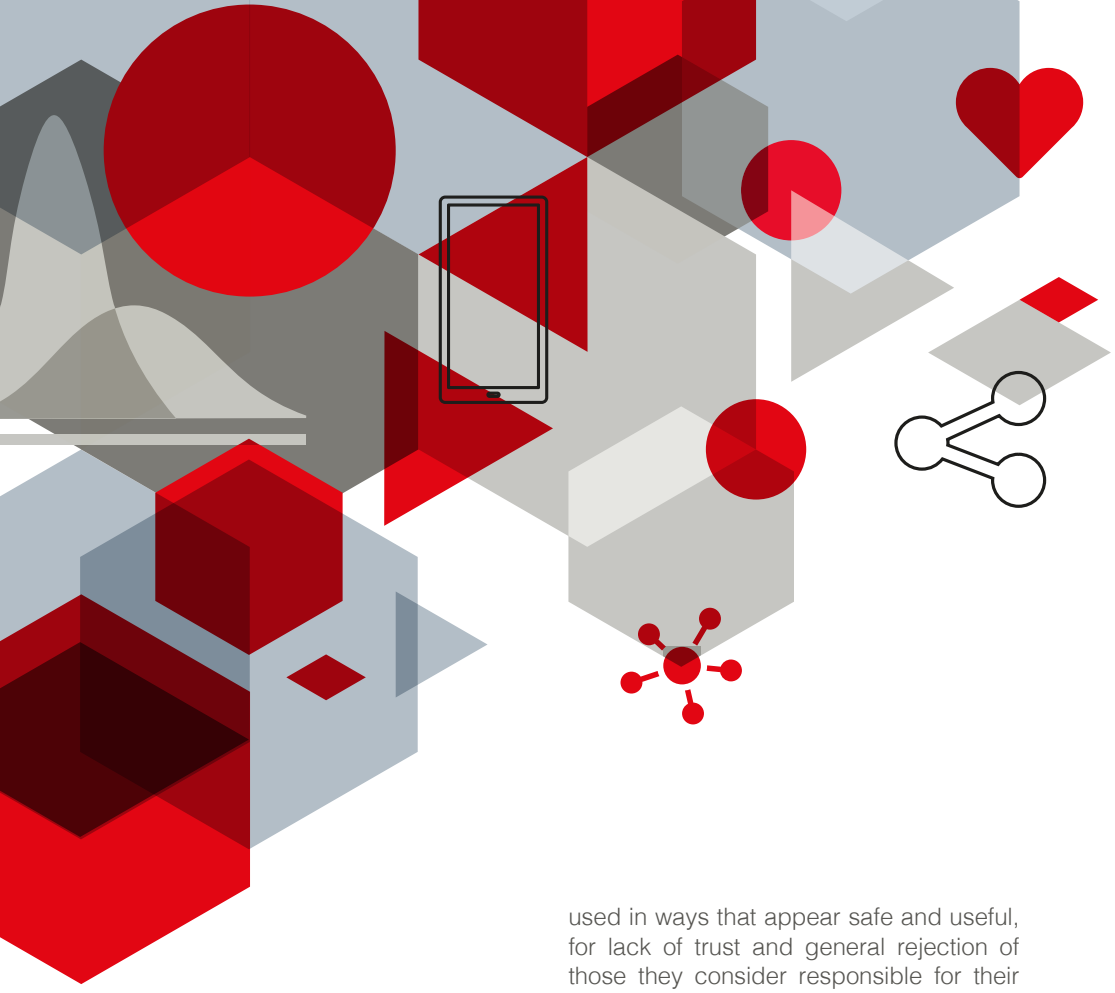
64 https://www.scss.tcd.ie/Doug.Leith/pubs/contact_tracing_app_traffic.pdf

65 https://arxiv.org/ftp/arxiv/papers/2004/2004.07463.pdf

66 Tracing. Testing. Tweaking. Approaches to data-driven Covid-19 management in China

67 https://www.politico.eu/article/european-commission-mobile-phone-data-thierry-breton-coronavirus-covid19

68 https://www.ine.es/covid/covid_movilidad.htm



While serious discussions must take place on the appropriateness of state surveillance to enforce quarantine measures or to obtain sensitive personal health data within the duration of the crisis, it is even more critical to assess and agree to democratic mechanisms the ways in which these data and technologies can and must be used in the future in a faster and reliable way.

4. Social and political

The above discussions point to recognized tensions and considerations that are political, societal and even sociological in nature. Private companies worry about data leaks revealing personal data or trade secrets but have little concern over crunching consumer data for their commercial interests or using crises as economic opportunities. Data rights advocates worry about accountability mechanisms and decision making around what data are collected, how they are stored and shared and who has access for what purposes; sometimes underestimating the expected benefits from such applications. Governments and lawmakers may be well intended in promoting their uses but may not fully grasp the potential for long-term intrusion and negative unintended effects on citizens' lives. Citizens may be alarmed at the idea of being monitored and "productified" by overbearing governments and big tech to the point of refusing data they generate used in ways that appear safe and useful, for lack of trust and general rejection of those they consider responsible for their hardships. Proponents of "data solutions" in private companies, public institutions, international and other mainstream organizations or bodies such as major universities and NGOs may be willing to use data to fight COVID-19, but too often fall short of questioning or understanding the root causes of the inequalities, along with their implications, that have been exposed and often exacerbated by the pandemic.

Exposing fault lines and exacerbating inequalities: pandemic or syndemic?

Many articles and commentators have argued that our societies' social, economic and political structures have been "laid bare" by the pandemic.[69] Far from hitting groups indifferently, the crisis has been hitting people along social, demographic and economic lines. As mentioned above, mobility models show how, in the US, poorer people reduced their mobility several days later than in richer areas. It can be hypothesized that key reasons for this included the ability to forgo income, being able to work from home, or moving out to a secondary residence. Similar patterns and trends have been observed all around the world, to the point that some argue that the pandemic is actually a syndemic, referring to a pandemic or epidemic with deep social origins and effects that need to be taken into account to respond properly.

Even so, a superficial conceptualization of the role of data would stop at a descriptive step, highlighting both general trends, key findings and differential

69 For example, https://www.theguardian.com/us-news/2020/aug/29/new-york-not-dead-coronavirus-pandemic-problems?

impacts. A more deeper conceptualization concerned with social justice would lead to investigating the proximate determinants and then the root drivers of such a difference—embracing the "syndemic" approach. The first approach can be described as merely technical; the second as political. Specific groups are also less able or willing to benefit from tech solutions. These groups include migrants and refugees who may fear being tracked, children of course, and the elderly who tend to be less connected. These differences must be taken into account when designing response systems.

Further, this differential impact is the result of subjective perceptions that are largely shaped by social media and peer pressures. As such, data and tech may have contributed to spreading—just as much as to curbing—the pandemic, since they were used voluntarily or involuntarily to diffuse misinformation or unproven allegations about lethality for young people, false remedies, ineffectiveness of face masks early on, just to name a few.

Data as a political tool

The current context is also an opportunity to reflect on the political nature and role of data. Some governments from Brazil to Florida have opted to selectively present data about morbidity, testing, co-morbidity, job losses, number of cases, deaths and more.[70] Most public health data are controlled by ministries of health that may not be willing to convey the full extent of the impacts. Many politicians, including some prominent ones, also seem to have consciously toyed with data for expected political gains.

This lack of transparency calls for renewed attention to the need to critically assess the production and use of official data—not to discard it and fuel conspiracy theories, but to improve its quality and reliability. This requires in-depth public discussions about what is measured, how, by whom and, ultimately, fundamentally, requires instilling a strong and healthy data culture and data literacy among citizens.

Reliability more broadly poses the question of who has control over decision making at different steps. Frameworks and best practices for this responsibility have put public health authorities at the center of control and supervision. Notwithstanding the central role of governments, stakeholders from the partnerships being formed and COVID-19 response technology developers will largely determine who will be the main gatekeepers to manage global (health) crises or essential services by defining how data is gathered, who has access to it and therefore what policy response is taken. These gatekeeping and control mechanisms being put in place through the design and deployment of current response have long-term implications for democracy, tech sovereignty and strategic autonomy for tech companies and government.

The importance of trust, and respect

A related point is that trust is needed for societies to function, especially in our polarized societies overloaded with data. Even if the best, most scientifically sound data and technological solutions were available, they would have no impact in the absence of trust among populations and between public and private institutions and citizens. Trust is built through regular interactions in which parties show competence, reliability, openness and, perhaps above all, respect, for one another.

The GDPR reflects and puts into practice the fundamental principle of informed consent. But anyone who has accepted terms and conditions about cookies knows that consent is given out of convenience and on the basis of trust, need or mere distraction, with little comprehension of what is actually done with the data. The risk is that trust be given unduly; or that lack of trust becomes so pervasive that many people stop providing consent. The latter option would mean that public good insights become increasingly slim, and thereby non-representative.

Notwithstanding the central role of governments, stakeholders from the partnerships being formed and COVID-19 response technology developers will largely determine who will be the main gatekeepers to manage global (health) crises

Returning to the question of uptake raised in the technical and commercial implications, none of these technological solutions would be successful without the buy-in from the societies where they are deployed.

What can we learn from all the information?

COVID-19 has given governments, technology companies, data stewards and individuals a unique opportunity to demonstrate the potential that data has to solve real life problems. Formidable innovations are surfacing that could potentially generate unprecedented amounts of pertinent information. If, upstream, policy makers had been aware and acted on the need for clear privacy protocols to ensure the wellbeing of their citizens, they could have had information in real time during the lockdowns. In most countries

70 https://www.hrw.org/news/2020/04/10/brazil-bolsonaro-sabotages-anti-covid-19-efforts; https://www.theguardian.com/world/2020/may/06/brazil-coronavirus-deaths-covid-19-bolsonaro; https://edition.cnn.com/2020/04/09/politics/florida-ron-desantis-false-claim-coronavirus/index.html

this was not the case. But COVID-19 is still rampant; this gives the opportunity to re-think the importance of data in society today. It is not something to think about in the future, but now. The current situation is a test bed that is providing the ideal conditions to move forward and learn from it as fast as possible.

Several considerations should be taken into account before moving forward:

1 When using technologies, policy makers need to be sure the functionalities of such technologies are suitable for the problem to be solved.

2 Adoption rates that depend on state-of-the-art technology is an issue that must be addressed. Not everyone will be counted or included so there needs to be provision for those not using it or not represented by the data. Moreover, non-technical factors have to be taken into account. It does not make sense to have a technical solution without having in mind health system capacity, for example.

3 Interoperability protocols are an issue in situations where borders are porous, for example.

4 What can incentivize people to engage with proposed tools? How can developers and owners of tools make sure the technical aspect they offer is actually in people's interest?

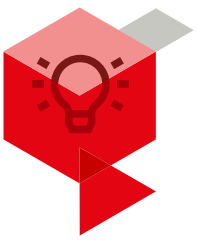
5 What are the commercial incentives for data owners? Shouldn't there be a business model for companies to share the data? Moreover, the stigma attached to companies which evaluate the use of their (extremely valuable) data, not only as a corporate responsibility goal, needs to be re-assessed. If these companies ensure that privacy concerns are addressed, surely a business model could be elaborated for them?

6 Is data regulation moving at the speed it needs to? Are there appropriate safeguards to make sure that privacy will be the most important concern when using data? Should short-term trade-offs be consented in the interest of long-term impacts for society? Will regulation allow companies willing to share their data to test their hypothesis faster in a "safe" space so that if they fail there is no reprimand on their decision?

7 COVID-19 responses have put the focus firmly on data, but its scope has evidenced that there are other issues that have to be further analyzed. Why do some people react the way they do? Is it because they have different needs? Shouldn't governments exploit data to explore previously unchartered situations?

8 As data is a non-neutral asset, what should be the role of governance? Should there be a data revolution in order to secure transparency on the information flow in countries where data is controlled by authoritarian leaders? What would it need for data to create trust between stakeholders? How can data users create incentives for people to believe in what the data shows?

It is not simply a challenge, but a unique opportunity to enhance the status of data and create a better life for all



These are some of the questions that point to the importance of using data to address the current situation. It is not simply a challenge, but a unique opportunity to enhance the status of data and create a better quality of life for all.

3. Key recommendations for a fairer post-COVID-19 world

The issues discussed in this paper are significant not only for COVID-19 response, but more widely for any digital technology application or intervention. They are also relevant to all governments and privacy enforcement authorities. Reflecting on lessons learned from the COVID-19 context, as well as history, six recommendations stand out:

1. Think and act boldly and decisively—now

A key message to retain is that this historical pandemic should be viewed as an opportunity to identify and implement the conditions to leverage data and technology to enable positive social transformation beyond short-term superficial solutionism. This may be a once-in-our-lifetime opportunity to "get it right" and "save (Big) Data and Technology from itself". [71] We should encourage and demand strategic, ambitious and long-term thinking, now.

The question is how we as societies, communities, individuals, wish to construct our world. When turning to digital technologies to collect and process data destined to instruct our collective response, we must ask ourselves which **purposes** and **whose interests** we want data and tech to serve. The bases on which data are collected all too often lack in transparency, let alone public consultation. Are we sure that our privacy is not being violated and our data misused for commercial gain or—under less democratic governments—for political gain? Too often, the "data for good" community and initiatives have failed to acknowledge political and economic factors.

71 http://ide.mit.edu/publications/saving-big-data-itself

The short-term calculation may differ from that of the long term. For example, it may be rational and indeed useful to some extent to partner with larger corporations such as a large bank or telco willing to share data with academics to inform the response, but from a long-term global perspective it feels like we are missing the forest for the trees. To be clear, corporations controlling and sharing data can and must foster social progress. But it is the time to think and talk extensively about, on the basis of facts and lessons from the past, about the kind of social progress we want in the 21st century, what the roles of different stakeholder groups are and what role digital technologies can play, since not all technological developments imply progress.

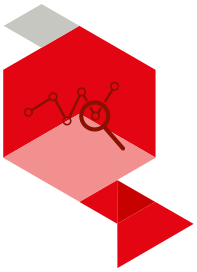
What we should aim for is progress, understood as an improvement of the quality of life – for all people, not just some. For years, global extreme poverty has, by most accounts, been steadily decreasing, driven in particular by the sharp drop in poverty in China and, to a lesser extent, India and some other large countries (i.e. Nigeria, parts of sub-Saharan Africa). At the same time, inequality has been rising to dramatic levels, and human development progress and prospects are, according to some sources, quite somber.[72] What is all but certain is that not only will the world economy shrink significantly but the pandemic is threatening years of gains painstakingly achieved in terms of gender equality, education and global poverty.

Data has shown that poor sectors of the population in the US had greater difficulty than their affluent neighbors in reducing mobility due to their socio-economic conditions; that the fatality rate of African Americans, Latinos and immigrant communities is about three times greater than that of White Americans. These are not random outcomes. Rather, they are deeply embedded in political and economic structures.

The Internet has proved to be a critical tool in the fight against COVID-19, providing populations with essential sources of information and remote learning options. But an estimated 3.6 billion people across the world—900 million of whom are in Africa—are not connected to the Internet. People who do not have access to it are at a distinct disadvantage, and the cost to governments of not collecting and exploiting data to create welfare is huge. It is in the interest of both parties to significantly expand coverage, both in the fight against the pandemic and to preserve economic development and democracy.

The pandemic has also brought to the surface some of the deep inequalities that persist in our societies. In the US, the Black Lives Matter movement has been re-ignited due to a spate of unjustifiable killings, but the explosion would probably have petered out as it had so many times before if it had not been for the heightened sensitivity to injustices exacerbated by COVID-19 and exposed unambiguously in data. As such it is as if feigned ignorance has been shredded by the pandemic—showing once again the power of data.

This historical pandemic is an opportunity to leverage data and technology for positive social transformation



Instead, thinking of a post COVID-19 world, societies, both rich and poor, should be guided by a "veil of ignorance" as coined by Rawls, whereby how one will fare in life is based on equality of opportunity rather than one's status at birth. In (more) just societies, law makers and citizens in particular should value and promote social protection mechanisms to ensure that the lowest paid are not so destitute that they cannot go without working for a few weeks in the case of a pandemic.

Translating this into concrete steps forward, opportunities lie in creating and nurturing intentional spaces such as European-wide discussions about the use of data and tech to reflect and promote the values enshrined in the Treaty of Lisbon. Intentional design of digital solutions must include social justice and inclusion within their scope through data policies and laws that hold governments and companies implementing digital technologies to account.

2. Only deploy data and technology fit for purpose and context

Balancing digital and non-digital technology solutions is of paramount importance. A good principle well known to EU stakeholders is that of subsidiarity: digital technology is used when other methods do not perform well. Moreover, technological solutions should be thought of as enablers, with a clearly stated rationale and purpose, e.g. "we are suggesting using this that way because we think it will help achieve X, Y, Z that are collectively deemed desirable outcomes, after having considered other options and variables in the equation". Next, every such project should be critically assessed regarding its impact versus its initially stated objective, and the result of the evaluation be made public.

72 https://amp-theguardian-com.cdn.ampproject.org/c/s/amp.theguardian.com/global-development/2020/jul/11/covid-19-has-revealed-a-pre-existing-pandemic-of-poverty-that-benefits-the-rich; https://chrgj.org/wp-content/uploads/2020/07/Alston-Poverty-Report-FINAL.pdf

As described in the *Business-to-Government (B2G) Data Sharing Report*,[73] authored by the European Commission's high level Expert Group on B2G Data Sharing, any (data sharing) project for social good would need to be socially acceptable, legally compliant, technically feasible and financially viable, in order for it to be sustainable over time.

It should be clear from the analysis presented in Section 1 that several initiatives, especially contact tracing, do not meet these criteria. And indeed, there are mounting calls to boycott them which are legitimate. Even as many digital solutions described here can help address the speed or cost-effectiveness of response in some cases, they need to be tightly integrated with local healthcare systems to be useful. One major stumbling block is who decides how public health experts can profile an app. In the case of the coronavirus response, this decision is falling on the lap of a few major tech companies.

To realize their full potential, the apps must be fully integrated in the response process undertaken by governments and public health authorities; this goes beyond just deploying an app. And as testing is proposed to close contacts there needs to be a human touch when notifying such people, as well as quarantine facilities for the large percentage of people who report not being able to self-isolate.[74]

In all cases, decision makers need to carefully consider and convey what they are trying to achieve and how their strategies will get them there. They also need to ensure that contextual factors and obstacles are adequately considered. Negotiating the balance of these criteria could be supported through developing resources and tools such as a checklist for each project that stipulates its goals, modalities, expected benefits and possible unintended consequences.

Context-agnostic tools (i.e. generalized so that they are interoperable among various systems) and diagnostic software could be complemented by context-specific human enablers bringing both technical/scientific and governance/procedural expertise.

73 https://ec.europa.eu/digital-single-market/en/news/experts-say-privately-held-data-available-european-union-should-be-used-better-and-more

74 https://covid19impactsurvey.org/results

3. Put and keep people at the center and in the know at all times

Privacy, autonomy and other human rights should be at the core of these initiatives. Human rights assessments which take into account unintended consequences from an ethical perspective should be carried out. Privacy, when properly safeguarded, can federate support, promote active awareness and engage data subjects.

Trust is the cornerstone upon which the legitimacy of public institutions depends. It is also critical to the success of digital solutions, notably as public-private data sharing and digital technology initiatives are taking shape and propel tech companies to the fore; they are becoming the main gatekeepers to manage this global crisis. It is important at this moment in time to examine the acceptance rates of new partnership models and COVID-19 response technology developers in order to gain deeper insights about trust. "Saving Big Data from itself" is based on legitimate, earned trust. Simply hammering "trust us" does not work.

Advocates for data protection and transparency rightly point out that COVID-19 often blurs existing boundaries between data protection and data sharing beyond the fight against the pandemic. In this context, low-threshold, transparent information and education for users becomes more important than ever to combat misinformation and avoid unintended consequences. Privacy-preserving solutions should be woven into the design of technology to minimize the risks inherent to the collection of personal data; records should be not retained beyond the duration of the crisis or its aftermath.

Attention to social and behavioral responses to digital technology interventions need to be evaluated, anticipated and embedded in the design of tools/apps. These factors need to be analyzed given that such tools, knowingly or inadvertently, create consequences and effects on human behavior, psychology and health.

In addition, technology interventions need to address structural inequities and biases and should provide inclusive solutions for a significant proportion of the older population. In general, it is and will be critical to design and implement these digital technology solutions with considerations for their longer-term implications for structural injustices and inefficiencies that the pandemic has exposed. As technology-enabled dis- and misinformation can lead to a polarization of society and an aggravation of inequalities, this situation urgently needs to be corrected and technology leveraged to fight the spread of fake news. In a non-distant future, this may require building digital public spaces or "online parks" where respectful discussions and disagreements can take place.[75]

We have made exceptions in our individual and collective freedoms and rights (e.g. freedom of movement) for the sake of group safety measures designed to save lives and slow down the pandemic. However, although such concessions may not disappear once the pandemic has been mastered, they will need to be reviewed. Thus, we would recommend a full human rights assessment of the systems, technologies, governance models and protocols post COVID-19.

This requires large-scale public consultation covering the objectives sought, data collection methods and the conservation of data. Beyond intentional spaces for discussion as suggested above, critical governance and accountability mechanisms that promote good practices can be implemented such as the establishment of ethics committees at different levels of society (city, regional, national), online portals to ensure that citizens are informed and can weigh in appropriately, and local forums via which citizens can find information and actively participate in outcomes.

4. Build "data literate" human and data systems

A major challenge and objective over the coming years will be to actively strengthen "data literacy" among both governmental agencies and citizens—defined as "the desire and ability to constructively engage in society through and about data".[76] What this means is to build data skills and develop the necessary technological infrastructures and a culture on which to base discussions and decisions on the basis of facts.

Key to this environment is fighting disinformation via both technological and non-technological solutions. The potential harmful impact of disinformation on individuals and society have become even clearer during the pandemic. Social media providers and platforms have a duty of care for public health and safety. As disinformation has become a real problem—as cited in Section 1 with the examples of attacks on 5G masts in the UK and dangerous health advice dispensed by the US and Brazilian presidents—we return to the importance of legal responsibility and accountability.

It is critical that social media platforms and tech companies be held accountable for the damaging consequences of digital interventions that are responsible for exacerbating structural or systemic violence. A related difficulty conditioning how information is communicated and by whom stems from the politicization of public interest decisions related to health, education, etc. We have witnessed this polarization in the US with the lockdown and contradictory sanitary guidance in different states or cities. There is an urgent need for multi-partisan, multi-country cooperation agreements to ensure the mid/long-term sustainability of these projects and initiatives.

We need human rights assessments of the systems, technologies, governance models and protocols post COVID-19

The COVID-19 pandemic has made evident the limited level of digitization of many public administrations worldwide, which has led to a lack of consistent, rigorous and systematic data collection and sharing. A culture of data sharing that values interoperability is also key, yet it is lacking. A pandemic requires fast decision making whereas public administrations and governments tend to have a slow modus operandi. Attention needs to be paid to how to integrate new technologies, expert teams and existing systems without going through complex and long-winded public tender processes. On the one hand, transparency and auditing are essential to ensure good use of public funds; on the other hand, decisions need to be made rapidly in crisis situations such as COVID-19. Synchronous digitization in public administrations needs to be put in place, data capacities enhanced, new technology developed, and efforts made to facilitate interoperability. The availability of indicators and high-quality data, captured, updated and shared systematically and regularly, is a must. It would allow analysts to make a diagnosis of where we are, analyze the causes, determine what has worked and what has not, and model where we are going. It is only via such a data culture that we would achieve evidence and knowledge-driven decision making.

75 https://www.wired.com/story/to-mend-a-broken-internet-create-online-parks/

76 DPA Data Literacy White Paper, 2015.

Moreover, in global human systems where data flows and fuels better decisions and sub-systems, interoperability across distributed networks and systems is key. Shifting to a culture of data sharing requires individual and institutional-level change through a focus on training, which many administrations tend to resist.

5. Test and scale sustainable business and partnerships models

Looking further at the ecosystem implications, the current COVID-19 pandemic and response context provides an unprecedented opportunity to think broadly and boldly about sustainable business models for public-private data sharing and use. Emerging from this context of a data boom and raised visibility of digital solutions are greater incentives for the private sector to allocate more resources into data sharing for public good and to formalize public-private partnerships e.g. in the form of data4good consortia at regional and global levels.

This objective also merits developing dedicated European-level discussions, building on the European Commission's high level Expert Group on B2G Data Sharing report[77] and other initiatives.[78] We also advocate for allocating more EU research funding to foster public-private-people-partnerships (PPPP) research consortia within the next EU Horizon 2027 program as a cross-cutting ambition. It must also be stressed that data sharing is no substitute for more ethical and sustainable behaviors on the part of large companies—in economic, social and environmental terms—which otherwise would merely be "data washing".

6. Consider and enforce regulation as an enabler

We should share and reuse data following certain enabling principles, such as: (1) encouraging data sharing through voluntary, market-driven mechanisms as opposed to mandatory, top-down requirements (except in limited circumstances, for example, when a firm is deemed to have significant or strategic market power); (2) sharing should only take place if it is legally compliant, ethical and socially acceptable, in line with the principles of trustworthiness and privacy by design; and (3) it should be subject to fair remuneration (which takes into account the significant upfront investment required to produce meaningful and accurate insights from large volumes of data), thereby creating the conditions for a sustainable European market for data-driven products and services.

Clear privacy regulations need to be implemented; in the case of Europe, they should be better aligned across member states. Privacy is key and there is a large amount of insight that can be extracted from privacy-preserving data analysis, protecting individuals and advancing society at the same time. It is key, moreover, that there be an effective harmonization of regulation so that initiatives can be scaled up faster. A number of initiatives to share best practices have been created in "data for social good" fields (e.g. from the GSMA) and their harmonization would leverage fast deployment.

Beyond privacy, data protection and transparency are a must. The pandemic has revealed clear tensions between data protection, privacy and data sharing, due to the urgency to combat the pandemic. This context has increasingly been used by central players and providers to capture and share data "for the sake of the pandemic". Given that misinformation can wreck havoc on societies, more then ever we need transparency regarding which data is captured, for which purposes and via what kinds of algorithms; transparency and reliable information regarding the actual state of affairs in the pandemic and, finally, citizen education. All become more important than ever, given the serious consequences that misinformation can have, as we have already experienced.

Technology is undoubtedly a key ally in the fight against the pandemic. Technology should always be fit for purpose, with guarantees of its efficacy to perform its intended purpose and be focused on the problem at hand. At the same time we, as citizens, should insist on analyses regarding the impact of such technology on the lives of people, not only during these exceptional times but in the years to come. After all, throughout history technological breakthroughs have often been precipitated by a crisis and then adapted and reused elsewhere, sometimes for good, sometimes not. Public discussions and a careful analysis of what are the parameters that matter must underpin the choice of technologies to be taken forward. Not all technological development implies progress, understood as an increase in the quality of life of people—of all people—and the planet itself. Regulation can and should play a crucial role in incentivizing the development of technologies for the progress of societies.

77 https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=64954.

78 OPAL, Vodafone-DPA paper.

Concluding remarks

COVID-19 brought most of the world to a standstill in the spring of 2020, and dangerously close to the brink of collapse in the second half of the year. None of us, apart perhaps from those living in countries ravaged by the HIV/AIDS epidemic two to three decades ago, had faced anything similar. Rapidly, divergences on how to respond emerged. Taken off guard, governments adopted varying and sometimes contradictory stances. Some leaders—including, alarmingly, in democracies—have been in denial about the realities of the virus, thereby creating cleavages and confusion amongst their citizens. Others took radical steps that plunged segments of their population into hardships. The reaction of populations has been mixed: some conformed willingly to restrictions; many proved to be, at best, wary of the methods imposed to try to curb its spread, while for some the enemy was not the virus, but overbearing public authorities infringing on civil liberties. Amid this rainbow of antagonisms, the impact in terms of deaths and job losses have been so severe that the Sustainable Development Goals may well be compromised.

Everywhere, politicians and citizens have been faced with a steep learning curve as they—we—scramble to cope. One of the best ways through which humans can learn is data. Today, we have at our disposal a wealth of digital data and technology that surpasses anything we have seen before. Some governments turned to these tools for help. But often this was done hastily, with little to no public consultation and perhaps even less consideration for the purposes they ought to serve, and how they could best serve those on the long run. COVID-19 has shed light, through data, on pre-existing inequalities that should not survive it. **Digital data and technology should be brought to the fight against COVID, but in ways that reflect and promote values of equity, efficiency, sustainability. In turn, the pandemic provides a moment in time when we should pause and reflect on our values and goals as societies and figure out how digital data and technology can help us achieve them.**

Now it is up to national governments and supranational organizations, science, business and citizens to apply learnings and do (even) better – right across the globe, and not just for privileged societies. The guiding questions for the next 6-12 months could be:

1 How can we leverage **data** and academic insights to enable governments, the private sector and citizens to make informed decisions as to how to cope best with the pandemic?

2 How can we ensure that **all citizens** have access either to digital solutions, or an effective alternative, to assess and alleviate the impact of COVID-19?

3 How can we foster **collaboration** between the public and private sectors and civil society to fight back and build back better?

4 How can we combat **disinformation** on COVID-19 and key challenges of our time to avoid the further polarization of societies?

USING DATA TO FIGHT COVID-19
And Build Back Better

Authors
Emmanuel Letouzé, PhD
Director and co-Founder, Data-Pop Alliance
Nuria Oliver, PhD
Chief Data Scientist, Data-Pop Alliance
Maria Antonia Bravo,
Program Officer, Data-Pop Alliance
Natalie Shoup,
Research Affiliate, Data-Pop Alliance

Editor
Vodafone Institute for Society and Communications
Behrenstraße 18
10117 Berlin
Germany

Editorial Management
Inger Paus, Managing Director, Vodafone Institute
Julia Ebert, Vodafone Institute

Assistance
Pedro Rente Lourenco, Vodafone Group
Matthew Allison, Vodafone Group
Juan Camilo Mejía, Data-Pop Alliance

Layout
Nick Böse
Robin Kupski

Data-Pop Alliance, NYC
October 2020

www.vodafone-institut.de
Twitter: @vf_institute
LinkedIn: www.linkedin.com/company/vodafone-institute/
Facebook: www.facebook.com/VodafoneInstitute