Enlisting Big Data in the Fight Against Coronavirus
Before the
Committee on Commerce, Science, and Transportation
April 9, 2020

Statement of
Michelle Richardson, Director, Privacy and Data Project
Center for Democracy and Technology

On behalf of the Center for Democracy & Technology (CDT), thank you for the opportunity to testify about enlisting big data in the fight against coronavirus. CDT is a nonpartisan, nonprofit 501(c)(3) charitable organization dedicated to advancing the rights of the individual in the digital world. CDT is committed to protecting privacy as a fundamental human and civil right and as a necessity for securing other rights such as access to justice, equal protection, and freedom of expression. CDT has offices in Washington, D.C., and Brussels, and has a diverse funding portfolio from foundation grants, corporate donations, and individual donations.[1]

We commend the committee for holding this paper hearing while usual congressional functions are suspended. Both governmental and corporate responses to the coronavirus are evolving quickly, and Congressional oversight will hopefully encourage best practices while deterring behavior that is unjustifiably risky. It is never too soon to ask whether data use violates privacy, treats people unfairly, or fails to solve the problem for which it was obtained.

Academics, public health officials, and advocacy organizations have already registered their concerns with coronavirus response efforts violating privacy or other human rights.[2] These

---

[1] Annual Report: Center for Democracy & Technology, https://annualreport.cdt.infohttps://cdt.org/wp-content/uploads/2020/03/2018-CDT-Annual-Report-compressed.pdf/.

[2] *See* Joint Civil Society Statement: States' Use of Digital Surveillance Technologies to Fight Pandemic Must Respect Human Rights (Apr. 2, 2020), https://www.accessnow.org/cms/assets/uploads/2020/04/Joint-statement-COVID-19-and-surveillance-FINAL1.pdf ; Estelle Massé, *Recommendations on Privacy and Data Protection in the Fight Against Covid-19*, (Access Now, 2020), https://www.accessnow.org/cms/assets/uploads/2020/03/Access-Now-recommendations-on-Covid-and-data-protection-and-privacy.pdf; Privacy International, *Bluetooth Tracking and COVID-19: A Tech Primer*, (Mar. 31, 2020), https://privacyinternational.org/explainer/3536/bluetooth-tracking-and-covid-19-tech-primer; Letter to Congress

concerns are only heightened when a government agency is the collector or recipient of the data, the data is individualized, or the data collection or use is mandatory. It's important to note that these entities and individuals do not categorically oppose all corporate or government collection or use of data in response to the coronavirus. They instead seek to ensure it is conducted in a way that does not risk privacy or human rights.

I. Big Data Can Contribute to the Fight Against the Coronavirus Without Jeopardizing Privacy

A. Initial U.S. Response

In the face of an unprecedented epidemic, American business has sprung into action. Many companies have donated money to relief efforts, and others have made in-kind donations of equipment, goods, or services that are critical to a successful response. For example, some have switched from producing clothes and alcoholic beverages to making personal protective equipment and hand sanitizer.[3] Others have provided transportation to redistribute goods to where they are most needed or donated software to help coordinate first responders.[4]

Several technology companies activated their "Data for Good" operations, which may generally be described as programs that use and share information to advance the public interest. While the data inputs of these programs may have been originally collected or accessed for commercial purposes, the use and output of these programs is often shared with nonprofits, governmental entities, or the public. In some instances, the technology companies provide data to nonprofits or academics under agreements requiring that the recipients will use data in ways that are privacy-protective. In other instances, companies will donate their computing power—either the hardware or the artificial intelligence to process the data. In some cases, companies provide all of these things.

---

from Fifteen Nonprofit Organizations to Congress (Mar. 20, 2020), https://www.citizen.org/wp-content/uploads/Covid-Response-Privacy-Protections-Letter-3-20-with-Signatories.pdf; Open Letter From Public Health and Legal Experts to Vice President Mike Pence and Public Health and Legal Experts (Mar. 2 2020), https://law.yale.edu/sites/default/files/area/center/ghjp/documents/final_covid-19_letter_from_public_health_and_legal_experts.pdf.

[3] *See* Lindsay Cates, *How 10 Small Companies Are Fighting Coronavirus in Creative Ways*, US Chamber of Commerce (Mar. 27, 2020), https://www.uschamber.com/series/above-the-fold/how-10-small-businesses-are-fighting-coronavirus-creative-ways.

[4] *See generally* US Chamber of Commerce Foundation, Corporate Aid Tracker: COVID-19 Business in Action, https://www.uschamberfoundation.org/aid-event/corporate-aid-tracker-covid-19-business-action.

At the time of this hearing, big data processing in the U.S. falls into several categories:

- Publication and analysis of existing research: Academic institutions, think tanks, and companies have published all or part of 29,000 academic articles in machine-readable text for data mining purposes.[5]

- Medical research: Computer processing is a staple of modern medical research, and companies are donating platform power to speed up analysis. For example, Microsoft has donated supercomputing resources to help map the coronavirus on an atomic level.[6] IBM, in partnership with the White House Office of Science and Technology Policy, is coordinating the COVID-19 High Performance Computing Consortium, which provides computing power to approved academic research projects.[7]

- Symptom trackers: Government agencies, private companies, and academic institutions are launching symptom trackers. These include the Centers for Disease Control, Verily, the University of Alabama, independent medical researchers, Microsoft, and Apple.[8] Some of these trackers help people determine whether they should seek medical treatment, but others collect information for research purposes and to identify areas that may have higher infection rates.

---

[5] *See* White House Office of Science and Technology Policy, Call to Action to the Tech Community on New Machine Readable COVID-19 Dataset (Mar. 16, 2020), https://www.whitehouse.gov/briefings-statements/call-action-tech-community-new-machine-readable-covid-19-dataset/.

[6] *See* Conor Hale, *Microsoft Lends Petaflops to ImmuityBio's Coronavirus Protein Modeling Efforts*, Fierce BioTech (Apr. 1, 2020), https://www.fiercebiotech.com/medtech/microsoft-lends-petaflops-to-immunitybio-s-coronavirus-protein-modeling-efforts.

[7] *See* COVID-19 High Performance Computing Consortium, https://covid19-hpc-consortium.org/.

[8] *See* Verily, Project Baseline, https://www.projectbaseline.com/; Apple, COVID-19 Screening Tool, https://www.apple.com/covid19/; ZOE, COVID Symptom Tracker, https://covid.joinzoe.com/us; Center for Disease Control, Coronavirus (COVID-19), https://www.cdc.gov/coronavirus/2019-ncov/index.html (providing a self-checker to assess symptoms); University of Alabama at Birmingham, COVID-19 Symptom Tracker, https://www.helpbeatcovid19.org/. The Apple App Store and Google Play are reportedly removing coronavirus symptom tracker and other apps that do not have a connection to government agencies or health professionals. *See* Thomas Brewster, *Google Bans Coronavirus Infection Trackers...But Not Before They Get 400,000 Downloads*, Forbes (Mar. 24, 2020), https://www.forbes.com/sites/thomasbrewster/2020/03/24/google-bans-coronavirus-apps-but-after-400000-downloads/#35c8400234c0.

- Heatmaps: Heatmaps are available from multiple sources, including platforms, newspapers, and media outlets.[9] They often represent the number of cases officially confirmed by test and usually report data at the city, county, or state level. Population density maps have also been used to determine where medical resources should be deployed.[10]

- Aggregate movement patterns and statistics: Location information has been processed to determine how much people are travelling and where they are going. For example, Google is releasing state mobility reports reflecting the increase or decrease in visits to grocery stories, entertainment venues, and more.[11] Facebook releases daily mobility data that reveals travel between populations and comparative rates of movement compared to pre-coronavirus levels.[12]

As one can see, there are a number of ways big data processing can advance the coronavirus response without unduly risking individual privacy. Some of this data does not reflect personal information at all—such as state level statistics that are aggregated and cannot be associated with specific individuals.

But there are also uses of data that are riskier. For example, if heat maps or case reporting become too granular, it may be easy to associate a positive coronavirus status with identifiable people.

Symptom trackers may also pose privacy risks if they collect personal information. Some of the symptom trackers here report that they do not collect or retain any personal health information at all.[13] Others are less clear. One notifies users that it uses a number of cookies and does not honor "Do Not Track" browser indicators.[14] It goes on to state that personal information will be shared broadly with medical and academic institutions, but recognizes three versions of "rights" that people are entitled to: European, Californian, and other Americans.

---

[9] Coronavirus Map: Tracking the Global Outbreak, N.Y. Times (last updated Apr. 6, 2020), https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html; Johns Hopkins University Coronavirus Resource Center, Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering, https://coronavirus.jhu.edu/map.html; Will Maddox, *The Nonprofit Mapping COVID-19 in Dallas*, D Magazine (Mar. 25, 2020), https://www.dmagazine.com/healthcare-business/2020/03/the-nonprofit-mapping-covid-19-in-dallas/.

[10] *See* Facebook Data for Good, Population Density Maps, https://dataforgood.fb.com/tools/population-density-maps/; Development Data Project, Facebook and Mapbox: Addressing COVID-19 through Public-Private Data Partnerships—Where Do We Put New Testing Facilities?, https://datapartnership.org/updates/covid19-and-public-private-data-partnerships/.

[11] Google, COVID-19 Community Mobility Reports, https://www.google.com/covid19/mobility/. Google uses data collected through "Location History," which reflects the location of devices that are signed into Google accounts.

[12] Facebook Covid-19 Response, https://dataforgood.fb.com/docs/covid19/.

[13] Apple Covid-19 Screening Tool, https://www.apple.com/covid19/.

[14] *See* ZOE, *supra* n.8.

B. Learning from Past Emergencies and Epidemics

It is important to note that epidemics and natural disasters are regrettably common, even if not usually seen in the current magnitude. As a result, corporate, academic, and government responses can and should draw on the extensive experience of epidemiologists, public health officials, and emergency response experts to be informed by what has worked in the past. The Centers for Disease Control, World Health Organization, academic institutions, think tanks, and human rights organizations have developed best practices for tracking and combating epidemics, for example. While they recognize the incomparable value of data, they also seek to use it in ways that support meaningful responses. Innovation should certainly continue, but Congress should be guided by these experts and not endorse novel or untested responses. There is simply too much at stake.

II. The Lack of a Comprehensive Privacy Law Can Undermine Trust in Health Services and Inhibit Disease Response

The United States does not have a comprehensive privacy law to protect Americans' personal information. Instead, we have a patchwork of federal, state, and local laws that regulate specific sectors or data sets like education records, financial records, children's information, and health records if they are held by certain entities. This has led to the explosion of risky and exploitive data-driven behaviors in the vast unregulated space in between. It has reduced public trust in technology companies, and as a result, may discourage people from using legitimate services or waste precious time and resources on untested products. This in turn may inhibit the coronavirus response.

A. Our health privacy statute covers only a small portion of the health data ecosystem, and does not apply to data in the consumer market

Most people are familiar with the Health Insurance Portability and Accountability Act (HIPAA),[15] which governs identifiable "protected health information" (PHI) and includes demographic and other information related to current or past health status. Most people do not understand that HIPAA only applies to data created, held, or transmitted by certain entities such as doctors,

[15] Health Insurance Portability and Accountability Act of 1996 (HIPAA), 29 U.S.C. § 1181 et seq., 42 U.S.C. § 1320d et seq. (1996). When HIPAA applies, its Privacy Rule regulates how PHI can be used and disclosed, such as for treatment, payment, public health, and research, and it specifies when an entity needs to obtain the prior authorization of the data subject. It also establishes rights for individuals, including the right to obtain a copy of PHI and to request amendments to this data. HIPAA permits states to supplement these protections with additional requirements and many states have done so. Importantly, when PHI leaves a covered entity to join the general consumer data ecosystem, it loses HIPAA protections.

health plans, and health care clearinghouses, and their contractors (known as "business associates").

Emerging technologies are collecting increasing amounts of data that reflect physical or mental health or status. The entities collecting the data include social media networks, mobile app developers, smartphone and home IoT manufacturers, and internet search engines. They may also be more traditional enterprises like life insurers, retailers, credit card companies, employers, and data brokers. While consumers are likely to understand that fitness trackers, DNA tests, diet apps, and similar services create health-related information, they do not foresee location data, search terms, and their online browsing history becoming the basis for the ads they see or the services they are offered. The average person just cannot anticipate that any data can become health data if a company chooses to use it for those purposes.

To be clear, this type of commercial data use is regulated under the Federal Trade Commission Act that prohibits unfair and deceptive practices.[16] While that statute does not have clear rules about the collection, use, and sharing of data, it has been enforced against companies that have failed to provide meaningful notice and choice to consumers or used data in ways that harm people. Case-by-case enforcement has slowly developed a body of case law but a lack of hard and fast rules, and limited resources have allowed reckless and exploitive behavior to flourish.

B. The unregulated consumer market has permitted broad sharing and use of health information

There is mounting evidence that non-HIPAA-covered entities are regularly collecting, using, and sharing consumer health information in ways that would surprise individuals and arguably exploit their sensitive data. Here are just a few examples where unregulated health data about consumers was collected and shared:

- A 2019 study reported that a range of health and wellness applications, including smoking cessation, fitness tracking, mental health, and period tracking applications had shared their users' sensitive information with third parties.[17]

[16] Federal Trade Commission Act of 1914 (FTCA), 15 U.S.C. §§ 41-58 (1914).
[17] *See* Kit Huckvale, John Torous & Mark E. Larsen, *Assessment of the Data Sharing and Privacy Practices of Smartphone Apps for Depression and Smoking Cessation*, J. Am. Med. Ass'n (2019), https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2730782?utm_source=For_The_Media&utm_medium=ref%20erral&utm_campaign=ftm_links&utm_term=041919%E2%80%8B; Kari Paul, *Fitness & Health Apps May Be Sharing The Most Private Details About Your Life*, MarketWatch (Mar. 5, 2019), https://www.marketwatch.com/story/fitness-and-health-apps-may-be-sharing-the-most-private-details-about-your-life-2019-02-26.

- A *Financial Times* report found that popular health websites share sensitive consumer information, including medical symptoms, diagnosis, drug names, and menstrual and fertility information with advertisers and data brokers.[18]

- The app GoodRX shared consumer information, including specific prescription information, with advertisers.[19]

- DNA testing company 23andMe recently sold the rights to a drug that it developed using its customers' data.[20]

- Voice recordings from consumer smart speakers are being studied to detect signs of dementia.[21]

- Consumer pregnancy apps are sharing users' fertility data with their employers.[22]

- Data brokers and advertisers have used inferred health information to target ads in a harmful and exploitative manner, such as serving treatment ads to someone facing challenges with addiction.[23]

---

[18] *See* Madhumita Murgia and Max Harlow, *How Top Health Websites are Sharing Sensitive Data with Advertisers*, Fin. Times (Nov. 12, 2019), https://www.ft.com/content/0fbf4d8e-022b-11ea-be59-e49b2a136b8d.

[19] *See* Shoshana Wodinsky, *GoodRx Shared My Prescriptions with Third Parties—And It's Perfectly Legal*, Gizmodo (Feb. 27, 2020), https://gizmodo.com/goodrx-shares-my-prescriptions-with-third-parties-and-i-1841772965?rev=1582832321935.

[20] *See* Jessica Hamzelou, *23andMe Has Sold the Rights to Develop a Drug Based on Its Users' DNA*, NewScientist (Jan. 10, 2020), https://www.newscientist.com/article/2229828-23andme-has-sold-the-rights-to-develop-a-drug-based-on-its-users-dna/.

[21] *See* Michael Casey, *Researchers at Dartmouth and UMass Boston Hope Voice Assistants Can Spot Signs of Dementia*, Bos. Globe (Feb. 1, 2020), https://www.boston.com/news/local-news/2020/02/01/researchers-at-dartmouth-and-umass-boston-hope-voice-assistants-can-spot-signs-of-dementia.

[22] *See* Drew Harwell, *Is Your Pregnancy App Sharing Your Intimate Data With Your Boss?*, Wash. Post (Apr. 10, 2019), https://www.washingtonpost.com/technology/2019/04/10/tracking-your-pregnancy-an-app-may-be-more-public-than-you-think/?arc404=true.

[23] *See What Information Do Data Brokers Have on Consumers, and How Do They Use It? Before the S. Comm. on Commerce, Science, and Transportation,* 113th Cong. (2013) (statement of Pam Dixon, Executive Director, World Privacy Forum), https://www.worldprivacyforum.org/wp-content/uploads/2013/12/WPF_PamDixon_CongressionalTestimony_DataBrokers_2013_fs.pdf. *See, e.g.*, Upturn, *Led Astray: Online Lead Generation and Payday Loans* (Oct. 2015), https://www.upturn.org/static/reports/2015/led-astray/files/Upturn_-_Led_Astray_v.1.01.pdf; Michael Corkery, *Google Sets Limits on Addiction Treatment Ads, Citing Safety*, N.Y. Times (Sept. 14, 2017), https://www.nytimes.com/2017/09/14/business/google-addiction-treatment-ads.html.

- Insurance companies are using data from their partners, such as purchase and browsing history and social media activity, to make decisions regarding eligibility, rates, and targeted marketing.[24]

C. Passing a comprehensive privacy law would provide consumers and corporations with clarity while protecting individual rights

Chairman Wicker, Ranking Member Cantwell, and several other committee members have introduced comprehensive privacy legislation that would provide heightened protections to currently unregulated health information.[25] Importantly, they also permit the sharing of information in emergency situations, in support of public interest research, and for broader uses of data that are anonymized, de-identified, or used in aggregate. Additional provisions are likely to deter abuses, such as individual rights to access, correct, and delete data, and obligations on companies to conduct risk assessments on their data use.

While some may argue that a privacy law would only hamper innovations around the coronavirus response, failure to impose reasonable protections may backfire. First, improper use of consumer health data leads to an erosion in consumer trust that may deter people from voluntarily sharing information with legitimate entities for important public health purposes. If consumers do not trust new technologies they will refrain from using them. Second, the race to collect as much information as possible only encourages scams, false cures, and "click bait" to entice people to share information that may be very revealing. To the extent that people rely on such products, they may do so in lieu of accessing real information or meaningful medical help. If the United States faces such a health crisis in the future, we will likely find that the clarity and trust that comes from a meaningful privacy law actually serves public health.

III. Urgent Use Case: Location and Proximity Tracking Around the World

While many types of data may eventually be used to gain insight about the coronavirus, location and proximity information are being used now by a wide array of public interest, government and corporate entities. There are also reports that the U.S. government may seek to expand its access to more individualized forms of this data or call on companies in the

[24] *See, e.g.*, Marshall Allen, *Health Insurers are Vacuuming Up Details About You—And It Could Raise Your Rates*, ProPublica (July 17, 2018), https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates; Jessica Baron, *Life Insurers Can Use Social Media Posts to Determine Premiums, As Long As They Don't Discriminate*, Forbes (Feb. 4, 2019), https://www.forbes.com/sites/jessicabaron/2019/02/04/life-insurers-can-use-social-media-posts-to-determine-premiums/.

[25] Staff discussion draft, United States Consumer Data Privacy Act of 2019, 116th Cong. (2019), https://aboutblaw.com/NaZ; Consumer Online Privacy Rights Act, S.2968 116th Cong. (2019).

private sector to more aggressively use location data they are already collecting in the course of business.

To that end, we note for the Committee that location data is unusually sensitive and difficult to anonymize. The use of this data in other countries has had serious implications for privacy while the value added by these programs is unclear.

A. Location data is uniquely difficult to de-identify or anonymize

Location data, because of its revealing nature, is very difficult to sufficiently anonymize. In two blockbuster investigations by *The New York Times*, reporters analyzed a database of de-identified location data (location data points not directly connected to an individual's name), and demonstrated that the data could become identifiable.[26] From the millions of data points, investigators were able to identify many individuals due to the uniqueness of their travel patterns.[27] For example, they were able to identify participants at a political protest by tracking them to their homes.[28] The *Times* is not alone. Researchers at MIT and a leading Belgian university, who studied 15 months of anonymized mobile phone location data of 1.5 million people, were able to uniquely identify 95 percent of the individuals in their study from just four data points each.[29] Comparatively, they note it takes 12 points on a fingerprint to identify an individual. Part of what makes location information difficult to anonymize is the wealth of information that can be combined with the dataset to enable conclusions about identity to be drawn. For example, an anonymized route of a morning commute can reveal where the commuter likely lives and works. This information almost always uniquely describes one person and is identifiable to that person. Publicly available housing records, or online information about who works where, for example, can likely reveal the commuter's identity. Given these risks, it is a best practice for companies to pursue solutions that center on aggregated data that illustrate trends and to provide those outputs to the U.S. government, as opposed to volunteering datasets.

---

[26] *See* Jennifer Valentino-DeVries et. al., *Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret*, N.Y. Times (Dec. 10, 2018), https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html; Stuart A. Thompson and Charlie Warzel, *Twelve Million Phones, One Dataset, Zero Privacy*, N.Y. Times (Dec. 19, 2019), https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html.

[27] *See* Valentino-DeVries et. al., *supra* n.26.

[28] *See* Stuart A. Thompson, *supra* n.26.

[29] *See* Yves-Alexandre de Montjoye et. al., *Unique in the Crowd: The Privacy Bounds of Human Mobility*, Nature (2013), https://www.nature.com/articles/srep01376.

B. States and the federal government have extended protections for location information

In 2012, the Supreme Court held in *United States v. Jones* that installing a GPS tracker on a vehicle and using the device to monitor the vehicle's movements constituted a search under the Fourth Amendment.[30] States responded by going even further and passing legislation requiring their law enforcement entities to get a warrant prior to accessing an individual's historical location information, or tracking them in real time.[31] Additionally, some state high courts determined that location information was protected under their constitution.[32] CDT believes this trend is compelling proof that there is an expectation of privacy in location information.

As Supreme Court Justice Sotomayor noted in *Carpenter v. United States,* a comprehensive record of a person's movements "reflects a wealth of detail about [the person's] familial, political, professional, religious, and sexual associations."[33] From location information one can infer sensitive medical, religious, or legal needs from visits to cancer centers, churches or mosques, or visits to an immigration legal clinic. And when the government collects and retains this information over a lengthy period of time, the government accesses "a category of information otherwise unknowable."[34] This type of information grants the government the ability to "travel back in time to retrace a person's whereabouts[.]"[35] As a result, location tracked over time is deeply personal information that is among the "privacies of life" protected by the Fourth Amendment.[36]

C. Countries are adopting location tracking and proximity-based contact tracing although the effectiveness and privacy impact are unclear

While the details of many location tracking programs are still unknown, researchers are able to provide basic information about how many of them collect and share data.[37] Two major

---

[30] *United States v. Jones*, 565 US 400 (2012).
[31] *See, e.g.*, Peter Cihon, *Status of Location Privacy Legislation in the States: 2015*, ACLU (Aug. 26, 2015), https://www.aclu.org/blog/privacy-technology/location-tracking/status-location-privacy-legislation-states-2015.
[32] *See, e.g.*, *Commonwealth v. Augustine*, 4 N.E.3d 846 (Mass. 2014) (holding that under the Massachusetts constitution an individual has a reasonable expectation of privacy in historical location information held by a service provider); *State v. Earls*, 214 N.J. 564 (N.J. 2013) (holding that under the New Jersey constitution an individual has a reasonable expectation of privacy in their cell phone location information).
[33] *Carpenter v. United States,* 138 S. Ct. 2206, 2217 (2018) *(quoting United States v. Jones, 132 S.Ct. 945, 955 (2012) (Sotomayor, J., concurring).*
[34] *Carpenter v. United States*, 138 S. Ct. 2206, 2218 (2018).
[35] Id. at 2214.
[36] Id.
[37] *See* Future of Privacy Forum, *Privacy & Pandemics; The Role of Mobile Apps (Chart)*, https://fpf.org/wp-content/uploads/2020/04/Privacy-Pandemics-The-Role-of-Mobile-Apps.pdf.

approaches have emerged: contact tracing through location comparisons and contact tracing through proximity detection.

Location tracking

In China, the government partnered with the private sector and produced a smartphone app that dictates whether a person could travel freely or be quarantined.[38] The app generated a green, yellow, or red code on the user's mobile phone indicating each person's contagion risk. When *The New York Times* analyzed the software's code, it reported that the app not only determines in real time whether someone poses a contagion risk, it also shares information with the police. Moreover, the *Times's* analysis also found that each time a person's color code is scanned — at a health checkpoint, for instance — his or her current location appears to be sent to the system's servers, possibly allowing the authorities to track people's movements over time.[39] Chinese officials have not explained how the app makes its determinations or what data the app utilizes.[40]

In South Korea, the government initially posted the detailed location history of individuals who tested positive for coronavirus online.[41] These public postings included a host of sensitive personal information including, the timing of individuals' daily routines, whether they wore masks in public, what public transit stops they visited, where they ate and socialized, and even the name of the clinics where they were tested for the virus.[42] Those whose information was released frequently found themselves stigmatized and attacked online.[43] In an effort to ensure that South Koreans' are not deterred from getting tested, the government has taken to reevaluating how it shares information about confirmed cases moving forward.[44]

In Israel, Prime Minister Netanyahu ordered his country's domestic spy agency, Shin Bet, to tap cell phone location information it had been secretly collecting in bulk since 2002 in Israel and

---

[38] *See* Paul Mozur, Raymond Zhong & Aaron Krolik, *In Coronavirus Fight, China Gives Citizens a Color Code, With Red Flags*, N.Y. Times (Mar. 1, 2020), https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html.
[39] Id.
[40] *See* Yuan Yang et al., *China, Coronavirus and Surveillance*, Fin. Times (Apr. 2, 2020), https://www.ft.com/content/760142e6-740e-11ea-95fe-fcd274e920ca.
[41] *See* Natasha Singer & Choe Sang-Hun, *As Coronavirus Surveillance Escalates, Personal Privacy Plummets,* N.Y. Times (updated Mar. 24, 2020), https://www.nytimes.com/2020/03/23/technology/coronavirus-surveillance-tracking-privacy.html.
[42] Id.
[43] *See* Min Joo Kim & Simon Denyer, *A 'Travel Log' of the Times in South Korea: Mapping the Movements of Coronavirus Carriers*, Wash. Post (Mar. 13, 2020), https://www.washingtonpost.com/world/asia_pacific/coronavirus-south-korea-tracking-apps/2020/03/13/2bed568e-5fac-11ea-ac50-18701e14e06d_story.html.
[44] *See* Singer, *supra* n.41.

the occupied territories.[45] Using this data, government officials text individuals who they believe may have been exposed to a person who tested positive for COVID-19 and direct them to self-quarantine.[46] Shin Bet states that this program has led to the isolation of more than 500 people who later tested positive for COVID-19.[47] The program has been criticized by privacy advocates and is also facing scrutiny by the Israeli Parliament.[48]

Proximity monitoring

In Singapore, the government was also interested in determining whether individuals have been exposed to people recently diagnosed with the coronavirus and developed an app called TraceTogether as a way to improve contact-tracing efforts.[49] As of April 6, the app had over 1,000,000 unique users.[50]

In April, India's government launched a coronavirus tracking app in conjunction with private sector partners.[51] The app is designed to use a phone's location and Bluetooth data to determine if the phone's owner has been within six feet of a person infected with COVID-19. According to reports, the anonymised data collected by the app is checked against a database of known cases and their movements. If an app user tests positive or has been in close contact with someone who has, the app will share that data with the Indian government.[52]

---

[45] *See,* David M. Halfinger, Isabel Kershner, and Ronen Bergman, *To Track Coronavirus, Israel Moves to Tap Secret Trove of Cellphone Data,* N.Y. Times (March 18, 2020), https://www.nytimes.com/2020/03/16/world/middleeast/israel-coronavirus-cellphone-tracking.html.

[46] *See* Daniel Estrin, *Israel Begins Tracking and Texting Those Possibly Exposed to the Coronavirus*, Nat'l Pub. Radio (Mar. 19, 2020), https://www.npr.org/2020/03/19/818327945/israel-begins-tracking-and-texting-those-possibly-exposed-to-the-coronavirus.

[47] *See* Daniel Estrin, *Israel's Defense and Spy Agencies Step Up Anti-Coronavirus Efforts*, Nat'l Pub. Radio (Apr. 2, 2020), https://www.npr.org/sections/coronavirus-live-updates/2020/04/02/825898681/israels-defense-and-spy-agencies-step-up-anti-coronavirus-efforts.

[48] Id.

[49] *See* Saheli Roy Choudhury, *Singapore Says It Will Make Its Contact Tracing Tech Freely Available to Developers,* CNBC (updated Mar. 25, 2020), https://www.cnbc.com/2020/03/25/coronavirus-singapore-to-make-contact-tracing-tech-open-source.html.

[50] *See* TraceTogether, a Singapore Government Agency Website, https://www.tracetogether.gov.sg/, *last visited April 6, 2020*., https://tracetogether.zendesk.com/hc/en-sg/articles/360045013734-24-Mar-600K-supporters-Advice-on-scams-and-more.

[51] *See* Mia Hunt, *India Launches App to Track Corona Cases and Tackle Misinformation*, Global Government Forum (Apr. 6, 2020), https://www.globalgovernmentforum.com/india-launches-apps-to-track-corona-cases-and-tackle-misinformation/.

[52] *See* Ivan Mehta, *Indian Government Officially Launches Its Coronavirus Tracking App*, The Next Web (Apr. 2, 2020), https://thenextweb.com/corona/2020/04/02/indian-government-officially-launches-its-coronavirus-tracking-app/.

Just this month, a group of technologists and scientists in the European Union unveiled that they are also working on a contact-tracing technology, similar to that deployed in Singapore, but designed to comply with the EU's stringent privacy rules, namely the EU's General Data Protection Regulation.[53] The European effort, dubbed the Pan-European Privacy-Preserving Proximity Tracing (PEPP-PT), was created to develop technical mechanisms and standards that protect privacy while also leveraging digital resources to bolster pandemic response efforts.[54] The PEPP-PT model does not collect location data, contact information, or identifiable features of the end devices. Instead, the app will generate temporary IDs. When two or more smartphones running the app come into proximity, they exchange these IDs, encrypt and save them locally on the device.[55]

The central idea behind both the Singapore, Indian, and EU approaches is leveraging the Bluetooth capabilities of smartphones to log certain interactions between devices that, in turn, help prevent the spread of the coronavirus by notifying individuals who have come into close contact with an infected person.[56] Neither reveals to the notified individual the identity of the infected person with whom they came into contact or the place where the contact occurred. These apps are voluntary, at least at inception: the user voluntarily downloads the app that records the unique identifiers of others who have voluntarily downloaded the app. They rely, with some degree of transparency, on Bluetooth data, rather than on less-precise location information surreptitiously compelled from a communications service provider that may produce more false positives.

D. Personally identifiable information is being used overseas in ways that may threaten health and safety

As discussed above, there are different location and proximity data sets available on the consumer market and they pose different accuracy, privacy and security challenges depending on how contact-tracing applications are engineered. We flag for the Committee that human rights organizations have documented risky, draconian, or opportunistic coronavirus responses overseas that grow from aggressive state-driven disease tracking.[57] While U.S. institutions may

---

[53] *See* Pan-European Privacy-Preserving Proximity Tracing, https://www.pepp-pt.org.

[54] *See* Natasha Lomas, *An EU Coalition of Techies is Backing a 'Privacy-Preserving' Standard for COVID-19 Contacts Tracing*, TechCrunch (Apr. 1, 2020), https://techcrunch.com/2020/04/01/an-eu-coalition-of-techies-is-backing-a-privacy-preserving-standard-for-covid-19-contacts-tracing/.

[55] *See* Pan-European Privacy-Preserving Proximity Tracing, *supra* n.53.

[56] *See* Lomas, *supra* n.54.

[57] *See* Kenneth Roth, *How Authoritarians Are Exploiting the COVID-19 Crisis to Grab Power*, Human Rights Watch (Apr. 3, 2020), https://www.hrw.org/news/2020/04/03/how-authoritarians-are-exploiting-covid-19-crisis-grab-power.

provide a check on the most egregious behavior, emergency situations like the one we are facing now can lead to overreach in ways that are uncommon in less urgent situations. We must ensure that if companies or government agencies in the U.S. choose to use location or proximity data for contact tracing, it is done in a manner that minimizes risk to infected individuals and communities that may have legitimate reasons for failing to self-quarantine.

IV. Considerations

The coronavirus epidemic continues to grow and companies, academics, and governments are searching for ways to leverage technology in response. When deciding what types of data practices are appropriate, Congress should remember that privacy is a balancing of equities. We no longer think of privacy as an on-off switch, or something that can be dismissed after a person agrees to a lengthy privacy policy. It instead weighs the intrusion of any product or program against the benefit of the data use, the secondary effects on individuals, and any mitigating steps that can be taken to minimize harms. As policymakers review data collection, use and sharing, they should:

- Focus on prevention and treatment, not punishment: Past epidemics have demonstrated that fear is not as effective as clear, meaningful information from a reliable source and the ability to voluntarily comply with medical and governmental directives. Successfully fighting the coronavirus will mean ensuring that a government response does not evolve into law enforcement and broad surveillance functions.

- Ensure accuracy and effectiveness: There does not appear to be a universally accepted definition of "accurate" or "effective" when it comes to predicting, preventing, or responding to the coronavirus. Nevertheless, if a tool or practice is unlikely to provide meaningful and measurable contributions to the coronavirus response, companies and governments should consider alternatives. This is not only because the privacy risks may not be justified but because people may rely on these measures in lieu of those that actually work.

- Provide actionable information: In a time of crisis, more information isn't always better. New data collection or novel data uses should inform individual, corporate, or government behavior in a constructive way. Symptom trackers, for example, may tell a person whether he or she should seek medical care. Contact tracing on the other hand, when it relies on insufficiently granular data, may result in unnecessary or unproductive quarantine, testing, and fear.

- Require corporate and government practices that respect privacy: People are reasonably fearful for their own health and the health of their loved ones. The burden for constructing privacy-protective products and responses must not be on concerned citizens but on companies and governments. That includes:
    - *A preference for aggregated data.* Individually identifiable information should not be used when less intrusive measures will suffice. If aggregated data will not do, industry best practices in anonymization and de-identification must be applied.
    - *Minimizing collection, use, and sharing.* When identifiable information is necessary, data processing should be limited when possible.
    - *Purpose limitations.* Data collected or used for the coronavirus response should not be used for secondary purposes. For corporate actors, this means advertising for commercial purposes or unrelated product development. For government actors, that means any function not directly related to their public health functions.
    - *Deletion.* Data should be deleted when it is no longer necessary for responding to the coronavirus epidemic or conducting public health research, especially if it is personally identifiable.

- Build services that serve all populations: Newly released data is confirming that minorities are contracting the coronavirus at a higher rate and are more likely to die from it.[58] There are also legitimate questions about how actionable mobility tracking data is for rural, poor, and working class communities that must travel for work or to secure food and medical care. As technology seeks to find solutions to the coronavirus, it is crucial that it does so in a way that serves all demographics and does not exacerbate existing inequalities.

- Empower individuals when possible: Epidemic response may not always allow for individualized opt-ins or opt-outs of data collection and use. To the extent possible, participation in data based programs should be voluntary and individuals should maintain traditional rights to control one's data.

- Be transparent to build trust: People will hesitate to participate in programs that involve their personal information but that are not transparent in how that information will be used. Companies that provide data, or inferences from data, and the governmental entities that use such information, must be transparent to users and residents about how data will be used.

[58] *See* Ibrahim X. Kendi, *Why Don't We Know Who the Coronavirus Victims Are?*, The Atlantic (Apr. 1, 2020), https://www.theatlantic.com/ideas/archive/2020/04/stop-looking-away-race-covid-19-victims/609250/.

- Be especially rigorous when considering government action: A coordinated government response is necessary for successfully fighting the coronavirus epidemic, but the United States has an important tradition of recognizing that the powers of the state pose unique threats to privacy and liberty.

V. Conclusion

The U.S. is facing an unprecedented epidemic that is straining our resources and testing our values. Congress is wise to consider right now how big data uses may infringe privacy, especially in ways that may not actually deliver relief, or do so in ways that are disproportionate to the public health threat or do not fairly serve different communities. As this crisis continues, we recommend that Congress also turn its oversight authority to evaluating how technology is meeting downstream challenges of large scale remote work and schooling, both of which are raising new questions about privacy and security of worker and student data.

We thank you for the opportunity to testify today and look forward to responding to questions from the committee.