Testimony and Statement for the Record

Stacey Gray

Senior Counsel, Future of Privacy Forum

Enlisting Big Data in the Fight Against Coronavirus

Senate Committee on Commerce, Science, and Transportation

April 9, 2020

Thank you for the opportunity to testify on enlisting big data in the fight against COVID-19. FPF is a non-profit organization promoting privacy leadership, scholarship, and principled data practices in support of emerging technologies. We are supported by leading foundations, as well as by more than 170 companies, with an advisory board representing academics, industry, and civil society.[1] We bring together privacy officers, academics, consumer advocates, and other thought leaders to explore the challenges posed by technological innovation and develop privacy protections, ethical norms, and workable business practices.

In this testimony, I describe how collection and uses of data, including personal data, to respond to a public health crisis like a pandemic can be compatible with privacy and data protection principles. I also highlight several FPF resources that provide greater detail regarding these important issues.

Specifically, I discuss below:

- In **Part 1**, the commercial sources and relative risks and benefits of precise location data generated by consumer devices;
- In **Part 2**, recommendations based on observations of how privacy experts in the United States and around the world are currently mitigating the risks of using data to combat the COVID-19 pandemic; and
- In **Part 3**, how this public health crisis highlights the ongoing need for baseline federal consumer privacy legislation. In addition to providing legal protections for individuals, a federal privacy law would also provide much-needed legal clarity for US companies to be able to respond quickly and understand what kind of data they may or may not share legally and ethically to support emergency public health initiatives.

[1] The views herein do not necessarily reflect those of FPF's supporters or Advisory Board.

Executive Summary

As the global response to COVID-19 continues, we have seen strong interest in recent weeks from local, state, and federal government entities, as well as academic institutions, in accessing commercial data held by tech companies. Data generated in high volume by smartphones and other consumer services and platforms, collectively referred to as "big data," has long been of interest to public health experts,[2] and can be valuable if used within the parameters of data protection and privacy law to safeguard civil liberties. Such data can originate from providers of telecommunications services (cell phone carriers), but also from mobile apps installed by users, fitness and wearable devices, and other Internet-connected consumer electronics (the Internet of Things or IoT).

In many cases, commercial data can be shared that ***does not reveal any information*** about identified or identifiable individuals. For example, private companies may process aggregated data about the use of public transportation or supply chain management in partnership with local governments.[3] In other cases, data originally collected from individuals can be transformed or de-identified to a sufficient extent that it only reveals aggregate trends, such as movements of people at the city, county, or state level.

Commercial sources of precise location data vary widely and include cell phone carriers, mobile operating systems, apps, app partners, and others with device location information that varies in its accuracy, precision, and volume (Part 1). Lawmakers should understand these sources to account for concerns around data quality, bias, and equity.

In order to mitigate the risks of processing location data and other consumer data for public health initiatives, FPF offers recommendations based on recent workshop with global experts (Part 2), including:

- *Follow the lead of public health experts.* Rather than leading the way with data that is already available, technology companies should play a supporting role to epidemiologists, established research partners, and public health experts and rely on their expertise in determining what data is useful to achieving specific, clear public health goals.
- *Ensure transparency and lawfulness*. In order to ensure public trust, including in the use of voluntary pandemic apps, companies should be as transparent as possible about data shared with government or public health officials.
- *Apply privacy enhancing technologies (PETs)*. Companies should take advantage of advances made by privacy engineers in recent years, and apply privacy enhancing technologies (PETs), such as differential privacy, in accordance with principles of data minimization and privacy by design.
- *Employ privacy risk assessments*. Companies should use well-established privacy and data protection impact assessment frameworks to help identify risks and find ways to mitigate or eliminate them.

---

[2] See Amy Wesolowski et al., *Connecting Mobility to Infectious Diseases: The Promise and Limits of Mobile Phone Data*, The Journal of Infectious Diseases (2016) 24 (4): S414–S420 (reviewing opportunities and challenges of mobile phone data, illustrated by analyses of two pathogens in Kenya); Yves-Alexandre de Montjoye et al., *Enabling Humanitarian Use of Mobile Data*, Brookings (2014), https://www.brookings.edu/wp-content/uploads/2016/06/BrookingsTechMobilePhoneDataWeb.pdf (exploring case studies using mobile data to understand and address infections diseases such as Ebola and recommending nuanced approaches to protecting privacy where data may be used to avoid serious harm to people).

[3] See, e.g., World Economic Forum, *Data Collaboration for the Common Good: Enabling Trust and Innovation Through Public-Private Partnerships* (2019), http://www3.weforum.org/docs/WEF_Data_Collaboration_for_the_Common_Good.pdf; Shannon Bouton et al., *Public–Private Collaborations for Transforming Urban Mobility*, McKinsey (2017), https://www.mckinsey.com/business-functions/sustainability/our-insights/public-private-collaborations-for-transforming-urban-mobility

- *Follow core purpose limitation principles*. Any personal data collection and use enlisted to fight the pandemic should be limited in time and limited to a specific, well-defined purpose identified in advance, with clear limitations on secondary uses.

Finally, we observe that the current public health emergency underscores the ongoing need for a baseline federal consumer privacy law (Part 3). Congress should address the gaps in existing legal protections for highly sensitive data –including precise location data and health and wellness data–and provide US companies with much-needed structure and clarity for when they may or may not share data ethically and legally in situations such as this.

1. Many Commercial Sources of Precise Location Data Exist

In recent weeks, there has been great interest in whether and how to use the precise location information of individuals inferred through the location of mobile devices. Governments and public health officials are primarily interested in tracking individuals affected by the virus ("contact tracing"), alerting individuals who might be affected based on their proximity to known cases, and better understanding the effectiveness of physical distancing measures. Data from private companies may also help in other areas, e.g., modelling the spread of the virus and resource planning of medical equipment.[4]

Location data involves information about how devices and people move through spaces over time. Most of this information comes from the devices we carry with us, with smartphones and fitness trackers acting as proxies for people.[5] Such location data is collected by many different types of companies in a variety of commercial contexts, with varying degrees of accuracy, precision, and representativeness.

The first question should be to examine whether the underlying data is *fit for the proposed uses* – i.e., will it really serve the intended objectives? If the data is not useful for addressing legitimate public health needs, it should not be used for those purposes.[6] In addressing this question, companies should follow the lead of epidemiologists, virologists, and other public health experts.

A. Location Data Varies in Accuracy, Precision, and Volume

Not all location data will be accurate or precise enough to achieve the desired ends, or be available in great enough volume for useful analysis.

- *Accuracy*. Accuracy is the closeness of a reported location measurement to a "true value," or the actual physical location of a specific device or person. Accuracy depends on a number of factors, including the sources of data (GPS, Wi-Fi, cell towers, or combinations of signals), the physical interference (for example, indoor accuracy of GPS-based signals tends to be lower than outdoor accuracy), and the density of signals (for example, in cities as compared to rural areas). A high level of accuracy can be reached by a mathematical calculation, i.e., by triangulation of multiple location measurements.[7]

---

[4] See Nuria Oliver et al., *Mobile Phone Data and COVID-19: Missing an Opportunity?* (2020), https://arxiv.org/ftp/arxiv/papers/2003/2003.12347.pdf.
[5] According to Pew, smartphone ownership in 2019 was near-universal at 81% of Americans. Pew Research Center, *Mobile Fact Sheet* (2019), https://www.pewresearch.org/internet/fact-sheet/mobile/.
[6] See Susan Landau, *Location Surveillance to Counter COVID-19: Efficacy Is What Matters*, Lawfare (2020), https://www.lawfareblog.com/location-surveillance-counter-covid-19-efficacy-what-matters.
[7] Wide Area Augmentation System (WAAS) improves the location accuracy by taking into account the GPS signal corrections provided by satellites and ground stations. See Federal Aviation Administration, Satellite Navigation - Wide Area Augmentation System (WAAS), https://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/navservices/gnss/waas/ (last accessed April 7, 2020).

- *Precision*. Precision refers to the level of specificity or granularity of a location measurement. Location is expressed in latitude and longitude, and the greater the number of digits following a decimal place, the greater the location measurement's precision.[8] In general, location data is considered to have privacy implications when it has a high enough level of precision to identify or single out a device or a person with a reasonable degree of specificity ("precise location data") – usually around two decimal points.[9] This depends in part on population density, because a lower level of precision might be more capable of singling out or identifying a person in a rural or remote area than if that same person were standing in Times Square.

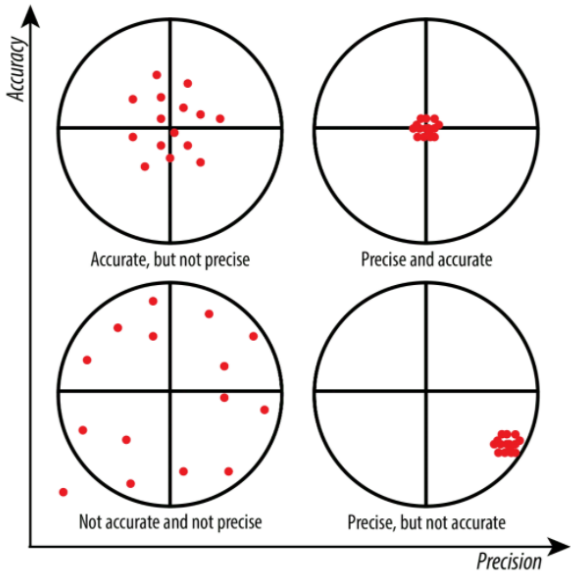

Figure 1. Data may be accurate, but not precise (top left); conversely, it may be very precise, but still inaccurate, if the coordinates do not represent the true location of the person (bottom right).[10]

- *Volume and Representativeness.* Even if data is both precise and accurate enough to provide useful information to public health authorities, it may not be fairly representative of the population. Certain types of devices and mobile apps, for example, are more likely to capture affluent communities.[11] Data processing involving location data that omits certain segments, especially from vulnerable communities, may lead to biased distribution

[8] See A.D. Chapman and J. Wieczorek, eds, *Guide to Best Practices for Georeferencing* (2006), https://research.nhm.org/pdfs/12847/12847.pdf; Aaron Schumacher, *Geolocation Precision by Digit* (2018), https://planspace.org/20180719-geolocation_precision_by_digit/.
[9] See Network Advertising Initiative, *Guidance for NAI Members: Determining Whether Data is Imprecise*, https://www.networkadvertising.org/sites/default/files/nai_impreciselocation.pdf. Recent legislative proposals have attempted to create strict cut-offs to achieve similar ends – for example, precision to within a 1,640 foot radius under the U.S. House and Commerce Discussion Draft, or an 1,850 foot radius under the California Privacy Rights Act ballot initiative of 2020.
[10] Source: *Precision vs. Accuracy*, St. Olaf College, https://wp.stolaf.edu/it/gis-precision-accuracy/.
[11] For example, the mobile app 'Street Bump' was released by a municipal authority in an attempt to crowdsource data to work out which roads it needed to repair. However, affluent citizens downloaded the app more than people in poorer neighborhoods. The system thus reported a disproportionate number of potholes in wealthier neighborhoods, and could have led the city to distribute or prioritize its repair services inequitably. *See* Executive Office of the President, *Big Data: Seizing Opportunities, Preserving Values* (2014), at 51, https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf. See also Marianne Bertrand and Emir Kamenica, Nat'l Bureau of Economic Research Working Paper No. 24771: *Coming Apart? Cultural Distances in the United States Over Time* (2018), https://www.nber.org/papers/w24771.pdf (stating, "Across all years in our data, no individual brand is as predictive of being high-income as owning an Apple iPhone in 2016.").

of public health resources or other unfair outcomes. This includes underrepresentation of the elderly, very young, or lowest income people who do not own cell phones, or anyone who does not own a cell phone for other reasons, such as refusal on religious grounds. The most prominent driver of the divide in cell phone ownership in emerging economies is cost.[12]

B. A Range of Companies Maintain Precise Location Data

Keeping these concepts in mind, there are a wide range of commercial entities that collect or process precise location data in different contexts. Depending on how that data was collected or inferred from the end user, it will vary in accuracy, precision, and volume.

- *Cell Phone Carriers*. Wireless telecommunications service providers, or cell phone carriers, generally are able to determine where phones are located because they direct calls and content to phones through local cell towers. In some cases, this data may be enhanced with GPS location data. In general, cell phone carriers can associate precise location data with known individuals (account names associated with cell phone subscriptions).

- *Mobile Operating Systems.* Providers of mobile operating systems (OS), such as Android (Google) and iOS (Apple), may know where devices are located as a result of providing services, improving functionality, or enabling opt-in location history. In addition, the mobile OS provides the technical permission layers through which apps request permission from the user to access a user's location (below). Some users may have also opted in to the use of cell tower and Wi-Fi data to improve location services.

- *Apps and App Partners.* Many people have installed apps with location-based features, such as weather alerts, ridesharing, or groceries deliveries. This location data is often shared with partners in order to provide personalized advertising, measure the effectiveness of marketing campaigns, or to support a free app. Many apps use Software Development Kits (SDKs), or code (libraries) developed by third party partners, to easily include features and allow partners to collect data. Apps request permission from the user to access precise location data through technical permissions provided by the Operating System (OS), for example to grant access "once," "always," or "only when in use."[13] Generally speaking, unlike cell phone carriers, apps and app partners collect precise location data associated with a device identifier, usually a (relatively stable) device identifier.

- *Location Analytics Providers.* Connected devices emit identifying information that allows them to be tracked, even when they are not actively connected to a network. This includes mobile phones (when Wi-Fi or Bluetooth are turned on), but also other Internet of Things (IoT) devices such as handsfree headsets, fitness trackers, smart toys, or vehicles. As a result, many airports, stadiums, and brick-and-mortar stores analyze Wi-Fi and Bluetooth radio signal data to better understand when their busiest hours are, where the highest in-store foot-traffic is, what products customers show an interest in, or how long people wait in lines. Depending on the number and quality of sensors in an indoor positioning system,[14] location analytics data can be highly accurate and precise.

---

[12] Laura Silver et al., *Mobile Divides in Emerging Economies*, Pew Research Center (2019), https://www.pewresearch.org/internet/2019/11/20/mobile-divides-in-emerging-economies/.
[13] See, e.g., Apple Developer Documentation, *Requesting Authorization for Location Services*, https://developer.apple.com/documentation/corelocation/requesting_authorization_for_location_services; Android Maps SDK, *Location Permissions*, https://developers.google.com/maps/documentation/android-sdk/location#location_permissions.
[14] See, e.g., G.M. Mendoza-Silva et al., *A Meta-Review of Indoor Positioning Systems*, Sensors (2019), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6832486/.

C. Location Data can be Inferred by Using Different Methods

Commercial entities use a wide variety of methods to collect or infer precise location data. Mobile operating systems, apps, app partners, and others commonly use GPS (Global Positioning System), Cell Towers, Wi-Fi Networks, and Beacons (among others). Each provides a different level of precision and can be used for different purposes.

- *GPS*. Smartphones and other devices can detect location via satellite GPS independently of any telephone or internet reception, although a phone's GPS chip is only one sensor among many. The accuracy of GPS signals varies widely, and can be affected by weather, or physical interference. For example, it is much less accurate in urban areas, and especially poor for detecting specific locations inside large buildings. As a result, modern cell phones use GPS in combination with other forms of location signal (Wi-Fi, Bluetooth) at various times to create a more accurate location determination.

- *Cell Towers*. The main function of cell towers is to be used by carriers to provide cell service. In addition, cell towers emit unique "Cell Tower IDs" that can be freely detected. There are many private and public databases of the Cell Tower IDs associated with mapped locations of known cell towers. As a result, the proximity of nearby cell towers (and the signal strength of their IDs) can be used to infer where a device is located.

- *Wi-Fi Networks*. Mobile devices can also infer their location by scanning for nearby Wi-Fi networks. Nearby networks or "access points" might include, for example, neighbors' Wi-Fi or the Wi-Fi available in cafes and shops. Large databases exist of the unique identifiers (MAC addresses and SSIDs)[15] of wireless routers and their known locations, with companies such as Mozilla and Combain reporting databases of millions of unique Wi-Fi networks. In 2011, Google created an opt-out approach for allowing a particular access point to avoid inclusion in its database, which involves appending the phrase "_nomap" to the end of the wireless router's SSID.[16] Mozilla similarly honors the _nomap method. Other databases do not, or they offer their own opt-outs.[17]

- *Bluetooth Beacons*. Many apps are designed to detect their proximity to "beacons," small radio transmitters that broadcast one-way Bluetooth signals. Beacons are inexpensive and can be attached to personal items such as a person's keys or wallet. They can also be installed at known locations, for example in a retail space or in front of a special display of products in a shop. In these cases, an app that a user has given permission to access Bluetooth can infer the device's location or send proximity-based alerts or other content.

---

[15] An SSID is a name assigned to a wireless access point that allows stations to distinguish one wireless access point from another. NIST Glossary, https://csrc.nist.gov/glossary/term/SSID.
A "Media Access Control" (MAC) address is a unique 48-bit value that is assigned to a particular wireless network interface by the manufacturer. NIST Glossary, https://csrc.nist.gov/glossary/term/Media_Access_Control. See also Latanya Sweeney, *My Phone at Your Service*, Tech@FTC (2014), https://www.ftc.gov/news-events/blogs/techftc/2014/02/my-phone-your-service; Ashkan Soltani, *Privacy Trade-Offs in Retail Tracking*, Tech@FTC (2015), https://www.ftc.gov/news-events/blogs/techftc/2015/04/privacy-trade-offs-retail-tracking.

[16] The investigation by the Dutch Data Protection Authority (DPA) of Google Streetview resulted in the worldwide _nomap opt-out solution. See Autoriteit Persoonsgegevens, Google announces opt-out option for collection of data about WiFi routers (15 November 2011), https://autoriteitpersoonsgegevens.nl/en/news/google-announces-opt-out-option-collection-data-about-wifi-routers

[17] See, e.g., *End User Opt-Out of Skyhook Products*, Skyhook, https://www.skyhook.com/opt-out-of-skyhook-products.

- *Combining Signals for Accuracy*. Modern smartphones can detect signals from many sources to create a more accurate location measurement than any one signal (such as GPS) would provide alone. For example, iOS and Android harness the signals from many different sensors on the device, such as the accelerometer, to provide a consolidated "Location Services" feature that offers highly precise location information to apps (with a user's permission) and that users can control in Settings.

D. Different Types of Location Datasets can Provide Different Uses for Public Health Decisions

Evaluating the usefulness of location data from these different sources can be highly dependent on context. Cell phone carriers and mobile operating systems may offer high volume and representativeness, within the limits of cell phone owners, and must be accurate and precise enough to deliver services. Cell phone carrier data, because it is based solely or primarily on cell tower triangulation, is likely not precise enough for effective COVID-19 contact tracing (although it may be useful for other public health purposes, such as population-level trends).

In contrast, mobile operating systems compile location data through the combinations of a number of sensors (such as GPS, Wi-Fi, and Bluetooth), and can therefore have a much higher accuracy for certain services. Similarly, mobile apps and app partners may have the potential to generate highly accurate and precise location data. However, in practice, mobile SDK location data may not always be high quality (i.e., accurate) when collected and re-used by third party intermediaries for advertising and marketing.[18]

When considering the usefulness of location data from different providers, government entities should carefully consider the limitations of these datasets and avoid "availability bias."[19] For their part, commercial entities should be transparent about the accuracy, precision, volume, and any other limitations of their datasets. Rather than seeking to lead the way, companies should look to public health experts for guidance on what is needed and what is useful. Only the expert should be in the position to decide on the utility of data from apps and smart devices for public health purposes.

2. Policymakers Must Consider how to Mitigate Privacy Risks when Implementing Public Health Initiatives

Amassing and using large volumes of consumer data in the fight against the spread of COVID-19 can pose risks to the rights of individuals and to their communities. Such risks vary from immediate risks, like discrimination, endangered physical security,[20] disproportionate loss of privacy, and unjustified limitations on freedom of movement, to long term risks to freedom, civil liberties and even to democracy. Some risks stem from secondary uses of data. For example, some apps may require an individual to take multiple pictures of themselves during the day and upload the selfies to a centralized database.[21] For example, with the current state of facial

[18] The lowest quality location data, for example, often comes from bidstream data, or data collected from information supplied by apps for real-time ad auctions. Josh Anton, *The Location Data Crisis of 2020*, AdExchanger (2019), https://www.adexchanger.com/data-driven-thinking/the-location-data-crisis-of-2020/.
[19] See American Psychological Association, *Availability Heuristic*, APA Dictionary, https://dictionary.apa.org/availability-heuristic; Amos Tversky and Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, Cognitive Psychology (1973) 5 (2): 207-232.
[20] Farah Stockman, *What It's Like to Come Home to the Stigma of Coronavirus*, New York Times (2020), https://www.nytimes.com/2020/03/04/us/stigma-coronavirus.html.
[21] See, e.g., Poland's official government mobile app *Home Quarantine*, https://apps.apple.com/us/app/home-quarantine/id1502997499; Kenneth Garger, *Polish Residents Can Send Government Selfies to Prove Quarantine Compliance*, New York Post (2020),

recognition technology it is not farfetched to foresee the use of the database as a training set for machine learning algorithms.

The Future of Privacy Forum convened a workshop[22] on March 26, 2020, with a dozen ethicists, scholars, government officials, and corporate leaders, and over 100 corporate attendees from the United States and Europe, to discuss responsible data sharing in times of crisis. Workshop participants discussed specific ways to mitigate risks to enable uses of commercial data with the potential to inform academic and public health discussions. These recommendations below are informed by best practices of "Data for Good" initiatives[23] as well as by long-established privacy and data protection principles from the United States[24] and the European Union.

The following considerations should inform public debate over the uses of commercial data, particularly highly sensitive data, for public health initiatives.

1) **Follow the needs of public health experts.** As a threshold matter, any commercial data to be used in the fight against the COVID-19 pandemic must respond to the needs of health experts.[25] As FPF's CEO, Jules Polonetsky, has recently written, technology solutions must follow, rather than lead, the way to combatting the pandemic.[26] They should seek to support virologists, epidemiologists, public health experts, and safety experts in every way they need, consistent with civil liberties. It is essential to work with medical and public health partners to understand their data needs, rather than merely provide analysis based on data available from commercial datasets. This recommendation is also aligned with the principles of necessity and proportionality in data protection law.[27] An essential part of this analysis is an assessment of whether the underlying data is accurate. As discussed in Part 1 of this submission, accuracy of certain types of location data is only one aspect of whether such data may be useful to achieving public health goals (precision and volume may be just as important in evaluating location datasets). However, all workshop experts agreed that datasets must be accurate in order to diminish false negatives and false positives that render public health measures ineffective.

2) **Work with established partners.** Companies with established Data for Good programs are already working with university partners to ensure review of data sharing arrangements. While review procedures vary by organization, the objectives of review include determination that the datasets are appropriate, anonymized, and aggregated as

---

https://nypost.com/2020/03/24/polish-residents-can-send-government-selfies-to-prove-quarantine-compliance/.

[22] Katelyn Ringrose, *Privacy and Pandemics: A Thoughtful Discussion*, Future of Privacy Forum (2020), https://fpf.org/2020/03/27/privacy-and-pandemics-a-thoughtful-discussion/.

[23] See, e.g., Jake Porway, *Using Collaboration to Harness Big Data for Social Good*, Stanford Social Innovation Review (2017), https://ssir.org/articles/entry/using_collaboration_to_harness_big_data_for_social_good; Lydia Clougherty Jones et al., *How to Use Data for Good to Impact Society*, Gartner (2018), https://www.gartner.com/en/documents/3880666.

[24] Many of the principles here stem from the US Fair Information Practices (FIPPs) articulated in the *Records, Computers, and Rights of Citizens* report of the U.S. Department of Health, Education and Welfare (1973), https://www.justice.gov/opcl/docs/rec-com-rights.pdf. These principles were enshrined in the federal Privacy Act of 1974 and helped shape OECD and EU principles that informed the development of data protection law.

[25] See Nuria Oliver et al., *Mobile Phone Data and COVID-19: Missing an Opportunity?* (2020), https://arxiv.org/ftp/arxiv/papers/2003/2003.12347.pdf.

[26] Jules Polonetsky, *Silicon Valley, Follow, Don't Lead*, LinkedIn (2020), https://www.linkedin.com/pulse/silicon-valley-follow-dont-lead-jules-polonetsky/?trackingId=NyNKmbJAfPuTlfLHeUi77A%3D%3D.

[27] See, e.g., the European Data Protection Supervisor (EDPS) toolkits on necessity and proportionality in data protection law. EDPS, *Assessing the Necessity of Measures That Limit the Fundamental Right to the Protection of Personal Data: A Toolkit* (2017), https://edps.europa.eu/sites/edp/files/publication/17-04-11_necessity_toolkit_en_0.pdf; EDPS, *Guidelines on Assessing the Proportionality of Measures That Limit the Fundamental Rights to Privacy and to the Protection of Personal Data* (2019), https://edps.europa.eu/sites/edp/files/publication/19-02-25_proportionality_guidelines_en.pdf.

much as possible and that the research question serves a public interest.[28] These existing arrangements and those established within university settings, such as independent research ethics review boards, can allow groups to act as trusted partners between companies and public agencies.

3) **Transparency and lawfulness are key to public trust.** Certain public health measures, such as voluntary testing, self-reporting, or use of contact tracing or symptom tracing apps, rely on individuals being willing to provide their information in support of a larger goal. In support of such measures, governments and companies seeking to share data should take extra steps to be transparent about the personal data used, how it is processed, who has access, and for what specific purposes. The purpose(s) must be clear, specific, granular, well-defined, and individuals within companies and institutions should be identified as responsible for these commitments. Information should be provided in an easy to read, intelligible way for individuals whose personal data are processed to understand the consequences of participating. Transparency should include making the source code of an app available to computer scientists and civil society, and as much as possible allowing outside scrutiny from public health and de-identification experts..

4) **Apply privacy enhancing technologies (PETs) (Data Minimization[29] and Privacy by Design[30]).** Generally, Privacy by Design means that privacy is proactively embedded into the design and operation of IT systems, networked infrastructure, and business practices. Any use of commercial datasets of personal information should take advantage of the advances made by privacy engineers in recent years, developing techniques that support effective de-identification of personal data sets, de-centralized data analysis, and privacy by design.[31] For example, the ongoing release of heat maps by Google, showing the effectiveness of "stay-at-home" orders across the world during the pandemic, relies on differential privacy.[32] Anonymization techniques and aggregating data are safeguards that should be applied,[33] depending on the state of the art techniques that are available. Centralization and decentralization are at the core of an emerging debate related to the infrastructure of COVID-19 contact tracing apps. Local processing on an individual's device, instead of in a centralized server, offers more privacy protection. For example, contact tracing using Bluetooth identifiers can be processed on a user's device.[34]Centralized "pools" of data elevate privacy and security risks and may enable surveillance creep if that pool of data is kept after the public health emergency.[35]

---

[28] Ronald Sandler and John Basl, *Building Data and AI Ethics Committees*, Accenture (2019), https://www.accenture.com/_acnmedia/PDF-107/Accenture-AI-And-Data-Ethics-Committee-Report-11.pdf#zoom=50.
[29] See U.S. Federal Trade Commission, *Internet of Things: Privacy & Security in a Connected World* (2015), https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf.
[30] See U.S. Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers* (2012), https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf. This final report sets forth best practices for businesses to protect the privacy of American consumers and give them greater control over the collection and use of their personal data.
[31] See, e.g., the achievements of the members of the Privacy Tech Alliance initiative here: https://fpf.org/privacy-tech-alliance/.
[32] Google, *COVID-19 Community Mobility Reports* (2020), https://www.google.com/covid19/mobility/.
[33] See also the Guidelines of the European Data Protection Board (EDPB) in relation to accessing telecommunications data to combat the COVID-19 pandemic: https://edpb.europa.eu/news/news/2020/statement-edpb-chair-processing-personal-data-context-covid-19-outbreak_hu.
[34] See Privacy International, *Bluetooth Tracking and COVID-19: A Tech Primer* (2020), https://privacyinternational.org/explainer/3536/bluetooth-tracking-and-covid-19-tech-primer.
[35] One example of a centralized pool of data is the "data lake" proposed by C3.ai. See https://c3.ai/covid/.

5) **Use existing privacy risk assessment (PIA) frameworks to identify risks and find ways to mitigate or eliminate them.** Given the volume of data-sets that seem to be required for the response to COVID-19, and the sensitivity of the data (information related to health, location data), existing tools for privacy risk assessments can be used to help mitigate risks before the deployment of a data-based measure to fight the pandemic.[36]

6) **Follow core purpose limitation principles.** Any personal data collection and use enlisted to fight the pandemic should be limited in time and limited to a specific, well defined purpose (purpose limitation).[37] History tells us that it is difficult to discontinue practices started in an emergency.[38] In the absence of clear systemic rules, organizations should establish an exit strategy up front to protect against continued "emergency" practices after the crisis. Companies must be clear that data shared now should not be kept forever or used for other purposes. The purpose for data collection should be specified at the time of data collection and that the subsequent use of data be limited to fulfilling that specified purpose. For example, public health officials should not be provided with swaths of individualized location and behavior data and given the choice of what to use. Rather, the purpose - the policy and health outcome desired - should be agreed upon and then only the data necessary and proportionate to fulfilling that purpose be used. To ensure that data is not repurposed and used for secondary or tertiary purposes by the government the following data protection principles should be followed: privacy by design, retention and deletion policies and technical mechanisms, and walled off or decentralized storage and processing, and data minimization methods powered by privacy-enhancing technologies (PETs).

3. The COVID-19 Crisis Highlights the Ongoing Need for a Comprehensive U.S. Federal Privacy Law

Finally, the current global public health crisis only draws attention to the ongoing need for Congress to draft and pass a federal comprehensive consumer privacy law that would fill in the gaps between existing federal sectoral regimes and provide much-needed clarity and guidance to enable ethical and responsible data sharing.

In comparison to the European Union and other governments with comprehensive data privacy laws,[39] the United States does not currently have a baseline set of legal protections that apply to all commercial data about individuals, regardless of the particular industry, technology, or user base. Instead, the United States has taken a sectoral approach that provides strong privacy and security protection for information collected in certain contexts, while leaving equally sensitive information about those same individuals largely unregulated aside from the FTC's generally applicable Section 5 authority. For example, health records held by hospitals and covered by the Health Insurance Portability and Accountability Act (HIPAA) are subject to clear privacy and security rules, whereas health data collected by wearable devices or consumer-facing platforms and services, or even highly precise location data collected through mobile apps, are largely unregulated.

---

[36] Many companies already perform risk assessments, which are commonplace in the US and EU for high risk processing (see GDPR, Article 35). As a result, many existing well-informed privacy risk assessment frameworks can be deployed in this context prior to using existing data for new uses or deploying a new strategy. See Future of Privacy Forum, *City of Seattle Open Data Risk Assessment* (2018), https://fpf.org/wp-content/uploads/2018/01/FPF-Open-Data-Risk-Assessment-for-City-of-Seattle.pdf.
[37] See GDPR, https://gdpr-info.eu/; OECD, *OECD Privacy Principles*, http://oecdprivacy.org/#purpose.
[38] See Peter Swire, *Security, Privacy and the Coronavirus: Lessons From 9/11*, Lawfare (2020), https://www.lawfareblog.com/security-privacy-and-coronavirus-lessons-911.
[39] See, e.g., Regulation (EU) 2016/679 General Data Protection Regulation, 2016 O.J. (L.119) (GDPR), https://www.privacy-regulation.eu/en/index.htm.

In addition to the lack of strong legal protections for much of the data discussed in this submission, the lack of a federal framework for data sharing has also led to uncertainty and trepidation from US companies about what they may and may not share, and with what protections. In contrast, part of the reason that public health authorities in the European Union have been able to respond comparatively rapidly to address this crisis using data and privacy preserving technological innovation is the existence of clear rules and a large body of guidance from Data Protection Authorities.[40]

FPF has long supported[41] comprehensive federal privacy legislation and observed that it should be flexible enough to support data-driven public health initiatives under the right safeguards and within limits consistent with privacy and civil liberties. We have previously recommended:

- **Protections for Sensitive Data –** A federal privacy law should create heightened legal protections for sensitive data, including for health information and precise geo-location data,[42] in line with global norms and legal standards.[43] These protections should include limits on collection of data in the first instance, which should generally be limited to affirmative, express consent.
- **Privacy Risk Assessments** - In addition, privacy risk assessments can be a useful mechanism for corporate accountability and risk management. They are a central element of data governance at responsible companies and a core component of existing privacy regimes in the United States[44] and Europe. In the EU, risk assessments are required when companies engage in high-risk data processing.[45] FPF has worked on risk assessments for many years, beginning with a project in 2014 which sought to help provide guidance for big data related risk assessments.[46] More recently, FPF conducted a privacy risk analysis of the City of Seattle's Open Data program.[47]
- **Independent Ethical Review Boards.** Although many institutions already conduct research under sectoral privacy laws (e.g. healthcare centers, hospitals, pharmaceutical development, or academic institutions abiding by the federal Common Rule), the current pandemic is demonstrating that a broad range of beneficial research currently falls outside of the scope of these regulations. In these cases, oversight from ethical review

[40] See Wojciech Wiewiorowski, *EU Digital Solidarity: A Call for a Pan-European Approach Against Pandemic* (2020), https://edps.europa.eu/press-publications/press-news/videos/eu-digital-solidarity-call-pan-european-approach-against_en. See also EDPB, *Statement on the Processing of Personal Data in the Context of the COVID-19 Outbreak* (2020), https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_statement_2020_processingpersonaldataandcovid-19_en.pdf (including guidance on mobile location data); EDPB, *European Data Protection Board to Issue Guidance on Data Processing in the Fight Against COVID-19* (2020), https://edpb.europa.eu/news/news/2020/european-data-protection-board-issue-guidance-data-processing-fight-against-covid-19_en (announcing that new guidance is forthcoming).
[41] See Future of Privacy Forum, *Long Overdue: Comprehensive Federal Privacy Law* (2018), https://fpf.org/2018/11/15/fpf-comments-on-a-national-baseline-consumer-privacy-law/.
[42] The FTC has long considered precise geolocation uniquely sensitive, among other types of data. See FTC report, *supra* note 29, at 14.
[43] The GDPR defines sensitive data broadly by recognizing special categories of personal data, including "personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation." GDPR, Art. 9, Recital 51–52.
[44] In the United States, the FTC has required comprehensive privacy oversight programs to include risk assessments in its long history of privacy-related consent decrees. *See e.g.,* Google, Inc., F.T.C. 1-2 3136 (2011), https://www.ftc.gov/sites/default/files/documents/cases/2011/03/110330googlebuzzagreeorder.pdf. Privacy risk assessments are the focus of a major ongoing effort by the National Institute of Standards and Technology (NIST). See *Privacy Risk Assessments: A Prerequisite to Privacy Risk Management*, NIST (2017), https://www.nist.gov/news-events/events/2017/06/privacy-risk-assessment-prerequisite-privacy-risk-management.
[45] GDPR Article 35, https://gdpr-info.eu/art-35-gdpr/.
[46] Jules Polonetsky et al., *Benefit-Risk Analysis for Big Data Projects*, Future of Privacy Forum (2014), https://fpf.org/wp-content/uploads/FPF_DataBenefitAnalysis_FINAL.pdf.
[47] See *City of Seattle Open Data Privacy Risk Assessment*, Future of Privacy Forum (2018), https://fpf.org/wp-content/uploads/2018/01/FPF-Open-Data-Risk-Assessment-for-City-of-Seattle.pdf.

boards can provide a useful governance mechanism to oversee research using data that was originally collected for other purposes, or where there are tensions between conducting the research and complying with privacy rights, such as access and deletion.[48]

- **Purpose Limitation (Secondary Uses of Data).** As discussed in Part 2 of this submission, purpose limitation is a core principle of data protection, and uniquely important for considerations of whether and how to use data for public health initiatives. Under the GDPR, scientific research conducted pursuant to strong privacy and data protection safeguards is considered *per se* "compatible" with the initial purpose for which the personal data was processed.[49]

We commend the Chairman and Ranking Member of the Committee for making privacy legislation a priority over the last year, and hope to see long-term momentum towards consensus on key issues as a result of the current heightened and well-deserved public attention to the ways in which existing commercial datasets are generated.

Conclusion and Resources

FPF has further explored these issues and proposed policy solutions based on privacy principles in the following resources:

- FPF Charts the Role of Mobile Apps in Pandemic Response – looking into the various objectives and methods of specific apps and software development kits.[50]
- A Closer Look at Location Data: Privacy and Pandemics –providing a brief explainer guide of the basics: (1) what is location data, (2) who holds it, and (3) how is it collected?[51]
- EU DPAs Issue Green and Red Lights for Processing Health Data During the COVID-19 Epidemic – exploring how various European Data Protection Authorities issued public interest guidance on the limits of collecting, sharing and using personal data relating to health in these exceptional circumstances.[52]

We hope this testimony is useful on the issue of leveraging big data in the fight against COVID-19, and look forward to engaging further on these important issues.

[48] Mike Hintze, *Science and Privacy: Data Protection Laws and Their Impact on Research*, Washington Journal of Law, Technology, & Arts (2019) 14: 103.
[49] GDPR, Recital 159, https://gdpr-info.eu/recitals/no-159/.
[50] See Pollyanna Sanderson, *FPF Charts the Role of Mobile Apps in Pandemic Response*, Future of Privacy Forum (2020), https://fpf.org/2020/04/03/fpf-charts-the-role-of-mobile-apps-in-pandemic-response-chart/.
[51] See Stacey Gray, *A Closer Look at Location Data: Privacy and Pandemics*, Future of Privacy Forum (2020), https://fpf.org/2020/03/25/a-closer-look-at-location-data-privacy-and-pandemics/.
[52] See Gabriela Zanfir-Fortuna, *EU DPAs Issue Green and Red Lights for Processing Health Data During the COVID-19 Epidemic*, Future of Privacy Forum (2020), https://fpf.org/2020/03/10/eu-dpas-issue-green-and-red-lights-for-processing-health-data-during-the-covid-19-epidemic/.