# Ms. Stacey Gray, Senior Counsel, Future of Privacy Forum

Answers to Questions Submitted by Members of the Senate Committee on Commerce, Science, and Transportation

*Enlisting Big Data in the Fight Against Coronavirus*

April 15, 2020

**Chairman Wicker**

1. Many national and local governments around the world are seeking to use new technology to combat this unprecedented pandemic. Earlier this week, the German government launched an app that allows users to "donate" personal data collected by their fitness trackers or other health devices to help authorities analyze the spread of COVID-19. Authorities in Moscow have launched an app intended to be downloaded by those who test positive for COVID-19. Yet this app raises privacy concerns, as it would allow officials to track residents' individual movements.
    - As governments seek to use new technologies in the fight against COVID-19, it is imperative that privacy rights be protected. Are there specific examples of app-based programs you can recommend to policymakers that are both useful in the fight against COVID-19 and respectful of individual privacy rights?

**Answer to Question 1:**

Mobile apps, if adopted voluntarily by a sufficient percentage of a population, can support public health initiatives by providing data that is precise and accurate enough for effective person-to-person contact tracing. Apps that have the potential to achieve these goals without sacrificing individual privacy typically: are based on user consent (voluntary); feature data minimization; and use decentralized device-to-device signaling, on-device processing, transparent source code, and technical and administrative safeguards to prevent abuse or mis-use. Accessing an individual's detailed location history may not be necessary if contact tracing and alerts can be adequately enabled through proximity-based solutions, such as Bluetooth.

At this time, there are several promising frameworks being developed globally, including the Pan-European Privacy-Preserving Contact Proximity Tracing (PEPP-PT), and particularly the Decentralized Privacy Preserving Proximity Tracing (DP-3T) Protocol developed under this initiative. This protocol relies on short-range Bluetooth technology and decentralized rotating identifiers. The changes to Bluetooth protocols recently announced by Apple and Google will also help to improve interoperability between iOS and Android devices and enable apps used by health authorities to comply with these frameworks.

While these developments are promising, it's important to note that contact tracing relies on the availability of testing. If the availability of testing is limited, the ability to rely on contract tracing is limited as well. Apps should also address risks related to potential mis-use (trolling or other false alerts) or abuse (spoofing). Public health authorities, rather than tech companies, must lead the way in helping shape the development of these apps. Healthcare professionals should also play a role in approving the triggering of alerts for individuals to self-quarantine.

Sources:

- DP-3T repository on GitHub, available at https://github.com/DP-3T.
- Apple Newsroom, *Apple and Google partner on COVID-19 contact tracing technology* (April 10, 2020), available at https://www.apple.com/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/.
- TraceTogether app, available at https://apps.apple.com/us/app/tracetogether/id1498276074.
- TraceTogether privacy policy, available at https://www.tracetogether.gov.sg/common/privacystatement.
- HaMagen app, available at https://play.google.com/store/apps/details?id=com.hamagen.
- HaMagen privacy policy, available at https://govextra.gov.il/ministry-of-health/hamagen-app/terms-and-conditions-of-use-en/.
- Future of Privacy Forum, *Privacy & Pandemics: The Role of Mobile Apps (Chart)* (April 2020), available at https://fpf.org/wp-content/uploads/2020/04/Privacy-Pandemics-The-Role-of-Mobile-Apps.pdf.

2. Much of the discussion surrounding the collection of private data to fight the spread of COVID-19 presents two goals – effectiveness and privacy protection – as mutually exclusive factors that need to be balanced. On one side of the balance, it is assumed that greater amounts of personal data, in more granular form, will allow authorities to track the spread of the virus more effectively. On the other side of the balance is protection of individual privacy, which is believed to be threatened by greater surveillance of individuals by the government.
    - Is this an accurate view of the situation? Are privacy and effectiveness always part of a trade- off, such that the most effective public health measures will come at the expense of privacy, and vice versa? Or do you believe that the most effective policies for combating COVID-19 can also respect individuals' privacy?

**Answer to Question 2:**

Privacy versus effectiveness of data-based solutions against the spread of COVID-19 is a false trade-off. Thoughtful, sophisticated solutions can provide effective solutions that also protect personal data. Particularly for technologies that depend on broad distribution and network effects, trust is key for effectiveness. So not only do privacy and effectiveness not conflict, but they also depend on and reinforce each other.

In addition, data is rarely able to be classified categorically as "anonymous." Rather, most data exists on a spectrum of identifiability, from explicitly personal (e.g. a person's name and address) to expressly non-personal (e.g. supply chain data), with a wide range of data types in

between with varying degrees of protection or anonymity, and varying types of administrative, technical, and legal controls. There are many methods, including privacy enhancing technologies (PETs), to reduce the identifiability of data.

Often, decreasing identifiability (more privacy) coincides with decreasing usefulness of the data (less value) for certain purposes on a sliding scale. For example, differential privacy systems can protect individual privacy by inserting random "noise" into a dataset, which decreases the overall accuracy of the dataset by decreasing the chances that any individual data point represents a real person. However, differential privacy systems are able to *measure and control* this tradeoff through a "privacy budget," a mathematical relationship between the noise added and the resulting accuracy. Thus, while differential privacy can significantly reduce the risk of re-identifiability when deployed correctly, it can also reduce the usefulness of the data for certain purposes, including to identify bias or discrimination in the underlying data.

In the context of public health efforts, however, "usefulness" depends on the purpose for which the data is being collected. For example, for analyzing population-level location trends, highly aggregated and anonymized data may be adequately tailored and useful without requiring access to the underlying data and with no corresponding tradeoff of privacy. Data protection principles, built on the U.S. Fair Information Practice Principles, exist to protect the rights and freedoms of individuals and society and recognize that the basis for data being accessed must always be necessary, proportionate, and limited. Around the world, these principles in law recognize that public health data processing is in the best interest of individuals and society -- when limited and constrained in order to respect the privacy of the individual.

Sources:

- Commission on Evidence-based Policymaking, *CEP Final Report: The Promise of Evidence-Based Policymaking* (September 7, 2017), available at https://www.cep.gov/cep-final-report.html
- United States Census Bureau, *Memorandum 2019.13: Disclosure Avoidance System Design Parameters and Global Privacy-Loss Budget for the 2018 End-to-End Census Test* (July 1, 2019), available at https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/memo-series/2020-memo-2019_13.html (documenting requirements the 2020 Census Program received from the Data Stewardship Executive Policy Committee [DSEP] regarding how to protect the information they collect from the public).
- Heng Xu and Nan Zhang, *Privacy in Health Disparity Research* (December 8, 2018), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3284780 (examining the interplay between data anonymization processes and identifying underlying biases in health disparity research).

3. Today, the United States has numerous federal laws governing different types of data, such as health-care data or financial data. However, there is currently no federal privacy

law that applies generally to all types of consumer data. As Chairman of the Commerce Committee, I have made it a priority to get a national data privacy law enacted as soon as possible.

- If the United States had a national data privacy law in place before the COVID-19 pandemic began, what would the effect have been on efforts to use data to combat the spread of the virus? Would Americans' privacy be more protected, and would companies be more incentivized to take privacy-protective approaches, if we had such a law?

**Answer to Question 3:**

Yes, a comprehensive national consumer privacy law with clear parameters for permissible data sharing for public health or in times of emergency would provide much-needed clarity. Today, without a national framework, we observe significant confusion within U.S. companies and health authorities around the legality and ethics of efforts to support public health emergency needs. In contrast, jurisdictions with comprehensive data protection laws, such as the European Union, have been able to move more quickly to guide appropropriate practices.

Guidance from European Data Protection Authorities (DPAs) and the European Commission in recent months has been critical both to protecting individual privacy rights against government overreach and providing companies with greater clarity on what they may legally share.

Sources:

- European Data Protection Board, *Letter concerning the European Commission's draft Guidance on apps supporting the fight against the COVID-19 pandemic* (April 14, 2020), available at https://edpb.europa.eu/sites/edpb/files/files/file1/edpbletterecadvisecodiv-appguidance_final.pdf.
- European Commission, *Commission Recommendation of 8.4.2020 on a common Union toolbox for the use of technology and data to combat and exit from the COVID-19 crisis, in particular concerning mobile applications and the use of anonymised mobility data* (April 8, 2020), available at https://ec.europa.eu/info/sites/info/files/recommendation_on_apps_for_contact_tracing_4.pdf.
- European Data Protection Board, *Statement on the processing of personal data in the context of the COVID-19 outbreak* (March 19, 2020), available at https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_statement_2020_processingpersonaldataandcovid-19_en.pdf.
- European Data Protection Supervisor, *Comments on DG CONNECT of the European Commission on monitoring of COVID-19 spread* (March 25, 2020), available at https://edps.europa.eu/sites/edp/files/publication/20-03-25_edps_comments_concerning_covid-19_monitoring_of_spread_en.pdf.
- European Data Protection Supervisor, *EU Digital Solidarity: a call for a pan-European approach against the pandemic* (April 6, 2020), available at

https://edps.europa.eu/sites/edp/files/publication/2020-04-06_eu_digital_solidarity_covid19_en.pdf.

- French Data Protection Authority (CNIL), *Les registres communaux d'alerte et d'information des populations* (*Township registries for alerting and informing the population)* (April 14, 2020), available at https://www.cnil.fr/fr/les-registres-communaux-dalerte-et-dinformation-des-populations.
- French Data Protection Authority (CNIL), *Guidance for research projects using health data to combat COVID-19* (March 26, 2020), available at https://www.cnil.fr/fr/recherches-sur-le-covid-19-la-cnil-se-mobilise.
- Spanish Data Protection Authority (AEPD), *Comunicado de la AEPD sobre Apps y Webs de autoevaluación del coronavirus* (March 26, 2020), available at https://www.aepd.es/es/documento/2020-0017.pdf.
- Irish Data Protection Authority (DPC), *Data protection and COVID19* (March 6, 2020), available at https://dataprotection.ie/en/news-media/blogs/data-protection-and-covid-19.
- G. Zanfir-Fortuna, *EU DPAs Issue Green and Red Lights for Processing Health Data During the COVID-19 Epidemic* (March 10), available at https://fpf.org/2020/03/10/eu-dpas-issue-green-and-red-lights-for-processing-health-data-during-the-covid-19-epidemic/.

4. In the United States, the mobile advertising industry and technology companies are collecting consumers' smartphone location data to track the spread of COVID-19 and compliance with social distancing measures. The location data is purported to be in aggregate form and anonymized so that it does not contain consumers' personally identifiable information.
    - How can the use of anonymized, de-identified, and aggregate location data minimize privacy risks to consumers? And, what additional legal safeguards should be imposed on the collection of this data to prevent it from being used or combined with other information to reveal an individual's identity?

**Answer to Question 4:**

Aggregated location data, when it does not reveal device-level information about individual behavior, can almost certainly be used safely and effectively to support public health officials. As long as the underlying data is of sufficiently high quality, and measures are taken to address representativeness, bias, and other group-level risks, including re-identification within very small or rural communities, aggregation of location data represents a useful way to balance public health needs with fully protecting individual privacy.

In contrast, individualized location data is highly sensitive information, even when the data is tied to a device identifier or other pseudonymous identifier, rather than a name or identity. This kind of data is often (incorrectly) referred to within consumer data industries as "anonymous," but in fact is very challenging to de-identify when the data is 1) individualized and 2) collected over time. Individual behavior patterns (especially, home and work locations) are relatively easy to re-identify, and can be highly revealing of our behavior, beliefs, and character. In most cases,

individualized location data will not be necessary, where device-to-device proximity information can be used instead.

Finally, there should be skepticism on the underlying question of whether commercial location data is of sufficient quality to be useful to meet public health needs. In aggregated datasets used to evaluate population trends, individual variation may not be as crucial; in other cases, such as contact tracing, margins of error of only a few meters (or more) could effectively eliminate the usefulness of the data. For example, data collected through cell tower triangulation (cell site location information or CSLI) is likely not precise enough to be effective in COVID-19 response efforts. In addition, there are serious concerns to be addressed regarding the accuracy and representativeness of individualized location data held by the mobile advertising industry. Although mobile apps have the potential to generate highly accurate location and proximity information (due to the number of hardware sensors in a typical smartphone), in practice, many data brokers and third party intermediaries in the mobile advertising industry receive data second-hand and may not have robust quality assurance mechanisms to avoid processing low-quality data. Such data should also be carefully evaluated for its volume and representativeness, given differences in mobile app usage between demographics and age groups.

5. As technology companies share anonymized location data with the U.S. government to support COVID-19 response efforts, to what extent should purpose limitation principles apply to the use and analysis of this data? And, when the pandemic finally passes, what should be done with any anonymized or de-identified data – and identifiable data, if applicable – collected by technology companies and the government for the purpose of addressing the public health crisis?

**Answer to Question 5:**

Purpose specification and use limitation (purpose limitation) are fundamental privacy and data protection principles that should apply to all entities that collect and use personal data. This applies equally to the U.S. government, to the extent they choose to either purchase commercial location data or make use of data shared voluntarily by companies and individuals to support COVID-19 response efforts. In practice, this means that the purpose of using the specific data at issue should be articulated in clear, specific, and granular terms before the data is collected. It also means having specific, public-facing plans for who will have access to the data, and whether and how it will be retained for any future uses.

If some or all data is not deleted or destroyed after its use for COVID-19 efforts, the government entity involved should clearly articulate the reasons for which it is retained, and how the data retained serves those needs while minimizing risks to privacy. For example, it may be possible to allow individuals to opt in to the use of their individualized data for future scientific research. Alternatively, underlying individual data could be deleted, while retaining aggregate or high-level statistical data for research on infectious diseases, or to help prepare response plans for future pandemics. Clearly beneficial research uses could also be supported through oversight

by ethical review boards, or limiting access to data enclaves within the federal government or data trusts overseen by trusted third parties.

**Sen. Thune**

6. More and more Americans all throughout the country are turning to online video services to conduct their jobs, education, and social interactions in an effort to practice social distancing. For instance, Zoom Communications had more than 200 million daily users last month. It was found that thousands of Zoom's calls and videos have been exposed to other users online and log-in information has been stolen resulting in many individuals' personal information being compromised.
    - Did Zoom's privacy policy clearly outline what types of information its platform would collect on individuals? If not, what transparency requirements should be in place for companies like Zoom?
    - Americans are connecting with each other via online services across all 50 states. Would a patchwork of state laws benefit consumers and better protect their privacy? Should the United States enact a national privacy standard to safeguard consumer's information?

**Answer to Question 6:**

The use of Zoom and other forms of collaboration software by Americans in all 50 states highlights the need for uniform standards for consumers to be able to understand their rights and to create clarity for businesses when complying with their obligations across state lines. Transparency is one obligation, and we are supportive of its inclusion in the leading federal privacy proposals in the Committee, including Chairman Wicker's Discussion Draft and Senator Cantwell's Consumer Online Privacy Rights Act (COPRA). Both proposals contain similar strong requirements that companies disclose in easy to read, accessible, public-facing privacy policies information related to the categories of data collected and the purposes for such processing, to whom it is transferred and for what purposes, how long the data will be retained and a detailed description of the covered entity's data minimization and data security policies. In addition, the identity and contact information of the covered entity and a representative thereof should be provided, along with information about how individuals can exercise their rights. Following public criticism from consumer organizations, Zoom updated its privacy policy in March 2020 to a version that provides greater clarity and is better aligned with these requirements.

While transparency is an important legal requirement, it is not a sufficient safeguard alone to rely on end-users to understand and make informed choices about privacy and security. Among other things, a federal privacy law must also require data minimization and privacy by design that would require companies like Zoom to consider privacy from the beginning of product development, rather than addressing it as an after-thought.

## Ms. Stacey Gray, Senior Counsel, Future of Privacy Forum

We are preparing guidance for assessing privacy and security features of video conferencing software and other forms of collaboration software. Factors include whether the company has internal accountability mechanisms, such as a Chief Privacy Officer or equivalent; whether and how the platform could be susceptible to misuse or abuse; what settings and controls are available to users for sharing audio, video, and messages, and under what "defaults;" powers and privileges of administrators or "hosts" over other meeting participants; what personal information and user-created content the platform collects, how it is used, and whether it is sold or shared; and whether the platform uses video data to train or implement biometric technologies (e.g. facial recognition or characterization).

In particular, Zoom and other collaboration software should be scrutinized when used in schools or educational settings, where federal privacy laws apply, including the Family Educational Rights and Privacy Act (FERPA) and the Children's Online Privacy Protection Act (COPPA).

Sources:

- Lisa Weintraub Schifferle, *COPPA Guidance for Ed Tech Companies and Schools during the Coronavirus*, Federal Trade Commission (April 9, 2020), available at https://www.ftc.gov/news-events/blogs/business-blog/2020/04/coppa-guidance-ed-tech-companies-schools-during-coronavirus.
- Future of Privacy Forum, *Online Learning Best Practices for Schools and Educators* (April 15, 2020), available at https://ferpasherpa.org/wp-content/uploads/2020/04/FPF_PP_Online-Learning-Best-Practices_final3.pdf.
- Anisha Reddy and Amelia Vance, *Social (Media) Distancing: Online Learning During a Pandemic* (March 31, 2020), available at https://ferpasherpa.org/social-media-distancing-covid19/.
- FERPA|Sherpa, *Student Privacy During the COVID-19 Pandemic: Resources* (April 2, 2020), available at https://ferpasherpa.org/covid19resources/.
- Future of Privacy Forum, *The Policymaker's Guide to Student Data Privacy* (April 2019), available at https://ferpasherpa.org/wp-content/uploads/2019/04/FPF-Policymakers-Guide-to-Student-Privacy-Final.pdf.
- FERPA|Sherpa, *Student Privacy During the COVID-19 Pandemic*, available at https://ferpasherpa.org/ (consisting of an Education Privacy Resource Center website, which includes a searchable student privacy resource database of over 500 resources from multiple organizations).

7. Without a federal privacy law in place, the American people must rely on the promises of tech companies that all have varying degrees of commitment to maintain consumers' privacy.
    - How do we ensure that organizations are actively engaging in data minimization and strategic deletion practices after data is used or transferred?

**Answer to Question 7:**

Ensuring that companies are actively engaging in data minimization and strategic deletion requires both legal requirements and strong enforcement and accountability mechanisms. For example, FPF supports the accountability provisions in federal proposals by Chairman Wicker and Ranking Member Cantwell that would require covered entities of a certain size to designate a chief privacy officer (CPO) to oversee a comprehensive privacy program, including: monitoring the data practices of an organization; advocating internally for policies and procedures that promote privacy; coordinating efforts to comply with legal privacy responsibilities; and serving as the point of contact between covered entities and regulators and enforcement authorities.

Regular privacy and security auditing with independent external third parties and privacy risk assessments may also provide important accountability mechanisms, especially if made available to the public or the relevant regulator in some cases. Other accountability mechanisms may include whistleblower protections, employee compliance training and education, executive responsibility, and annual data protection reports to the Commission. Along with fear of enforcement for bad behavior, accountability mechanisms such as these provide incentives for covered entities to not only comply, but also to engage in responsible management of personal data throughout the data lifecycle.

8. The country of Israel, through its internal security service, has reportedly used smart-phone location-based contact tracing to notify citizens via text that they have been in close proximity to someone infected with COVID-19, and ordering them to self-isolate for 14 days. A recent opinion piece in the Scientific American urged democratic governments to quickly follow Israel's lead (see "As COVID-19 Accelerates, Governments Must Harness Mobile Data to Stop Spread").
    - Please provide your thoughts on smart-phone location-based contact tracing in light of the extraordinary privacy and other civil liberties concerns such an approach raises for U.S. citizens.
    - According to the Wall Street Journal, MIT is developing a contact tracing app for COVID-19 patients and others who have not been infected by COVID 19 that can be voluntarily downloaded to a person's smart-phone. Please provide your views on this approach to contact tracing.

**Answer to Question 8:**

Israel conducts two distinct programs of relevance for contact tracing. One program relies on the capability of the Shin Bet (the Israeli internal security service) to access cell phone network data to provide the Ministry of Health with information about the location of individuals who were in some proximity to infected individuals, and supports triggering of alerts to those individuals. This program, which is classified and based on new emergency regulations that expire in April

2020, involves large-scale mandatory data collection with no ability for users to opt-out. As a result, it has been criticized for raising serious civil liberties concerns.

However, a second program launched by the Ministry of Health has been supported by leading privacy academics in Israel. This program involves an app, "HaMagen," which individuals may use voluntarily, and leverages GPS data, Wi-Fi data, Google Timeline history (upon separate consent), and Bluetooth data to enable alerts to users who have been in the proximity of a known infected person. Alerts trigger a recommendation for users to voluntarily self-quarantine. HaMagen is open source, voluntary, and according to the Ministry of Health has been adopted by approximately 1.4 million people, or 25% of the desired population.

Overall, mobile apps can support public health initiatives by providing data that is precise and accurate enough for effective person-to-person contact tracing. Apps that have the potential to achieve these goals without sacrificing individual privacy are ones that are based on user consent (voluntary); feature data minimization; decentralized device-to-device signaling; on-device processing; transparent source code; and technical and administrative safeguards to prevent abuse or mis-use.

At this time, there are several promising frameworks being developed globally, including the Pan-European Privacy-Preserving Contact Proximity Tracing (PEPP-PT), and particularly the Decentralized Privacy Preserving Proximity Tracing (DP-3T) Protocol developed under this initiative. This protocol relies on short-range Bluetooth technology and decentralized rotating identifiers. The Future of Privacy Forum supports this approach, which does not rely on bulk collection of precise location histories, whether from existing sources (cell phone carriers or technology providers) or from individuals directly. The changes to Bluetooth protocols recently announced by Apple and Google will also help to improve interoperability between iOS and Android devices and enable apps used by health authorities to comply with these frameworks.

While these developments are promising, it's important to note that contact tracing relies on the availability of testing. If the availability of testing is limited, the ability to rely on contract tracing is limited as well. Apps should also address risks related to potential mis-use (trolling or other false alerts) or abuse (spoofing). Public health authorities should play a role in approving the triggering of alerts for individuals to self-quarantine.

Sources:


- Israeli Ministry of Health, *HaMagen* (English), available at https://govextra.gov.il/ministry-of-health/hamagen-app/download-en/ (providing information on Israel's contact tracing app).
- DP-3T repository on GitHub, available at https://github.com/DP-3T.
- Apple Newsroom, *Apple and Google partner on COVID-19 contact tracing technology* (April 10, 2020), available at https://www.apple.com/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/.
- Future of Privacy Forum, *Privacy and Pandemics: The Role of Mobile Apps (Chart)* (April 2020), available at

https://fpf.org/wp-content/uploads/2020/04/Privacy-Pandemics-The-Role-of-Mobile-Apps.pdf (comparing the different approaches to contact tracing apps and their privacy implications).

- Limor Shmerling-Magazanik, *Brief on Digital Means Employed by the Government of Israel Re: COVID-19* (April 15, 2020), available at https://techpolicy.org.il/brief-on-digital-means-employed-by-the-government-of-israel-re-covid-19/.

9. COVID-19 has caused private companies to seek out and utilize health data in an effort to protect users, employees, and the general public from the spread of the virus. Both Apple and Alphabet have released websites to help users self-screen for exposure to COVID-19. This data will be used to help public health officials. However, these tools also allow technology companies access to user's health information which the companies could in turn profit from in the future.
    - How are technology companies balancing the need for timely and robust reporting to prevent the spread of the virus with the confidentiality and privacy of the participants?
    - What safeguards are in place to ensure data collected as part of the fight against COVID-19 are not sold to business partners or used for the development of other commercial products?

**Answer to Question 9:**

In addition to leveraging existing datasets, several large technology companies are developing new consumer-facing websites, apps, and surveys within existing platforms to assist individuals with early self-screening for symptoms of COVID-19. Examples of these include: Alphabet's Verily, Apple's COVID-19 Screening Tool; and Facebook's Survey for COVID-19 Public Health Research. Whereas data collected by healthcare professionals using symptom screening software or other patient monitoring tools would be subject to the Health Insurance Portability and Accountability Act (HIPAA), most data collected through these consumer-facing services is currently not subject to a national privacy law.

As a result, it is uniquely important that these companies make strong, clear commitments that go beyond their existing privacy policies and specify who will receive health data and how it will be used. It should be clear if, and when, data is shared with collaborating government entities. It should also be clear that the data will not be used for any other purposes than fighting the public health crisis. For example, Verily's Project Baseline provides an FAQ that answers questions about how data will be used and how it is protected. Apple's COVID-19 Screening Tool states that Apple does not collect answers from the screening tool. Facebook's COVID-19 survey notifies users that health data is not collected by Facebook, but rather by health researchers at the Carnegie Mellon University Delphi Research Center.

Sources:

- Verily's Project Baseline, *FAQ* (2020), available at https://www.projectbaseline.com/faq/.
- Apple, *COVID-19 Screening Tool* (2020), available at https://www.apple.com/covid19/.
- Facebook Newsroom, *Data for Good: New Tools to Help Health Researchers Track and Combat COVID-19* (April 6, 2020), available at https://about.fb.com/news/2020/04/data-for-good/.

10. Anonymization techniques are also critical for safeguarding consumers' privacy. Truly anonymized data can protect a consumer's personal information, like their geolocation, political opinions, or religious beliefs.
    - How do companies guarantee that every dataset they are storing contains truly anonymous data? And is the ability to re-identify data a part of the discussion in data-sharing arrangements?

**Answer to Question 10:**

Data rarely categorically falls into the category of either "personal" or "anonymous." Rather, most data exists on a spectrum of identifiability, from explicitly personal (e.g., a person's name and address) to expressly non-personal (e.g. data from jet engines or factory equipment sensors), with a wide range of data types in between with varying degrees of protection or anonymity, and varying types of administrative, technical, and legal controls.

In particular, it is very challenging to fully "anonymize" individualized device location data, even when it is tied to a device identifier or other pseudonymous identifier (rather than a name or identity). This kind of data is often (incorrectly) referred to within consumer data industries as "anonymous," but in fact, individuals can often be re-identified based on their home and work locations or by cross-referencing the dataset with an individual's known locations and times.

Nonetheless, there are ways to reduce the identifiability or risk of a location dataset. These can be applied alone or in combination, and include, for example: differential privacy (perturbing or adding noise to the dataset to reduce the risk of any specific individual being capable of identification); removing or obscuring data (such as from home and work locations); applying administrative controls (such as limiting access to the dataset and placing contractual limitations on its use); and aggregating the data so that it reveals only movements of groups of people of a certain size. Aggregated location data can almost certainly be used safely and effectively to support public health officials.

In applying privacy enhancing technologies (PETs), companies can also look to existing guidance from U.S. federal agencies, including resources such as:

- NIST Special Publication 800-188 (2nd Draft), *De-Identifying Government Datasets* (December 2016), available at https://csrc.nist.gov/csrc/media/publications/sp/800-188/draft/documents/sp800_188_draft2.pdf.

- Federal Data Strategy, *2020 Action Plan*, available at https://strategy.data.gov/action-plan/ (announcing the Federal Committee on Statistical Methodology's *Data Protection Toolkit* and update to the *2005 Report on Statistical Disclosure Limitation Methodology*, both forthcoming in December 2020, which will provide relevant tools and guidance).
- US Census Bureau, *Disclosure Avoidance and the 2020 Census* (last revised March 27, 2020), available at https://www.census.gov/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html (describing differential privacy methods used to protect census data).
- NIST, *Collaboration Space*, available at https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/collaboration-space; and *2018 Public Safety Communications Research Differential Privacy Synthetic Data Challenge*, available at https://www.nist.gov/ctl/pscr/open-innovation-prize-challenges/past-prize-challenges/2018-differential-privacy-synthetic.
- Federal Statistical Research Data Centers, available at https://www.census.gov/fsrdc.
- National Academies of Science, *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy* (2017), available at http://mitsloan.mit.edu/shared/ods/documents/?DocumentID=4438; and *Federal Statistics, Multiple Data Sources, and Privacy Protection* (2017), available at https://www.ncbi.nlm.nih.gov/books/NBK475779/pdf/Bookshelf_NBK475779.pdf (recommending use of tiered access models that would increase restrictions on data access depending on the sensitivity and identifiability of the data accessed).

**Sen. Blunt**

As you know, this committee has prioritized drafting federal privacy legislation for the purpose of creating clear, baseline definitions and standards for data collection, storage, and use across industry sectors. Similarly, the bills before this committee attempt to create definitions to meet appropriate levels of consent and transparency for protecting consumers' privacy and security.

In relation to COVID-19, the end users of specific data sets, like location data, are more likely to be governmental entities than commercial entities. Big data can be an incredible tool to better understand the spread of the virus, and the impact on communities across the country. Data can help identify resource deficits, inform governments and health care professionals to employ countermeasures at the appropriate time, and provide insight to the downstream economic effects of this pandemic.

However, U.S. commercial entities that would likely be collecting this data have very few guardrails on the collection and distribution of this data. Similarly, there are few requirements or regulations at federal and state levels which guide methodologies for anonymizing or

pseudonymizing data. De-identifying data may result in greater data privacy and data security for consumers or individual citizens but relies heavily on all of the entities involved in the collection and storage of that data making decisions based on best practices.

11. What efforts do you recommend that federal agencies undertake to ensure that data being used to track viral spread are upholding the highest possible standards for individual privacy and security?

**Answer to Question 11:**

Personal information exists on a spectrum of identifiability, from explicitly personal (e.g. a person's name and address) to expressly non-personal (e.g. data from jet engines or factory equipment sensors). There are a wide range of data types in between with varying degrees of protection or anonymity, and varying types of administrative, technical, and legal controls.

Federal agencies should continue to promote and apply clear guidance and standards for robust privacy and security controls, and uniform terminology, such as those in NIST Draft Special Publication 800-53 (*Security and Privacy Controls for Information Systems and Organizations*) and NISTIR 8085 (*De-Identification of Personal Information*); promoting guidance and standards for the use of Privacy Enhancing Technologies (PETs) for disclosure avoidance, for example in accordance with detailed guidance from NIST and the Census Bureau; leveraging the expertise of the Federal Privacy Council, the Federal CIO Council, the Interagency Council on Statistical Policy; and the Federal Statistical Research Data Centers. Agencies can also encourage or mandate public commitments to maintain data in de-identified form, and promote public accountability by requiring organizations to be transparent about de-identification methodologies used (for example, Google's technical documentation for the anonymization methods used in generating COVID-19 Mobility Maps are publicly available).

Sources:

- NIST Special Publication 800-53 Revision 5 (Draft), *Security and Privacy Controls for Information Systems and Organizations* (March 2020), available at https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/draft.
- NIST Special Publication 800-188 (2nd Draft), *De-Identifying Government Datasets* (December 2016), available at https://csrc.nist.gov/csrc/media/publications/sp/800-188/draft/documents/sp800_188_draft2.pdf.
- Federal Privacy Council, available at https://www.fpc.gov/federal-privacy-council/.
- Federal CIO Council, available at https://www.cio.gov/.
- Interagency Council on Statistical Policy, available at https://www.census.gov/fsrdc.
- Federal Data Strategy, *2020 Action Plan*, available at https://strategy.data.gov/action-plan/ (announcing the Federal Committee on Statistical Methodology's *Data Protection Toolkit* and update to the *2005 Report*

*on Statistical Disclosure Limitation Methodology*, both forthcoming in December 2020, which will provide relevant tools and guidance).

- US Census Bureau, *Disclosure Avoidance and the 2020 Census* (last revised March 27, 2020), available at https://www.census.gov/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html (describing differential privacy methods used to protect census data).
- NIST, *Collaboration Space*, available at https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/collaboration-space; and *2018 Public Safety Communications Research Differential Privacy Synthetic Data Challenge*, available at https://www.nist.gov/ctl/pscr/open-innovation-prize-challenges/past-prize-challenges/2018-differential-privacy-synthetic.
- National Academies of Science, *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy* (2017), available at http://mitsloan.mit.edu/shared/ods/documents/?DocumentID=4438; and *Federal Statistics, Multiple Data Sources, and Privacy Protection* (2017), available at https://www.ncbi.nlm.nih.gov/books/NBK475779/pdf/Bookshelf_NBK475779.pdf (recommending use of tiered access models that would increase restrictions on data access depending on the sensitivity and identifiability of the data accessed).
- Office of Management and Budget (OMB), *Circular A-130: Managing Federal Information as a Strategic Resource*, available at https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/circulars/A130/a130revised.pdf.
- NIST, *Privacy Framework* (last revised January 16, 2020), available at https://www.nist.gov/privacy-framework.
- Future of Privacy Forum, *City of Seattle Open Data Risk Assessment* (January 2018), available at https://fpf.org/wp-content/uploads/2018/01/FPF-Open-Data-Risk-Assessment-for-City-of-Seattle.pdf.
- International Association of Privacy Professionals, *The Skill Set Needed to Implement a Privacy Risk Management Framework* (March 19, 2020), available at https://www.nist.gov/privacy-framework/iapp-cipm-crosswalk.

12. Does data lose any utility when it is de-identified or anonymized? Is it possible to have large data sets that are not tied to individual's identities, but which would still be useful for governments or public health-related end users?

**Answer to Question 12:**

Data is rarely able to be classified categorically as either "personal" or "anonymous." Rather, most data exists on a spectrum of identifiability, and there are methods, including privacy enhancing technologies (PETs), to reduce the identifiability of data. Often, decreasing

identifiability (more privacy) coincides with decreasing usefulness of the data (less value) for certain purposes on a sliding scale.

For example, differential privacy systems can protect individual privacy by inserting random "noise" into a dataset, which decreases the overall accuracy of the dataset by decreasing the chances that any individual data point represents a real person. However, differential privacy systems are able to *measure and control* this tradeoff through a "privacy budget," a mathematical relationship between the noise added and the resulting accuracy. Thus, while differential privacy can thus significantly reduce the risk of re-identifiability when deployed correctly, it can also reduce the usefulness of the data certain purposes, including to identify bias or discrimination in the underlying data.

In the context of public health efforts, "usefulness" depends on the purpose for which the data is being used. For example, for analyzing population-level location trends, highly aggregated and anonymized data may be adequately tailored and useful without requiring access to the underlying data and with no corresponding tradeoff of privacy.

As a result, FPF strongly believes that the framing of "privacy versus effectiveness" of data-based solutions against the spread of COVID-19 is a false trade-off. Data protection principles, built on the U.S. Fair Information Practice Principles, exist to protect the rights and freedoms of individuals and society and recognize that the basis for data being accessed must always be necessary, proportionate, and limited. Around the world, these principles in law recognize that public health data processing is in the best interest of individuals and society -- when limited and constrained in order to respect the privacy of the individual.

Sources:

- Commission on Evidence-based Policymaking, *CEP Final Report: The Promise of Evidence-Based Policymaking* (September 7, 2017), available at https://www.cep.gov/cep-final-report.html.
- United States Census Bureau, *Memorandum 2019.13: Disclosure Avoidance System Design Parameters and Global Privacy-Loss Budget for the 2018 End-to-End Census Test* (July 1, 2019), available at https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/memo-series/2020-memo-2019_13.html (documenting requirements the 2020 Census Program received from the Data Stewardship Executive Policy Committee [DSEP] regarding how to protect the information they collect from the public).
- Heng Xu and Nan Zhang, *Privacy in Health Disparity Research* (December 8, 2018), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3284780 (examining the interplay between data anonymization processes and identifying underlying biases in health disparity research).

13. It is important to me that as government entities access commercially collected or publicly available data, that those efforts are giving reasonable consideration to protecting individual privacy and security.

*(No reply.)*

14. Are there any technologies that offer the opportunity to collect data that would be useful to governmental pandemic response efforts, without resorting to surveillance methods that jeopardize individual privacy – like those which have been used recently by foreign governments?

**Answer to Question 14:**

At this time, the Future of Privacy Forum supports the approach of several promising frameworks being developed globally, including the Pan-European Privacy-Preserving Contact Proximity Tracing (PEPP-PT), and particularly the Decentralized Privacy Preserving Proximity Tracing (DP-3T) Protocol developed under this initiative. This protocol relies on short-range Bluetooth technology and decentralized rotating identifiers. This approach does not rely on bulk collection of precise location histories, whether from existing sources (cell phone carriers or technology providers) or from individuals directly. The changes to Bluetooth protocols recently announced by Apple and Google will also help to improve interoperability between iOS and Android devices and enable apps used by health authorities to comply with these frameworks.

While these developments are promising, it's important to note that contact tracing relies on the availability of testing. If the availability of testing is limited, the ability to rely on contract tracing is limited as well. Apps should also address risks related to potential mis-use (trolling or other false alerts) or abuse (spoofing). Public health authorities should play a role in approving the triggering of alerts for individuals to self-quarantine.

Sources:

- DP-3T repository on GitHub, available at https://github.com/DP-3T.
- Apple Newsroom, *Apple and Google partner on COVID-19 contact tracing technology* (April 10, 2020), available at https://www.apple.com/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/.
- Future of Privacy Forum, *Privacy and Pandemics: The Role of Mobile Apps (Chart)* (April 2020), available at https://fpf.org/wp-content/uploads/2020/04/Privacy-Pandemics-The-Role-of-Mobile-Apps.pdf (comparing the different approaches to contact tracing apps and their privacy implications).

**Sen. Cruz**

## Ms. Stacey Gray, Senior Counsel, Future of Privacy Forum

15. A little over two weeks ago, the Johns Hopkins Center for Health Security published a report titled *"Modernizing and Expanding Outbreak Science to Support Better Decision Making During Public Health Crises: Lessons for COVID-19 and Beyond."* Although full of thought-provoking ideas, one of the most notable was a recommendation to establish a "National Infectious Disease Forecasting Center," similar to the National Weather Service. Much like the National Weather Service, this new infectious disease forecasting center would have both an operational role—providing the best modeling and forecasting to policy makers and public health professionals before, during, and after a disease outbreak—as well as a research role—providing a venue for academic, private sector, and governmental collaboration to improve models and encourage innovation.
    - What do you all think of this idea, and what do you all think the positives and negatives would be if such a concept was operationalized?

16. One of the big reasons weather forecasting works, if not the biggest, is how many observations—things like water temperature, barometric pressure, radio profiles of the atmosphere, etc.—are fed into the weather model. Now while collecting ocean temperatures from buoys, or pressure readings from weather balloons, doesn't really raise privacy concerns, collecting health observations almost certainly would.
    - How can we thread the needle—either in this concept or private sector modeling—of getting enough of the right kind of data to accurately model infectious disease outbreaks while still protecting the privacy and security of individuals?

**Answer to Questions 15 and 16:**

There are clear benefits to the U.S. federal government playing a greater role in addressing public health needs, advancing scientific knowledge, and avoiding future pandemics. Specifically, there could be benefits to a federal agency that encourages academic, private sector, and government collaboration for scientific research relating to infectious diseases, if conducted in line with data protection and privacy law.

Around the world, data protection law recognizes that using data to advance scientific research and public health is in the best interest of both individuals and society, so long as the data is necessary, proportionate, and limited to protect the rights and freedoms of individuals. When privately held data is responsibly shared with academic and government researchers, it can support significant progress in medicine, public health, education, social science, and other fields. The Future of Privacy Forum has encouraged preservation of privacy in academic-industry research collaborations on data-driven research, including through an Award for Research Data Stewardship, and the formation of the Corporate-Academic Data Stewardship Research Alliance (CADRA), a peer-to-peer network of private companies who share the goal of facilitating privacy-protective data sharing between businesses and academic researchers.

Sources:

- Future of Privacy Forum, *Understanding Corporate Data Sharing Decisions* (November 2017), available at https://fpf.org/wp-content/uploads/2017/11/FPF_Data_Sharing_Report_FINAL.pdf
- Sara R. Jordan, *Designing an Artificial Intelligence Research Review Committee* (October 2019), available at https://fpf.org/wp-content/uploads/2019/10/DesigningAIResearchReviewCommittee.pdf.

17. To date the State of Texas has reported thousands of cases of coronavirus, and hundreds of deaths related to complications from infection. To mitigate the risk of infection in Texas and across the country, the administration has restricted international travel, provided more access to medical supplies by involving the powers of the Defense Production Act, and cut red tape to expand access to testing. Congress also passed the CARES Act which provided $377 billion in emergency loans for small businesses and directed $100 billion to hospitals and healthcare providers. However, I believe much still needs to be done to finish this fight and recover once this is behind us.
    - In your expert opinions, what more needs to be done to beat this virus, and how can federal, state, and local governments work with private companies to both mitigate spread of the virus—both now and later this summer or fall—and recover quickly once the threat of this virus has passed?

**Answer to Question 17:**

In addition to economic measures, we believe that data and technology can play an important role in supporting public health efforts, including now and later this summer and fall as people return to their normal routines of work and life. For example, mobile apps, if adopted voluntarily by a sufficient percentage of a population, can support public health initiatives by providing data that is precise and accurate enough for effective person-to-person contact tracing. Apps that have the potential to achieve these goals without sacrificing individual privacy are ones that are based on user consent (voluntary); feature data minimization; decentralized device-to-device signaling; on-device processing; transparent source code; and technical and administrative safeguards to prevent abuse or mis-use. Accessing an individual's detailed location history may not be necessary if contact tracing and alerts can be adequately enabled through proximity-based solutions, such as Bluetooth.

At this time, there are several promising frameworks being developed globally, including the Pan-European Privacy-Preserving Contact Proximity Tracing (PEPP-PT), and particularly the Decentralized Privacy Preserving Proximity Tracing (DP-3T) Protocol developed under this initiative. This protocol relies on short-range Bluetooth technology and decentralized rotating identifiers. The changes to Bluetooth protocols recently announced by Apple and Google will also help to improve interoperability between iOS and Android devices and enable apps used by health authorities to comply with these frameworks.

While these developments are promising, it's crucial to note that contact tracing relies on sufficient voluntary adoption and the availability of testing and adequate medical resources. If

the availability of testing is limited, the ability to rely on contract tracing is limited as well. Apps should also address risks related to potential mis-use (trolling or other false alerts) or abuse (spoofing). Public health authorities, rather than tech companies, must lead the way in helping shape the development of these apps. Healthcare professionals should also play a role in approving the triggering of alerts for individuals to self-quarantine.

Sources:

- DP-3T repository on GitHub, available at https://github.com/DP-3T.
- Apple Newsroom, *Apple and Google partner on COVID-19 contact tracing technology* (April 10, 2020), available at https://www.apple.com/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/.

**Sen. Moran**

18. Many of the discussed proposals related to utilizing "big data" to fight against the spread against coronavirus rely upon the concepts of anonymized and aggregated data to protect the personal identity of individuals that this information pertains to and prevent consumer harms that could result. As such, many members on this Committee have spent significant time and energy drafting federal privacy legislation that tries to account for practices such as these that prevent harmful intrusions into consumers' privacy while also preserving innovative processing practices that could utilize such information responsibly without posing risks.
    - That being said, do the witnesses have any policy recommendations for the Committee as it relates to effectively defining technical criteria for "aggregated" and "anonymized" data, such as requiring companies to publicly commit that they will refrain from attempting to re-identify data to a specific individual while adopting controls to prevent such efforts?

**Answer to Question 18:**

Federal proposals should reflect the fact that data is rarely able to be classified categorically as either "personal" or "anonymous." Rather, most data exists on a spectrum of identifiability, and there are methods, including privacy enhancing technologies (PETs), to reduce the identifiability of data. In defining these terms, policymakers can follow existing federal guidance and standards for robust privacy and security controls, and uniform terminology, such as those in NIST Draft Special Publication 800-53 (*Security and Privacy Controls for Information Systems and Organizations*) and NISTIR 8085 (*De-Identification of Personal Information*); promoting guidance and standards for the use of Privacy Enhancing Technologies (PETs) for disclosure avoidance, for example in accordance with detailed guidance from NIST and the Census Bureau; leveraging the expertise of the Federal Privacy Council, the Federal CIO Council, the Interagency Council on Statistical Policy; and the Federal Statistical Research Data Centers.

Policymakers can also encourage or mandate public commitments to maintain data in de-identified form, and promote public accountability by requiring organizations to be transparent about de-identification methodologies used (for example, Google's technical documentation for the anonymization methods used in generating COVID-19 Mobility Maps are publicly available).

Sources:

- NIST Special Publication 800-53 Revision 5 (Draft), *Security and Privacy Controls for Information Systems and Organizations* (March 2020), available at https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/draft.
- NISTIR 8053: *De-Identification of Personal Information*: https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf or ISO/IEC 27701 https://www.iso.org/obp/ui/#iso:std:iso-iec:27701:ed-1:v1:en.
- NIST Special Publication 800-188 (2nd Draft), *De-Identifying Government Datasets* (December 2016), available at https://csrc.nist.gov/csrc/media/publications/sp/800-188/draft/documents/sp800_188_draft2.pdf.
- Federal Data Strategy, *2020 Action Plan*, available at https://strategy.data.gov/action-plan/ (announcing the Federal Committee on Statistical Methodology's *Data Protection Toolkit* and update to the *2005 Report on Statistical Disclosure Limitation Methodology*, both forthcoming in December 2020, which will provide relevant tools and guidance).
- US Census Bureau, *Disclosure Avoidance and the 2020 Census* (last revised March 27, 2020), available at https://www.census.gov/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html.
- NIST, *Collaboration Space*, available at https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/collaboration-space; and *2018 Public Safety Communications Research Differential Privacy Synthetic Data Challenge*, available at https://www.nist.gov/ctl/pscr/open-innovation-prize-challenges/past-prize-challenges/2018-differential-privacy-synthetic.
- National Academies of Science, *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy* (2017), available at http://mitsloan.mit.edu/shared/ods/documents/?DocumentID=4438; and *Federal Statistics, Multiple Data Sources, and Privacy Protection* (2017), available at https://www.ncbi.nlm.nih.gov/books/NBK475779/pdf/Bookshelf_NBK475779.pdf.
- Federal Privacy Council, available at https://www.fpc.gov/federal-privacy-council/. =
- Federal CIO Council, available at https://www.cio.gov/.
- Interagency Council on Statistical Policy, available at https://www.census.gov/fsrdc.
- Ahmet Aktay et al., *Google COVID-19 Community Mobility Reports: Anonymization Process Description (version 1.0)* (last revised April 9, 2020), available at https://arxiv.org/abs/2004.04145.

19. Consumer data has tremendous benefits to society, as is clearly evident in the fight against the COVID-19 outbreak. Big data and the digitized processes and algorithms that technology companies are developing have led to an entirely new sector of the global economy.
    - Are you satisfied that the technology industry is striking an appropriate balance between producing services that better our ability to solve problems, as is clear in the fight against COVID-19, versus their production of products that increase their bottom line and generate profit? Are you satisfied that the United States government is striking an appropriate balance between supporting these companies in addressing COVID-19 versus ensuring we conduct adequate oversight of the industries' activities?

**Answer to Question 19:**

The COVID-19 outbreak has underscored the need for increased oversight of consumer data industries through comprehensive federal privacy legislation in the United States. Today, without a national framework, we observe significant confusion within U.S. companies and health authorities around the legality and ethics of efforts to support public health emergency needs. In contrast, jurisdictions with comprehensive data protection laws, such as the European Union, have been able to move more quickly to guide appropropriate practices. Comprehensive privacy legislation should incorporate the fundamental privacy and data protection principles of purpose specification and use limitation (purpose limitation).

These principles apply equally to the U.S. government, to the extent they choose to either purchase commercial data or make use of data shared voluntarily by companies and individuals to support COVID-19 response efforts. In practice, this means that the purpose of using the specific data at issue should be articulated in clear, specific, and granular terms before the data is collected. It also means having specific, public-facing plans for who will have access to the data, and whether and how it will be retained for any future uses.

If some or all data is not deleted or destroyed after its use for COVID-19 efforts, the government entity involved should clearly articulate the reasons for which it is retained, and how the data retained serves those needs while minimizing risks to privacy. For example, it may be possible to allow individuals to opt in to the use of their individualized data for future scientific research. Alternatively, underlying individual data could be deleted, while retaining aggregate or high-level statistical data for research on infectious diseases, or to help prepare response plans for future pandemics. Clearly beneficial research uses could also be supported through oversight by ethical review boards, or limiting access to data enclaves within the federal government.

20. Consumer trust is essential to both the United States government and to the companies whose products we use every day. We need to work to maintain that trust and ensuring that the big data being used to analyze the COVID-19 outbreak was collected and processed in a manner that aligns with our principles is important to my constituents.

- How can we adequately ensure that the data being used to address COVID-19 is sourced and processed in a manner that ensures consumer trust is not being violated, while allowing the innovation and success we've seen continue to grow?

**Answer to Question 20:**

Gaining and maintaining the trust of individuals is key for the effectiveness of technologies deployed to tackle COVID-19 that depend on broad distribution and network effects. For instance, mobile apps, if adopted voluntarily by a sufficient percentage of a population, can support public health initiatives by providing data that is precise and accurate enough for effective person-to-person contact tracing. Thoughtful, sophisticated solutions can be effective and also protect personal data. The U.S. government and companies can maintain trust by only collecting and using personal data to the extent that it is necessary to tackle COVID-19.

For instance, accessing an individual's detailed location history may not be necessary if contact tracing and alerts can be adequately enabled through proximity-based solutions, such as Bluetooth. In addition, many users would be dissuaded from voluntarily adopting an app that collected precise, persistent location history. To ensure that data being used to address COVID-19 is sourced and processed in a manner that ensures consumer trust is not being violated, solutions should be based on user consent (voluntary), feature data minimization, use decentralized device-to-device signaling, on-device processing, transparent source source code, and technical and administrative safeguards to prevent abuse or mis-use.

21. It is important to remember that the internet is a global network and that no matter how secure we make our networks, they remain vulnerable to bad actors, corruption, and misguided influence from around the world. Can you comment on the practices we've seen used by companies and international partners to ensure the data used to address COVID-19 is both accurately sourced and stored in a manner that is secure?

**Answer to Question 21:**

Aggregate location data from consumer smartphones, when it does not reveal device-level information about individual behavior, is currently being used in the United States and most other countries to safely and effectively to support public health officials in measuring population-level trends. As long as the underlying data is of sufficiently high quality, and measures are taken to address representativeness, bias, and other group-level risks, including re-identification within very small or rural communities, aggregation of location data represents a useful way to balance public health needs with fully protecting individual privacy.

There are also several promising frameworks being developed globally for "contact tracing," including the Pan-European Privacy-Preserving Contact Proximity Tracing (PEPP-PT), and particularly the Decentralized Privacy Preserving Proximity Tracing (DP-3T) Protocol developed under this initiative. This protocol relies on short-range Bluetooth technology and decentralized

rotating identifiers. The changes to Bluetooth protocols recently announced by Apple and Google will also help to improve interoperability between iOS and Android devices and enable apps used by health authorities to comply with these frameworks.

While these developments are promising, it's important to note that contact tracing relies on the availability of testing. If the availability of testing is limited, the ability to rely on contract tracing is limited as well. Apps should also address risks related to potential mis-use (trolling or other false alerts) or abuse (spoofing). Public health authorities, rather than tech companies, must lead the way in helping shape the development of these apps. Healthcare professionals should also play a role in approving the triggering of alerts for individuals to self-quarantine.

**Sen. Blackburn**

It's time that we align consumer expectations with reality. That holds true whether we are discussing the latest in wearables or the hot new videoconferencing app that helps people work remotely. Consumers have a reasonable expectation that their information will be kept private, and that the platforms they interact with will maintain adequate levels of security to bolster that effort.

We need to pass federal privacy legislation to set a national standard that will allow companies to innovate while protecting consumers. HIPAA was not designed to work with 21st century systems, yet consumers expect that all of their health-related information will be protected by those same standards. I'm afraid that the COVID-19 pandemic will only exacerbate these issues. Corona points out the need to update HIPAA, not to allow tech companies to exploit a crisis to gather even more personal data.

22. How do you see HIPAA interacting with your worldview of the tech industry?

**Answer to Question 22:**

HIPAA supports very wide sharing of data across many companies that provide support in the provision of healthcare services, but with relatively strict standards for data de-identification and the sharing of identifiable health information. HIPAA also restricts many uses of data, such as for marketing and (at least in spirit) offers patients unrestricted access to their personal health records.

However, a large amount of equally sensitive health and wellness data currently falls outside of the auspices of HIPAA when it is not collected by a doctor or healthcare institution -- for example, data from fitness and wearable devices, or self-reported health data generated by current consumer-facing COVID-19 screening tools and surveys. There are also rapidly emerging strategies and tools to give patients easier access to their electronic health records, such as through mobile apps. In order to more effectively protect patients engaging on or with these platforms, the US needs a broad comprehensive law that can assure that all companies provide

baseline privacy rights and comply with strict privacy and security standards when handling health information.

23. How do you envision working with the CDC to develop the updated surveillance system (which was given $500 million in the recently passed CARES Act) while protecting health information and thereby allow CDC to use their expertise – epidemiology that inherently seeks to protect health information – with big tech's powerful data collection and analysis tools?

*(No reply.)*

24. Today we are giving into state surveillance for the sake of saving thousands of lives that might otherwise be lost to coronavirus. The CDC is already relying on data analytics from mobile ad providers to track the spread of the disease. How can we ensure the data collection will only be done for the limited purposes of the emergency, with safeguards to ensure anonymity? On retention time, when should the data be deleted? Who has the right to that deletion – the federal government or the individuals themselves? Most importantly, what duty do tech companies owe to protect consumer privacy, even during a global pandemic?

**Answer to Question 24:**

Maintaining consumer trust in the use of sensitive data in a public health emergency is critical, especially when it relies on the voluntary adoption of consumer-facing apps and screening tools. As a threshold matter, companies and government agencies can help build trust through transparency, by being clear about the collection, use, and sharing of personal data, and sharing technical specifications for de-identification methods. If the public has no awareness of the data sharing, there can be no scrutiny or review to ensure it is appropriate.

Company and government agencies can also conduct Privacy Impact Assessments (PIAs) to evaluate and mitigate risks; promote internal accountability through privacy programs and oversight; and publicly commit to deletion or other limits on retention. The PIA should specify the purposes for the data being collected and how it will be used to fulfill the clear, specific, articulated goals. The types of data and the data sources should also be described in detail.

We also recommend that safeguards such as pseudonymisation, aggregation, encryption and decentralization be used when appropriate. Most data exists on a spectrum of identifiability, from explicitly personal (e.g. a person's name and address) to expressly non-personal (e.g. supply chain data), with a wide range of data types falling in between with varying degrees of anonymity. There are many methods, including privacy enhancing technologies (PETs), to reduce the identifiability of data. Some data, such as precise location history, is uniquely challenging to de-identify and will rarely be capable of being considered one hundred percent

"anonymous," and thus special consideration should be given to whether data should be collected to begin with and whether it should be shared with others or released to the public. We recommend that the company or government agency consult de-identification experts and use cutting edge and best in class technologies.

Finally, retention and data deletion must be appropriate and clearly communicated. For example, it has been reported that Apple and Google recently committed publicly to ensure data collected at the device level for COVID-19 efforts will be deleted in 14 days. In addition, the two companies publicly committed to terminate COVID-19 tracking tools once the pandemic ends.

Due to the lack of comprehensive privacy legal obligations in the US, it is currently up to companies or government agencies to voluntarily erase the data collected as part of the COVID19 response. Individuals do not generally have the right to submit deletion requests in relation to their data. One exception is the California Consumer Privacy Act (CCPA), which provides a right to deletion to California residents, but only in relation to data processing that falls under this Act. The CCPA does not apply to government agencies or any public bodies. We recommend that the accountable company or government agency set the retention period and deletion requirements (in consultation with stakeholders) and that the accountable entity be responsible for the secure deletion.

Sources:


- CNET, *Apple, Google to terminate COVID-19 tracking tools once pandemic ends* (April 14, 2020), available at https://www.cnet.com/news/apple-and-google-say-they-will-shut-down-covid-19-tracking-tools-once-pandemic-ends/.
- Bureau of Justice Assistance, US Department of Justice, *Guide to Conducting Privacy Impact Assessments for State, Local, and Tribal Justice Entities* (June 2012), available at https://it.ojp.gov/documents/d/Guide%20to%20Conducting%20Privacy%20Impact%20Assessments_compliant.pdf..
- US Department of Transportation, *Privacy Impact Assessments* (last updated April 7, 2020), available at https://www.transportation.gov/individuals/privacy/privacy-impact-assessments (offering sample PIAs, including original PIAs and updates to previously published PIAs).
- CNIL, *Privacy Impact Assessment (PIA) Application to IoT Devices* (February 2018), available at https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-piaf-connectedobjects-en.pdf (providing a sample of a PIA for IoT devices).
- CNIL, *Privacy Impact Assessment (PIA) Templates* (February 2018), available at https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-2-en-templates.pdf.
- Information Commissioner's Office (ICO), *Data Protection Impact Assessments*, available at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/ (giving examples of Data Protection Impact Assessments, which are similar to PIAs).

25. Foreign countries like South Korea, Taiwan, Singapore, and Israel swiftly mobilized collection of cell phone location data to track the spread of the virus and map out infection hot zones. Israel just released an app that allows the public to track whether they have may visited a location that put them into contact with an infected individual. Is it even possible to adopt similar measures while still balancing protections for privacy and civil liberties?

**Answer to Question 25:**

Privacy versus effectiveness of data-based solutions against the spread of COVID-19 is a false trade-off. Thoughtful, sophisticated solutions can provide effective solutions that also protect personal data. Particularly for technologies that depend on broad distribution and network effects, trust is key for effectiveness. So not only do privacy and effectiveness not conflict, but they also depend on and reinforce each other.

For example, Israel conducts two distinct programs of relevance for contact tracing. One program relies on the capability of the Shin Bet (the Israeli internal security service) to access cell phone network data to provide the Ministry of Health with information about the location of individuals who were in some proximity to infected individuals, and supports triggering of alerts to those individuals. This program, which is classified and based on new emergency regulations that expire in April 2020, involves large-scale mandatory data collection with no ability for users to opt-out. As a result, it has been criticized for raising serious civil liberties concerns.

However, a second program launched by the Ministry of Health has been supported by leading privacy academics in Israel. This program involves an app, "HaMagen," which individuals may use voluntarily, and leverages GPS data, Wi-Fi data, Google Timeline history (upon separate consent), and Bluetooth data to enable alerts to users who have been in the proximity of a known infected person. Alerts trigger a recommendation for users to voluntarily self-quarantine. HaMagen is open source, voluntary, and according to the Ministry of Health has been adopted by approximately 1.4 million people, or 25% of the desired population.

Overall, mobile apps can support public health initiatives by providing data that is precise and accurate enough for effective person-to-person contact tracing. Apps that have the potential to achieve these goals without sacrificing individual privacy are ones that are based on user consent (voluntary); feature data minimization; decentralized device-to-device signaling; on-device processing; transparent source code; and technical and administrative safeguards to prevent abuse or mis-use.

At this time, there are several promising frameworks being developed globally, including the Pan-European Privacy-Preserving Contact Proximity Tracing (PEPP-PT), and particularly the Decentralized Privacy Preserving Proximity Tracing (DP-3T) Protocol developed under this initiative. This protocol relies on short-range Bluetooth technology and decentralized rotating identifiers. The Future of Privacy Forum supports this approach, which does not rely on bulk collection of precise location histories, whether from existing sources (cell phone carriers or

technology providers) or from individuals directly. The changes to Bluetooth protocols recently announced by Apple and Google will also help to improve interoperability between iOS and Android devices and enable apps used by health authorities to comply with these frameworks.

While these developments are promising, it's important to note that contact tracing relies on the availability of testing. If the availability of testing is limited, the ability to rely on contract tracing is limited as well. Apps should also address risks related to potential mis-use (trolling or other false alerts) or abuse (spoofing). Public health authorities should play a role in approving the triggering of alerts for individuals to self-quarantine.

- DP-3T repository on GitHub, available at https://github.com/DP-3T.
- Apple Newsroom, *Apple and Google partner on COVID-19 contact tracing technology* (April 10, 2020), available at https://www.apple.com/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/.
- TraceTogether app, available at https://apps.apple.com/us/app/tracetogether/id1498276074.
- TraceTogether privacy policy, available at https://www.tracetogether.gov.sg/common/privacystatement.
- HaMagen app, available at https://play.google.com/store/apps/details?id=com.hamagen.
- HaMagen privacy policy, available at https://govextra.gov.il/ministry-of-health/hamagen-app/terms-and-conditions-of-use-en/.
- Future of Privacy Forum, *Privacy & Pandemics: The Role of Mobile Apps (Chart)* (April 2020), available at https://fpf.org/wp-content/uploads/2020/04/Privacy-Pandemics-The-Role-of-Mobile-Apps.pdf.

**Sen. Capito**

26. If the government were to utilize a combination of public and private consumer data to create a COVID-19 Public Health Initiative, what barriers to implementation could arise with a patchwork of State privacy laws?

**Answer to Question 26:**

Although there is currently only one state that has passed a comprehensive, baseline consumer privacy law (the California Consumer Privacy Act of 2018) (CCPA), we envision several possible ways that multiple state privacy laws might hinder the advancement of federal public health initiatives if those laws were to conflict with each other or diverge in significant ways. For example, state privacy laws might create different definitions of key terms, such as "personal information," "pseudonymization," or "de-identified," rather than following uniform federal standards for de-identification, such as those promoted in NISTIR 8053 or ISO/IEC 27701.

States might also create conflicting exemptions for when data may be lawfully disclosed for public health purposes. Most models for comprehensive privacy legislation, including the California Consumer Privacy Act and leading proposals in the Senate Commerce Committee, contain exemptions for using data for public health research. If states were to differ in their requirements for such exemptions -- for example, if some states were to freely allow scientific research as a compatible use of data, other states were to require individual opt-in consent for research, and still other states were to require approval or oversight by ethical review boards -- this divergence could create significant hurdles for companies processing data across state lines. In particular, companies with very large volumes (e.g., from millions of individuals) of data that is not easily identifiable (e.g., linked to devices from unknown geographic locations) can be limited in their ability to disambiguate data in order to comply with varying state requirements.

Sources:

- NISTIR 8053, *De-Identification of Personal Information*, https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf.
- ISO/IEC 27701, https://www.iso.org/obp/ui/#iso:std:iso-iec:27701:ed-1:v1:en.

**Sen. Lee**

27. To date, what specific data (or types of data) are companies (or your company) currently collecting for COVID-19 related purposes? What specific data (or types of data) are governments and health officials seeking for COVID-19 related purposes?

**Answer to Question 27:**

Many companies process individualized location data tied to a device identifier or other pseudonymous identifier, rather than a name or identity. This kind of data is often (incorrectly) referred to within consumer data industries as "anonymous," but in fact is very challenging to de-identify when the data is 1) individualized and 2) collected over time. Individual behavior patterns (especially, home and work locations) are relatively easy to re-identify, and can be highly revealing of our behavior, beliefs, and character.

In most cases, however, individualized location data will not be necessary or useful to COVID-19 efforts, where aggregated data or device-to-device proximity information can be used instead. Aggregated location data, when it does not reveal device-level information about individual behavior, can almost certainly be used safely and effectively to support public health officials. As long as the underlying data is of sufficiently high quality, and measures are taken to address representativeness, bias, and other group-level risks, including re-identification within very small or rural communities, aggregation of location data represents a useful way to balance public health needs with fully protecting individual privacy.

In addition, mobile apps can support public health initiatives by providing data that is precise and accurate enough for effective person-to-person proximity-based contact tracing. Apps that have the potential to achieve these goals without sacrificing individual privacy are ones that are based

on user consent (voluntary); feature data minimization; decentralized device-to-device signaling; on-device processing; transparent source code; and technical and administrative safeguards to prevent abuse or mis-use.

At this time, there are several promising frameworks being developed globally, including the Pan-European Privacy-Preserving Contact Proximity Tracing (PEPP-PT), and particularly the Decentralized Privacy Preserving Proximity Tracing (DP-3T) Protocol developed under this initiative. This protocol relies on short-range Bluetooth technology and decentralized rotating identifiers. The Future of Privacy Forum supports this approach, which does not rely on bulk collection of precise location histories, whether from existing sources (cell phone carriers or technology providers) or from individuals directly. The changes to Bluetooth protocols recently announced by Apple and Google will also help to improve interoperability between iOS and Android devices and enable apps used by health authorities to comply with these frameworks.

While these developments are promising, it's important to note that contact tracing relies on the availability of testing. If the availability of testing is limited, the ability to rely on contract tracing is limited as well. Apps should also address risks related to potential mis-use (trolling or other false alerts) or abuse (spoofing). Public health authorities should play a role in approving the triggering of alerts for individuals to self-quarantine.

Sources:

- DP-3T repository on GitHub, available at https://github.com/DP-3T.
- Apple Newsroom, *Apple and Google partner on COVID-19 contact tracing technology* (April 10, 2020), available at https://www.apple.com/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/.
- Future of Privacy Forum, *Privacy & Pandemics: The Role of Mobile Apps (Chart)* (April 2020), available at https://fpf.org/wp-content/uploads/2020/04/Privacy-Pandemics-The-Role-of-Mobile-Apps.pdf.

28. Most tech companies currently claim that the data being gathered is being "anonymized" so that a specific person is not identifiable.
    - What specific steps are companies (or your company) taking to anonymize this data?
    - Certain data may not necessarily be considered personally identifiable, but with enough data points, you could identify a specific person. How can we ensure that data is truly anonymous and is not traceable back to an individual person?
    - Can effective contact tracing be conducted with "anonymized data"? Or will it require personally identifiable information?

**Answer to Question 28:**

## Ms. Stacey Gray, Senior Counsel, Future of Privacy Forum

Data rarely categorically falls into the category of either "personal" or "anonymous." Rather, most data exists on a spectrum of identifiability, from explicitly personal (e.g., a person's name and address) to expressly non-personal (e.g. data from jet engines or factory equipment sensors), with a wide range of data types in between with varying degrees of protection or anonymity, and varying types of administrative, technical, and legal controls.

In particular, it is very challenging to fully "anonymize" individualized device location data, even when it is tied to a device identifier or other pseudonymous identifier (rather than a name or identity). This kind of data is often (incorrectly) referred to within consumer data industries as "anonymous," but in fact, individuals can often be re-identified based on their home and work locations or by cross-referencing the dataset with an individual's known locations and times.

Nonetheless, there are ways to reduce the identifiability or risk of a location dataset. These can be applied alone or in combination, and include, for example: differential privacy (perturbing or adding noise to the dataset to reduce the risk of any specific individual being capable of identification); removing or obscuring data (such as from home and work locations); applying administrative controls (such as limiting access to the dataset and placing contractual limitations on its use); and aggregating the data so that it reveals only movements of groups of people of a certain size. Aggregated location data can almost certainly be used safely and effectively to support public health officials.

Effective contact tracing also does not necessarily require a person's entire location history. At this time, there are several promising frameworks being developed globally, including the Pan-European Privacy-Preserving Contact Proximity Tracing (PEPP-PT), and particularly the Decentralized Privacy Preserving Proximity Tracing (DP-3T) Protocol developed under this initiative. This protocol relies on short-range Bluetooth technology and decentralized rotating identifiers. The Future of Privacy Forum supports this approach, which does not rely on bulk collection of precise location histories, whether from existing sources (cell phone carriers or technology providers) or from individuals directly. The changes to Bluetooth protocols recently announced by Apple and Google will also help to improve interoperability between iOS and Android devices and enable apps used by health authorities to comply with these frameworks.

Sources:


- DP-3T repository on GitHub, available at https://github.com/DP-3T.
- Apple Newsroom, *Apple and Google partner on COVID-19 contact tracing technology* (April 10, 2020), available at https://www.apple.com/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/.
- Future of Privacy Forum, *Privacy & Pandemics: The Role of Mobile Apps (Chart)* (April 2020), available at https://fpf.org/wp-content/uploads/2020/04/Privacy-Pandemics-The-Role-of-Mobile-Apps.pdf.

29. Since the beginning of this COVID-19 crisis, has a federal agency, a state government, or local government requested a company or association to gather any specific consumer data?
    - To your knowledge, are there any current COVID-19 related data sharing agreements in place between governments and private sector organizations? To your knowledge, has any federal, state, or local law enforcement used private sector collected data to enforce any COVID-19 related government orders or requirements?

*(No reply.)*

30. Ms. Gray, throughout your testimony you emphasize that the commercial data collected should be limited to data requested by public health officials. What line(s) should be drawn as to what type(s) of data should not be obtainable by public health officials? What underlying principles should inform the drawing of those lines?

**Answer to Question 30:**

In principle, there is no data that solely due to its nature should be off limits. Rather, the processing of data must be lawful, transparent, necessary, proportional, and limited in ways that respect the rights and freedoms of individuals. For example, in medical contexts, even very highly sensitive data about medical conditions can be collected with consent and appropriate privacy and security protections.

At times, sensitive data related to location, health, and race may be necessary to understand issues of structural bias and discrimination, including whether people of a particular race are being hit particularly hard by the effects of the virus; or the extent to which individuals are receiving unequal or biased treatment or access to healthcare resources. What is essential in the context of COVID-19 is that any data collected or accessed is truly necessary to achieve a goal set by public health specialists and epidemiologists. No personal data should be collected or given access to beyond what is necessary. It is also essential that companies and government entities make clear, strong public commitments to use data only to achieve those goals and to retain data only as long as necessary to achieve them (retention limitation).

**Sen. Johnson**

31. How does the pandemic shift the landscape for crafting and passing legislation? What would you say are the top 3 principles for striking the right balance between using consumer data for good during the pandemic and ensuring consumer's privacy rights are not violated? What do you see as the #1 priority (as far as using big data for COVID response) in a federal privacy bill? How do we ensure that pandemic-related data privacy provisions sunset appropriately?

**Answer to Question 31:**

FPF supports the enactment of federal comprehensive privacy law that would strike a balance between protecting individual privacy rights and using data for socially beneficial purposes, such as advancing scientific research. The pandemic has illustrated the need for comprehensive privacy legislation to provide clear parameters for U.S. companies to react in an efficient, lawful, and principled manner in times of crisis. Jurisdictions that have comprehensive data protection laws, such as the European Union, have been able to move more quickly than the U.S. to guide appropriate practices.

In striking these balances, we believe the top priorities ought to include the principles of lawfulness (i.e., the existence of lawful grounds to process data, such as an individual's consent, the existence of a contract, or necessity to address a public emergency, such as a pandemic); data minimization (i.e., the collection and use of only the minimal necessary information that is proportional to achieving a clear, specific, lawful purpose); and purpose limitation (or protections against secondary uses without the application of robust de-identification methods or consent).

Without embedding these principles of lawfulness, data minimization, and purpose limitation into comprehensive privacy legislation, entities that handle personal information are likely to continue to struggle to gain or maintain the trust of individuals necessary to effectively deploy data-driven technologies which rely on widespread adoption. For example, voluntary mobile apps for contact tracing applications will only be effective if a sufficient percentage of the population adopts them as a voluntary measure of civic participation.

**Sen. Scott**

32. For months, Communist China lied about the Coronavirus data, the spread of the virus, and their response. They silenced critics and those trying to alert the Chinese people to this public health crisis. The lack of usable data coming out of Communist China cost lives and put the world behind on response efforts, including here in the United States.
    - As we work to keep American families healthy, how can we follow the lead of countries with low case counts, like South Korea, using technology and data collection, without infringing on our citizens' rights and privacy?

**Answer to Question 32:**

Aggregate location data, when it does not reveal device-level information about individual behavior, can almost certainly be used safely and effectively to support public health officials. As long as the underlying data is of sufficiently high quality, and measures are taken to address representativeness, bias, and other group-level risks, including re-identification within very small

or rural communities, aggregation of location data represents a useful way to balance public health needs with fully protecting individual privacy.

In contrast, it is very challenging to fully "anonymize" individualized device location data, even when it is tied to a device identifier or other pseudonymous identifier (rather than a name or identity). This kind of data is often (incorrectly) referred to within consumer data industries as "anonymous," but in fact, individuals can often be re-identified based on their home and work locations or by cross-referencing the dataset with an individual's known locations and times.

There are ways to reduce the identifiability or risk of a location dataset, including, for example: differential privacy (perturbing or adding noise to the dataset to reduce the risk of any specific individual being capable of identification); removing or obscuring data (such as from home and work locations); applying administrative controls (such as limiting access to the dataset and placing contractual limitations on its use); and aggregating the data so that it reveals only movements of groups of people of a certain size.

Importantly, many use cases, including effective contact tracing, do not require data about an individual's precise location history. At this time, there are several promising frameworks being developed globally, including the Pan-European Privacy-Preserving Contact Proximity Tracing (PEPP-PT), and particularly the Decentralized Privacy Preserving Proximity Tracing (DP-3T) Protocol developed under this initiative. This protocol relies on short-range Bluetooth technology and decentralized rotating identifiers. The Future of Privacy Forum supports this approach, which does not rely on bulk collection of precise location histories, whether from existing sources (cell phone carriers or technology providers) or from individuals directly. The changes to Bluetooth protocols recently announced by Apple and Google will also help to improve interoperability between iOS and Android devices and enable apps used by health authorities to comply with these frameworks.

While these developments are promising, it's important to note that contact tracing relies on the availability of testing. If the availability of testing is limited, the ability to rely on contract tracing is limited as well. Apps should also address risks related to potential mis-use (trolling or other false alerts) or abuse (spoofing). Public health authorities, rather than tech companies, must lead the way in helping shape the development of these apps. Healthcare professionals should also play a role in approving the triggering of alerts for individuals to self-quarantine.

Sources:


- DP-3T repository on GitHub, available at https://github.com/DP-3T.
- Apple Newsroom, *Apple and Google partner on COVID-19 contact tracing technology* (April 10, 2020), available at https://www.apple.com/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/.
- Future of Privacy Forum, *Privacy & Pandemics: The Role of Mobile Apps (Chart)* (April 2020), available at https://fpf.org/wp-content/uploads/2020/04/Privacy-Pandemics-The-Role-of-Mobile-Apps.pdf.

**Ranking Member Cantwell**

33. Science and technology will be critical drivers of our response to COVID-19, and we have seen many examples of data being used in positive ways – from the University of Washington's forecasts of hospital needs to Johns Hopkins' maps of disease spread. These are leading examples of how firms can innovate while protecting other equities, like privacy.
    - What recommendations do you have to encourage further innovation to fight the virus? How do we encourage technologists to help people transition to regular life while preparing for future pandemic incidents? What are the best practices you have seen in innovating in the fight against COVID-19 that support privacy rights?

**Answer to Question 33:**

At this time, there are several promising frameworks being developed globally, including the Pan-European Privacy-Preserving Contact Proximity Tracing (PEPP-PT), and particularly the Decentralized Privacy Preserving Proximity Tracing (DP-3T) Protocol developed under this initiative. This protocol relies on short-range Bluetooth technology and decentralized rotating identifiers. The Future of Privacy Forum supports this approach, which does not rely on bulk collection of precise location histories, whether from existing sources (cell phone carriers or technology providers) or from individuals directly. The changes to Bluetooth protocols recently announced by Apple and Google will also help to improve interoperability between iOS and Android devices and enable apps used by health authorities to comply with these frameworks.

While these developments are promising, it's important to note that contact tracing relies on the availability of testing. If the availability of testing is limited, the ability to rely on contract tracing is limited as well. Apps should also address risks related to potential mis-use (trolling or other false alerts) or abuse (spoofing). Public health authorities, rather than tech companies, must lead the way in helping shape the development of these apps. Healthcare professionals should also play a role in approving the triggering of alerts for individuals to self-quarantine.

Similarly, addressing the pandemic will require marshalling the best evidence to make data driven decisions. While technology and the responsible use of consumer data no doubt has a meaningful role to play in addressing this public health emergency, in many cases there may be much more important priorities, such as ensuring adequate and fair distribution of medical resources, tests, and personal protective equipment (PPE).

Innovation and enabling our communities and economy to return to the pre-pandemic regular life requires us to look at the larger picture. For example, along with the privacy risks, there are other considerations for technologists, innovators, and policymakers, such as whether contact tracing apps will reinforce existing social biases, stigmatizing already marginalized communities. The over-reliance on apps may cause people to over-trust the app's ability to keep them safe, which

may increase social contact and undo the efforts made in flattening the curve.

We also encourage innovations for testing, tracing, and forecasting future movements that do not require individuals or companies to provide more data than is necessary. For example, allowing individuals to use web-based platforms with data local to their own computers is an example of a privacy protecting best practice. We are also supportive of recent efforts, such as by Apple and Google, to publicly commit to clear data retention practices and to end COVID-19 tracking efforts after the pandemic is over.

Sources:

- Jason Millar, Policy Options, *Five ways a COVID-19 contact-tracing app could make things worse* (April 15, 2020), available at https://policyoptions.irpp.org/magazines/april-2020/five-ways-a-covid-19-contact-tracing-app-could-make-things-worse/.
- DP-3T repository on GitHub, available at https://github.com/DP-3T.
- CNET *Apple, Google to Terminate COVID-19 Tracking Tools Once Pandemic Ends*, https://www.cnet.com/google-amp/news/apple-and-google-say-they-will-shut-down-covid-19-tracking-tools-once-pandemic-ends/

34. Frequently, data used to combat COVID-19 is described as "anonymized" or "aggregated" or "de-identified," and these terms are meant to convey that data will be used or shared in a privacy-protective manner.
    - How do you define "anonymized," "aggregated," and "de-identified" data? What are the best practices to ensure that the data remains anonymous?

**Answer to Question 34:**

Data rarely categorically falls into the category of either "personal" or "anonymous." Rather, most data exists on a spectrum of identifiability, from explicitly personal (e.g., a person's name and address) to expressly non-personal (e.g. data from jet engines or factory equipment sensors), with a wide range of data types in between with varying degrees of protection or anonymity, and varying types of administrative, technical, and legal controls.

In particular, it is very challenging to fully "anonymize" individualized device location data, even when it is tied to a device identifier or other pseudonymous identifier (rather than a name or identity). This kind of data is often (incorrectly) referred to within consumer data industries as "anonymous," but in fact, individuals can often be re-identified based on their home and work locations or by cross-referencing the dataset with an individual's known locations and times.

Nonetheless, there are ways to reduce the identifiability or risk of a location dataset. These can be applied alone or in combination, and include, for example: differential privacy (perturbing or adding noise to the dataset to reduce the risk of any specific individual being capable of identification); removing or obscuring data (such as from home and work locations); applying administrative controls (such as limiting access to the dataset and placing contractual limitations

on its use); and aggregating the data so that it reveals only movements of groups of people of a certain size. Aggregated location data can almost certainly be used safely and effectively to support public health officials.

Sources:

- Future of Privacy Forum, *Visual Guide to Practical De-identification*, available at https://fpf.org/wp-content/uploads/2017/06/FPF_Visual-Guide-to-Practical-Data-DeID.pdf
- NIST Special Publication 800-188 (2nd Draft), *De-Identifying Government Datasets* (December 2016), available at https://csrc.nist.gov/csrc/media/publications/sp/800-188/draft/documents/sp800_188_draft2.pdf.
- US Census Bureau, *Disclosure Avoidance and the 2020 Census* (last revised March 27, 2020), available at https://www.census.gov/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html (describing differential privacy methods used to protect census data).
- National Academies of Science, *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy* (2017), available at http://mitsloan.mit.edu/shared/ods/documents/?DocumentID=4438; and *Federal Statistics, Multiple Data Sources, and Privacy Protection* (2017), available at https://www.ncbi.nlm.nih.gov/books/NBK475779/pdf/Bookshelf_NBK475779.pdf (recommending use of tiered access models that would increase restrictions on data access depending on the sensitivity and identifiability of the data accessed).

**Sen. Blumenthal**

35. Privacy for America, a coalition of advertising associations including IAB and NAI, have proposed a federal privacy framework that is focused on a set of prohibited data uses, transparency measures, and a limited subset of data rights found under the GDPR and CCPA. However, the Privacy for America framework also provides wide discretion for companies to use particular types of data or engage in particular activities without consent, as well as a self-regulatory safe harbor and broad state preemption.
    - Would the Privacy for America framework provide Americans the full set of consumer rights and protections necessary to guarantee the privacy, security, and equitable use of their personal data, and the enforcement regime necessary to deter and punish the misuse of their information? Please elaborate on why or why not.

*(No reply.)*

## Ms. Stacey Gray, Senior Counsel, Future of Privacy Forum

**Sen. Schatz**

36. Americans are concerned about the use of their data to track the COVID-19 pandemic, and today's hearing highlights yet another reason why we need a robust federal privacy law. Although many entities are making available anonymized and aggregated data to help officials study and forecast the pandemic, current law is limited in its ability to ensure that companies are protecting their customers' data as they use this data to innovate new methods to track the virus's course.

The Data Care Act, which was reintroduced in December, imposes a requirement on companies that they not use customers' data in a manner that is harmful to those customers. This will ensure that companies carefully balance the reasonable expectations of their customers about the use of their data with other important interests.

How should we balance important prospects for using consumer data to address pressing public health concerns with protecting the privacy of individuals?

**Answer to Question 36:**

As well as imposing a requirement on companies that they not use customers' data in a manner that is harmful to the end user, the Data Care Act establishes duties of care, loyalty, and confidentiality for covered entities, and disallows uses of data that would be "unexpected and highly offensive to a reasonable end user." These are important elements of any federal comprehensive privacy law. In addition, we recommend that any leading federal privacy legislation include principles and legal protections for data minimization and regulatory flexibility for socially beneficial research, including scientific research and to support responsible data sharing for public health emergencies.

In balancing the need for consumer data to address public health concerns, beneficial research uses could also be supported through oversight by ethical review boards, or limiting access to data enclaves within the federal government or data trusts overseen by trusted third parties. In addition, the principles of purpose specification and use limitation (i.e., purpose limitation) should guide companies and government entities to the extent they choose to either purchase commercial location data or make use of data shared voluntarily by companies and individuals to support COVID-19 response efforts. In practice, this means that the purpose of using the specific data at issue should be articulated in clear, specific, and granular terms before the data is collected. It also means having specific, public-facing plans for who will have access to the data, and whether and how it will be retained for any future uses.

If some or all data is not deleted or destroyed after its use for COVID-19 efforts, companies and government entities involved should clearly articulate the reasons for which it is retained, and how the data retained serves those needs while minimizing risks to privacy. For example, it may be possible to allow individuals to opt in to the use of their individualized data for future scientific research. Alternatively, underlying individual data could be deleted, while retaining

aggregate or high-level statistical data for research on infectious diseases, or to help prepare response plans for future pandemics.

**Sen. Markey**

37. After reports emerged that federal government officials are engaging with technology companies to consider using location data collected from Americans' smartphones to track the coronavirus, I wrote a letter to the White House Office of Science and Technology Policy demanding answers about any plans it has to leverage information about individuals' location. Since then, new reports have emerged indicating that the Trump Administration is considering creating a vast surveillance network that may include health data relevant to the coronavirus pandemic. In response, I wrote to the White House with questions about this proposal and its potential to infringe on individuals' privacy. I urge this Committee to include meaningful transparency requirements for federal projects or partnerships that use individuals' location data or health data in any upcoming coronavirus legislation. Such requirements should apply to both the federal government and whatever private companies participate. And they should apply to the use of anonymized or aggregated data in addition to data that is linked to individuals.
    - Ms. Gray, do you agree that at the very least the public and experts should know what types of data the federal government is using —either on its own or in partnership with private companies— during this crisis, what form that data is in, and how the federal government is using that data?

**Answer to Question 37:**

Yes. Legal protections for lawfulness and transparency are critical to government collection and use of commercially available consumer data. While the Privacy Act of 1974 has long required federal agencies to comply with privacy and security standards for their own records on US persons, we have observed a concerning trend in recent years of government entities purchasing, or requiring as a condition of local permits, "anonymous" precise location information collected from consumer devices.

Individualized device location data, when tied to a device identifier or other pseudonymous identifier (rather than a name or identity) is often incorrectly referred to as "anonymous," but in fact, can often still easily be used to identify or reveal information about individuals and groups. When this data is collected for one set of consumer purposes (for example, to enable weather apps or use a CityBike), and ends up being used by the government for very different purposes (e.g., creating comprehensive historical location profiles of individuals), it violates fundamental rights and data protection principles of data minimization and purpose limitation.

In contrast, aggregated location data or data that has been properly de-identified using robust de-identification methods may sometimes be used by government agencies in ways that are appropriate and do not compromise individual rights, such as by the US Census Bureau. In order to lawfully distinguish between these kinds of uses, transparency is absolutely essential to shine light on government collection of data and open it up to scrutiny from public data scientists and de-identification experts.

Sources:

- Byron Tau and Michelle Hackman, *Federal Agencies Use Cellphone Location Data for Immigration Enforcement*, Wall Street Journal (February 7, 2020), available at https://www.wsj.com/articles/federal-agencies-use-cellphone-location-data-for-immigration-enforcement-11581078600.
- Over 80 U.S. cities are using the Mobility Data Specification (MDS), which is a set of Application Programming Interfaces (APIs) focused on dockless e-scooters, bicycles, mopeds and carshare data, some of which includes near real-time location data. Each city has its own rules and guidelines for data collection and use. See Open Mobility Foundation, *Mobility Data Specification*, https://github.com/openmobilityfoundation/mobility-data-specification.
- US Census Bureau, *Getting Started with OnTheMap*, available at https://lehd.ces.census.gov/doc/help/onthemap/GettingStartedwithOnTheMap.pdf (explaining a privacy-preserving online mapping application that shows where people work and where workers live).

**Sen. Peters**

38. The one thing that has been absent from this discussion is that neither the federal government nor the private sector have adequately anticipated nor met the demands for personal protective equipment. Even basic things like masks and gloves have been inaccessible. Our nation has unparalleled resources in the supply chain and manufacturing space.
    - From a data perspective—where have failures been and what improvements do you recommend?

**Answer to Question 38:**

The pandemic has brought to light many existing challenges to data-driven allocation of healthcare resources. As the National Academy of Medicine highlighted as recently as 2019, the most important challenges involve access to quality data, common and portable data structures, interoperability of systems, and ensuring replication of key studies that promise to change healthcare delivery.

For example, the lack of easily discovered, quality, and well organized data is a significant barrier to generating data-driven insights, whether in healthcare, government administration, or

machine learning and artificial intelligence. Methods to overcome these challenges are often summarized in the acronym FAIR-- Findable, Accessible, Interoperable, and Reusable-- which describe ideals for datasets and data repositories. Establishing common data definitions and models would also reduce time necessary to build systems that bring together multiple data sources. Enhancing interoperability is essential. As the National Academy of Medicine described it: "True interoperability is the ability to seamlessly and automatically deliver data when and where it is needed under a trusted network without political, technical, or financial blocking" ("Procuring Interoperability: Achieving High-Quality, Connected, and Person-Centered Care", National Academy of Medicine 2019, pg. xix).

Policymakers can also improve the quality of evidence based medicine by explicitly supporting reproducibility of studies and funding replication analyses to ensure that lessons learned from this pandemic response become validated and verified knowledge for the next public health emergency. In their consensus report on "Reproducibility and Replicability in Science," the National Academies of Sciences, Engineering, and Medicine point out that replication of key studies is essential to reducing confusion and concern about scientific results. Improving the credibility of the science behind healthcare, regardless of its position in a pandemic response, stands to strengthen trust and confidence in healthcare and governing systems.

Sources:

- National Academy of Medicine, *Accelerating Medical Evidence Generation and Use* (2017), available at https://nam.edu/wp-content/uploads/2017/06/Accelerating-Medical-Evidence-Final-Book-011918.pdf.
- National Academy of Medicine, *Procuring Interoperability: Achieving High-Quality, Connected, and Person-Centered Care* (2018), available at https://nam.edu/wp-content/uploads/2019/08/Interop_508.pdf.
- National Academy of Sciences, *Reproducibility and Replicability in Science* (2019), available at https://www.nap.edu/catalog/25303/reproducibility-and-replicability-in-science.

39. Despite many structural challenges, Taiwan has fared better than many countries in dealing with the COVID-19 pandemic. Stanford Medical School documented 124 distinct interventions that Taiwan implemented with remarkable speed including community initiatives, hackathons, etc. Their "Face Mask Map" a collaboration initiated by an entrepreneur working with government helped prevent the panicked buying of facemasks, which hindered Taiwan's response to SARS by showing where masks were available and providing information for trades and donations to those who most needed them, which helped prevent the rise of a black market.
    - What specific initiatives like this should we be implementing here?

*(No reply.)*

## Ms. Stacey Gray, Senior Counsel, Future of Privacy Forum

**Sen. Baldwin**

40. Emerging reports from many localities demonstrate that COVID-19 is having a disproportionate impact on African Americans and communities of color. For example, in my home state of Wisconsin, Milwaukee County reports that approximately 70% of those killed by coronavirus are African American, despite that community making up only 26% of the county's population.

We know this about Milwaukee County because the local government is proactive about collecting and reporting data on race and ethnicity. Reporting indicates that this disproportionate impact exists in places with significant African American communities, including Chicago, New Orleans, and Detroit. But a lack of consistent, quality data nationwide means we do not yet know just how sizable this disparity is, and what we can do about it.

While I am encouraged that we are drawing on the massive amount of data about Americans held by the private sector to support the COVID-19 response, I worry that it may not include and represent all communities equally. For example, if we use mobility data from mobile phones or particular apps to inform our understanding of adherence to social distancing requirements, I am concerned how it might affect the usefulness of the dataset if members of certain minority communities are less likely to own such a device or utilize such an app.

- For the members of our panel: how do you think "big data" can support efforts to strengthen our public health knowledge around COVID-19 and race, and how can we ensure that the methods and models through which "big data" supports our understanding of the epidemic take into account differences among communities?

**Answer to Question 40:**

Racial disparities and other forms of bias have to be considered in the context of evaluating the quality and representativeness of underlying data sources. For example, many commercial sources of precise location information may not be sufficiently high quality or representative of all populations to justify its use for public health objectives. There are well-known differences in mobile app usage between demographics and age groups. Similarly, data from high-end consumer electronics (such as smart thermometers and other devices) should be evaluated for whether it represents only an affluent subset of users.

Similarly, we believe that identifying and addressing bias and structural discrimiation is one of the most important roles for data-driven research. For example, we are alarmed by the early reports of COVID-19-related death disparities in African American communities. Understanding how and why these disparities exist is only possible with the collection of sensitive data combined with health information reflecting racial demographics. For example, voluntary contact tracing apps must be adopted by sufficient numbers of app users within high-risk populations, including those who cannot afford the latest mobile technology. To the extent

possible, mobile apps should be designed so they are not unduly limited to users of only the newest or more sophisticated devices that can accomodate the recent updates to iOS and Android operating systems.

Sources:

- Future of Privacy Forum and Anti-Defamation League, *Big Data: A Tool for Fighting Discrimination and Empowering Groups* (September 2014), available at https://fpf.org/wp-content/uploads/Big-Data-A-Tool-for-Fighting-Discrimination-and-Empowering-Groups-Report1.pdf.
- Federal Trade Commission, *Big Data: A Tool for Inclusion or Exclusion?* (January 2016), available at https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.

41. I am also concerned about the impact of "big data" informing our COVID-19 response on rural communities. Again, I worry that some of these data sources may not be well-utilized in rural America – where connectivity is still a significant challenge – and thus may not reflect the reality of the pandemic in those communities. But I recognize that this information is vital to developing better predictive models that can inform our current response to COVID-19 and help us prepare for the future.
    - For the members of our panel: how does "big data" ensure that the different experiences of rural, suburban and urban communities are taken into account when informing models that may guide the COVID-19 response?

**Answer to Question 41:**

While there are many potential benefits to using data to address connectivity and bring resources to rural or other underserved populations, the Federal Trade Commission has observed in recent years a number of risks to using Big Data without adequately considering its accuracy, completeness, and representativeness. These concerns are particularly prevalent during a pandemic crisis, and include: uncorrected hidden biases in the underlying consumer data; inaccurate predictions about where healthcare resources are needed; and failure to bring sufficient medical or other healthcare resources to rural communities.

Source:

- Federal Trade Commission, *Big Data: A Tool for Inclusion or Exclusion?* (January 2016), available at https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.

42. It is important that public health, and local public health departments in particular, have the data they need to map and anticipate hotspots for infectious disease outbreaks such as

COVID-19 or overdose patterns in a community, including data that may be generated by the private sector. It is also important that local health departments have the capability to leverage this information together with that available through traditional public health surveillance efforts.

- For the members of our panel: how can the private sector coordinate data efforts with public health and ensure that local health departments have the necessary capabilities to make full use of these efforts?

**Answer to Question 42:**

It is critical that the private sector follow, rather than lead, and tailor the availability of data and new platforms and services (such as mobile apps) to the needs of public health experts and local health departments. For example, mobile apps developed in the private sector should avoid risks of abuse or mis-use by requiring confirmation of a COVID-19 diagnosis from a medical professional before enabling alerts to nearby users based on their proximity that recommend self-quarantining.

Mobile apps, if adopted voluntarily by a sufficient percentage of a population, can support public health initiatives and local health departments by providing data that is precise and accurate enough for effective person-to-person contact tracing. Apps that have the potential to achieve these goals without sacrificing individual privacy typically: are based on user consent (voluntary); feature data minimization; and use decentralized device-to-device signaling, on-device processing, transparent source code, and technical and administrative safeguards to prevent abuse or mis-use. Accessing an individual's detailed location history may not be necessary if contact tracing and alerts can be adequately enabled through proximity-based solutions, such as Bluetooth.

It's important to note that effective use of data for contact tracing relies on the availability of testing and adequate healthcare resources. If the availability of testing is limited, the ability to rely on contract tracing is limited as well. Apps should also address risks related to potential mis-use (trolling or other false alerts) or abuse (spoofing). To emphasize - public health authorities, rather than tech companies, must lead the way in helping shape the development of these apps. Healthcare professionals should also play a role in approving the triggering of alerts for individuals to self-quarantine.

43. In speaking with experts in Wisconsin working on developing and refining predictive models around COVID-19, I heard that while there is a significant number of both public sector and private sector data sources to inform models, the data is not consistently easy to obtain and incorporate. As we rely on real-time models to inform the COVID-19 effort, as well as look to prepare for future infectious disease outbreaks, it is important that data-sharing be as seamless as possible.
    - For the members of our panel: what are ways we can strengthen the data-sharing infrastructure for government, public health, academic and private sector sources?

## Ms. Stacey Gray, Senior Counsel, Future of Privacy Forum

**Answer to Question 43:**

Policymakers should follow the recommendations of the National Academy of Medicine (NAM) in their recent reports describing the importance of interoperable systems that create bridges between sources of quality, well-organized data to reduce barriers to data-driven healthcare insights. As described by the NAM, this will require reducing political, financial, and technical barriers to sharing data at the micro level of healthcare institutions, the meso level of healthcare systems, and the macro level of the full healthcare economy in the United States. System interoperability is also key to making data sharing as seamless as possible.

In the long-term, Congress could also strengthen data-sharing infrastructure for government, public health, academic, and private sector sources through federal policies that implement these suggestions and ensure adequate funding for research partnerships. However, quality and safety checkpoints along conveyors moving data across interoperable or interfacing systems are needed to uphold the highest standards of data quality, utility, and fairness for data subjects.

Source:

- National Academy of Medicine, *Procuring Interoperability: Achieving High-Quality, Connected, and Person-Centered Care* (2018), available at https://nam.edu/wp-content/uploads/2019/08/Interop_508.pdf.

**Sen. Sinema**

44. This virus affects communities across our country. If a small community reports a single positive case, it is important to both inform the community and protect the privacy of the infected individual. Technology can play a role in helping us map the virus, but it is more difficult to sufficiently anonymize personal health data in smaller populations.
    - How do we ensure public health officials in underserved and unserved communities, especially in rural communities and Indian Country, are able to provide first responders and EMT dispatch with valuable information about the potential for exposure when firefighters or local law enforcement are responding to a call, while maintaining patient privacy?

**Answer to Question 44:**

Data privacy and anonymization, such as through data aggregation, certainly poses challenges for smaller populations or populations secluded to or dispersed across wide and less-populated geographic areas (like Indian reservations and rural areas, respectively). Digital connectedness is also a challenge in such areas, due to fewer cell towers and other infrastructure for precise geolocation measurement, which creates additional challenges for leveraging communication technology for virus mapping and predicting resource deployment needs (like first responders). Therefore, analog communication methods, like land-line telephone calls, to quickly communicate with local public health authorities can be particularly useful in such cases.

In these communities, additional data may be necessary to effectively address risks of bias and inadequate data in rural communities. As the Federal Trade Commission has observed, there are risks to using consumer data when it does not adequately represent rural communities, considering its accuracy, completeness, and representativeness. These concerns are particularly prevalent during a pandemic crisis, and include: uncorrected hidden biases in the underlying consumer data; inaccurate predictions about where healthcare resources are needed; and failure to bring sufficient medical or other healthcare resources to rural communities.

Source:

- Federal Trade Commission, *Big Data: A Tool for Inclusion or Exclusion?* (January 2016), available at https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.

45. Some states, including Arizona have limited testing capabilities and therefore limited testing. It is also widely reported that tests around the world have produced inaccurate results. How can we mitigate against inaccurate assumptions related to disease trends in situations in which we have limited or inaccurate data?

**Answer to Question 45:**

Commercial data can support the needs of public health officials only if it is accurate and useful, a question which requires considering bias in underlying data as well as potential biased outcomes that can result from inadequate or limited data. It is important to acknowledge that if there is limited testing available, the usefulness of many technological solutions (such as contact tracing apps) will be correspondingly limited.

These concerns are particularly prevalent during a pandemic crisis, and include: inaccurate predictions about where healthcare resources are needed; and failure to bring sufficient medical or other healthcare resources to rural communities. These concerns can sometimes be mitigated by cross-referencing self-reported data or commercial data with data collected by local public health authorities before making any decisions based on that data and to identify possible discrepancies in data collection methods across data sources.

In addition, public health experts should be skeptical of the potential lack of representativeness and potential biases inherent in many commercially available location datasets. For example, although mobile apps have the potential to generate highly accurate location and proximity information (due to the number of hardware sensors in a typical smartphone), in practice, many data brokers and third party intermediaries in the mobile advertising industry receive data second-hand and may not have robust quality assurance mechanisms to avoid processing low-quality data. In light of differences in mobile app usage between demographics and age groups, there is also a risk that such data underrepresents low-income populations, the elderly, or anyone who does not carry a cell phone or use particular apps.

46. Many point to travel as a key factor in the spread of COVID-19. Contact tracing for travelers, specifically by plane, is a mechanism that can slow the spread of the virus. The data collected (full name, address while in U.S., email address, and two phone numbers) enables the government to contact individuals who may have come into contact with an individual who has tested positive. Once contact is established, individuals can start self-quarantining.
    - What is the best way to balance the need for this information to slow the spread of the virus and privacy rights?

**Answer to Question 46:**

Mobile apps can support public health initiatives by providing data that is precise and accurate enough for effective person-to-person contact tracing, including for travelers. Apps that have the potential to achieve these goals without sacrificing individual privacy are ones that are based on user consent (voluntary); feature data minimization; decentralized device-to-device signaling; on-device processing; transparent source code; and technical and administrative safeguards to prevent abuse or mis-use.

At this time, there are several promising frameworks being developed globally, including the Pan-European Privacy-Preserving Contact Proximity Tracing (PEPP-PT), and particularly the Decentralized Privacy Preserving Proximity Tracing (DP-3T) Protocol developed under this initiative. This protocol relies on short-range Bluetooth technology and decentralized rotating identifiers. The Future of Privacy Forum supports this approach, which does not rely on bulk collection of precise location histories, whether from existing sources (cell phone carriers or technology providers) or from individuals directly. The changes to Bluetooth protocols recently announced by Apple and Google will also help to improve interoperability between iOS and Android devices and enable apps used by health authorities to comply with these frameworks.

While these developments are promising, it's important to note that contact tracing relies on the availability of testing. If the availability of testing is limited, the ability to rely on contract tracing is limited as well. Apps should also address risks related to potential mis-use (trolling or other false alerts) or abuse (spoofing). Public health authorities should play a role in approving the triggering of alerts for individuals to self-quarantine.

- DP-3T repository on GitHub, available at https://github.com/DP-3T.
- Apple Newsroom, *Apple and Google partner on COVID-19 contact tracing technology* (April 10, 2020), available at https://www.apple.com/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/.
- TraceTogether app, available at https://apps.apple.com/us/app/tracetogether/id1498276074.
- TraceTogether privacy policy, available at https://www.tracetogether.gov.sg/common/privacystatement.

- HaMagen app, available at https://play.google.com/store/apps/details?id=com.hamagen.
- HaMagen privacy policy, available at https://govextra.gov.il/ministry-of-health/hamagen-app/terms-and-conditions-of-use-en/.
- Future of Privacy Forum, *Privacy & Pandemics: The Role of Mobile Apps (Chart)* (April 2020), available at https://fpf.org/wp-content/uploads/2020/04/Privacy-Pandemics-The-Role-of-Mobile-Apps.pdf.

47. How can big data help resolve challenges within the manufacturing supply chain to spur increased production and distribution of needed testing, personal protective equipment, and other resources to address this pandemic?

*(No reply.)*

48. This pandemic has caused serious economic harm. Businesses of all sizes and their employees suffer as sales drastically fall or disappear altogether. State, tribal and local governments are under enormous strain as response costs increase and revenues drop.
    - How can big data assist in the better creation and execution of economic assistance programs like the Paycheck Protection Program, Treasury's lending facilities, business interruption or pandemic risk insurance, and state, tribal and local stabilization funds?

*(No reply.)*

**Sen. Rosen**

49. Germany's national disease control center recently asked their citizens to donate data collected by their fitness tracker. This voluntary initiative has consumers download an app on their phones and contribute health information such as pulse rates and temperature that is collected by fitness tracking devices anonymously. Using machine learning, epidemiologists can analyze this data to better understand the spread of the coronavirus across the country and detect previously unknown clusters.
    - What are the advantages and pitfalls in using voluntarily donated data to improve responses during a pandemic?
    - How can we use donated data to support our response to this pandemic and future similar public health issues?
    - What privacy guardrails are needed to ensure that this data is collected and analyzed safely and anonymously?
    - What are the gaps we need to consider when analyzing such data?

## Ms. Stacey Gray, Senior Counsel, Future of Privacy Forum

**Answer to Question 49:**

Regarding the "data donation" app created by the Robert Koch Institute (RKI) in Germany, it is relevant to mention that the Federal Data Protection Commissioner of Germany was involved in clearing the use of the app and published a Statement on April 7. Before the launch of the app, the Commissioner required RKI to clearly inform citizens about the data the app is collecting and for what purpose; to specify how long the data will be stored; and to re-evaluate the app on a regular basis to determine whether it is effective; if it is not effective, to end the processing of data.

The Federal Data Protection Commissioner also advised the RKI against labelling it as a "data donation" app, highlighting that even if individuals agree to participate and voluntarily transmit their data to the RKI, they do not relinquish their rights over their personal data and can revoke consent at any time. In response, the RKI guaranteed that following a revocation of consent, all collected data will indeed be deleted.

With this background in mind, to reply to your specific questions:

- The benefits of using voluntarily donated data to improve responses during a pandemic should be determined on the basis of public health authorities, epidemiologists and other relevant experts' advice. Such benefits might include facilitating large-scale research about how the pandemic is spreading, or how different populations are being affected. The answer will largely depend on what type of solution is being created and for what purposes. This is why it is important for the efficiency of the voluntary solution proposed to be reviewed periodically and for the data collection to end if efficiency is not proved.
- FPF agrees that it is not ideal to refer to this kind of data sharing as "donated data," considering that personal data is not transferable property over which individuals should be asked to permanently renounce their rights.
- Privacy guardrails can include: user consent; data minimization; decentralized device-to-device signaling; on-device processing; transparent source code; and technical and administrative safeguards to prevent abuse or mis-use. Rather than relying on location data, they could rely on short-range Bluetooth technology and decentralized rotating identifiers. The Future of Privacy Forum supports this approach, which does not rely on bulk collection of precise location histories, whether from existing sources (cell phone carriers or technology providers) or from individuals directly. The changes to Bluetooth protocols recently announced by Apple and Google will also help to improve interoperability between iOS and Android devices and enable apps used by health authorities to comply with these frameworks. Generally speaking, we must also consider accuracy of data, differences in digital literacy, differences in technology adoption between younger and older populations and other factors that may create bias in the sought results.

Source:

- *Statement of the BfDI [Federal Commissioner for Data Protection and Freedom of Information] on the Coronavirus Data App*, (April 7, 2020) (in German), available at https://www.bfdi.bund.de/SiteGlobals/Modules/Buehne/DE/Startseite/Kurzmeldung_Link/HP_Text_Kurzmeldung.html.

50. The National Science Foundation (NSF) is the only federal agency whose mission includes supporting all fields of fundamental science and engineering. The research and educational programs backed by NSF are integral to the continued success of our country's innovation, supporting scientific discoveries that have led to new industries, products, and services. Since 2012, NSF has funded research on the emerging field of data science through its BIG DATA program. Now, NSF's larger program – "Harnessing the Data Revolution" – will support research, educational pathways, and advanced cyberinfrastructure in the field of data science.
    - Given NSF's leadership in data science research and development, what role do you think NSF can play in leading public-private partnerships for increased research on big data that could help address the COVID-19 crisis or future pandemics?

**Answer to Question 50:**

Leadership of the National Science Foundation (NSF) is essential to US leadership in health and science around the world. The NSF can lead the way in preparedness for future pandemics by using its vast resources to improve the quality of data necessary to power healthcare, public health decision-making, and public policy choices. Specific endeavors by the NSF for future pandemic preparedness include: encouraging interoperability of the many data sources that inform data driven healthcare decision-making; creating FAIR (findable, accessible, interoperable, and reusable) data repositories; encouraging a next generation of data scientists and life-long education in data literacy; facilitating multi-institution sharing of data science education and research expertise; and accelerating programs for ethical sharing of data across industry-academic collaboratives.

The Future of Privacy Forum has first-hand experience with NSF's leadership in developing public-private partnerships for increased research on big data and accelerating practical, real-world applications. With NSF's support, FPF established the Privacy Research and Data Responsibility Research Coordination Network (RCN) to foster industry-academic collaboration on priority research issues and inform the public debate on data privacy. The RCN was organized and draws upon FPF's relationships with industry chief privacy officers, academic researchers, and government officials and promotes discussion of issues under the National Privacy Research Strategy (NPRS). A second Applied Privacy Research Coordination Network creates "research showcases" for industry and matches privacy scholars directly with industry leaders to transition academic privacy research to commercial practice.

NSF's Convergence Accelerator (C-Accel) is designed to accelerate data-driven research in areas of national importance through partnerships between industry, academics, nonprofits, and

government entities. The C-Accel program has supported FPF's efforts to define the future of privacy technology that will impact how society balances the need for personal health data during the COVID-19 crisis or future pandemics without sacrificing privacy and individual rights. In association with this work, FPF has launched the Privacy Tech Alliance, a collaborative global community bringing together industry, researchers and other stakeholders to define and advance the market for privacy enhancing technologies. Other NSF-supported FPF projects have brought together researchers, industry, civil society and government to discuss ethical considerations for big data research, and created a Civic Data Privacy Leaders Network to help city and municipal actors better understand, communicate and collaboratively address data privacy issues and principles.

A common thread of all of these projects is the goal of purposefully integrating knowledge and expertise from multiple disciplines and sectors to achieve real-world impact. NSF's leadership and existing role in promoting cross-disciplinary public-private partnerships will continue to have impact for COVID-19 and future public health initiatives.

Sources:


- Future of Privacy Forum, Future of Privacy Forum, Privacy and Data Responsibility Research Coordination Network (Research Coordination Network (RCN), available at https://rcn.fpf.org/.
- Future of Privacy Forum, *FPF Research Coordination Network Helps Academic Stars Connect With Private Sector Privacy Pros at IAPP* (May 10, 2019), available at Future of Privacy Forum, FPF Research Coordination Network Helps Academic Stars Connect With Private Sector Privacy Pros at IAPP (May 10, 2019), available at https://fpf.org/2019/05/10/fpf-research-coordination-network-helps-academic-stars-connect-with-private-sector-privacy-pros-at-iapp/.
- Jules Polonetsky and Jeremy Greenberg, *NSF Convergence Accelerator: The Future of Privacy Technology (C-Accel 1939288)*, available at https://fpf.org/wp-content/uploads/2020/03/NSF_FPF-REPORT_C-Accel1939288_Public.pdf.
- Future of Privacy Forum, *Privacy Tech Alliance*, available at https://fpf.org/privacy-tech-alliance/.
- Future of Privacy Forum, *Municipal Leaders Joining Network to Advance Civic Data Privacy* (March 28, 2019), https://fpf.org/2019/03/28/municipal-leaders-joining-network-to-advance-civic-data-privacy/.