Questions Submitted by Members of the Senate Committee on Commerce, Science, and Transportation
Enlisting Big Data in the Fight Against Coronavirus
April 9, 2020

Answers of Graham Dufault, Senior Director for Public Policy, ACT | The App Association

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?

Massachusetts Institute of Technology (MIT) developed a Bluetooth-enabled system that helps conduct contact tracing in a privacy protective manner. Contact tracing in particular presents some privacy challenges because it involves associating a positive COVID-19 diagnosis with a specific device. But MIT's system would avoid some of the privacy risks by anonymously associating a positive COVID-19 diagnosis to certain Bluetooth identifiers (or "chirps"), which make it more difficult for a human to associate the diagnosis with the person who owns the device. In practice, the MIT method would have participants' smartphones send out periodic chirps through Bluetooth. If a chirp determined that the participant's smartphone was close to someone with a positive COVID-19 diagnosis, then it would be known that that person was exposed to infection. But the system itself would not do anything more than match the positive COVID-19 diagnosis to the anonymous string of numbers associated with an individual's Bluetooth connection, so theoretically it would protect participants' privacy. Examples like this generally do a better job of preventing unforeseen privacy or security risks because they are designed not to collect unnecessary data about people.

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 combatting COVID-19 can also respect individuals' privacy?

The Bluetooth-enabled example described above illustrates that minimal data can be leveraged to enable targeted and effective policy responses. However, necessary ingredients to this solution likely include platform-level participation and sensible limits on government access and control over the data itself. Privacy experts expressed legitimate and well-placed concerns over whether government agencies may access databases of personally identifiable information, including precise geolocation data, in order to carry out unrelated or adjacent policy or enforcement goals. The possible availability of geolocation datasets, accumulated mainly through smart device usage, in readable format is perhaps a tempting notion for government agencies. But the law rightfully requires such data to be used in limited ways, consistent with the expectations of those to whom it pertains (with some exceptions). Emergency situations like the COVID-19 pandemic require swift and flexible responses, but unlocking sensitive and revealing data about American citizens for ill-defined governmental uses would be a virtually intractable mistake.

I also noted in my testimony that differential privacy and federated learning are two techniques private sector companies have developed to harness the benefits of large datasets while minimizing risks to privacy. To the extent that location and other personal data are needed to develop models to stop the spread of COVID-19, tools like this could help ensure privacy without sacrificing effectiveness.

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?

The short answer is, yes. It is true that under current law, competition has produced innovative approaches to privacy protection. But as companies, health systems, and other stakeholders begin to use a broader set of data for healthcare purposes--and more healthcare data is produced or transferred outside the Health Insurance Portability and Accountability Act (HIPAA) umbrella--there is a heightened need to adopt a national set of rules that account for the elevated privacy and security risks it poses. As states step into the void and adopt privacy laws that stop at their borders, regulatory uncertainty is the new normal for App Association members. Unfortunately, uncertainty around legal requirements and expectations makes it difficult to quickly create data-driven products and services, including those that respond to the COVID-19 pandemic. The clarity a national privacy law could provide would help the rapid development of privacy protective apps and other tech-driven products and service. Lastly, a strong, national privacy law would help instill a higher baseline level of trust in tech-driven tools. Both consumers and policymakers have met some of the responses from large tech firms with skepticism and myriad questions. A strong and *vigorously enforced national privacy law would obviate the need for many of these questions, along with the expenditure of the time and resources needed to ask and answer them.*

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?

As you proposed in the United States Consumer Data Privacy Act (USCDPA), companies should submit to data minimization, transparency, access, and correction requirements. Effectively anonymizing or de-identifying data should provide a partial shield from these requirements, but the law should be carefully crafted so as not to shield data that is readily re-identifiable. Notably, companies have been using ephemeral device identification methods for some time, in a manner that makes it extremely difficult to reassociate them with a device's owner. The law should keep pace with these techniques and avoid accidentally treating such identifiers as "personally identifiable information," or else the law would remove the incentive for the development of features like this.

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?

A national privacy law should appropriately restrict companies from further, unexpected use of these datasets with use limitations and data minimization requirements. To the extent companies collect sensitive personal data for the purpose of addressing the COVID-19 crisis, it should be up to the person to whom it pertains (if applicable) whether the data continues to have a use beyond addressing the crisis. It may be useful for the person (e.g., as a patient interested in their own health) to continue to have the company analyze it for them. However, government agencies should also be under strict limitations on further use of data collected to address the crisis. For the most part, the data should cease to be available to government actors for purposes beyond those for which it was collected.

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?

As of March 29, 2020, Zoom's privacy policy includes some relevant information about how it processes personal data and the purposes for processing. However, we would support a federal privacy law requiring the disclosure of the elements that appear in the proposed USCDA (see Sec. 102(b).

- 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?

Yes, Congress should enact a single set of rules that apply across the nation.

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?

A national privacy law should appropriately restrict companies from further, unexpected use of these data sets with use limitations and data minimization requirements. To the extent companies collect sensitive personal data for the purpose of addressing the COVID-19 crisis, it should be up to the person to whom it pertains (if applicable) whether the data continues to have a use beyond addressing the crisis. It may be useful for the person (e.g., as a patient interested in their own health) to continue to have the company analyze it for them. However, government agencies should also be under strict limitations on further use of data collected to address the crisis. For the most part, the data should cease to be available to government actors for purposes beyond those for which it was collected.

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.

Bluetooth-enabled contact tracing options have emerged as an alternative to location-based tracking. Privacy advocates seem to prefer the Bluetooth option because it avoids having a central database of location data, which can reveal much more about an individual than a collection of

Bluetooth beacon information. Instead of tracking all of a person's movements, the Bluetooth method only records whether they came into contact with someone with a positive COVID-19 test, a much less sweeping picture of that person's life and travels. However, smartphone makers and carriers have been protecting their databases of anonymized location data with a variety of methods (e.g., ephemeral device IDs) for some time now, and the idea that apps could call those databases in a privacy protective manner for the purpose of contact tracing is not necessarily outside the realm of possibility. However, government should use restraint in how it puts these tools to use, and we would not recommend taking measures as extreme as Israel's to track individuals and enforce quarantines.

- 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.

In my testimony, I noted that this app, Private Kit, takes some privacy protective steps. However, I also noted that unless a single, popular app or platform emerges, contact tracing using methods like this could be limited in their usefulness. Fortunately, after testimonies were submitted, Apple and Google announced a partnership to enable apps to make use of Bluetooth for contact tracing. The partnership will help ameliorate the limitations of a single app by making Bluetooth data available securely and anonymously to qualifying apps. Notably, this could include health system apps, which a user may have downloaded as part of a medical visit, including to test for or treat 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?

Apple's COVID-19 information app does not collect healthcare data in a way that makes it available to itself or other companies. Developers have independently verified this. Companies that make tech-driven tools are, therefore, taking steps to minimize the data they collect to develop a given app or service. They are also finding ways to ensure that large datasets are populated in a way that protects the anonymity of those whose information is plugged into the set. Apple's and Google's recently announced partnership to enable Bluetooth contact tracing is an example of a system that anonymizes data without sacrificing its usefulness. Not only that, the partnership proposes a set of cryptography specifications to protect the data it collects. Strong encryption is an important technical measure companies need to be able to use to protect data, especially when a wide variety of actors are expected to input and take out data. In a closed system, with very few participants, encryption is less necessary because the actors are presumably trusted or malfeasance is easily traced and stopped. But in an open, or "low trust" environment, encryption becomes a more important mechanism. Harnessing large datasets to create useful tools to *combat COVID-19 requires strong encryption, and policymakers should resist calls for government agencies to have built-in vulnerabilities to these technical protection mechanisms.*

- 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?

While the California Consumer Privacy Act requires companies to secure proactive, "opt-in" consent from consumers before selling data, a federal law should impose similar restrictions on companies with respect to the sale of identifiable consumer data. It should be up to the consumer whether the company that has collected data to fight COVID-19 is able to use it to develop further commercial products. Those products could help consumers in unexpected ways, so it should not be assumed that the development of further products using data in a manner to which a consumer consents is necessarily harmful or should categorically be banned.

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?

The research community and security professionals constantly test the privacy and security measures companies take to protect datasets, including anonymization. This market-driven discipline complements enforcers like the Federal Trade Commission (FTC) and state attorneys general (AGs) to hold companies to the promises they make, including as part of data-sharing arrangements, about the anonymization or possibility of re-anonymization of data.

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?

The Office of Personnel Management (OPM) breach announced in June 2015 is emblematic of government agencies' weakness when it comes to securing personally identifiable information. Security experts pinpointed OPM's aging information technology systems and a failure to use adequate encryption to protect the data that was stolen. Using up-to-date hardware and software that is able to combat the latest threats is important. Similarly, agencies should also use the latest privacy protection measures. The National Institute of Standards and Technology (NIST) Cybersecurity Framework as well as the NIST Privacy Framework are good places to start and should be adapted to the specific agency's needs, capabilities, and purposes. Moreover, proposals that would require companies to build vulnerabilities into encryption to enable law enforcement investigators to access otherwise encrypted materials are gravely concerning and would put private sector and federal actors at a totally unnecessary disadvantage against bad actors.

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?

Companies are finding ways to ensure that large datasets are populated in a way that protects the anonymity of those whose information is plugged into the set. Apple's and Google's recently announced partnership to enable Bluetooth contact tracing is an example of a system that anonymizes data without sacrificing its usefulness. Not only that, the model involves the storage of Bluetooth keys on a person's device, which for the most part are encrypted with strong technical measures that prevent access by unauthorized persons. Strong encryption is an important measure companies need to be able to use to protect data, especially when it is sensitive personal information that could be used to trace a person's medical or movement history. Harnessing large datasets to create useful tools to combat COVID-19 requires strong encryption and policymakers should resist calls for government agencies to have built-in vulnerabilities to these technical protection mechanisms.

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.

To the extent companies collect sensitive personal data for the purpose of addressing the COVID-19 crisis, it should be up to the person to whom it pertains (if applicable) whether the data continues to have a use beyond addressing the crisis. It may be useful for the person (e.g., as a patient interested in their own health) to continue to have the company analyze it for them. However, government agencies should also be under strict limitations on further use of data collected to address the crisis. For the most part, the data should *cease to be available to government actors for purposes beyond those for which it was collected.*

13. 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?

Bluetooth-enabled contact tracing options have emerged as an alternative to location-based tracking. Privacy advocates seem to gravitate to the Bluetooth option because it avoids having a central database of location data, which can reveal much more about an individual than a collection of Bluetooth beacon information. However, smartphone makers and carriers have been protecting their databases of anonymized location data with a variety of methods (e.g., ephemeral device IDs) for some time now, and the idea that apps could call those databases in a privacy protective manner for the purpose of contact tracing is not necessarily outside the realm of possibility. However, government should use restraint in how it puts these tools to use, and we would not recommend taking measures as extreme as Israel's or Taiwan's to track individuals and enforce quarantines.

Sen. Cruz

14. 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?

One benefit of a National Infectious Disease Forecasting Center would be to make a connection between epidemiology, the most useful datasets (including from the private sector where appropriate), and policymakers. Such an entity could lead the way in developing data visualization tools, for example, which turn unwieldy amounts of data into actionable information. A potential drawback would be access by a federal entity to sensitive personal data. Federal agencies have a poor track record of data stewardship and less of an incentive than commercial counterparts to develop innovative means of protecting it. Use of such data, where identifiable, should not be a surprise to individuals to whom it pertains and should not serve unrelated policy goals.

15. 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?

I noted in my testimony that differential privacy and federated learning are two techniques the private sector has developed to harness the benefits of large datasets while minimizing risks to privacy. Federated learning, in particular, allows a model to be trained in a way that eliminates the need for a single, central database. To the extent that location and other personal data are needed to develop models to stop the spread of COVID-19, tools like this could help ensure privacy without sacrificing effectiveness.

16. 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?

Experts and policymakers generally agree that reopening the economy must take place in stages and only where the risk of spreading and the concomitant danger to human life and strain on the healthcare system are controlled. Tools that show immunity rates and contacts are likely an important part of reaching this point. However, in addition to data-driven monitoring and modeling, Congress should also ensure that small companies that are part of investor portfolios can access Paycheck Protection Program funds. Some of our member companies are nearing a point where they will need to let go of employees or seek emergency funds. But the fact that they are part of investment portfolios is blocking their ability to benefit from a program that Congress likely intended for them to be able to access on behalf of their employees.

17. Mr. Dufault, as I'm sure you're aware, on March 23, 2020, the administration launched the COVID-19 High Performance Computing Consortium. This consortium is a public-private-partnership lead by the White House Office of Science and Technology Policy, the U.S. Department of Energy, and IBM. The goal of the consortium is to bring together government agencies, such as NASA and the National Science Foundation, academic institutions such as the University of Texas at Austin, and private companies such as Microsoft, Amazon, and Google, to use their resources and expertise to conduct research on the coronavirus. It is my understanding, however, that the consortium does not include any small-and-medium-sized tech companies who may have ideas and knowledge that the larger tech companies do not have.

- How can smaller tech companies help in the fight against coronavirus, and should the Federal Government do more to enlist their help?

Small companies, including ACT | The App Association members, are helping out in important ways. The examples I alluded to in my testimony, including the appendix, are just a few. Some, like Rimidi, are creating screening tools for health systems while others, like Particle Health, are creating platforms to create secure access for patients to complete healthcare records in usable formats. Last year, the National Science Foundation (NSF) awarded another of our member companies, BadVR, a grant to fund the development of a product that uses virtual reality to provide immersive data visualization experiences. The grant comes from the NSF Seed Fund, and we encourage the federal government to use this and resources like it to fund similar small business projects to further our capabilities in the fight against COVID-19 and future national health emergencies. Now more than ever, thanks to the entry of large software platforms and the wide availability of fast, reliable internet connectivity, small businesses are well positioned to play a role in cutting-edge responses to emergent societal and health problems.

Sen. Fischer

18. In your testimony, you mentioned an example of a new app being developed by researchers at MIT that tracks patients who have and have not tested positive for COVID-19, while anonymizing their location data. How can federal, state, and local governments become better equipped and more aware of reliable digital tools – many of which are still emerging or have recently emerged – on a short-term basis, to leverage big data benefits and combat the spread of the virus? In this rapid effort, how can government officials ensure that new apps are secure and mitigate privacy risks?

In my testimony, I noted that this app, Private Kit, takes some privacy protective steps. However, I also noted that unless a single, popular app or platform emerges, contact tracing using methods like this could be limited in their usefulness. Fortunately, after testimonies were submitted, Apple and Google announced a partnership to enable apps to make use of Bluetooth for contact tracing. The partnership will help ameliorate the limitations of a single app by making Bluetooth data available securely and anonymously to qualifying apps. Notably, this could include health system apps, which a user may have downloaded as part of a medical visit, including to test for or treat COVID-19. Earlier in the crisis, Apple and Google both updated their developer terms of service clarifying that apps developed in "highly-regulated" fields like healthcare should be submitted by the "legal entity that provides the services, and not by an independent developer." (Rule 5.1.1(ix)). The guideline helps weed out independently developed apps that do not provide reliable healthcare information about COVID-19, but purport to do so. Software platforms are, therefore, helping sort these out so that government entities have more reliable options to choose from on the app stores, including those developed in partnership with a healthcare system (a model several of our member companies have chosen).

Sen. Moran

19. 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?

Requiring companies to commit publicly not to re-identify data can be an effective method of requiring data to remain de-identified. Requiring a company to make a representation and then showing that the representation is false is an easier task for the FTC or other enforcers than less bright-line prohibitions or requirements. Another option to consider is HIPAA's two allowable means of de-identifying protected health information. Option 1: "expert determination" – have an expert apply a statistical model or scientific principles with a very small risk that an anticipated recipient could identify an individual. Or Option 2: "safe harbor" - remove 18 specific identifiers (if applicable) and have no actual knowledge of any residual information that could identify the individual. Among these options, we have a slight preference for the first approach, similar to the "de-identify" provision in your legislation, the Consumer Data Privacy and Security Act (S. 3456).

20. 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?

It is true that under current law, competition has produced innovative approaches to privacy protection. But as companies, health systems, and other stakeholders begin to use a broader set of data for healthcare purposes--and more healthcare data is produced or transferred outside the Health Insurance Portability and Accountability Act (HIPAA) umbrella--there is a heightened need to adopt a national set of rules that account for the elevated privacy and security risks it poses. As states step into the void and adopt privacy laws that stop at their borders, regulatory uncertainty is the new normal for App Association members. Unfortunately, uncertainty around legal *requirements and expectations makes it difficult to quickly create data-driven products and services, including those that respond to the COVID-19 pandemic. The clarity a national privacy law could provide would help the rapid development of privacy protective apps and other tech-driven products and service. Lastly, a strong, national privacy law would help instill a higher baseline level of trust in tech-driven tools. Both consumers and policymakers have met some of the responses from large tech firms with skepticism and myriad questions. A strong and vigorously enforced national privacy law would obviate the need for many of these questions, along with the expenditure of the time and resources needed to ask and answer them.*

21. 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?

While the California Consumer Privacy Act (CCPA) requires companies to secure affirmative, "opt-in" consent from consumers before selling data, a federal law should impose similar restrictions on companies with respect to processing activities that are incompatible with the stated purposes for which data was initially collected. It should be up to the consumer whether the company that has collected data to fight COVID-19 is able to use it to develop further commercial products. Those products could help consumers in unexpected ways, so it should not be assumed that the development of further products using data in a manner to which a consumer consents is necessarily harmful or should categorically be banned.

22. 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?

I am probably not familiar enough with the issue to comment on the practices companies use to ensure that the data they receive from foreign governments, for example, is accurate. However, the App Association is keenly aware of the importance of protecting such data from unauthorized access. For example, one of the important ways big data is playing a role in addressing the global COVID-19 pandemic is by helping enable contact tracing. Whether the digital mechanism is based on location data or Bluetooth interactions, the models involve the storage of sensitive data on a people's devices, which for the most part are encrypted with strong technical measures that prevent access by unauthorized persons. Strong encryption is an important measure companies need to be able to use to protect data, especially when it is sensitive personal information that could be used to trace a person's medical or movement history. Harnessing large datasets to create useful tools to combat COVID-19 requires strong encryption and policymakers should resist calls for government agencies to have built-in vulnerabilities to these technical protection mechanisms.

Sen. Gardner

23. Mr. Dufault, ACT | The App Association has numerous member companies in Colorado, and I appreciate your association's work with them to bolster Colorado's technology footprint and your members' efforts to help combat COVID-19. In your testimony, you express support for recently issued rules from the Department of Health and Human Services (HHS) regarding patient requests to transfer electronic health data to third parties. I strongly support patients having control over their own data, particularly sensitive healthcare data.

    You also criticize electronic health records companies (EHRs) as "a chokepoint for healthcare data that patients should otherwise be able to use as they wish." Numerous technology companies have been criticized for the lack of transparency into their data policies and confusing privacy regimes that leave users without a firm understanding of how their data will be used and maintained. That kind of transparency and control is critical, especially in times of crisis.

    - Will your members commit to ensuring that patients and users of their services have clear and transparent individual access to policies explaining how their data will be used, including any sale or procurement of it from other third parties?

This is a fair point and one we think deserves the oversight of both HHS and the Federal Trade Commission (FTC). As HHS' Office of the National Coordinator for Health IT (ONC) developed its "information blocking" rules, we urged it to require any third-party app developer receiving protected health information from a covered entity to attest to three things: 1) whether the app provides a model privacy notice; 2) whether the developer designed the app with consensus-driven privacy guidelines; and 3) whether the app's data use policies adhere to industry-developed best practices. These important questions do not unduly limit a patient's management of their own data, but they do provide simple and actionable information about an app's privacy and security practices. These disclosures in turn help patients make more informed privacy choices while simultaneously helping the federal enforcement agency responsible for policing non-covered entities' data practices—the FTC—with relevant information. ONC adopted our proposal in its final rule, requiring third-party app developers (who are not subject to HIPAA, and are not Business Associates [BAs] of HIPAA covered entities) to make a series of attestations with respect to their privacy policies. This is an important measure to inform patients as they consider options with respect to their healthcare information. However, it is not a substitute for a strong, national privacy framework. We strongly encourage Congress to develop a single, national set of privacy rules to govern the processing and transfer of data under the FTC's broad jurisdiction. As companies, health systems, and other stakeholders begin to use a broader set of data for healthcare purposes--and more healthcare data is produced or transferred outside the HIPAA umbrella--there is a heightened need to adopt a national set of rules that account for the elevated privacy and security risks it poses.

24. Mr. Dufault, in the interest of patients using their data "as they wish", will your member companies engaged in the Connected Health Initiative commit to permanently delete a patient's electronic data if the patient requests its deletion?

Connected Health Initiative (CHI) members that are either covered entities or BAs under HIPAA must dispose of protected health information in specific ways as outlined at 45 CFR 164.310(d)(2)(i). In addition, HIPAA requires entities to respond to requests from patients to request a "restriction of uses and disclosures," but permanent deletion does not appear to be one of those restrictions. 45 CFR 164.522. HIPAA does not require covered entities and BAs to retain protected health information, but state laws generally do, in order to facilitate compliance audits (for example, Florida requires covered entities to maintain medical records for five years after the last patient contact). CHI members under HIPAA could broadly commit to deleting data when requested by a patient, but state and federal legal requirements associated with HIPAA would likely impose a series of caveats on such a commitment.

As for CHI members outside HIPAA's scope, many of them are subject to the California Consumer Privacy Act (CCPA), which includes a requirement to respond to requests to delete information about a given verified individual. While the App Association cannot bind these companies to a commitment (we are not a self-regulatory body), these member companies generally do commit to deleting personally identifiable information pertaining to an individual pursuant to a verifiable consumer request from that person, consistent with other legal requirements.

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 HIPPA, not to allow tech companies to exploit a crisis to gather even more personal data.

25. How do you see HIPPA interacting with your worldview of the tech industry?

As HHS' Office of the National Coordinator for Health IT (ONC) developed its "information blocking" rules, we urged it to require any third-party app developer receiving protected health information from a HIPAA covered entity to attest to three things: 1) whether the app provides a model privacy notice; 2) whether the developer designed the app with consensus-driven privacy guidelines; and 3) whether the app's data use policies adhere to industry-developed best practices. These important questions do not unduly limit a patient's management of their own data, but they do provide simple and actionable information about an app's privacy and security practices. The disclosures in turn help patients make more informed privacy choices while simultaneously helping the federal enforcement agency responsible for policing non-covered entities' data practices—the Federal Trade Commission (FTC)—with relevant information. ONC adopted our proposal in its final rule, requiring third-party app developers (who are not subject to HIPAA, and are not Business Associates of HIPAA covered entities) to make a series of attestations with respect to their privacy policies. This is an important measure to inform patients as they consider options with respect to *their healthcare information. However, it is not a substitute for a strong, national privacy framework. We strongly encourage Congress to develop a single, national set of privacy rules to govern the processing and transfer of data under the FTC's broad jurisdiction. As companies, health systems, and other stakeholders begin to use a broader set of data for healthcare purposes--and more healthcare data is produced or transferred outside the HIPAA umbrella--there is a heightened need to adopt a national set of rules that account for the elevated privacy and security risks it poses.*

26. 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?

Small companies, including ACT | The App Association members, are helping out in important ways. Last year, the National Science Foundation (NSF) awarded one of our member companies, BadVR, a grant to fund the development of a product that uses virtual reality to provide immersive data visualization experiences. The grant comes from the NSF Seed Fund, and we encourage the federal government to use this and resources like it to fund similar small business projects to further our capabilities in the fight against COVID-19 and future national health emergencies.

In my testimony, I noted that contact tracing app Private Kit takes important privacy protective steps to safeguard location data. However, I also noted that unless a single, popular app or platform emerges, contact tracing using methods like this could be limited in their usefulness. As you pointed out, larger companies have the comparative advantage in bulk data collection and we should find ways to leverage this capability while limiting privacy risks. Fortunately, after testimonies were submitted, Apple and Google announced a partnership to enable apps to make use of Bluetooth for contact tracing. The partnership will help ameliorate the limitations of a single app by making Bluetooth data available securely and anonymously to qualifying apps. Notably, this could include health system apps, which a user may have downloaded as part of a medical visit, including to test for or treat COVID-19. Earlier in the crisis, Apple and Google both updated their developer terms of service clarifying that apps developed in "highly-regulated" fields like healthcare should be submitted by the "legal entity that provides the services, and not by an independent developer." (Rule 5.1.1(ix)). The guideline helps weed out independently developed apps that do not provide reliable healthcare information about COVID-19, but purport to do so. Software platforms are therefore helping sort these out so that users have more reliable (and inherently privacy protective) options to choose from on the app stores, including those developed in partnership with a healthcare system (a model several of our member companies have chosen).

27. 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?

A national privacy law should appropriately restrict companies from further, unexpected use of these datasets with use limitations and data minimization requirements. To the extent companies collect sensitive personal data for the purpose of addressing the COVID-19 crisis, it should be up to the person to whom it pertains (if applicable) whether the data continues to have a use beyond addressing the crisis. It may be useful for the person (e.g., as a patient interested in their own health) to continue to have the company analyze it for them. However, government agencies should also be under strict limitations on further use of data collected to address the crisis. For the most part, the data should cease to be available to government actors for purposes beyond those for which it was collected.

28. 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?

Bluetooth-enabled contact tracing options have emerged as an alternative to location-based tracking. Privacy advocates seem to gravitate to the Bluetooth option because it avoids having a central database of location data, which can reveal much more about an individual than a collection of Bluetooth beacon information. However, smartphone makers and carriers have been protecting their databases of anonymized location data with a variety of methods (e.g., ephemeral device IDs) for some time now, and the idea that apps could call those databases in a privacy protective manner for the purpose of contact tracing is not necessarily outside the realm of possibility. However, government should use restraint in how it puts these tools to use, and we would not recommend taking measures as extreme as Israel's or Taiwan's to track individuals and enforce quarantines.

Sen. Capito

29. One of West Virginia's greatest challenges remains lack of reliable broadband, but a separate challenge lies in convincing a portion of those that do have access to purchase it. This has been common among elderly West Virginians.
    - Since the elderly population remains particularly vulnerable to contract severe COVID-19 symptoms, how can we utilize various types of data to ensure that those who lack access to the internet, or a mobile device will be accurately represented?

The most effective way of reaching elderly populations who do not use smart devices during the COVID-19 crisis may be through their healthcare provider or retirement community. Accurate reporting by community health centers (RHCs), federally qualified community health centers (FQHCs), and skilled nursing facilities (SNFs) as to infection rates, symptoms, and resource needs is therefore critically important.

Sen. Lee

30. 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?

There are a few different types of responses companies are developing to COVID-19. Some are information apps, many of which avoid collecting data. For example, the Apple COVID-19 app does not collect healthcare data in a way that makes it available to itself or other companies. Developers have independently verified this. Companies that make tech-driven tools are, therefore, taking steps to minimize the data they collect to develop a given app or service. They are also finding ways to ensure that large datasets are populated in a way that protects the anonymity of those whose information is plugged into the set. Apple's and Google's recently announced partnership to enable Bluetooth contact tracing is an example of a system that anonymizes data without sacrificing its usefulness. Not only that, the model involves the storage of Bluetooth keys on a person's device, which for the most part are encrypted with strong technical measures that prevent access by unauthorized persons. Strong encryption is an important measure companies need to be able to use to protect data, especially when it is sensitive personal information that could be used to trace a person's medical or movement history. Harnessing large datasets to create useful tools to combat COVID-19 requires strong encryption and policymakers should resist calls for government agencies to have built-in vulnerabilities to these technical protection mechanisms.

31. 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?

I noted in my testimony that differential privacy and federated learning are two techniques the private sector has developed to harness the benefits of large datasets while minimizing risks to privacy. Federated learning, in particular, allows a model to be trained in a way that eliminates the need for a single, central database. To the extent that location and other personal data are needed to develop models to stop the spread of COVID-19, tools like this could help ensure privacy without sacrificing effectiveness.

- 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?

Requiring companies to commit publicly not to re-identify data--a requirement that appears in Chairman Wicker's USCDPA draft--can be an effective method of requiring data to remain de-identified. Requiring a company to make a representation and then showing that the representation is false is an easier task for the Federal Trade Commission (FTC) or other enforcers than less bright-line prohibitions or requirements. Another option to consider is HIPAA's two allowable means of de-identifying protected health information. Option 1: "expert determination" – have an expert apply a statistical model or scientific principles with a very small risk that an anticipated *recipient could identify an individual. Or Option 2: "safe harbor" - remove 18 specific identifiers (if applicable) and have no actual knowledge of any residual information that could identify the individual.*

- Can effective contact tracing be conducted with "anonymized data"? Or will it require personally identifiable information?

I am aware of two primary ways of conducting digital contact tracing: Bluetooth-enabled and location-based. In my testimony, I noted that Private Kit, a location-based app, takes some privacy protective steps to ensure the privacy of a user's identity and location. However, I also noted that unless a single, popular app or platform emerges, contact tracing using methods like this could be limited in their usefulness. Meanwhile, after testimonies were submitted, Apple and Google announced a partnership to enable apps to make use of Bluetooth for contact tracing. The partnership will help ameliorate the limitations of a single app by making Bluetooth data available securely and anonymously to qualifying apps. Both of these methods involve potentially sensitive information (location data or Bluetooth information that may identify a device), using an architecture that limits the transfer of that data in identifiable form to a central authority or database. Ideally, the resulting application programming interface (API) enabling access to data generated by devices across the nation allows developers to harness actionable data while blocking their or others' ability to reidentify individuals.

It should also be noted that the Apple and Google partnership described above involves the storage of Bluetooth keys on people's devices, which are mainly protected with strong encryption. Without this technical measure, unauthorized bad actors could to retrace those keys and reidentify individuals or leak a set of stored keys to the public in readable format. Therefore, encryption plays an important role in the development of data-driven tools to combat COVID-19. We commend you for your steadfast defense of strong encryption and share your concerns with proposals that would require companies to build encryption vulnerabilities specifically for law enforcement investigators. These vulnerabilities would ultimately weaken encryption that protects a wide range of data, creating an opening for bad actors to exploit COVID-19 data--not just a set of keys for well-intentioned actors like American law enforcers.

32. 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?

One example that comes to mind is the COVID-19 High Performance Computing Consortium. This consortium is a public-private-partnership led by the White House Office of Science and Technology Policy, the U.S. Department of Energy, and Microsoft, IBM, and several other private sector and academic entities. The Consortium may share data, but the purpose is not necessarily to facilitate enforcement, but rather, to pool together computing capabilities to assist with research in bioinformatics, epidemiology, and molecular modeling. I am unaware of instances in the United States where government agencies have sought data from private companies in order to enforce *orders or requirements. Although many of our member companies are actively working to assist with the broad response to the COVID-19 pandemic, we would be very concerned with data requests for personally identifiable information that do not observe proper due process, especially for enforcement purposes.*

Sen. Johnson

33. What are your members that are part of ACT's Connected Health Initiative saying about medical equipment and PPE shortage? Is there a way for healthcare-related apps to leverage data to better address medical supply chain needs?

ACT | App Association members often help other companies digitize processes to make back-end functions more efficient and mobile. For example, our most recent State of the App Economy Report points to Convoy, which helps companies book shipping trucks to deliver products. The platform allows companies to optimize their supply chains by drawing on relevant data and insights. Digitization of supply chain management is a key to success for a number of industries, and medical supply chains are no different in that respect.

Similarly, another one of our members, the Medical Society of Northern Virginia, has partnered with Model Space to 3D print face shields for physicians in northern Virginia. Partnerships like this are cropping up all over the nation thanks to the availability of open source face shield designs for 3D printers. These are a complement to a robust and flexible supply chain for medical equipment as opposed to a replacement for it, but they are an important option as supply chains fail to deliver personal protection equipment.

Sen. Young

34. Mr. Dufault, sometimes consumers trust companies with their data more than they trust the government. What assurances should the government provide for any data it obtains from the private sector?

To the extent companies collect sensitive personal data for the purpose of addressing the COVID-19 crisis, it should be up to the person to whom it pertains (if applicable) whether the data continues to have a use beyond addressing the crisis. It may be useful for the person (e.g., as a patient interested in their own health) to continue to have the company analyze it for them. However, government agencies should also be under strict limitations on further use of data collected to address the crisis. For the most part, the data should cease to be available to government actors for purposes beyond those for which it was collected.

35. Mr. Dufault, how might the public and private sector work together to develop model data sharing agreements for use during a pandemic?

The purposes for proposed data sharing should drive the attributes of data sharing agreements. For example, the COVID-19 High Performance Computing Consortium is a public-private-partnership led by the White House Office of Science and Technology Policy, the U.S. Department of Energy, Microsoft, IBM, and several other private sector and academic groups. The Consortium may share data, but the purpose is not necessarily to facilitate enforcement. Instead, it seeks to pool together computing capabilities to assist with research in bioinformatics, epidemiology, and *molecular modeling. Arrangements that provide government entities with access to data collected in other settings should carefully constrain what agencies can do with that data, including setting limits on how it can be used to enforce government orders or regulations.*

Sen. Scott

36. 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?

Bluetooth-enabled contact tracing options have emerged as an alternative to location-based tracking. Privacy advocates seem to be less uncomfortable with the Bluetooth option because it avoids having a central database of location data, which can reveal much more about an individual than a collection of Bluetooth beacon information. However, smartphone makers and carriers have been protecting their databases of anonymized location data with a variety of methods (e.g., ephemeral device IDs) for some time now, and the idea that apps could call those databases in a privacy protective manner for the purpose of contact tracing is not necessarily outside the realm of possibility. However, government should use restraint in how it puts these tools to use, and we would not recommend taking measures as extreme as Israel's or Taiwan's to track individuals and enforce quarantines.

Ranking Member Cantwell

37. 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?

Small companies, including ACT | The App Association members, are poised to contribute in important ways, with the help of federal resources. Last year, the NSF awarded one of our member companies, BadVR, a grant to fund the development of a product that uses virtual reality to provide immersive data visualization experiences. The grant comes from the NSF Seed Fund, and we encourage the federal government to use this and programs like it to fund similar small business projects to further our capabilities in the fight against COVID-19 and future national health emergencies.

I am aware of two primary ways of conducting digital contact tracing: Bluetooth-enabled and location-based. In my testimony, I noted that Private Kit, a location-based app, takes some privacy protective steps to ensure the privacy of a user's identity and location. However, I also noted that unless a single, popular app (or platform) emerges, contact tracing using methods like this could be limited in their usefulness. Meanwhile, after testimonies were submitted, Apple and Google announced a partnership to enable apps to make use of Bluetooth for contact tracing. The partnership will help ameliorate the limitations of a single app by making Bluetooth data available securely and anonymously to qualifying apps. Both of these methods involve potentially sensitive information (location data or Bluetooth information that may identify a device), using an architecture that limits the transfer of that data in identifiable form to a central authority or database. Ideally, the resulting application programming interface (API) enabling access to data generated by devices across the nation allows developers to harness actionable data while blocking their or others' ability to reidentify individuals.

38. Some witness testimony today refers to apps applying innovative decentralized architectures. Ms. Richardson described a technology known as Pan-European Privacy-Preserving Proximity Tracing (PEPP-PT), which was "created to develop technical mechanisms and standards that protect privacy while also leveraging digital resources to bolster pandemic response efforts." Mr. Dufault described Private Kit, an app designed so that "location data stays on the user's phone and does not go to a centralized server."
    - These are both examples of the emerging decentralization movement in software and Internet design, whose proponents claim the benefits of greater efficiency and enhanced privacy protection. Are you monitoring this movement? Do you believe decentralized applications can inspire greater trust in users, leading to more rapid voluntary adoption?

Yes, developers are starting to use decentralization techniques to enhance the security and privacy of data, and it can be an effective tool. The privacy protective capacity of decentralized architectures is demonstrable and hopefully does help inspire greater trust in users and in turn, wider and more rapid voluntary adoption. Device capabilities are critical for decentralization--they must be capable of storing data securely while putting it to use to inform a single model or a set of distributed models. The availability of these features to broad swaths of the general public has been relatively recent so we are hopeful that developers will continue to find new and beneficial ways of employing decentralized models rapidly both in response to this crisis and going forward.

39. 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?

As you proposed in the Consumer Online Privacy Rights Act (COPRA, S. 2968), companies should submit to data minimization, transparency, access, and correction requirements. Effectively anonymizing or de-identifying data should provide a partial shield from these requirements (especially where compliance would involve re-identification), but the law should be careful not to shield data that is not effectively de-identified. COPRA's de-identified data definition would likely work well as a deterrent against companies taking inadequate steps to de-identify or anonymize data. Public commitments to take certain steps have the added benefit of providing an *enforcement hook for the FTC and state AGs when those commitments are broken. Another option to consider is HIPAA's two allowable means of de-identifying protected health information. Option 1: "expert determination" – have an expert apply a statistical model or scientific principles with a very small risk that an anticipated recipient could identify an individual. Or Option 2: "safe harbor" - remove 18 specific identifiers (if applicable) and have no actual knowledge of any residual information that could identify the individual. These concepts may be more difficult to enforce than COPRA's, however. And while the expert option may be out of reach for companies with less resources, the second option may not be strong enough to prevent re-identification.*

Notably, companies have been using ephemeral device identification methods for some time, in a manner that makes it extremely difficult to reassociate them with a device's owner. The law should keep pace with these techniques and avoid accidentally treating such identifiers as "personally identifiable information," or else it might remove the incentive for the development of features like this.

Sen. Blumenthal

40. In late March, StatSocial, a data broker that caters to advertisers, announced a "Crisis Insights" product, which claims to monitor consumer sentiments regarding the COVID-19 pandemic. StatSocial boasts that its analysis is "based on StatSocial's unique Silhouette social identity platform with its 85,000-segment taxonomy across 1.3 billion social accounts that connect to more than 70% of US households." StatSocial further elaborated that its dataset is sourced from "1.2 billion profiles sourced from 60 different platforms and more than a million websites, 300 million individual email accounts, and more than one billion IP addresses and mobile devices."
    Clearly, StatSocial has not obtained consent, nor even awareness, from 70% of U.S. households to monitor their concerns and their families' wellbeing from their social media accounts and inboxes during the Coronavirus pandemic.
    - Please list which members of your association harvest, process, or disclose to third parties (such as through SDKs or data sharing agreements) information gathered from social media accounts or email inboxes for the purposes of audience segmentation, building profiles, ad targeting, advertising analytics, or other commercial purposes not technically necessary for the provision of a service or product that an individual has requested?

Privacy policies typically describe which partners or companies a user's data is shared with, so users can determine whether their data is being shared with marketing or social media companies. We have not surveyed our members about their specific use of software development kits (SDKs) or other sharing arrangements, so we cannot provide detailed information about this. However, both Apple and Google require every app that collects personally identifiable data to have a privacy policy--and California requires the publication of privacy policies--so our member companies generally disclose sharing arrangements publicly.

Most companies that provide an SDK publish the SDK's source code, so developers can inspect the code before including the SDK into their app. It is a best practice to only include open source

SDKs in apps, but not all app makers follow this. In the case of Zoom, I believe they could have reviewed the Facebook SDK to make sure that it met privacy expectations because the source code is available here. They didn't do so, and they really should have. The problem was that the Facebook SDK was, regardless of whether the user logged in using Facebook or not, sending information to Facebook. Zoom would know this by reviewing the SDK code, and they could have either removed the feature from the SDK or informed the users that this would happen (typically as part of their privacy policy). Similarly, many app makers include the "Sign in with Facebook" SDK for the convenience of their users. Reports last year suggested that the SDK was funneling personal data collected through these apps to Facebook to fuel targeted ads to people using the apps, but it was hardly clear to the app makers that this was happening. Much of the responsibility for clarity is on the provider of the SDK to explain how it plans to use this data so that the app developer knows and can either decline to use the SDK or clarify the use in its own privacy policy.

SDKs also serve a number of purposes beyond shuttling data to a third party in order to build a consumer profile or feed targeted advertising. There are SDKs for visualization, data processing, e-commerce, financial services, user authentication, analytics, and many other uses. They should not be trojan horses to secretly shuttle away data to feed ad targeting or profiles.

Sen. Schatz

41. Companies' datasets have been used to create models to forecast the spread of the pandemic. However, according to a recent Pew study, only 80% of Americans have access to a smartphone. A lot fewer use smart thermometers.

    This Committee is well acquainted with the digital divide and the discriminatory impact caused by the lack of availability and access to broadband and smart technologies. Accordingly, can you assure the Committee that the datasets of your companies or member companies' are truly non-biased representations of the population, and will you commit to have these datasets audited by independent experts to ensure we are not making critical-decisions regarding the pandemic based on biased data?

The federal government should adopt policies that help close the digital divide by supporting the deployment of cost-effective and affordable broadband alternatives to rural and urban areas alike. The Federal Communications Commission's (FCC's) Lifeline program is only one piece of this puzzle. The FCC should also adopt final rules before the end of the year clarifying how operators can use vacant television channels or "TV white spaces." Unlicensed use of the TV band is an important part of closing the digital divide. We are also strong supporters of your legislation, the CONNECT for Health Act and advocated for the inclusion of a complete waiver from Section 1834(m) of the Social Security Act, which imposes draconian restrictions on Medicare payment for telehealth (live voice or video) services to Medicare beneficiaries, in the latest COVID-19 legislation. This is not enough, however. Inclusion of items like smart thermometers in Flex Spending Accounts (FSAs) would make them more affordable, but the IRS does not consider many wearables and other connected devices as "medical care" for the purpose of FSAs. This is an outdated policy approach, and we recommend that Congress immediately update the statute to include wearable devices and software apps and platforms in the definition of "medical care" for the purpose of FSAs so that more Americans can access digital health tools to manage or prevent chronic or acute conditions. Americans in rural areas and in disadvantaged socio-economic conditions suffer chronic illnesses at higher rates than Americans in urban, suburban areas and with higher incomes.

Therefore, they may be at higher risk of becoming more sick as a result of COVID-19 and should be a primary focus of federal policies to combat the virus.

While our member companies make use of large datasets, generally they are not the owners or collectors of bulk data. However, they are keenly aware of the dangers of biased datasets. Our member companies know that the output of an algorithm is usually only a suggestion. The same is true with simple conclusions from a dataset. Ultimately, the company using the dataset or the algorithm is responsible for the impact it has on the company's clients and customers.

Sen. Peters

42. 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?

While we are not experts in supply chain issues, we observe that a global pandemic presented a unique challenge to American PPE supply chains. In a national or regional outbreak, the inability to provide surge capacity could be offset by outsourcing supply to China or Mexico. A national stockpile is therefore likely a wise course of action to properly prepare for another public health emergency like COVID-19. Meanwhile, healthcare systems, manufacturers, and tech companies are pitching in to find ways around the shortage. One of our members, the Medical Society of Northern Virginia, has partnered with Model Space to 3D print face shields for physicians in northern Virginia. Partnerships like this are cropping up all over the nation thanks to the availability of open source face shield designs for 3D printers. These complement a robust and flexible supply chain for medical equipment as opposed to replacing it, but they are an important option as supply chains fail to deliver personal protection equipment.

43. 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?

The 3D printable face mask example I mentioned above is one option along these lines. The other possible digital interventions try to solve the contact tracing problem rather than the PPE problem. I am aware of two primary ways of conducting digital contact tracing: Bluetooth-enabled and location-based. In my testimony, I noted that Private Kit, a location-based app, takes some privacy protective steps to ensure the privacy of a user's identity and location. However, I also noted that unless a single, popular app or platform emerges, contact tracing using methods like this could be limited in their usefulness.

44. Earlier this week, The European Commission called for greater coordination between countries by creating a shared toolbox for the development of coronavirus smartphone apps. Other countries such Poland, France, Ireland and the U.K. have done the same or plan to in the near future.
    - Should the United States create its own digital toolbox and coordinate with other countries doing the same and how do you envision the coordination effort being executed?

The private sector is leading the charge on contract tracing in the U.S. and these digital toolboxes are being developed, appropriately, at the software platform level. For example, last week, Apple and Google announced a partnership to enable apps to make use of Bluetooth for contact tracing. The partnership will help ameliorate the limitations of a single app by making Bluetooth data available securely and anonymously to qualifying apps. With this system, public health agencies or health systems would be able to make use of the Bluetooth keys to alert users who have opted into the system if they had previously come into contact with someone who registered a positive COVID-19 test through an app provided by the authority. Both the location-based and Bluetooth methods involve potentially sensitive information (location data or Bluetooth information that may identify a device), using an architecture that limits the transfer of that data in identifiable form to a central authority or database. Ideally, the resulting application programming interface (API) enabling access to data generated by devices across the nation allows developers to harness actionable data while blocking their or others' ability to reidentify individuals.

It should also be noted that the Apple and Google partnership described above involves the storage of Bluetooth keys on people's devices, which are mainly protected with strong encryption. Without this technical measure, unauthorized bad actors could to retrace those keys and reidentify individuals or leak a set of stored keys to the public in readable format. Therefore, encryption plays an important role in the development of data-driven tools to combat COVID-19. We commend you for your steadfast defense of strong encryption and share your concerns with proposals that would require companies to build encryption vulnerabilities specifically for law enforcement investigators. These vulnerabilities would ultimately weaken encryption that protects a wide range of data, creating an opening for bad actors to exploit COVID-19 data--not just a set of keys for well-intentioned actors like American law enforcers.

Sen. Baldwin

45. 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 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?

Americans in rural areas and in disadvantaged socio-economic conditions suffer chronic illnesses at higher rates than Americans in urban and suburban areas and with higher incomes. The data clearly shows that this is true. And yet, the use of digital health tools to monitor and manage chronic conditions also indicates a major impact on both the health of these patients and healthcare costs. A study by our Connected Health Initiative steering committee member University of Mississippi Medical Center (UMMC) showed that including just 20 percent of the state's Medicaid population with diabetes in UMMC's digital health program could save the state $189 million per year, primarily by reducing rehospitalizations. These patients are among the most socioeconomically disadvantaged and may be at higher risk of becoming more sick as a result of COVID-19. They should be a primary focus of federal policies to combat the virus.

While our member companies make use of large datasets, generally they are not the owners or collectors of bulk data. However, they are keenly aware of the dangers of biased datasets. Our member companies know that the output of an algorithm is usually only a suggestion. The same is true with simple conclusions from a dataset. Ultimately, the company using the dataset or the algorithm is responsible for the impact it has on the company's clients and customers.

46. 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?

The most effective way of reaching rural populations (including elderly Americans) who do not use or cannot access smart devices during the COVID-19 crisis may be through their healthcare provider or retirement community. Accurate reporting by community health centers (RHCs), federally qualified community health centers (FQHCs), and skilled nursing facilities (SNFs) as to infection rates, symptoms, and resource needs is, therefore, critically important. That said, *controlling the spread of COVID-19 in rural areas likely requires different measures than those that should be deployed in urban and suburban areas. For example, lower population densities may present a lower risk of spreading, but rural areas also generally have a far more limited capacity to handle a surge in healthcare needs presented by a local outbreak. High quality data is important, but perfect data is usually impossible. We should focus on making big data-driven tools available to local authorities and healthcare systems so that they can manage their regional situations armed with the best tools available.*

47. 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?

Local transportation departments, especially in urban areas, are well-heeled now at forging partnerships with private, data-driven companies. For example, cities across the nation are issuing permits to scooter companies in return for access to mobility data, some of them adopting the Mobility Data Specification (MDS) forged initially in Los Angeles. However, MDS has come under fire for paying too little attention to data governance and privacy. The lessons learned by public transportation divisions should be quickly understood and internalized by sister organizations, including the local public health departments, as they consider public-private data partnerships to address COVID-19 and related public health issues.

48. 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?

In addition to the considerations described above, a swift response requires the buy-in of large software platforms. The Google-Apple partnership announced last week is one example and offers tremendous potential value for public and private purposes alike in responding to COVID-19. Ultimately, a system that ensures data sharing between and among public and private entities should offer adequate measures of both standardization and privacy protection. For research data, it is not enough for it to be available, there have to be tools developed specifically to analyze a body of research for specific findings. That's why the COVID-19 High Performance Computing Consortium, which seeks the development of tools and computing capacity to analyze research about COVID-19 and related coronaviruses, is a positive step. For Bluetooth or location data, standardization, privacy, and encryption are all important elements and platform-level stewardship and API development (as we see with the Google-Apple partnership) are advantageous.

Sen. Tester

49. You speak highly of the privacy protections built into apps like Private Kit. But especially in some of the small towns in my state, it's pretty easy to guess someone's identity based on clues about where they've been or who they've talked to. Is there an inherent tradeoff between the usefulness and the anonymity of this kind of data?

There are a few different types of responses companies are developing to COVID-19 that aim to provide adequate measures of both privacy and usefulness. Some are information apps, many of which avoid collecting data. For example, the Apple COVID-19 app does not collect healthcare data in a way that makes it available to itself or other companies. Developers have independently verified this. Companies that make tech-driven tools are therefore taking steps to minimize the data they collect to develop a given app or service.

They are also finding ways to ensure that large datasets are populated in a way that protects the anonymity of those whose information is plugged into the set. Apple's and Google's recently announced partnership to enable Bluetooth contact tracing is an example of a system that anonymizes data without sacrificing its usefulness. The system involves the issuance of new Bluetooth keys every so often to make tracing them to an individual more difficult, while storing them only on the person's device. Not only that, the model involves the storage of Bluetooth keys on a person's device, which for the most part are encrypted with strong technical measures that prevent access by unauthorized persons. Strong encryption is an important measure companies need to be able to use to protect data, especially when it is sensitive personal information that could be used to trace a person's medical or movement history. Harnessing large datasets to create useful tools to combat COVID-19 requires strong encryption and policymakers should resist calls for government agencies to have built-in vulnerabilities to these technical protection mechanisms.

50. You've described a number of ways data analytics can help solve problems that are new or newly urgent in the face of COVID-19. I am very concerned about the spread of false information during this vulnerable time. Is big data hurting or helping with the spread of misinformation and how do we hold these companies accountable?

Big data is helping retail platforms remove fraudulent offerings of vaccines, for example. It is also helping social media platforms detect, tag, and remove harmful misinformation, especially when bots are spreading the false news, as is apparently the case with a significant proportion of the misinformation on COVID-19.

Sen. Sinema

51. 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?

Our member company Beyond Lucid serves first responders and has developed a COVID-19 tool. It is an adaptation of its Medview Trips software and is a "rapidly deployable inversion of CP/MIH charting software, turned inward to track crew and staff health over time for self-monitoring, active observation, and even quarantine through hospitalization and return to work." I encourage you to share this with your local first responders, which may benefit from software like Beyond Lucid's.

52. 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?

High quality data is important, but perfect data is usually impossible. The best analyses of available data adequately account for its shortcomings, false assumptions, omissions, and biases.

53. 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?

I am aware of two primary ways of conducting digital contact tracing: Bluetooth-enabled and location-based. In my testimony, I noted that Private Kit, a location-based app, takes some privacy protective steps to ensure the privacy of a user's identity and location. However, I also noted that unless a single, popular app (or platform) emerges, contact tracing using methods like this could be limited in their usefulness. Meanwhile, after testimonies were submitted, a single platform did emerge in the form of Apple and Google announcing a partnership to enable apps to make use of Bluetooth for contact tracing. The partnership will help ameliorate the limitations of a single app by making Bluetooth data available securely and anonymously to qualifying apps. Both of these methods involve potentially sensitive information (location data or Bluetooth information that may identify a device), using an architecture that limits the transfer of that data in identifiable form to a central authority or database. Ideally, the resulting application programming interface (API) enabling access to data generated by devices across the nation allows developers to harness actionable data while blocking their or others' ability to reidentify individuals. Relevantly, the API could enable app makers that serve first responders, Beyond Lucid Technologies, to access the data and plug it into their COVID-19 tool.

It should also be noted that the Apple and Google partnership described above involves the storage of Bluetooth keys on people's devices, which are mainly protected with strong encryption.

Without this technical measure, unauthorized bad actors could to retrace those keys and reidentify individuals or leak a set of stored keys to the public in readable format. Therefore, encryption plays an important role in the development of data-driven tools to combat COVID-19. We commend you for your steadfast defense of strong encryption and share your concerns with proposals that would require companies to build encryption vulnerabilities specifically for law enforcement investigators. These vulnerabilities would ultimately weaken encryption that protects a wide range of data, creating an opening for bad actors to exploit COVID-19 data--not just a set of keys for well-intentioned actors like American law enforcers.

54. 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?

While we are not experts in supply chain issues, we observe that a global pandemic presented a unique challenge to American PPE supply chains. In a national or regional outbreak, the inability to provide surge capacity could be offset by outsourcing supply to China or Mexico. A national stockpile is therefore likely a wise course of action to properly prepare for another public health emergency like COVID-19. As for help from digital tools and big data, App Association members often help other companies digitize processes to make back-end functions more efficient and mobile. For example, our most recent State of the App Economy Report points to Convoy, which helps companies book shipping trucks to deliver products. The platform allows companies to optimize their supply chains by drawing on relevant data and insights. Digitization of supply chain management is a key to success for a number of industries, and medical supply chains are likely no different in that respect.

Meanwhile, healthcare systems, manufacturers, and tech companies are pitching in to find ways around the shortage. One of our members, the Medical Society of Northern Virginia, has partnered with Model Space to 3D print face shields for physicians in northern Virginia. Partnerships like this are cropping up all over the nation thanks to the availability of open source face shield designs for 3D printers. These are a complement to a robust and flexible supply chain for medical equipment as opposed to a replacement for it, but they are an important option as supply chains fail to deliver personal protection equipment.

55. 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?

Many of our member companies have had frustrating and fruitless experiences with the Paycheck Protection Program (PPP) application process. In one case, the local lender received the application and simply never responded. In another case, a member company was simply unable to get through to anyone at the local bank. Large datasets could help identify areas where demand for assistance will likely be highest so that resources can be deployed to assist lenders with the application process to those parts of the country.

Sen. Rosen

56. 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?

Voluntarily donated data has the benefit of being consistent with legal and societal expectations that its use be sanctioned by the person to whom it pertains. The problem we encounter is that not everyone is in a position or willing to volunteer data about themselves. Therefore, the data is incomplete and ultimately biased.

- How can we use donated data to support our response to this pandemic and future similar public health issues?

Donated data could show the general health of those to whom it pertains. It could, therefore, be used to track how COVID-19 impacts people with varying levels of health and fitness; across different parts of the country; with different ethnic backgrounds; which underlying chronic conditions puts patients at greatest risk of getting very sick from COVID-19; and a variety of other inferences.

- What privacy guardrails are needed to ensure that this data is collected and analyzed safely and anonymously?

Congress should pass a comprehensive set of federal privacy rules to impose a set of guardrails. A national privacy law should appropriately restrict companies from further, unexpected use of these data sets with use limitations and data minimization requirements. To the extent companies collect sensitive personal data for the purpose of addressing the COVID-19 crisis, it should be up to the person to whom it pertains (if applicable) whether the data continues to have a use beyond addressing the crisis. It may be useful for the person (e.g., as a patient interested in their own health) to continue to have the company analyze it for them. However, government agencies should also be under strict limitations on further use of data collected to address the crisis. For the most part, the data should cease to be available to government actors for purposes beyond those for which it was collected.

- What are the gaps we need to consider when analyzing such data?

There may be gaps in data regarding Americans who are unable to access smartphones or internet connections. The most effective way of reaching populations, including elderly Americans, who do not use smart devices during the COVID-19 crisis may be through their healthcare provider or retirement community. Accurate reporting by community health centers (RHCs), federally *qualified community health centers (FQHCs), and skilled nursing facilities (SNFs) as to infection rates, symptoms, and resource needs is therefore critically important.*

57. Location tracking services serve as a powerful tool in understanding the movement of the coronavirus. Anonymized, aggregated data from GPS, Wi-Fi, and Bluetooth technology on our mobile devices can provide insights into how social distancing and shelter-in-place measures are changing people's behavior. A number of companies have come forward to help in the fight against the coronavirus, working to analyze and share these insights with governments on the local, state, and country level. They have stressed that the data collected is stripped of personally identifiable information.

    But according to recent news investigations, researchers have developed a machine learning model that can correctly re-identify 99.98% of individuals in anonymized data sets with just 15 demographic attributes. In other studies, researchers used credit card meta data and with four random pieces of information were able to re-identify 90% of the customers.

    - What data security steps are your member companies/your company taking to ensure anonymized and aggregated data remain anonymized?

I am aware of two primary ways of conducting digital contact tracing: Bluetooth-enabled and location-based. In my testimony, I noted that Private Kit, a location-based app, takes some privacy protective steps to ensure the privacy of a user's identity and location. However, I also noted that unless a single, popular app or platform emerges, contact tracing using methods like this could be limited in their usefulness. Meanwhile, after testimonies were submitted, Apple and Google announced such a platform in the form of a partnership to enable apps to make use of Bluetooth for contact tracing. The partnership will help ameliorate the limitations of a single app by making Bluetooth data available securely and anonymously to qualifying apps. Both of these methods involve potentially sensitive information (location data or Bluetooth information that may identify a device), using an architecture that limits the transfer of that data in identifiable form to a central authority or database. Ideally, the resulting application programming interface (API) enabling access to data generated by devices across the nation allows developers to harness actionable data while blocking their or others' ability to reidentify individuals.

It should also be noted that the Apple and Google partnership described above involves the storage of Bluetooth keys on people's devices, which are mainly protected with strong encryption. Without this technical measure, unauthorized bad actors could to retrace those keys and reidentify individuals or leak a set of stored keys to the public in readable format. Therefore, encryption plays an important role in the development of data-driven tools to combat COVID-19. We commend you for your steadfast defense of strong encryption and share your concerns with proposals that would require companies to build encryption vulnerabilities specifically for law enforcement investigators. These vulnerabilities would ultimately weaken encryption that protects a wide range of data, creating an opening for bad actors to exploit COVID-19 data--not just a set of keys for well-intentioned actors like American law enforcers.

58. 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?

The National Science Foundation (NSF) has an opportunity to support science-driven responses to COVID-19 through small companies like ACT | The App Association members, which are helping out in important ways. The examples I alluded to in my testimony, including the appendix, are just a few. Some, like Rimidi, are creating screening tools for health systems while others, like Particle Health, are creating platforms to create secure access for patients to complete healthcare records in usable formats. Last year, the NSF awarded another of our member companies, BadVR, with a grant to fund the development of a product that uses virtual reality to provide immersive data visualization experiences. The grant comes from the NSF Seed Fund, and we encourage the federal government to use this and resources like it to fund similar small business projects to further our capabilities in the fight against COVID-19 and future national health emergencies. Now more than ever, thanks to the entry of large software platforms and the wide availability of fast, reliable internet connectivity, small businesses are well positioned to play a role in cutting-edge responses to emergent societal and health problems.