CENTER FOR DEMOCRACY & TECHNOLOGY

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

Answers by Michelle Richardson, Director, Privacy and Data Project at the Center for Democracy and Technology

April 15, 2020

Chairman Wicker

1. As the chairman points out, app-based programs are proliferating and will likely draw on increasingly large or diverse datasets. Regarding privacy, apps that do not transfer personal information are best of class. Those that need personal information should be subject to strict purpose limitations so data cannot be used for non-coronavirus applications. As for effectiveness, there is no reliable data available at this time. Even though location and proximity tracing apps have been deployed in other countries, their impact has not been disentangled from contemporaneous efforts like widespread testing, compulsory quarantines, public information on the movement of infected individuals, and other responses.

2. We do not believe that privacy and effectiveness are inversely proportional. Given the extraordinary resources that U.S. companies are investing in the coronavirus response, it is not a tradeoff we need to accept. In fact, excess data collection can often hide useful 'signals' behind a lot of data 'noise'. Data collectors should have a clear idea of what data they want and why. This will encourage minimal data collection, strong data limitations, and result in the best health outcomes.

3. A comprehensive privacy law would have likely had several effects. First, it would have encouraged companies to conduct research in privacy protective ways. For example, the Chairman's draft bill includes protections for public interest research that is necessary, proportionate and limited in purpose. It also excludes aggregate and de-identified data from its scope altogether. To maximize data use while receiving liability protection, companies would be more likely to commit to these methods. Second, under these protections people would likely feel more comfortable sharing their personal information. Knowing that there are clearer and more meaningful rules - including a way to enforce them - would encourage people to take part in voluntary data sharing that may currently feel too risky.

4. We recommend that location tracking use aggregated and anonymized data whenever possible. Less stringent de-identification tactics - such as creating a pseudonymous identifier -

are not sufficient for such a sensitive data set. Because it is so easy to re-identify individual location data, it's collection should be strongly disfavored. We are still working to understand how to effectively use anonymized, de-identified, and aggregate location data, but one area of benefit is allowing public health officials to identify and compare, in aggregate, the effectiveness of social distancing measures.

5. Data collected or shared during this health emergency should only be used to inform the response to the COVID-19 pandemic. The data should not be repurposed or retained for any other reason. Once the immediate public health crisis has passed, data collected by companies and the government should only be used by researchers for the sole purpose of learning from this episode and planning for future occurrences. Otherwise, the data should be destroyed. This is crucial for maintaining public trust and hence public health. Without these controls the public is less likely to share data or work to actively subvert data collection methods.

Senator Thune

6. Now more than ever, technology use is just not optional. Consumer-facing products are facilitating work, learning, and social connections that would not be possible while complying with self-quarantine efforts. It is regrettable that only those of us who live in California or Nevada have statutory protections for the vast amount of data that is being created by and about us right now. Avoiding a patchwork approach would not only guard against inconsistent laws, but provide protection to people who may be waiting for years for action at the state-level otherwise. It is important that all laws include transparency provisions so consumers can have a clear understanding of data collection practices of companies like Zoom. But it's more important that any law also put substantive limits on data use, and not put the burden on the public to understand data collection.

7. Oversight of data practices in the coronavirus response will be difficult. One way Congress may encourage the minimization and deletion of data after the immediate response winds down is to continue inquiring directly with companies about their practices and receiving a public response. Another crucial component will be to increase the funding of regulators such as the FTC so that they can provide meaningful oversight.

8. As your question notes, there are several different location and proximity tracking proposals percolating through the U.S. technology sector now. The opinion-editorial you reference encourages the U.S. to adopt any and all of them. We would recommend against directly and surreptitiously collecting location information from telecommunications or technology companies. And while we strongly prefer apps or programs that are voluntary, there is a very real risk that employment, education, housing, or other opportunities may be informally

conditioned on the use of these apps, even if they are voluntary. With those considerations, the MIT app does appear to provide comparatively better privacy protections, but we continue to flag there are unresolved questions about the effectiveness of such apps.

9. It is commendable that tech companies like Apple and Google have developed public facing tools to help individuals screen for COVID-19. Both of these companies also accompanied their offerings with privacy policies. However, in the absence of a national privacy law, the key privacy safeguards for consumers who chose to use these websites are set by the companies in their terms of service and use. In the event that consumer data is subsequently inappropriately sold or misused (in violation of one of those terms), it may be possible for the FTC to step in.

10. N/A

Senator Blunt

11-14. Identifiable personal information may be needed in some instances to track the spread of the coronavirus, conduct medical research, or otherwise marshall resources to the individuals who need it. Disease surveillance conducted by epidemiologists and medical professionals to those ends is conducted under long standing medical privacy and ethical principles that do not encourage the mass collection of data. Guidance issued by the Center for Disease Control, the Word Health Organization, and academics instead focus on tried and true tracing techniques, which may include the use of technology but do not count on it to create proxies for effective public health measures.

When considering tracking and tracing efforts, we recommend that U.S. officials and companies be more discerning with app-based location and proximity tracing. It is possible or even likely that some of the activities conducted overseas would be found to violate the Fourth Amendment here. We acknowledge that some of the emerging proposals for voluntary bluetooth proximity tracking and notification appear to incorporate significant privacy protections and are preferable to efforts that compel companies and individuals to turn over data.

We caution, however, that there is no reliable data about such apps' effectiveness at this time. Even though location and proximity tracing apps have been deployed in other countries, their impact has not been disentangled from contemporaneous efforts like widespread testing, compulsory quarantines, public information on the movement of infected individuals, and other responses. Some may argue "effectiveness" is not important if at least *some* people are being notified of a possible interaction with an infected person. But these apps are being represented as a proxy for one's exposure and may result in people relying on data that is both

over- and underinclusive. The result may be creating an unnecessary panic, a burden on the health care system, or a false sense of security that the problem is under control.

Senator Cruz

15-17. N/A

Senator Moran

18. There is no one agreed upon definition of de-identification, but it includes technical and administrative processes to prevent an individual's identity from being associated with specific information. We refer you to NIST's 2015 paper, *De-Identification of Personal Information*, as a useful source, and note that anonymization and aggregation are usually considered the most privacy protective measures. Different privacy bills have adopted different definitions of de-identification and have assigned different consequences for data being de-identified. We recommend that at the least, your definition require entities to make public promises about not re-identifying data, contractual obligations with third parties and service providers to ensure data will not be re-identified, and ongoing due diligence requirements for first parties that share de-identified data.

19. We commend the Committee for holding this hearing and encourage it to continue oversight efforts in the months ahead. New proposals for data collection and use are emerging quickly and it will only be through continued inquiry that Congress will be able to understand corporate and government practices. Our written testimony lists eight factors that should be used to evaluate any tool and on a larger scale, the sum of technological efforts.

20-21. We recommend that U.S. officials and companies be more discerning with app-based location and proximity tracing. It is possible or even likely that some of the activities conducted overseas would be found to violate the Fourth Amendment here. We acknowledge that some of the emerging proposals for voluntary bluetooth proximity tracking and notification appear to incorporate significant privacy protections and are preferable to efforts that compel companies and individuals to turn over data.

We caution, however, that there is no reliable data about such apps' effectiveness at this time. Even though location and proximity tracing apps have been deployed in other countries, their impact has not been disentangled from contemporaneous efforts like widespread testing, compulsory quarantines, public information on the movement of infected individuals, and other responses. Some may argue "effectiveness" is not important if at least *some* people are being notified of a possible interaction with an infected person. But these apps are being

represented as a proxy for one's exposure and may result in people relying on data that is both over- and underinclusive. The result may be creating unnecessary panic, burden on the health care system or a false sense of security that the problem is under control.

Senator Blackburn

22. The United States does not have a federal privacy law that covers all health data. HIPAA only offers privacy protections to data in the possession of specific covered entities related to the provision and distribution of health care and related services. To the extent the administration recently relaxed HIPAA enforcement, we encourage Congress to monitor the situation closely. As for the countless data sets that also contain health data, or serve as reliable proxies for personal health data, outside of HIPAA, CDT believes that the best way to protect that information is with a comprehensive national privacy law. This law should treat health information as sensitive data that can only be used for the purpose for which it is collected, and not shared, sold, or repurposed by the data holder.

23. N/A

24. To the extent that some of the data is aggregated and anonymized, deletion is not as crucial. Personal information should not be retained for longer than necessary, however, and purpose limitations should always prevent data created or collected for the coronavirus response from secondary uses. These rules should apply to both corporate and government actors.

25. We appreciate the Senator's recognition that there are many ways for big tech to contribute to the coronavirus response, and the consortium is a good example. At this time, the consortium has authorized 16 research projects at universities across the country. We encourage technology companies to find additional ways to meet the stated needs of health professionals.

26. We recommend that U.S. officials and companies be more discerning with app-based location and proximity tracing. It is possible or even likely that some of the activities conducted overseas would be found to violate the Fourth Amendment here. We acknowledge that some of the emerging proposals for voluntary bluetooth proximity tracking and notification appear to adopt significant privacy protections and are preferable to efforts that compel companies and individuals to turn over data. However, there are still unanswered questions about whether such apps give individuals meaningful and actionable information, and we encourage the U.S. to better understand the apps accuracy in real world conditions before endorsing any one approach.

Senator Capito

27. As we discuss more thoroughly in our written statement, we encourage the adoption of privacy engineering principles that will help ensure privacy is baked into a product, instead of consumers and users being subject to lengthy privacy policies and individual controls. Such approaches include a preference for anonymization and aggregation, minimizing the collection, use, and sharing of information, purpose limitations, and eventually, deletion.

Senator Lee

28. From publicly available information, data processing for the coronavirus response ranges from the use of irreversibly aggregated data to more invasive sharing of more easily identifiable information. To the extent that privacy policies often distinguish between anonymized and personal information, some of the mobility analysis may fall under anonymized use of data. As for whether anyone consented to these specific uses, it appears to vary depending on the company and use case. For example, some companies have chosen to use location data sets that are affirmatively opt-in and provide explicit notice to users that their information may be used for public interest research. Others appear to be pulling data through application programming interfaces (APIs) that have been collecting and using data for other purposes without any meaningful consent from users.

This moment perfectly illustrates how notice and consent does not work in our modern tech landscape, and frankly never will again. Technology use is not optional, and individual management of the dozens or hundreds of apps, accounts, and websites we use everyday is impossible. As this committee works on federal privacy legislation, we recommend protecting sensitive data like location or health information with strict use limitations, subject to exceptions for public interest research that is conducted pursuant to field-recognized ethical and privacy guidelines. It would be a universal way to ensure that legitimate research goes forward without sanctioning riskier behavior that may have nothing meaningful to contribute in times like these.

29. It is incredibly difficult to discern which apps and tools are being developed, and which ones are just thought experiments. It is also next to impossible to track what coronavirus-related personal information is being created, used, and shared on the commercial market. We note that the headlines of late have focused on large American companies that are processing data in ways that seem to be privacy-protective, such as aggregated mobility data. We encourage Congress to continue its direct inquiries of companies that are less transparent too. Congressional inquiry often obtains information that isn't available otherwise.

30. We only note here that "contact tracing" is now being used to describe a number of very different activities. To the extent public health officials must use personally identifiable information to trace the contacts of those who have tested positive in order to monitor and offer services to down-chain individuals, this use of data is necessary and to be encouraged. More recently proposed methods of on-phone proximity tracking - such as bluetooth proximity - are potentially pro-privacy, but their effectiveness is unclear at this time. Repurposing personally identifiable location information, collected for purposes other than contact tracing, raises very serious and unresolved privacy questions. We urge the committee to strongly disfavor those uses and question whether they are in fact beneficial for public health.

31. No, not to our knowledge. Because reporting in this area is unclear, we encourage the Committee to directly ask the administration and federal agencies what requests are pending and what future requests are under consideration.

32. As we discuss more thoroughly in our written statement, we encourage the adoption of privacy engineering principles that will help ensure privacy is baked into a product, instead of consumers and users being subject to lengthy privacy policies and individual controls. Such approaches include a preference for anonymization and aggregation, minimizing the collection, use, and sharing of information, purpose limitations, and eventually, deletion.

We also encourage the Committee to inquire about whether these tools or tactics provide accurate and actionable information, whether they are representative of diverse populations, and whether they are operating in a transparent manner.

Senator Johnson

33. The Federal Trade Commission has oversight of commercial data practices, and may bring enforcement actions against practices that are unfair or deceptive. But we also commend Congress for asking important questions and demanding public answers from those who are processing our data. In many ways, our system is still stuck in a notice and choice model, and getting clear promises about data use is a way to empower users and permit the FTC to sanction those who do not follow through on them. When Congress considers comprehensive privacy legislation, we urge that it also include requirements for data security.

34. N/A

Senator Scott

35. When considering tracking and tracing efforts, we recommend that U.S. officials and companies be more discerning with app-based location and proximity tracing. It is possible or even likely that some of the activities conducted overseas would be found to violate the Fourth Amendment here. We acknowledge that some of the emerging proposals for voluntary bluetooth proximity tracking and notification appear to incorporate significant privacy protections and are preferable to efforts that compel companies and individuals to turn over data.

We caution, however, that there is no reliable data about such apps' effectiveness at this time. Even though location and proximity tracing apps have been deployed in other countries, their impact has not been disentangled from contemporaneous efforts like widespread testing, compulsory quarantines, public information on the movement of infected individuals, and other responses. Some may argue "effectiveness" is not important if at least some people are being notified of a possible interaction with an infected person. But these apps are being represented as a proxy for one's exposure and may result in people relying on data that is both over- and underinclusive. The result may be creating an unnecessary panic, a burden on the health care system, or a false sense of security that the problem is under control.

Ranking Member Cantwell

36. N/A

37. We recommend that U.S. officials and companies be more discerning with app-based location and proximity tracing. It is possible or even likely that some of the activities conducted overseas would be found to violate the Fourth Amendment here. We acknowledge that some of the emerging proposals for voluntary bluetooth proximity tracking and notification appear to incorporate significant privacy protections and are preferable to efforts that compel companies and individuals to turn over data.

We caution, however, that there is no reliable data about such apps' effectiveness at this time. Even though location and proximity tracing apps have been deployed in other countries, their impact has not been disentangled from contemporaneous efforts like widespread testing, compulsory quarantines, public information on the movement of infected individuals, and other responses. Some may argue "effectiveness" is not important if at least *some* people are being notified of a possible interaction with an infected person. But these apps are being represented as a proxy for one's exposure, and may result in people relying on data that is both

over- and underinclusive. The result may be creating an unnecessary panic, a burden on the health care system, or a false sense of security that the problem is under control.

38. Decentralized coronavirus responses come with benefits and drawbacks. One benefit of this model is that different apps or services may appeal to different communities, and treat data with a sophistication that a general purpose tool may not be able to. The University of Alabama launched a symptom tracking tool targeted at southern, rural, and underserved communities, for example, which is quite intentionally trying to fill gaps in existing programs. On the downside, decentralization makes oversight more difficult and often invites individuals with no relevant expertise to offer their own solutions. While this may be tolerable for low stakes commercialism or other common data use, it is absolutely not during a global pandemic. The platforms that are serving as a conduit for this information will have to thoroughly vet entities for legitimacy, and we encourage Congress to keep asking these crucial gatekeepers questions about how they are allowing consumer data to be used.

39. There is no one agreed upon definition of de-identification, but it includes technical and administrative processes to prevent an individual's identity from being associated with specific information. We refer you to NIST's 2015 paper, *De-Identification of Personal Information*, as a useful source, and note that anonymization and aggregation are usually considered the most privacy protective measures. Different privacy bills have adopted different definitions of de-identification, and have assigned different consequences for data being de-identified. We recommend that at the least, your definition require entities to make public promises about not re-identifying data, contractual obligations with third parties and service providers to ensure data will not be re-identified, and ongoing due diligence requirements for first parties that share de-identified data.

Senator Klobuchar

40. The U.S. does not have a federal privacy law that covers all health data. HIPAA only offers privacy protections to data in the possession of specific covered entities related to the provision and distribution of health care and related services. As for the countless commercial data sets that also contain health data or serve as reliable proxies for personal health data outside of HIPAA, CDT believes that the best way to protect that information is with a comprehensive national privacy law. Many of the concepts in your bill are consistent with a comprehensive regime that will regulate all commercial data, and CDT recommends that Congress lean heavily on purpose limitations when it comes to sensitive data like health information. This law should treat health information as sensitive data that can only be used for the purpose for which it is collected, and not shared, sold, or repurposed by the data holder.

41. Using data that does not accurately reflect all communities can cause resources to be distributed in a way that underserves certain demographics. There is already significant reporting that people of color are being infected by the coronavirus at higher rates, and are more likely to die from it. There is also data confirming that the ability to shelter in place, work from home, and participate in remote learning is closely correlated with race, socioeconomic status, and other demographic categories. While tech companies may not have created the underlying and systemic disparities that are making the coronavirus more dangerous for certain groups, all of these disparities are widely known and should be accounted for in the use of big data.

Senator Blumenthal

42. N/A

Senator Shatz

43. We recommend that U.S. officials and companies be more discerning with app-based location and proximity tracing. It is possible or even likely that some of the activities conducted overseas would be found to violate the Fourth Amendment here. We acknowledge that some of the emerging proposals for voluntary bluetooth proximity tracking and notification appear to incorporate significant privacy protections and are preferable to efforts that compel companies and individuals to turn over data.

As we mentioned elsewhere, we request Congress pay particular attention to how the notice and consent model has failed us in far less urgent scenarios, and be mindful that we cannot presume the "voluntariness" of app tracing as a meaningful check. There is a very real risk that employment, education, housing, or other opportunities may be conditioned on the use of these apps.

Senator Markey

44. We agree that the coronavirus epidemic underscores the need for a comprehensive privacy law in the U.S. It is regrettable that only those of us who live in California or Nevada have statutory protections for the vast amount of data that is being created by and about us right now. Your bill would discourage unnecessary collection, use, and sharing of sensitive personal information, limit the purposes for which data can be used, and empower individuals to control their data after it is collected by companies. We strongly encourage the committee to move forward with legislation this year, and look forward to working with you and your staff on synthesizing the many meaningful proposals before the committee now.

Senator Peters

45. N/A

46. We appreciate the many possible corporate and government activities that may contribute to the coronavirus response, many of which can be low tech or frankly no tech at all. While we do not have suggestions about what other tactics could be deployed here, we support thinking about the coronavirus response in holistic terms. Technology has a huge role to play here, but knowing its limitations will ensure it is a smart one.

Senator Baldwin

47-48. Data aggregation is an important tool for protecting privacy. However, it can also obscure important differences about certain demographics. We encourage companies to do additional processing of data where it is necessary to understand local variances, and the how and why of the coronavirus spread and response. For example, reporting that on average, an entire state or county has poor self-quarantine rate provides only so much actionable information. *Why* are they traveling - are they essential workers, or do they have to travel by car to reach a grocery store? *How* do we address those needs so that these communities can self quarantine? These are far more complicated questions but they are ones we need to answer if we are going to build a coronavirus response that serves everyone, regardless of where they live, their race, age, or socioeconomic status.

49-50. N/A

Senator Sinema

51-52.

53. Identifiable personal information may be needed in some instances to track the spread of the coronavirus, conduct medical research, or otherwise marshall resources to the individuals who need it. Disease surveillance conducted by epidemiologists and medical professionals to those ends is conducted under long standing medical privacy and ethical principles that do not encourage the mass collection of data. Guidance issued by the Center for Disease Control, the Word Health Organization, and academics instead focus on tried and true tracing techniques - which may include the use of technology, but do not count on it to create proxies for effective public health measures.

When considering tracking and tracing efforts, we recommend that U.S. officials and companies be more discerning with app-based location and proximity tracing. It is possible or even likely that some of the activities conducted overseas would be found to violate the Fourth Amendment here. We acknowledge that some of the emerging proposals for voluntary bluetooth proximity tracking and notification appear to incorporate significant privacy protections, and are preferable to efforts that compel companies and individuals to turn over data.

We caution, however, that there is no reliable data about such apps' effectiveness at this time. Even though location and proximity tracing apps have been deployed in other countries, their impact has not been disentangled from contemporaneous efforts like widespread testing, compulsory quarantines, public information on the movement of infected individuals, and other responses. Some may argue "effectiveness" is not important if at least *some* people are being notified of a possible interaction with an infected person. But these apps are being represented as a proxy for one's exposure and may result in people relying on data that is both over- and underinclusive. The result may be creating an unnecessary panic, a burden on the health care system, or a false sense of security that the problem is under control.

54-55. N/A

Senator Rosen

56. N/A

57. Data aggregation is an important tool for protecting privacy. However, it can also obscure important differences about certain demographics. We encourage companies to do additional processing of data where it is necessary to understand local variances, and the how and why of the coronavirus spread and response. For example, reporting that on average, an entire state or county has poor self-quarantine rate provides only so much actionable information. *Why* are they traveling - are they essential workers, or do they have to travel by car to reach a grocery store? *How* do we address those needs so that these communities can self quarantine? These are far more complicated questions, but they are ones we need to answer if we are going to build a coronavirus response that serves everyone, regardless of where they live, their race, age, or socioeconomic status.

Using data that does not accurately reflect all communities can cause resources to be distributed in a way that underserves certain demographics. There is already significant reporting that people of color are being infected by the coronavirus at higher rates, and are more likely to die from it. There is also data confirming that the ability to shelter in place, work

from home, and participate in remote learning is closely correlated with race, socioeconomic status, and other demographic categories. While tech companies may not have created the underlying and systemic disparities that are making the coronavirus more dangerous for certain groups, all of these disparities are widely known and should be accounted for in the use of big data.

58. N/A