740 15th St NW
Washington, DC 20005

3 April 2019

Thank you for the opportunity to respond to additional questions associated with the March 12, 2019 Senate Judiciary Committee hearing entitled *GDPR & CCPA: Opt-ins, Consumer Control, and the Impact on Competition and Innovation*. It was a privilege to share Mapbox's perspective on these important issues during the initial hearing, and we are grateful for the subsequent opportunity to address your office's questions and concerns in greater depth in this document.

Answers to the questions you posed have been added inline to the original document, in italicized text, in the hope that this will maximize readability and clarity. Please let me know if a different format would be preferable.

In response to question 6 I have also included the text of Mapbox's March 8, 2019 comment to the California Department of Justice in connection with the CCPA rulemaking process. This document encapsulates our perspective on the CCPA, and can be found at the end of this PDF.

Thomas Lee
Policy Lead, Mapbox
[redacted]

Questions for the Record from Senator Lindsey O. Graham
To Mr. Thomas Lee
U.S. Senate Committee on the Judiciary
GDPR & CCPA: Opt-ins, Consumer Control, and
the Impact on Competition and Innovation"
Submitted on March 15, 2019

1. What was your estimated initial cost (both time and expense) to become GDPR compliant?

   We spent approximately 400 hours of in-house attorney time in preparation. In addition, our outside counsel expenses were approximately $25,000. The engineering and business teams do not track their work on an hourly basis, but we believe they spent approximately a similar number of hours in preparation.

2. What are your estimated recurring annual GDPR compliance costs (both time and expense)?

   We estimate approximately 100-200 hours of ongoing compliance costs per year, with that number to grow as the company grows.

3. What is your estimated initial costs (both time and expense) to become CCPA compliant?

   Because the California Attorney General has not yet issued guidelines for CCPA, and because there remains the possibility that the California legislature will make substantive changes to CCPA before the implementation date that may or may not make our GDPR efforts sufficient, it is very difficult to estimate the potential costs of CCPA compliance. However, we would look to our GDPR costs as a ballpark.

4. What are your estimated recurring annual CCPA compliance costs (both time and expense)?

   Please refer to our answer to question 3.

5. Are you differentiating your products based on consumers or businesses in the EU and California?

   Mapbox does not differentiate its products based on whether a customer is in the EU or in California. However, we have found that dealing with potential customers who are impacted by GDPR requires an additional and more burdensome contracting process.

6. What are the specific areas of the CCPA that could have a negative impact on competition and innovation? What areas of the CCPA need more clarity, improvement,

or removal?

We believe that the CCPA could benefit from several improvements and clarifications, including its definition of "personal information" and how it relates to deidentified information; the inclusion of a grace period for data deidentification so as not to empower platform owners at the expense of smaller businesses; its definition of "consumer" and whether it includes data created during work-for-hire; its data deletion and portability requirements and their potential for enabling identity theft; and the practical utility of some disclosure requirements, such as its mandate that businesses set up toll-free phone numbers for user requests.

Mapbox submitted comments explaining our perspective on these issues to the California Department of Justice as part of the CCPA rulemaking process. We are attaching those comments in the hope that they will prove to be of interest.

50 Beale Street, Ninth Floor
San Francisco, CA 94105

8 March 2019

The following comments are submitted on behalf of Mapbox, a leading provider of map and location services, in response to a call for comments by the California Department of Justice regarding rulemaking associated with the California Consumer Privacy Act of 2018 (CCPA).

Mapbox considers the responsible stewardship of the data in our possession to be among our most important duties. The privacy of our customers' and users' personal data shapes our engineering, business and legal decisions on a daily basis. Unfortunately, this commitment is not shared by all parties in our industry. We therefore welcome California's leadership on this issue and your office's efforts to craft regulations that offer strong privacy guarantees without unduly burdening businesses that collect and use data ethically.

CCPA was drafted in haste, and although it has been improved by subsequent legislation, we believe the law still contains a number of provisions that are unclear, unwise or dangerous. It is our hope that your rulemaking process will address and ameliorate some of the following concerns.

The definition of "personal information" requires clarification

The statutory definition of "information that...is capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household" is vague and appears self-contradictory to Section 1798.145(a), which states: "(a) The obligations imposed on businesses by this title shall not restrict a business's ability to . . . (5) Collect, use, retain, sell, or disclose consumer information that is deidentified or in the aggregate consumer information."

However, there is no explicit carve-out from the definition of "personal information" for deidentified, pseudonymized, or in the aggregate consumer information, even though each of these terms is defined in the legislation. Instead, subsection (K)(2) of the definition of "personal information" states it "does not include publicly available information" but that "'Publicly available' *does not include* consumer information that is deidentified or aggregate consumer information" (emphasis added).

This confusing definition leaves companies to guess at how to comply. For example, it is routine for a website to keep a log of Internet Protocol (IP) addresses and access times for security purposes to detect malicious behavior. If the website stores the log of IP addresses separately from a log of visitor activity on the website (e.g., the website knows a user visits the homepage of the website, then the product page followed by the contact information page, but does not know the IP address, name, or any other information about the identity of the user), then is the log of the visitor activity "deidentified" information? Does it matter if the company randomly assigns each visitor a session ID for each visit? Does it matter if the company discards the IP addresses after 30 days?

When Mapbox worked to comply with GDPR, one point on which we needed outside guidance was the definition of personal information under the law. Various European regulatory authorities issued clarifying guidance ahead of the implementation date.

As companies prepare to comply with CCPA, it would be very helpful for your office to issue examples of specific scenarios in which information is not "personal information" and examples of specific deidentification techniques your office views as sufficient to qualify information as "deidentified."

Data deidentification often requires a grace period

On a typical day, Mapbox collects over one hundred and fifty million miles of anonymized telemetry data from users of our maps. This information allows us to offer real-time traffic predictions, detect unmapped roads, and informs many other tasks we perform to improve the services we offer.

We are able to collect this data in part because of the anonymization practices that we employ. Shortly after data is collected it is stripped of permanent identifiers and broken into smaller pieces, and the beginnings and ends of journeys are discarded. Through these measures we produce a dataset that is useful for improving our maps, but useless for identifying individuals. We do not collect end users' names, email addresses, phone numbers or similar personal information (we do receive IP addresses and related information in the course of providing our services, but we take steps to separate such information from other user information and to minimize its retention). We do not resell individuals' data, and all phases of our processing pipeline, including the anonymized data, are subject to strong encryption and access control policies.

We believe that these practices confer robust privacy protections and represent the objectives of the CCPA. The law attempts to lower the burden associated with possession of deidentified data. This is a wise and laudable goal: deidentified data typically poses substantially fewer risks to users than data that has not been so processed.

Unfortunately, even our systems might fail to meet the law's standards for deidentified data. There are two reasons why.

First, unlike some of our competitors, Mapbox does not control a major mobile operating system. This introduces technical limitations which necessitate that some deidentification processes occur on our servers rather than on users' devices.

Second, the collection of any data over the internet requires the disclosure of IP addresses. This is a fundamental aspect of how communication occurs on modern networks. The CCPA identifies IP addresses as a type of personal information that triggers the law's protections.

The CCPA defines "collection" but does so with insufficient precision. It identifies IP addresses as personal information, but the law's structure makes it implausible that its authors meant to identify all internet-transmitted information as triggering CCPA's strongest protections. These ambiguities will at minimum induce considerable uncertainty among those working to comply with the law. At worst, they leave open the possibility of tendentious readings of the statute that could make it difficult or impossible for smaller businesses like Mapbox to make good-faith efforts to deidentify user data in a way that comports with CCPA.

This situation could be improved both by clearer definitions and by identifying a reasonable grace period for processing and deidentification of data rather than tying it to the act of collection. We believe that long-term storage and/or resale of personal data represent the overwhelming majority of the data risk that concerns consumers and that motivated the authors of the CCPA.

The definition of "consumer" requires clarification

Mapbox recognizes the the importance of consumer privacy and the legislature's motivation in passing CCPA. At the same time, the definition of "consumer" as "a natural person who is a California resident" captures many situations involving persons we would not conversationally refer to as "consumers."

For example, we do not believe that concerns about employee information led the legislature to take up CCPA, and this leads to potential for confusion or business hardships. Employees have very different privacy interests than customers, and there are already existing regulations regarding employees that address the privacy interests of those individuals while acknowledging businesses' need to record and retain certain information about those individuals. Reconciling them with CCPA when the laws do not make direct reference to each other will impose considerable compliance burdens on businesses or hamper their effectiveness.

For example, a delivery service has a business need to track the movement and timing of deliveries made by its drivers. At the same time, the delivery service also has a business need to *avoid* disclosing that information in a convenient electronic format to a former driver who has left to work for a competitor. A business might also track the salary information of all employees, including past employees, for various financial and planning purposes. It would be unreasonable for an ex-employee to demand deletion of that information. Even requiring businesses to

analyze and respond to such requests is an unnecessary burden outside of the consumer-protection purposes of CCPA.

Similarly, we do not believe information about routine business contacts was among the concerns that motivated the CCPA's authors. Information like business phone number, business email, and business address are relevant to a business relationship. Businesses should not need to justify collecting and storing such information, and an employee should not be able to request deletion of business contact information, which may be contrary to the wishes or needs of the employer on whose behalf the business relationship was pursued.

We believe this office should issue a guideline that excludes employees or contractors of a business acting in their roles as employees or contractors from the definition of consumer in Section 3. This is the path taken by the Washington State legislature in its version of CCPA, Senate Bill 5376.

Data portability and deletion requirements pose risk to both businesses and consumers
In its original form as a ballot measure, the CCPA required the disclosure of the types of data that are collected and how they will be used. In statutory form, the CCPA requires the disclosure of the specific data collected. The law also extends a mechanism by which consumers can request that their data be deleted. These requests may also be made by an agent authorized to act upon the consumer's behalf.

Providing a means by which personal data may be deleted or disclosed substantially increases the risk faced by consumers relative to the simple disclosure of what kinds of data have been collected. Identity thieves and vandals are sure to make use of these new capabilities. Perhaps most worryingly, the CCPA defines the scope of "personal information" to include an entire household, creating the chilling possibility that the law could be turned against victims of domestic abuse.

These potentially dire consequences make the task of confirming a requester's identity a serious responsibility. This is likely to impose a substantial burden on businesses. This is doubly true in the case of requests made by an authorized agent: in such cases the businesses may be responsible for confirming both the user's identity and the veracity of the delegation of authority.

The problem of verifying requests is also likely to induce businesses to collect more sensitive data than they otherwise might. A business might have little need for a driver's license or social security number except in order to verify a user's identity upon receipt of a CCPA request. This will make the consequences of data breaches more severe, a result that is clearly at odds with the CCPA's objectives.

In the case of business-to-business ("B2B") companies like Mapbox, it is not clear that the CCPA's data export and deletion scheme is workable at all. Mapbox has relationships with customers and provides services to those customers' users. We typically do not have a direct

relationship with those users--they do not have Mapbox user accounts and we have no means of contacting them. We also do not have information like names, birth dates, phone numbers, and addresses that could be used for verification purposes. This makes the problem of identity verification all the harder.

We suggest four measures to address this problematic dynamic:

1. Create a safe harbor for businesses when they have a good faith suspicion that a request is illegitimate. In such circumstances a business should be empowered to deny the request or ask for more information in order to confirm its legitimacy.
2. Create a licensure regime for all agents authorized to make requests on consumers' behalf. When a request is made through such an agent, that agent should bear the legal responsibilities and risks associated with ensuring the request is legitimate.
3. If a business holds a consumer's personal information in connection with an account registered with that business by the consumer, the business should be entitled to require the consumer to log in to the account as a means of confirming a CCPA data export or deletion request's legitimacy.
4. If a business holds a consumer's personal information as the result of a consumer's interaction with another service--such as in the case of a shipping company ("the secondary vendor") holding a consumer's address in the course of the fulfillment of an online order with another business ("the primary point of contact")--the secondary vendor should be empowered to require that CCPA data export or deletion requests be filed with the primary point of contact. While a secondary vendor who elects this form of verification should be required to verify the identity of the primary point of contact, this should be an option for businesses legitimately attempting to minimize the amount of personal information they collect. This will reduce the risk of fraudulent requests being filed en masse against B2B companies; and will reduce the need of such businesses to retain additional personal data in order to comply with CCPA requests.

Disclosure requirements will be more useful if matched to their context

We welcome CCPA's enhanced disclosure requirements. Its authors' efforts to make the law accessible to all Californians are laudable. We understand this aim to be the motivation behind CCPA's requirement that businesses offer a toll-free number by which consumers may file requests.

For some businesses this requirement might make sense. In the case of our own business, it seems likely to confuse consumers. Mapbox does not typically communicate with customers or users by phone--not for sales and not for support. Our services and the ways in which users interact with them are fundamentally mediated by interfaces like smartphones and computers. In this context a telephone interface seems unhelpful at best, and perhaps even confusing.

We acknowledge the need to ensure that CCPA's guarantees are made apparent to users of services, but we believe that consumers will be better served by notice mechanisms that are harmonized with the nature of the services to which they apply.

In closing

We realize that in some cases the issues we have identified might require statutory changes. However, we understand that your office is in dialogue with the California legislature as that body continues to improve the CCPA. We therefore offer these suggestions in the hopes that they might inform the goal we all share: producing the best privacy law possible.

We welcome the Department's attention to this matter and thank you for your consideration of these comments. We look forward to working with the Department as it proceeds toward implementation of the CCPA.

Thomas Lee
Policy Lead, Mapbox
[redacted]

Kathleen Lu
IP and Open Data Counsel, Mapbox
[redacted]

740 15th St NW
Washington, DC 20005

3 April 2019

Thank you for the opportunity to respond to additional questions associated with the March 12, 2019 Senate Judiciary Committee hearing entitled *GDPR & CCPA: Opt-ins, Consumer Control, and the Impact on Competition and Innovation*. It was a privilege to share Mapbox's perspective on these important issues during the initial hearing, and we are grateful for the subsequent opportunity to address your office's questions and concerns in greater depth in this document.

Answers to the questions you posed have been added inline to the original document, in italicized text, in the hope that this will maximize readability and clarity. Please let me know if a different format would be preferable.

Thomas Lee
Policy Lead, Mapbox
[redacted]

Questions for the Record from Senator Chuck Grassley of Iowa

GDPR & CCPA: Opt-ins, Consumer Control, and the Impact on Competition and Innovation

Questions for the First Panel

1. Please briefly explain the importance of transparency and ensuring that consumers can make informed decisions about the information they share.

   Irresponsible use of consumer data can put people at risk and imperil their privacy. We believe that users deserve to know how data is collected and used by the apps and services they interact with.

2. Transparency is critical in ensuring that consumers can make informed decisions. That can become more complicated, however, as our lives are increasingly connected to the technologies around us, like autonomous vehicles. According to one report, by 2025 each person will have at least one data interaction every 18 seconds – or nearly 5,000 times per day.[1]

   a. How do we balance the need for transparency and informed consent with the reality of our increasingly data-connected daily lives?

      In our non-digital lives, we do not need to sign anything to be sure that medicine we buy is safe, that a cashier will make accurate change, or that a car won't break as it leaves the dealership.

      A similar approach can and should be applied to our digital lives. Data interchange that is common and safe should be governed by well-understood rules that can be counted on without the need for endless granular scrutiny. Bespoke agreements should be reserved for circumstances where data sharing is particularly sensitive or unusual.

   b. Should consumers have to consent to every data interaction throughout their day?

      No. The vast majority of our data interactions should be governed by predictable rules that users can count on to ensure their safety and privacy.

3. If Congress enacts federal data privacy legislation, how do we ensure that companies are still incentivized to innovate in their privacy and data protections, rather than just 'check the box' of regulatory compliance?

   While there is exciting work being done in fields like differential privacy protection and homomorphic encryption, we believe that the fundamentals of privacy and data protection have remained unchanged for some time. Practices like minimizing data collection, using widely-adopted encryption standards, and enforcing strong access control policies have been well-known best practices for decades.

   Many companies like Mapbox compete in the marketplace at least in part on the basis of privacy. Further advancement in security technology is desirable and can be incentivized by

[1] David Reinsel et al., *The Digitization of the World—From Edge to Core*, IDC (Nov. 2018).

providing carve-outs or otherwise immunizing companies that use industry best practices (or best practices designated by a regulator), such as anonymizing data . Providing incentives that discourage the collection of unnecessary data is even more effective and important, but given the pace of innovation it would be unwise for Congress to mandate specific technologies. But, speaking generally, we believe that data security and privacy requirements that are clear, pragmatic and strong are unlikely to inhibit innovation in these fields.

740 15th St NW
Washington, DC 20005

3 April 2019

Thank you for the opportunity to respond to additional questions associated with the March 12, 2019 Senate Judiciary Committee hearing entitled *GDPR & CCPA: Opt-ins, Consumer Control, and the Impact on Competition and Innovation*. It was a privilege to share Mapbox's perspective on these important issues during the initial hearing, and we are grateful for the subsequent opportunity to address your office's questions and concerns in greater depth in this document.

Answers to the questions you posed have been added inline to the original document, in italicized text, in the hope that this will maximize readability and clarity. Please let me know if a different format would be preferable.

Thomas Lee
Policy Lead, Mapbox
[redacted]

Questions for the Record for Tom Lee
From Senator Mazie K. Hirono

1. During the hearing, I mentioned that there is significant evidence that a consumer's privacy settings are "sticky," with consumer's rarely altering their default privacy settings.

   Do you agree that the vast majority of consumers rarely change their default privacy settings?

Yes. In general, a default value for a software configuration option--whether privacy-related or otherwise--will represent the most frequently observed value for that option in real-world use. The magnitude of this effect can vary depending on the specifics of the software and the option in question

2. In view of the "sticky" nature of privacy settings, my inclination is to have a system in which, by default, a consumer is considered to have opted out of data collection and a company can only collect that consumer's data if the consumer expressly opts in to data collection. I understand from the hearing that you do not support such an "opt-in" privacy regime.

   Please explain why you do not think an "opt-in" privacy regime is the right approach and how you propose to ensure that each consumer is aware that his or her data is being collected and that the consumer consents to that collection.

Opt-in approaches are superficially appealing, but in practice they offer few improvements from the dysfunctional notice and consent status quo.

Most apps and services require some user data to function, whether it's a user's location for a ridesharing pickup or an email address for password reset requests. Under privacy rules like the GDPR and CCPA, user data includes IP addresses, which are part of every transmission of information across the internet.

Today, users are typically presented with lengthy privacy policies and terms of service to which they must agree before establishing an account and beginning to use a service. These agreements are unreadable to most non-lawyers, and their length and ubiquity makes careful reading impractical even for those with the skills necessary to comprehend them. For users, it's a take-it-or-leave-it dynamic: if they wish to use the service they must enter into an agreement they do not fully understand.

Users routinely click "I accept" or sign contracts, whether online or offline, to open an email account, buy a new phone, or sign up at a gym, all without reading the lengthy terms. An opt-in approach to privacy offers few improvements. Because some user data is necessary for most or all services to function, under an opt-in regime users will still be presented with ubiquitous take-it-or-leave-it notices that they will habitually agree to without real understanding.

Moving data collection to opt-in could also have a chilling effect on the collection of data that poses little or no privacy risk and which has desirable uses. Mapbox's own collection of anonymized telemetry data could be an example of this. We have developed a rigorous methodology to prevent location data from ever being connected back to the individuals who contribute it via apps that use our mapping software. By collecting this anonymized telemetry data we are able to provide users with accurate directions ETAs, find unmapped roads, detect and prevent vandalism of our maps, contribute to open data resources, and sustain the success of our business--all without putting users' privacy at risk.

If given the chance to make this case to individual users--to explain our security practices and the socially beneficial outcomes that result from our collection of anonymized data--we believe that the vast majority would opt-in to participate. But we will never have that chance. An app using Mapbox technology typically depends not only on our software but that of many other companies. Users already face an impossible problem in managing and understanding their contractual relationships with each company that makes apps they use. It is not tenable to ask them to extend such efforts to those companies' vendors, too. We also note that some of our competitors do not face this challenge: because of their prominence, unique positions as owners of the Android and iOS platforms, user-facing business models and/or ability to induce users to opt-in via cross-subsidies, their data collection practices would be less impacted than ours under such a scheme. We believe that opt-in would tilt the playing field toward the handful of tech giants that dominate the market for user-facing apps.

When paired with strong privacy rules, opt-out is a better alternative. It can provide users with a single and understandable set of guarantees about how their data will be used, enabling users to feel secure about their privacy without having to keep track of dozens of different agreements. And it can enable companies that collect data ethically, safely and securely to do so on a level playing field, regardless of their business strategy or name recognition.

Opt-out is useless without strong privacy rules. Creating such rules will be a substantial undertaking for policymakers. But we believe that if this challenge can be met it will deliver a better outcome both for users and for businesses.

740 15th St NW
Washington, DC 20005

3 April 2019

Thank you for the opportunity to respond to additional questions associated with the March 12, 2019 Senate Judiciary Committee hearing entitled *GDPR & CCPA: Opt-ins, Consumer Control, and the Impact on Competition and Innovation*. It was a privilege to share Mapbox's perspective on these important issues during the initial hearing, and we are grateful for the subsequent opportunity to address your office's questions and concerns in greater depth in this document.

Answers to the questions you posed have been added inline to the original document, in italicized text, in the hope that this will maximize readability and clarity. Please let me know if a different format would be preferable.

In response to question 4.b. I have also included the text of Mapbox's March 8, 2019 comment to the California Department of Justice in connection with the CCPA rulemaking process. This document encapsulates our perspective on the CCPA, and can be found at the end of this PDF.

Thomas Lee
Policy Lead, Mapbox
[redacted]

Thomas Lee
Policy Lead
Mapbox
Questions for the Record
Submitted March 19, 2019

QUESTIONS FROM SENATOR BOOKER

1. Marginalized communities, and specifically communities of color, face a disproportionate degree of surveillance and privacy abuses. This has been the case since the Lantern Laws in eighteenth-century New York City (requiring African Americans to carry candle lanterns with them if they walked unaccompanied in the city after sunset) up through the stop-and-frisk initiatives of more recent years.

   There are echoes of this tradition today in the digital realm as marginalized communities suffer real harm from digital discrimination. For example, in recent years we have seen many instances of housing discrimination and digital redlining, employment discrimination through digital profiling and targeted advertising, exploitation of low tech literacy through misleading notice and choice practices, discriminatory government surveillance and policing practices, and voter suppression and misinformation targeting African Americans and other minorities.

   I am concerned that—rather than eliminating the bias from our society—data collection, machine learning, and data sharing may actually augment many of the kinds of abuses we fought so hard to eliminate in the Civil Rights Movement. We need privacy legislation that is centered around civil rights.

   a. In your view, is a private right of action critical to protecting the civil rights of individuals affected by data collection and disclosure practices?

      While private rights of action have played an important role in advancing civil rights in our country, we believe they are not the only tool by which data privacy concerns may be addressed, including those with disparate impacts on vulnerable groups. Policies that minimize data collection or which encourage anonymization, for instance, may substantially ameliorate these concerns without necessitating some of the less desirable features associated with a PRA (e.g. larger impact on small firms).

   b. How easy is it for seemingly non-sensitive information like a ZIP Code to become a proxy for protected class or other sensitive information? How can that information be used to discriminate?

      Information that is or seems to be non-sensitive is often highly correlated with personal attributes that carry legal protection or demand enhanced sensitivity and discretion. Today the history of redlining and the HOLC maps that enabled it are

well known, thanks to the scholarship of Richard Rothstein and others. It is likely that a substantial amount of those maps' spatial specificity could be recreated using only a cruder measure like ZIP Code.

Unfortunately, it is not possible to create a definitive list of attributes that are safe and those that are sensitive, since it is often through those attributes' combination that statistical conclusions about an individual's characteristics can be drawn. In isolation facts like the type of car I drive, the app I most recently downloaded and the song I listened to last might reveal little about me, but in aggregate they could easily suggest a demographic portrait upon which prohibited discriminatory behavior could be based. The prospects for preventing such behavior via prohibitions on information collection are consequently poor: it is difficult to envision a plausible set of policies sufficiently expansive to eliminate these concerns. It may be that this challenge has to be tackled in other ways.

None of this is meant to suggest that privacy reform is not needed or that it will not help to address these concerns. But it is unlikely that any policy can be crafted that prevents information collection in such a way that such information can never be used to deduce protected characteristics.

c. Significant amounts of data about us are gathered by companies most people have never heard of. Do we need a registry of data brokers, similar to what Vermont established last year?

We think a registry of data brokers might be a wise national policy, but because Mapbox does not sell or license brokered personal information, we are not subject to the Vermont law and consequently do not have deep familiarity with its design or specific provisions.

However, we believe it is worth discussing the "companies most people have never heard of" criterion included in your question. Mapbox is such a company. This is by virtue of our business model: we sell software to businesses, not to consumers. In order to provide our services, we collect anonymized data from those businesses' users, including depersonalized telemetry data. We do so ethically, with stringent security and privacy practices.

We believe that users are overwhelmed by the number of entities offering them terms of service and privacy policies. Users deserve to know about the data collection methods that affect them. But they also deserve to know that their data is being collected responsibly and without putting them at risk, even if they do not take the time to investigate the minute details of the data ecosystem in which they live and work.

For this reason we encourage your office not to consider a company's lack of popular familiarity as a strike against them--such a policy would only further empower a handful of user-facing tech giants at the expense of startups and

business-to-business companies that are trying to compete. Rather, it is whether the data is collected and used responsibly that should be the determinant of whether a company is subject to registration, regulation or censure.

2. The tech journalist Kashmir Hill recently wrote a widely circulated article on her efforts to leave behind the "big five" tech companies—Facebook, Google, Apple, Microsoft, and Amazon. Using a VPN, she blocked all of the IP addresses associated with each company and then chronicled how her life changed. She experimented first by blocking individual companies, and then, at the end of the series, she blocked all five at once. Ms. Hill found that—to varying degrees—she could not get away. Repeatedly, her efforts to intentionally block one company created unpredictable ripple effects for engaging with other, seemingly unrelated, companies and services. Ms. Hill's article spoke to how pervasive these companies are and how much data they capture about us when we're not even (knowingly) using their services.[1]

    a. How would you respond to the following argument? "If people are uncomfortable with the data practices of certain tech companies, they simply shouldn't use their services."

    This proposition is a fundamental tenet of much of U.S. commercial and technology policy, which envisions most of our digital lives as a series of voluntary commercial exchanges. As Ms. Hill's reporting makes clear, this vision is increasingly at odds with the practical realities of our society thanks to the concentrated power of some companies; and the development of race-to-the-bottom dynamics in some lightly- or unregulated disciplines.

    b. What does providing consent mean in a world where it's extremely difficult to avoid certain companies?

    User consent remains an important element of our system: it establishes practices and guarantees, and in some cases it provides the basis for accountability. At the same time, it is reasonable to say that users' freedom, in a practical sense, has significantly diminished as the inconvenience of withholding consent has risen and the practicality of thoroughly reviewing user agreements has shrunk. These are some of the reasons why Mapbox requires our customers to offer their users the ability to opt-out from our data collection practices.

3. It would take each of us an estimated 76 working days to read all the digital privacy policies we agree to in a single year.[2] Most people do not have that much time. They might prefer

---

[1] Kashmir Hill, *I Cut the 'Big Five' Tech Giants from My Life. It Was Hell*, GIZMODO (Feb. 7, 2019), https://gizmodo.com/i-cut-the-big-five-tech-giants-from-my-life-it-was-hel-1831304194.

[2] Alexis C. Madrigal, *Reading the Privacy Policies You Encounter in a Year Would Take 76 Work Days*, ATLANTIC (Mar. 1, 2012), https://www.theatlantic.com/technology/archive/2012/03/reading-the-privacy-policies-you-encounter-in-a-year-would-take-76-work-days/253851.

something simple, easy, and clear—something much like the Do-Not-Track option that has been featured in most web browsers for years.

However, there is a consensus that Do-Not-Track has not worked, because despite the involvement and engagement of stakeholders across the industry, only a handful of sites actually respect the request. A 2018 study showed that a quarter of all adult Americans were using Do-Not-Track to protect their own privacy—and yet 77 percent of Americans were unaware that Google, Facebook, and Twitter don't respect Do-Not-Track requests.[3] Just last month, Apple removed the feature from its Safari browser because, ironically, Do-Not-Track was being used for browser fingerprinting, i.e., having the feature turned on was used to distinguish individual users and track them across the web.[4]

a. What purpose does a notice-and-consent regime serve if the most prominent consent mechanism is only regarded as a suggestion at best?

   In such a case it would serve very little purpose. However, we do not believe that Do-Not-Track (DNT) is the most prominent consent mechanism. Although an admirable effort, DNT never achieved widespread salience. Terms of service, user agreements, privacy notices and related disclosures, as well as the processes by which users agree to them, are far more ubiquitous and obtrusive from a user experience perspective. These agreements carry their own problems, but they are vastly more important for protecting users and communicating businesses' policies than DNT ever was.

b. How much faith should the failure of Do-Not-Track give us in the ability of the industry stakeholders to regulate themselves?

   The failure of DNT can reasonably be seen as cause for skepticism of industry's ability to self-regulate, particularly when self-regulation schemes are in tension with industry's economic interests. But DNT's failure can also reasonably be seen as cause for skepticism of technological solutions to the problems of data privacy and consent.

c. In your view, should this approach be abandoned, or would federal legislation requiring companies to respect the Do-Not-Track signal breathe new life into the mechanism?

   It is clear that DNT has failed in the absence of regulatory power behind it. But it is not clear that it would succeed if given such power. DNT would remain difficult to enforce, especially prospectively; would likely privilege users with the

[3] *The "Do Not Track" Setting Doesn't Stop You from Being Tracked*, DUCKDUCKGO BLOG (Feb. 5, 2018), https://spreadprivacy.com/do-not-track.
[4] Ahiza Garcia, *What Apple Killing Its Do Not Track Feature Means for Online Privacy*, CNN (Feb. 13, 2019), https://www.cnn.com/2019/02/13/tech/apple-do-not-track-feature/index.html.

time and education to learn about the feature, enable it, and avoid entreaties to disable it; and could easily inhibit the development of new companies and technologies that rely on data collection that poses little or no privacy risk. Although we understand the appeal of delegating choices about privacy entirely to the user, both to maximize freedom and to minimize the scope of an enforcement agency's mandate, we believe there are reasons to doubt the viability of such an approach, and that these reasons extend beyond a lack of industry compliance in similar efforts' recent history.

4. Given that California has enacted its own privacy legislation that will take effect next year, much of the discussion at the hearing centered on how a federal data privacy law will affect state-level efforts to regulate in the same space. However, most of our existing privacy statutes do not include provisions to overrule stricter protections under state law.[5] These preemption provisions are the exception rather than the rule, and became more prevalent starting in the 1990s in statutes like the Children's Online Privacy Protection Act of 1998, the CAN-SPAM Act of 2003, and the 1996 and 2003 updates to the Fair Credit Reporting Act.

    a. In your view, should a federal data privacy law preempt state data privacy laws? Why?

    Yes. Americans deserve consistent protections for their data and privacy. And businesses deserve a set of rules of the road that is unified and minimizes legal costs and compliance burden. As we have previously stated, we believe that America's privacy law should be strong and that it should be uniform.

    b. In your view, should a federal data privacy law implement the requirements of the California Consumer Privacy Act as a floor? If not, please explain the most significant change you would suggest.

    We believe that the California Consumer Privacy Act (CCPA) includes many good ideas and could serve as a productive starting point for a federal data privacy law. However, we note that the CCPA contains several elements that require or could benefit from improvement or clarification, including its definition of "personal information" and how it relates to deidentified information; the inclusion of a grace period for data deidentification so as not to empower platform owners at the expense of smaller businesses; its definition of "consumer" and whether it includes data created during work-for-hire; its data deletion and portability requirements and their potential for enabling identity theft; and the practical utility of some disclosure requirements, such as its mandate that businesses set up toll-free phone numbers for user requests.

[5] The following statutes do not preempt stricter protections under state law: the Electronic Communications Privacy Act, the Right to Financial Privacy Act, the Cable Communications Privacy Act, the Video Privacy Protection Act, the Employee Polygraph Protection Act, the Telephone Consumer Protection Act, the Drivers' License Privacy Protection Act, and the Telemarketing Consumer Protection and Fraud Prevention Act.

We are eager to do our part to improve the CCPA, and consequently we have submitted comments to the California Department of Justice as part of the CCPA rulemaking process. Those comments expand on the above concerns at length. We are including a copy of them in the hopes that they will prove to be of interest.

c. The specific wording of a proposed preemption provision will invite considerable debate in Congress and, ultimately, will still require courts to interpret and clarify the provision's scope. Should the Federal Trade Commission have notice-and-comment rulemaking authority to aid in the statute's interpretation and to clarify which types of state laws are preempted? Or, alternatively, is case-by-case adjudication of multiple state privacy laws preferable? Would rulemaking authority obviate the need for Congress to solve each and every preemption issue in drafting the text?

Given the pace of technological progress and the relative infrequency of revisions to our nation's privacy laws, we believe that empowering a regulatory body with rulemaking authority is probably wise. We do not have a perspective on FTC rulemaking's relationship to state law preemption; but the demonstrated uncertainty associated with judicial refereeing of other tech policy issues makes adjudication of competing state mandates unappealing. We do not have a perspective on the extent to which rulemaking authority would obviate the need for Congress to solve preemption issues statutorily.

d. The preemption language in, for example, the amendments to the Fair Credit Reporting Act was included as part of a heavily negotiated process in which consumers received a package of new rights in exchange for certain preemption provisions.[6] Rather than centering the federal privacy bill debate on the existence of a preemption provision, shouldn't our starting point be: "Preemption in exchange for what?" In other words, what basic consumer protections should industry stakeholders be willing to provide in exchange for preemption? Do the requirements of the California Consumer Privacy Act represent a good floor for negotiating preemption?

We do not believe that data privacy needs to be approached as an antagonistic, zero-sum negotiation. With the exception of a few bad actors, businesses and consumers can all expect to benefit from clearly defined expectations and rules of the road. Consumers will benefit from clearer guarantees and new rights related to data privacy, and businesses will benefit from a stable policy environment and respite from race-to-the-bottom dynamics that let unscrupulous competitors gain market advantages.

[6] The 1996 and 2003 amendments included, for example: new obligations on businesses to ensure the accuracy of reports, increased civil and criminal penalties, remedial rights for identity theft victims, and the right to free annual credit reports.

As stated above, we do believe that many of the ideas in the CCPA represent productive starting points for a federal privacy law.

5. At the hearing, several witnesses indicated that opt-out requirements that permit users to tell companies not to process and sell their data are more protective of data privacy and more conducive to the user experience, since they do not impose the "take it or leave it" dynamic that opt-ins tend to create. In your view, are opt-outs preferable to opt-ins in terms of both data privacy and user experience? Why?

Opt-in approaches are superficially appealing, but in practice they offer few improvements from the dysfunctional notice and consent status quo.

Most apps and services require some user data to function, whether it's a user's location for a ridesharing pickup or an email address for password reset requests. Under privacy rules like the GDPR and CCPA, user data includes IP addresses, which are part of every transmission of information across the internet.

Today, users are typically presented with lengthy privacy policies and terms of service to which they must agree before establishing an account and beginning to use a service. These agreements are unreadable to most non-lawyers, and their length and ubiquity makes careful reading impractical even for those with the skills necessary to comprehend them. For users, it's a take-it-or-leave-it dynamic: if they wish to use the service they must enter into an agreement they do not fully understand.

Users routinely click "I accept" or sign contracts, whether online or offline, to open an email account, buy a new phone, or sign up at a gym, all without reading the lengthy terms. An opt-in approach to privacy offers few improvements. Because some user data is necessary for most or all services to function, under an opt-in regime users will still be presented with ubiquitous take-it-or-leave-it notices that they will habitually agree to without real understanding.

Moving data collection to opt-in could also have a chilling effect on the collection of data that poses little or no privacy risk and which has desirable uses. Mapbox's own collection of anonymized telemetry data could be an example of this. We have developed a rigorous methodology to prevent location data from ever being connected back to the individuals who contribute it via apps that use our mapping software. By collecting this anonymized telemetry data we are able to provide users with accurate directions ETAs, find unmapped roads, detect and prevent vandalism of our maps, contribute to open data resources, and sustain the success of our business--all without putting users' privacy at risk.

If given the chance to make this case to individual users--to explain our security practices and the socially beneficial outcomes that result from our collection of anonymized data--we believe that the vast majority would opt-in to participate. But we will never have that chance. An app using Mapbox technology typically depends not only on our software but

that of many other companies. Users already face an impossible problem in managing and understanding their contractual relationships with each company that makes apps they use. It is not tenable to ask them to extend such efforts to those companies' vendors, too. We also note that some of our competitors do not face this challenge: because of their prominence, unique positions as owners of the Android and iOS platforms, user-facing business models and/or ability to induce users to opt-in via cross-subsidies, their data collection practices would be less impacted than ours under such a scheme. We believe that opt-in would tilt the playing field toward the handful of tech giants that dominate the market for user-facing apps.

When paired with strong privacy rules, opt-out is a better alternative. It can provide users with a single and understandable set of guarantees about how their data will be used, enabling users to feel secure about their privacy without having to keep track of dozens of different agreements. And it can enable companies that collect data ethically, safely and securely to do so on a level playing field, regardless of their business strategy or name recognition.

Opt-out is useless without strong privacy rules. Creating such rules will be a substantial undertaking for policymakers. But we believe that if this challenge can be met it will deliver a better outcome both for users and for businesses.

6. At the hearing, several witnesses also indicated that the Federal Trade Commission, and perhaps state attorneys general, should have primary enforcement authority for data privacy violations. In your view, what additional authority and/or resources would the FTC need to perform this function effectively?

   We do not have a perspective on new FTC authorities or the resources required to support them, except insofar as we believe an engaged and empowered regulator will be better for both business and consumers than a combination of statute and slowly-developing caselaw.

50 Beale Street, Ninth Floor
San Francisco, CA 94105

8 March 2019

The following comments are submitted on behalf of Mapbox, a leading provider of map and location services, in response to a call for comments by the California Department of Justice regarding rulemaking associated with the California Consumer Privacy Act of 2018 (CCPA).

Mapbox considers the responsible stewardship of the data in our possession to be among our most important duties. The privacy of our customers' and users' personal data shapes our engineering, business and legal decisions on a daily basis. Unfortunately, this commitment is not shared by all parties in our industry. We therefore welcome California's leadership on this issue and your office's efforts to craft regulations that offer strong privacy guarantees without unduly burdening businesses that collect and use data ethically.

CCPA was drafted in haste, and although it has been improved by subsequent legislation, we believe the law still contains a number of provisions that are unclear, unwise or dangerous. It is our hope that your rulemaking process will address and ameliorate some of the following concerns.

The definition of "personal information" requires clarification

The statutory definition of "information that...is capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household" is vague and appears self-contradictory to Section 1798.145(a), which states: "(a) The obligations imposed on businesses by this title shall not restrict a business's ability to . . . (5) Collect, use, retain, sell, or disclose consumer information that is deidentified or in the aggregate consumer information."

However, there is no explicit carve-out from the definition of "personal information" for deidentified, pseudonymized, or in the aggregate consumer information, even though each of these terms is defined in the legislation. Instead, subsection (K)(2) of the definition of "personal information" states it "does not include publicly available information" but that "'Publicly available' *does not include* consumer information that is deidentified or aggregate consumer information" (emphasis added).

This confusing definition leaves companies to guess at how to comply. For example, it is routine for a website to keep a log of Internet Protocol (IP) addresses and access times for security purposes to detect malicious behavior. If the website stores the log of IP addresses separately

from a log of visitor activity on the website (e.g., the website knows a user visits the homepage of the website, then the product page followed by the contact information page, but does not know the IP address, name, or any other information about the identity of the user), then is the log of the visitor activity "deidentified" information? Does it matter if the company randomly assigns each visitor a session ID for each visit? Does it matter if the company discards the IP addresses after 30 days?

When Mapbox worked to comply with GDPR, one point on which we needed outside guidance was the definition of personal information under the law. Various European regulatory authorities issued clarifying guidance ahead of the implementation date.

As companies prepare to comply with CCPA, it would be very helpful for your office to issue examples of specific scenarios in which information is not "personal information" and examples of specific deidentification techniques your office views as sufficient to qualify information as "deidentified."

Data deidentification often requires a grace period

On a typical day, Mapbox collects over one hundred and fifty million miles of anonymized telemetry data from users of our maps. This information allows us to offer real-time traffic predictions, detect unmapped roads, and informs many other tasks we perform to improve the services we offer.

We are able to collect this data in part because of the anonymization practices that we employ. Shortly after data is collected it is stripped of permanent identifiers and broken into smaller pieces, and the beginnings and ends of journeys are discarded. Through these measures we produce a dataset that is useful for improving our maps, but useless for identifying individuals. We do not collect end users' names, email addresses, phone numbers or similar personal information (we do receive IP addresses and related information in the course of providing our services, but we take steps to separate such information from other user information and to minimize its retention). We do not resell individuals' data, and all phases of our processing pipeline, including the anonymized data, are subject to strong encryption and access control policies.

We believe that these practices confer robust privacy protections and represent the objectives of the CCPA. The law attempts to lower the burden associated with possession of deidentified data. This is a wise and laudable goal: deidentified data typically poses substantially fewer risks to users than data that has not been so processed.

Unfortunately, even our systems might fail to meet the law's standards for deidentified data. There are two reasons why.

First, unlike some of our competitors, Mapbox does not control a major mobile operating system. This introduces technical limitations which necessitate that some deidentification processes occur on our servers rather than on users' devices.

Second, the collection of any data over the internet requires the disclosure of IP addresses. This is a fundamental aspect of how communication occurs on modern networks. The CCPA identifies IP addresses as a type of personal information that triggers the law's protections.

The CCPA defines "collection" but does so with insufficient precision. It identifies IP addresses as personal information, but the law's structure makes it implausible that its authors meant to identify all internet-transmitted information as triggering CCPA's strongest protections. These ambiguities will at minimum induce considerable uncertainty among those working to comply with the law. At worst, they leave open the possibility of tendentious readings of the statute that could make it difficult or impossible for smaller businesses like Mapbox to make good-faith efforts to deidentify user data in a way that comports with CCPA.

This situation could be improved both by clearer definitions and by identifying a reasonable grace period for processing and deidentification of data rather than tying it to the act of collection. We believe that long-term storage and/or resale of personal data represent the overwhelming majority of the data risk that concerns consumers and that motivated the authors of the CCPA.

The definition of "consumer" requires clarification

Mapbox recognizes the the importance of consumer privacy and the legislature's motivation in passing CCPA. At the same time, the definition of "consumer" as "a natural person who is a California resident" captures many situations involving persons we would not conversationally refer to as "consumers."

For example, we do not believe that concerns about employee information led the legislature to take up CCPA, and this leads to potential for confusion or business hardships. Employees have very different privacy interests than customers, and there are already existing regulations regarding employees that address the privacy interests of those individuals while acknowledging businesses' need to record and retain certain information about those individuals. Reconciling them with CCPA when the laws do not make direct reference to each other will impose considerable compliance burdens on businesses or hamper their effectiveness.

For example, a delivery service has a business need to track the movement and timing of deliveries made by its drivers. At the same time, the delivery service also has a business need to *avoid* disclosing that information in a convenient electronic format to a former driver who has left to work for a competitor. A business might also track the salary information of all employees, including past employees, for various financial and planning purposes. It would be unreasonable for an ex-employee to demand deletion of that information. Even requiring businesses to analyze and respond to such requests is an unnecessary burden outside of the consumer-protection purposes of CCPA.

Similarly, we do not believe information about routine business contacts was among the concerns that motivated the CCPA's authors. Information like business phone number, business email, and business address are relevant to a business relationship. Businesses should not need to justify

collecting and storing such information, and an employee should not be able to request deletion of business contact information, which may be contrary to the wishes or needs of the employer on whose behalf the business relationship was pursued.

We believe this office should issue a guideline that excludes employees or contractors of a business acting in their roles as employees or contractors from the definition of consumer in Section 3. This is the path taken by the Washington State legislature in its version of CCPA, Senate Bill 5376.

Data portability and deletion requirements pose risk to both businesses and consumers
In its original form as a ballot measure, the CCPA required the disclosure of the types of data that are collected and how they will be used. In statutory form, the CCPA requires the disclosure of the specific data collected. The law also extends a mechanism by which consumers can request that their data be deleted. These requests may also be made by an agent authorized to act upon the consumer's behalf.

Providing a means by which personal data may be deleted or disclosed substantially increases the risk faced by consumers relative to the simple disclosure of what kinds of data have been collected. Identity thieves and vandals are sure to make use of these new capabilities. Perhaps most worryingly, the CCPA defines the scope of "personal information" to include an entire household, creating the chilling possibility that the law could be turned against victims of domestic abuse.

These potentially dire consequences make the task of confirming a requester's identity a serious responsibility. This is likely to impose a substantial burden on businesses. This is doubly true in the case of requests made by an authorized agent: in such cases the businesses may be responsible for confirming both the user's identity and the veracity of the delegation of authority.

The problem of verifying requests is also likely to induce businesses to collect more sensitive data than they otherwise might. A business might have little need for a driver's license or social security number except in order to verify a user's identity upon receipt of a CCPA request. This will make the consequences of data breaches more severe, a result that is clearly at odds with the CCPA's objectives.

In the case of business-to-business ("B2B") companies like Mapbox, it is not clear that the CCPA's data export and deletion scheme is workable at all. Mapbox has relationships with customers and provides services to those customers' users. We typically do not have a direct relationship with those users--they do not have Mapbox user accounts and we have no means of contacting them. We also do not have information like names, birth dates, phone numbers, and addresses that could be used for verification purposes. This makes the problem of identity verification all the harder.

We suggest four measures to address this problematic dynamic:

1. Create a safe harbor for businesses when they have a good faith suspicion that a request is illegitimate. In such circumstances a business should be empowered to deny the request or ask for more information in order to confirm its legitimacy.
2. Create a licensure regime for all agents authorized to make requests on consumers' behalf. When a request is made through such an agent, that agent should bear the legal responsibilities and risks associated with ensuring the request is legitimate.
3. If a business holds a consumer's personal information in connection with an account registered with that business by the consumer, the business should be entitled to require the consumer to log in to the account as a means of confirming a CCPA data export or deletion request's legitimacy.
4. If a business holds a consumer's personal information as the result of a consumer's interaction with another service--such as in the case of a shipping company ("the secondary vendor") holding a consumer's address in the course of the fulfillment of an online order with another business ("the primary point of contact")--the secondary vendor should be empowered to require that CCPA data export or deletion requests be filed with the primary point of contact. While a secondary vendor who elects this form of verification should be required to verify the identity of the primary point of contact, this should be an option for businesses legitimately attempting to minimize the amount of personal information they collect. This will reduce the risk of fraudulent requests being filed en masse against B2B companies; and will reduce the need of such businesses to retain additional personal data in order to comply with CCPA requests.

Disclosure requirements will be more useful if matched to their context

We welcome CCPA's enhanced disclosure requirements. Its authors' efforts to make the law accessible to all Californians are laudable. We understand this aim to be the motivation behind CCPA's requirement that businesses offer a toll-free number by which consumers may file requests.

For some businesses this requirement might make sense. In the case of our own business, it seems likely to confuse consumers. Mapbox does not typically communicate with customers or users by phone--not for sales and not for support. Our services and the ways in which users interact with them are fundamentally mediated by interfaces like smartphones and computers. In this context a telephone interface seems unhelpful at best, and perhaps even confusing.

We acknowledge the need to ensure that CCPA's guarantees are made apparent to users of services, but we believe that consumers will be better served by notice mechanisms that are harmonized with the nature of the services to which they apply.

In closing

We realize that in some cases the issues we have identified might require statutory changes. However, we understand that your office is in dialogue with the California legislature as that body

continues to improve the CCPA. We therefore offer these suggestions in the hopes that they might inform the goal we all share: producing the best privacy law possible.

We welcome the Department's attention to this matter and thank you for your consideration of these comments. We look forward to working with the Department as it proceeds toward implementation of the CCPA.

Thomas Lee
Policy Lead, Mapbox
[redacted]

Kathleen Lu
IP and Open Data Counsel, Mapbox
[redacted]