March 12, 2019

GDPR & CCPA: Opt-ins, Consumer Control, and the Impact on Competition and Innovation
Dirksen 226 Senate Office Building

Testimony of Thomas Lee (Mapbox) before the Senate Judiciary Committee

Chairman Graham, Ranking Member Feinstein, and members of the committee, thank you for the opportunity to appear before you. My name is Tom Lee. I'm an engineer by training and I now lead policy at Mapbox. Today I'd like to talk about how our company approaches privacy, and how privacy reform should approach smaller companies like ours.

We make maps. Our customers are developers. From weather forecasts to messaging tools to major news sites, you've probably used our maps. In total, we serve over 520 million monthly active users.

Our users benefit from our maps, and we benefit from their use. By collecting GPS data we make our maps more accurate, detect traffic jams, and give better directions. We built these features knowing that we had a responsibility to put user privacy first. We work to minimize the information we collect. We anonymize what we do collect. We require customers to let users opt out. We encrypt data in transit and at rest. We apply strong access control policies. And we only use the data to make our products better. We're in the business of selling maps, not information about the people using them.

Our success proves that you can build a valuable business and protect user privacy at the same time. We are glad to see growing attention to this issue from lawmakers in this body, in state legislatures, and around the world. We think it's time for some rules of the road--common-sense ethical standards for anyone that asks users to trust them with personal data.

New regulations inevitably carry costs and risks, especially for smaller businesses like ours, which aren't among the names we're all used to seeing in headlines about privacy. I'd like to highlight some issues that deserve attention as you consider how to craft reform without harming competitiveness or innovation.

First, the burden imposed by a proliferation of varying privacy standards is real. Our small but mighty legal team has to handle customer contracts, patents, employee policies, vendor agreements, and scores of other issues. Proceeding through the GDPR compliance process cost us hundreds of hours of effort initially, and continues to introduce additional time and complexity as we negotiate deals with customers. Startups can not afford to multiply that cost by dozens of additional jurisdictions--especially if some of those future regulatory regimes prove to be in conflict with one another.

We believe that our nation's privacy laws should be strong and that they should be unified: we favor a single national standard. Avoiding a patchwork of state rules will not only help smaller businesses like ours, but will give Americans assurances that don't change when they cross a state border.

Second, a jumble of state privacy laws risks creating loopholes, oversights, and errors. It's easy to see why when you consider how much of the conversation on privacy has focused on the tech giants whose apps are used directly by billions of end users. The California Consumer Privacy Act (CCPA) is a good example of a law that contains many good ideas, but fails to fully imagine businesses like Mapbox. Some of our customers run vehicle delivery fleets, and they use our technology to monitor how efficiently those deliveries are being made. Under the CCPA, the drivers in those fleets could request data about their employers' operations, even if they've since left to work for a competitor. Exposing trade secrets isn't what the CCPA intended, and we're hopeful that this problem can be fixed. But we worry that similar oversights will be inevitable if state laws proliferate in the absence of a clear federal standard.

Third, poorly-designed reform could entrench big business and harm smaller companies. Unlike some of our competitors, we don't own a major mobile operating system. We collect anonymous data when people use maps in our customers' apps. The platform owners can collect data at a much lower level than this. Reform that fails to adequately protect secure and ethical data collection like ours risks creating uncertainty among our customers and a chilling effect. If that happens, the accuracy of our maps and driving directions will suffer. That would make it much harder to compete with those platform owners.

Finally, some well-meaning reforms could put users at greater risk by forcing the collection of more user data. Data export and deletion requirements, in particular, often fail to envision businesses like ours. We rarely have a direct relationship with end users. We don't know their names, emails, phone numbers or other personal details. All we collect is anonymized data; and metadata like IP addresses, which are part of any internet request. But many privacy reform efforts, including the CCPA and GDPR, name IP addresses as personal information, and include data export and deletion provisions. This combination opens the possibility of requests filed by identity thieves, vandals, and abusers. To reliably detect illegitimate requests we might need to collect much more personal information about users, putting us and them at greater risk. It would be ironic if privacy reform led to more collection of personal data rather than less.

I know some of these concerns are technical and specific. I mention these details only to make a larger point. We agree that Americans deserve stronger privacy guarantees. But the details are critically important, and it will be easy to get them wrong. This work needs to be pursued in a unified and careful manner, in a way that minimizes the opportunities for mistakes. Businesses subject to new rules will deserve detailed guidance about how to comply, and the people depending on those rules will deserve a system with the flexibility to respond to new problems in the future.

We think the work of privacy reform can bring real benefits to Americans, and we're eager to do whatever we can to support it. I thank you for the opportunity to appear before you today, and look forward to any questions you might have.

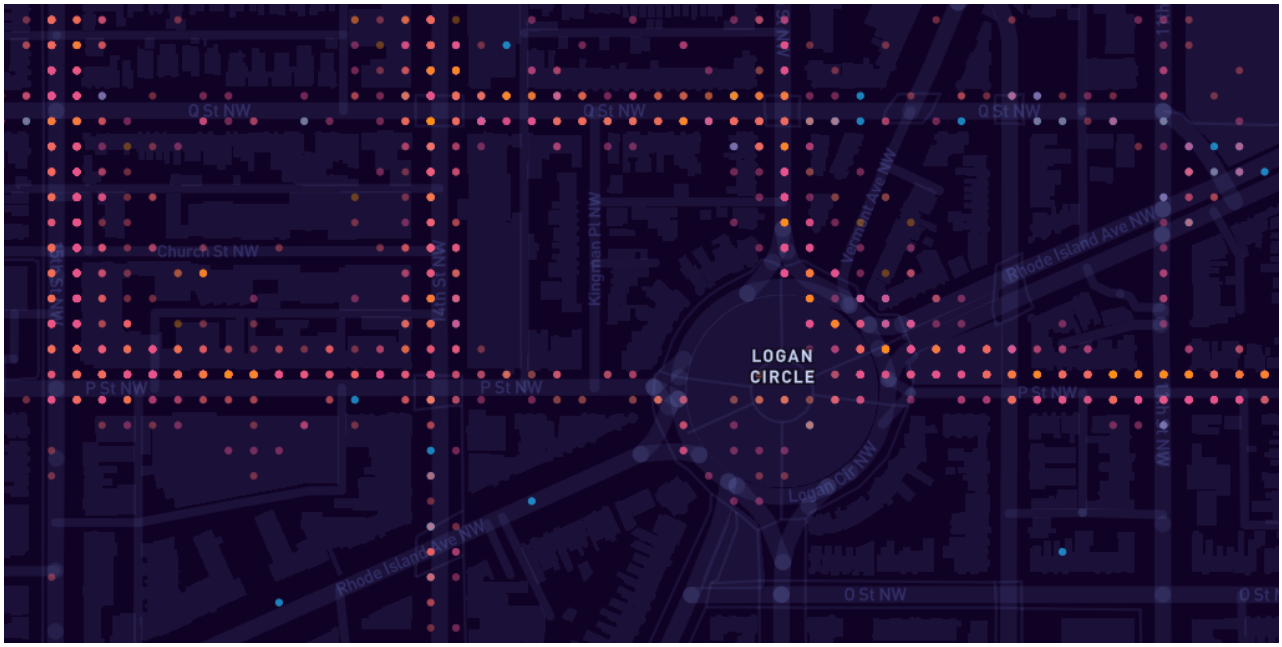
Prioritizing privacy when using location in apps

blog.mapbox.com/prioritizing-privacy-when-using-location-in-apps-f31cdec85fc9

Mapbox January 29, 2019

5 Recommendations for developers

Mapbox
Jan 29



By: Tom Lee

Location is finally a prominent part of the data privacy conversation. This is overdue. The physical location of your body in a given moment, and over time, is uniquely personal.

The future of location is exciting: connecting our digital lives to the real world is going to let us solve old problems and uncover new insights. But we have to approach that future responsibly. As more developers and companies experiment with these capabilities, it's essential that they also consider the ethical obligations that come with handling personal data—not only by their own teams, but by the third party services they depend on.

The privacy and dignity of our customers and users has been a north star for us since before we ever collected our first byte of user data. Our team has spent a lot of time considering these questions. Here are 5 privacy recommendations for developers and companies building with location:

1. De-identification & anonymization

Location can reveal a lot about an individual. Even after removing obvious identifiers like IP addresses or session tokens, a pattern of travel between specific places can contain private details about an individual's identity. The risk of data being connected back to an individual can be reduced by breaking location data into shorter segments that can't be linked back together. At Mapbox we also discard the beginning and end of each trace, as well as data that looks like it's from residential dwellings. This process leaves us with short segments that are useful for detecting traffic conditions, but useless for identifying individuals.

2. Fuzzing & aggregation

Anonymizing the data you collect can substantially reduce risk—but not eliminate it. That's because of what privacy researchers sometimes call the "Mosaic Effect." Put simply, it says that the privacy implications of a piece of data can't be fully understood in isolation. You must also consider how your data can interact with other data. Think of a Sherlock Holmes story, where the hero combines many seemingly innocuous details to reach an unexpected revelation. Because it's impossible to anticipate all of the datasets that might be combined with your data, it's hard to ever declare data conclusively safe.

But it's possible to reduce risk substantially by discarding, attenuating or obfuscating the signal that data contains. Aggregation is one way of doing this, and the US Census is a good example of how it can be put into action. The Census collects highly personal data from individuals, including details like race and income. But this data is only published in aggregate form, by tract, block, block group and so on.

If aggregation isn't an option, it may suffice to reduce the data's fidelity. Statistical techniques like differential privacy get researchers and computer scientists excited (trust us, we know). But techniques as simple as rounding geographic coordinates to a few decimal places can substantially reduce risk. Whether this is viable depends on what you need the data for: if you're showing users a weather forecast, you might require less precision than if you're helping them find a coffee shop. And of course this works both ways: being intentional about your business strategy lets you offer stronger privacy guarantees. Before Mapbox began collecting location data for our traffic product we carefully considered whether we would ever sell the data to advertisers. Our decision not to has let us put user privacy first.

3. Standardized encryption at rest & in transit

Data should be encrypted both as it's transmitted and when it's stored. This should always be done using widely adopted libraries that implement modern standards and which have been independently audited. Unless you employ a PhD cryptographer you should never think about using homegrown ideas or implementations (and even then you should probably think about it carefully).

The specifics of how to best implement encryption depend on your use case and threat model, but there are a variety of techniques that can harden a typical implementation, from certificate pinning to hardware security modules.

4. Access control

Scrubbing data's content and securing its form are essential, but your first and most important line of defense must be controlling access to what you collect. Implementing the principle of least privilege—by which staff may only access the resources they need—is a priority for any top-notch security team. And having such a team in place is a prerequisite to handling location data responsibly. Without carefully designed access control, both your users and your business will be at risk. This should include both procedures for onboarding and offboarding those who need access, and instrumentation to detect unexpected attempts at access or privilege escalation.

5. Give users choice

Your users deserve to know how their location data is being collected and used. This is not just the right thing to do: in more and more places, it's the law. Providing clear, unambiguous details about how data is shared and monetized lets users make an informed decision about whether to use your service—or, in some cases, whether to simply opt out of data collection (an option that we require all Mapbox customers to offer their users).

Is a user's data going to be scrubbed and used to build better maps? Analyzed in aggregate by urban planners? Or is their location going to be used to do things they don't like or approve of? The answers matter a lot.

The above list isn't meant to be exhaustive. But it reflects some of the answers that we've arrived at as we've considered how to do right by our customers and their users. And, like location services themselves, these techniques are both powerful and accessible.