Google’s Sidewalk Labs Plans to Package and Sell Location Data on Millions of Cellphones

Ava Kofman

Most of the data collected by urban planners is messy, complex, and difficult to represent. It looks nothing like the smooth graphs and clean charts of city life in urban simulator games like “SimCity.” A new initiative from Sidewalk Labs, the city-building subsidiary of Google’s parent company Alphabet, has set out to change that.

The program, known as Replica, offers planning agencies the ability to model an entire city’s patterns of movement. Like “SimCity,” Replica’s “user-friendly” tool deploys statistical simulations to give a comprehensive view of how, when, and where people travel in urban areas. It’s an appealing prospect for planners making critical decisions about transportation and land use. In recent months, transportation authorities in Kansas City, Portland, and the Chicago area have signed up to glean its insights. The only catch: They’re not completely sure where the data is coming from.

Typical urban planners rely on processes like surveys and trip counters that are often time-consuming, labor-intensive, and outdated. Replica, instead, uses real-time mobile location data.

As Nick Bowden of Sidewalk Labs has explained, “Replica provides a full set of baseline travel measures that are very difficult to gather and maintain today, including the total number of people on a highway or local street network, what mode they’re using (car, transit, bike, or foot), and their trip purpose (commuting to work, going shopping, heading to school).”

To make these measurements, the program gathers and de-identifies the location of cellphone users, which it obtains from unspecified third-party vendors. It then models this anonymized data in simulations — creating a synthetic population that faithfully replicates a city’s real-world patterns but that “obscures the real-world travel habits of individual people,” as Bowden told The Intercept.

The program comes at a time of growing unease with how tech companies use and share our personal data — and raises new questions about Google’s encroachment on the physical world.

> If Sidewalk Labs has access to people’s unique paths of movement prior to making its synthetic models, wouldn’t it be possible to figure out who they are, based on where they go to sleep or work?

Last month, the New York Times revealed how sensitive location data is harvested by third parties from our smartphones — often with weak or nonexistent consent provisions. A Motherboard investigation in early January further demonstrated how cell companies sell our locations to stalkers and bounty hunters willing to pay the price.

For some, the Google sibling's plans to gather and commodify real-time location data from millions of cellphones adds to these concerns. "The privacy concerns are pretty extreme," Ben Green, an urban technology expert and author of "The Smart Enough City," wrote in an email to The Intercept. "Mobile phone location data is extremely sensitive." These privacy concerns have been far from theoretical. An Associated Press investigation showed that Google's apps and website track people even after they have disabled the location history on their phones. Quartz found that Google was tracking Android users by collecting the addresses of nearby cellphone towers even if all location services were turned off. The company has also been caught using its Street View vehicles to collect the Wi-Fi location data from phones and computers.

This is why Sidewalk Labs has instituted significant protections to safeguard privacy, before it even begins creating a synthetic population. Any location data that Sidewalk Labs receives is already de-identified (using methods such as aggregation, differential privacy techniques, or outright removal of unique behaviors). Bowden explained that the data obtained by Replica does not include a device's unique identifiers, which can be used to uncover someone's unique identity.

However, some urban planners and technologists, while emphasizing the elegance and novelty of the program's concept, remain skeptical about these privacy protections, asking how Sidewalk Labs defines personally identifiable information. Tamir Israel, a staff lawyer at the Canadian Internet Policy & Public Interest Clinic, warns that re-identification is a rapidly moving target. If Sidewalk Labs has access to people's unique paths of movement prior to making its synthetic models, wouldn't it be

possible to figure out who they are, based on where they go to sleep or work? “We see a lot of companies erring on the side of collecting it and doing coarse de-identifications, even though, more than any other type of data, location data has been shown to be highly re-identifiable,” he added. “It’s obvious what home people leave and return to every night and what office they stop at every day from 9 to 5 p.m.” A landmark study uncovered the extent to which people could be re-identified from seemingly-anonymous data using just four time-stamped data points of where they’ve previously been.

There are also lingering questions about how Sidewalk Labs sets limits about the type and quality of consent obtained. As the past year’s tsunami of privacy breaches has shown, many users do not understand how closely they are being tracked and how often their data is being resold to advertisers or third parties or programs like Replica. “We need to do a better job in ensuring the type of express consent commensurate with sensitivity of data is actually being enforced when data is collected,” Israel noted. Consent has historically been defined by broad and vague terms of service, leveraging companies’ knowledge of intricate technical details at the expense of users too pressed for time to read — let alone understand — their jargon-laden privacy policies. The Times investigation found, for instance, that “the explanations people see when prompted to give permission are often incomplete or misleading.” Even while they may retain a broad right to sell or share location data in an opaque privacy policy, many apps do not explicitly tell their users that they are doing so.

It's difficult to evaluate who might be consenting when it's not clear where the data comes from. Sidewalk Labs explains that Replica's data is purchased from telecommunications companies and companies that aggregate mobile location data from different apps. "We audit their practices to ensure they are complying with industry codes of conduct," said Bowden. "No Google data is used. This extensive audit process includes regular reporting, interviews, and evaluation to ensure vendors meet specified requirements around consent, opt-out, and privacy protections."

Yet because the exact sources of data have not been revealed, it is unclear whether Replica draws from the ranks of unregulated apps that profit from indefinite privacy policies to continuously collect users' precise whereabouts. Publicly available documents from cities piloting or purchasing Replica offer conflicting information about Replica's exact sources of data. A document from the Illinois Department of Transportation describes Replica's data sources as "mobile carrier data, location data from third-party aggregators and Google location data, to generate travel data for a region." This data sample, it adds, "is not limited to Android devices" and "is collected from individuals for months at a time, allowing for a complete picture of individual travel patterns." In Portland, documents filed with its city council state that the data is sourced from "Android Phones and Google apps." Officials at the Portland Bureau of Transportation told Oregon Public Broadcasting that some of the sources of Sidewalk Lab's

mobile location data may also come from other sources, not yet known to them. Minutes from a regional transit planning meeting for Kansas City suggest that it's possible for Replica "to get data on things like Uber & Lyft," while a city PowerPoint states that the tool is "based off of Google data."

At stake with Replica is the value that can be produced by aggregating data about our movements and then selling it back to governments. The program was originally pitched by Sidewalk Labs "to support the development" of Quayside, the controversial "smart" city planned for Toronto's eastern waterfront. (A Sidewalk Labs spokesperson told The Intercept that there are no plans to bring Replica to Toronto.) Yet Torontonians have been watching Replica's plans closely. Some see the project as an example of the way the proprietary tools and techniques developed by Sidewalk Labs at Quayside might be exported — or imported — to other cities, without creating any additional economic benefits for the residents who have produced this data.

"Replica is a perfect example of surveillance capitalism, profiting from information collected from and about us as we use the products that have become a part of our lives," said Brenda McPhail, director of the Canadian Civil Liberties Association's Privacy, Technology, and Surveillance Project. "We need to start asking, as a society, if we are going to continue to allow business models that are built around exploiting our information without meaningful consent."