GOVLAB

cuebiq

What is Mobility Data? Where is it Used?
GovLab Brief: April 2021

Mobility data is data about the geographic location of a device passively produced through normal activity. Throughout the pandemic, public health experts and public officials have used mobility data to understand patterns of COVID-19's spread and the impact of disease control measures. However, privacy advocates and others have questioned the need for this data and raised concerns about the capacity of such data-driven tools to facilitate surveillance, improper data use, and other exploitative practices.

This piece attempts to allow for an informed discussion of the opportunities, risks, and challenges associated with mobility data by describing the notable types of mobility data. It uses the taxonomy provided by The Use of Mobility Data for Responding to the COVID-19 Pandemic, a report published by The GovLab and Cuebiq, supported by the Open Data Institute (ODI). Each data type includes an example of how the data has been used. As these explanations and examples suggest, mobility data refers not to one type of data but, many different ways to describe geographic location. It varies by collection method, format, and context. Consequently, mitigating risks and governing use requires organizations to develop similarly varied approaches.

Telecommunications companies (telecoms) are frequent providers of mobility data due to the wide availability of portable cellular devices and the wide coverage of their networks. Two notable types of telecom data include:

TELECOMMUNICATIONS DATA

Detail Records (CDR) are records of each time a mobile phone connects to a network when sending or receiving a voice call or SMS. A typical CDR contains the timestamp, call duration, caller identifier, recipient identifier, and the origin and destination of cell towers when a call between devices occured. CDR's specificity can be limited because the coverage of cell towers tends to be large. In the TELUS Data for Good initiative, a Canadian national telecom company gave supervised and guided access to its network mobility data to help the Natural Sciences and Engineering Research Council of Canada study the COVID-19 crisis.

x-Data Records (x-DRs) refers to internet connection detail records generated when a device connects to a mobile internet. x-DRs provide more limited information than CDRs, often only describing when the connection happened and the coordinates of the telecom antenna used to download the content. However, it tends to be more time-specific. In Santiago, Chile, researchers at the Universidad del Desarrollo, University of Turin, and University of Greenwich used x-DRs provided by the telecom provider Telefónica to assess mobility as the country responded to COVID-19.

SOFTWARE DEVELOPMENT KIT DATA

A Software Development Kit (SDK) is a piece of software integrated into a smartphone application and enables the app to collect location data by using the device's hardware, such as GPS, Bluetooth and Wi-Fi. It can either be collected directly or indirectly. Two notable data types are:

First-party, SDK-derived data is collected directly from a smartphone application without intermediary brokers. As it is data about one's own audience that an organization owns and manages itself, first-party SDK-derived data allows a company to exercise more control over how it collects data (including whether it allows individuals to opt in or out) and what it collects. However, it is limited to the scale and breadth of the organization's own operations and audience. In May 2020, a research team that included the University of Southampton and the Wuhan Center for Disease Control and Prevention used first-party, SDK-derived data from the Chinese search engine Baidu to measure the effects of non-pharmaceutical interventions in China in response to COVID-19. The team released the models it used for its analysis to inform future research and decision-making.

Third-party, SDK-derived data is data collected from smartphone apps by many sources aggregated into one dataset. This data can be rich (in that it describes all the items collected by the source organizations) and broad (in that it includes all organizations' audiences). However, the source organization has no little to no control over collection practices and may be unable to trace back data to a reliable source that collected data with the knowledge and consent of data subjects. In the Teralytics project the Zurich-based mobility data company provided dashboards to health authorities, emergency responders, mobility providers, and transportation planners based on its collected third-party SDK-derived data. These dashboards informed decision-makers on how people move through their communities.

GPS AND OTHER RESOURCES

Mobility data can also come from other sources, such as those that use geographic positioning service technology or in-device Bluetooth. Three types of data have generated public interest:

Vehicle GPS data describes the location of an automobile, ship, airplane, or other vehicle as indicated by portable devices placed in vehicles or hardware built into them. Geotab, a Canadian company that connects commercial vehicles to the internet and provides tools to analyze fleet activity, manages a COVID-19 Mobility Impact dataset that depicts changes in mobility due to the pandemic. This dataset supports decision makers planning to transport goods.

Bluetooth allows for short-range wireless communication between devices. It can be used to determine when and for how long a Bluetooth device came into proximity with another, though distance, object interference, and signal strength affect accuracy and precision. In April 2020, Google and Apple announced they would support health agencies by building cross-device Bluetooth-based contact tracing functionality into their devices.

Geotagged social media data refers to time and location data connected to videos, photographs, messages, and other content posted on social media. Social media data has been used to assess containment measures amid the COVID-19 pandemic Huang et al. analyzed over 580 million tweets worldwide using Twitter's public Streaming API to understand how global measures to reduce human mobility have impacted social media users

Authors: Andrew J. Zahuranec, Stefaan Verhulst, Andrew Young, Aditi Ramesh, and Brennan Lake