

Mobility Data Sharing: Challenges and Policy Recommendations

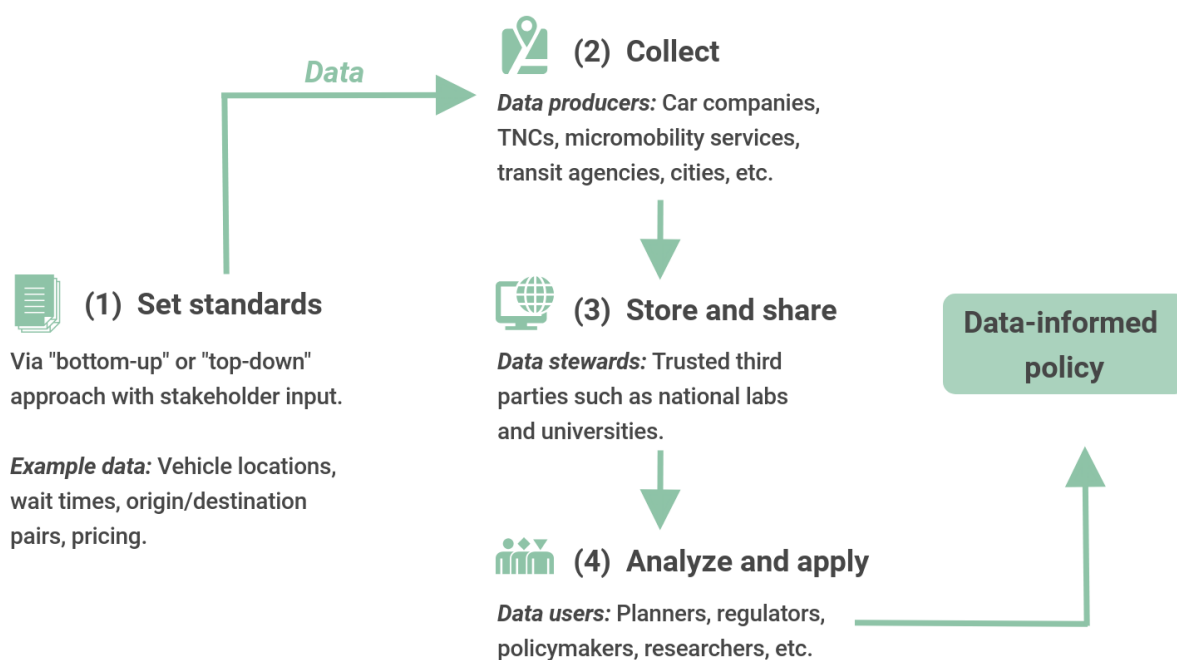
ISSUE PAPER
August 2019

Authors: Mollie D'Agostino, Paige Pellaton, and Austin Brown

Contributors: Hannah Safford, Kelly Fleming, and Cassidy Craford

Summary

Dynamic and responsive transportation systems are a core pillar of equitable and sustainable communities. Achieving such systems requires comprehensive mobility data, or data that reports the movement of individuals and vehicles. Such data enable planners and policymakers to make informed decisions and enable researchers to model the effects of various transportation solutions. However, collecting mobility data also raises concerns about privacy and proprietary interests. Oversharing and undersharing mobility data are both problematic. We argue that a middle-ground approach, in which data are shared in specific contexts and managed by a trusted third party, can capture the benefits of data sharing while minimizing risks.



A framework for effective mobility data sharing.

In this issue paper, we provide an overview of the top needs and challenges surrounding mobility data sharing. We present four policy strategies—at multiple levels of governance—to support needed data sharing while respecting the aforementioned concerns. These are:

- (1) Foster voluntary agreement among mobility providers for a set of standardized data specifications. (Cities, Regions)

- (2) Develop clear data-sharing requirements designed for transportation network companies (TNCs)¹ and other mobility providers.² (Cities, Regions, States)
- (3) Establish publicly held big-data repositories, managed by third parties, to securely hold mobility data and provide structured access by states, cities, and researchers. (States, Federal)
- (4) Leverage innovative land-use and transportation-planning tools. (Cities, Regions)

These recommendations are connected and should be developed in an integrated fashion. Standardized specifications will make data more interoperable and data sharing less expensive. Clear data-sharing requirements will ensure that data are used in a way that benefits society and supports mobility goals without imposing overly burdensome requirements on mobility providers. Well-designed structures for data storage and sharing are needed to protect sensitive and personal information. Existing tools can help facilitate all of the above.

¹ The California Public Utilities Commission defines transportation network companies as organizations that “provide prearranged transportation services for compensation using an online-enabled application or platform (such as smart phone apps) to connect drivers using their personal vehicles with passengers.” TNCs are also referred to as ridehailing services/providers.

² In this paper, we use the term “mobility providers” to refer to TNCs as well as companies that offer or manage emerging mobility services (such as scooter- and bike-sharing). Public transit is not included in this context.

Key Challenges in Mobility Data Sharing

It may seem surprising that mechanisms for mobility data are not already well established. On the one hand, there is intense unmet demand for mobility data from state, city, and regional planners to support better decision making. For instance, comments submitted jointly in 2019 by multiple San Francisco transportation agencies to the California Public Utilities Commission stated that “Without access to granular TNC data by zip code, census tract, or at the neighborhood level, [the San Francisco Municipal Transportation Agency] is constrained in its ability to engage in comprehensive transportation planning as required under the City’s Charter.”³

On the other hand, mobility providers, telecommunications companies, and mapping providers are engaged in large-scale mobility-data collection efforts. One might expect that these two sides would be well matched and that mobility-data partnerships would be ubiquitous. But sharing of new mobility data between the public and private sectors is still relatively limited. Those tasked with solving transportation problems (e.g., city planners) find it difficult to access and use useful data that exists, while those who hold such data may be concerned about sharing proprietary and personally identifiable information (PII)⁴ that the data contains. This dilemma speaks to a need for public policy that can foster collaboration and find mutually beneficial solutions.

The solution is not as simple as mandating that data providers “just share the data.” Mobility data sharing entails legitimate challenges. Overly burdensome data-sharing requirements can deter companies from investing in mobility or expanding into new markets. Costs associated with meeting these requirements could be passed onto consumers. Perhaps most importantly, data sharing could jeopardize consumer privacy. In a response to updates proposed by the CPUC for data-sharing requirements for TNCs in California,⁵ Uber argued that

“[d]issemination of sensitive, personal information to parties beyond the Commission not only undermines a user’s privacy rights, but also heightens the risk of unintended disclosures through data breaches or mishandling of the data... In the wrong hands, TNC data could allow stalkers to track their victims, undermine the confidentiality of media journalists, and assist burglars [*sic*] from monitoring target homes.”⁶

Whether or not to share mobility data is often presented as a binary choice. This framing is overly simplistic. Mobility data sharing is not an all-or-nothing proposition—it is a goal that requires a thoughtful, informed approach to achieve. This issue paper discusses five of the most important issues surrounding mobility data sharing. In brief, these are:

- (1) cost of data collection and storage;
- (2) lack of data standardization;
- (3) difficulty of anonymizing mobility data;
- (4) high levels of expertise needed for data analysis and visualization; and
- (5) the proprietary nature of mobility data.

Each of these issues is explained further below.

First, data sharing involves multiple logistical challenges. Collecting and storing data can be complicated and expensive. Specifics of these challenges depend on the nature of the data being collected and the time period over which data are gathered, maintained, and (ultimately) deleted. For the public sector, the high cost of data collection and safe storage often means that public data-collection efforts are limited in number and scope. This in

³ SFCTA and Joe Castiglione, “Reply Comments of [SFCTA].”

⁴ “Rules and Policies – Protecting PII – Privacy Act,” U.S. General Services Administration. The U.S. GSA defines Personally Identifiable Information as “information that can be used to distinguish or trace an individual’s identity, either alone or when combined with other personal or identifying information that is linked or linkable to a specific individual.”

⁵ See Recommendation 2 for more information on California’s existing data-sharing requirements and proposed updates.

⁶ Yee et al., “Reply Comments of Uber.”

turn makes it difficult for public officials to conduct informed transportation and land-use planning. For the private sector, investments of time, money, and personnel are needed to meet data-reporting requirements. Inconsistent and/or excessive requirements can result in the magnitude of these investments being unreasonably high.

Second, a lack of standardization contributes to inconsistent data collection and sharing practices. Different mobility companies collect data using different units of analysis (e.g., by trip, by driver, by passenger), measurement scales, timescales, and so on. This is not a problem when the data is used only internally, but inconsistency across companies makes it difficult for public agencies to aggregate data from multiple sources for planning and regulatory purposes.

Third, mobility data is challenging to anonymize. Fine-grained mobility data—such as trip-level data containing the longitude and latitude of trip origin and destination—is often personally identifiable, even when time and identity information is removed. This is especially true of data on origin-destination pairs that originate from granular trip traces, such as GPS traces of individual riders on mobility apps, or traces from cell-phone GPS data. Over multiple trips, granular trace data can easily be linked to an individual person by evaluation of common travel patterns (e.g., home-to-work routes). This raises privacy concerns for individuals using mobility services, including ride providers (for ridehailing services) as well as riders. Anonymity is generally less of a concern for data that represent aggregated trips at the block or census-tract level. However, aggregation can blunt data utility. There is a tension between achieving aggregation sufficient to protect privacy and preserving enough specificity that the data remain useful.⁷

Fourth, challenges associated with data analysis and visualization can prevent policymakers from extracting useful insights. Managing large datasets can be expensive, computationally intensive, and time consuming. Public agencies may lack in-house data-science experts and/or may lack the resources to hire such experts. Thus even if widespread data collection and sharing is achieved between public and private stakeholders, the value of the raw data to public decision-makers may remain limited.

Fifth and finally, private companies have legitimate proprietary, financial, and fairness concerns about sharing data, especially data captured in highly competitive markets. Proprietary concerns could arise if, for example two companies are responsible for the bulk of ridehailing trips in a given region. Requiring these companies to share their trip data publicly could enable each to study their main competitor's trips and possibly gain a competitive advantage from doing so. Financial concerns arise from the high costs of data collection and storage, which may make private organizations more reticent to share data without reasonable compensation. Fairness concerns arise from the possibility of increasing data-reporting requirements for mobility providers without imposing similarly stringent requirements on the automakers or private motorists who account for a larger share of travel. Unbalanced data-reporting requirements could also hinder development and adoption of new and sustainable mobility alternatives—either by discouraging automakers from venturing into new mobility markets, increasing costs of mobility services, and motivating consumers to stick to conventional transportation options for which privacy concerns do not appear to be as salient.

Policy Recommendations

Communities with ready access to mobility data may be better able to make dynamic and responsive planning decisions. For example, transportation planners rely on models to predict how travel behavior and transportation-mode choice will evolve. These predictions are used to direct infrastructure investments and guide decisions. Robust mobility data is needed to power these models. Regulators also require mobility data to ensure compliance with codes and statutes and assess if and how regulations should be modified to better achieve societal objectives.

In this section, we recommend four policies to unlock the substantial societal benefits of mobility data sharing

⁷ See Recommendation 2 for more on this topic and what types of aggregated data states and cities are collecting from TNCs.

while addressing the challenges presented in the previous section. The recommended policies—and the level(s) of governance for which each is best suited—are as follows:

- (1) Foster voluntary agreement among mobility providers for a set of standardized data specifications. (Cities, Regions)
- (2) Develop clear data-sharing requirements designed for transportation network companies (TNCs) and other mobility providers. (Cities, States, Regions)
- (3) Establish publicly held big-data repositories, managed by third parties, to securely hold mobility data and provide structured access by states, cities, and researchers. (States, Federal)
- (4) Leverage innovative transportation-planning tools. (Cities, Regions)

Recommendation 1. Foster voluntary agreement among mobility providers for a set of standardized data specifications. (Cities, Regions)

Standardizing data-sharing specifications is a critical first step to improving data comparability and reducing administrative burdens associated with data reporting and sharing among mobility providers. Standardization implies that there is a consistent, comprehensive, and agreed-upon set of parameters for how data is reported. The details of the standards selected matter less than the extent to which the standards achieve consensus and are broadly adopted within an industry. An overview of the data standards addressed in this section can be found in Table 1 below.

Data standards can emerge via several pathways. They can be introduced from the “top down” by a single large and influential actor (e.g., a governmental agency or a major private company), or from the “bottom up” by multiple smaller actors who agree to adopt the same specifications. Regardless of origin, data specifications should be compelling enough to spread until eventually there is tacit or explicit agreement among most or all stakeholders to adopt those specifications as the go-to standard. Considering how and why existing data standards have emerged is useful for understanding how to create and encourage adoption of additional data standards in the future.

Example 1: The General Transit Feed Specification

The origin of data standardization in transportation planning began with the **General Transit Feed Specification (GTFS)**, originally called the Google Transit Feed Specification.⁸ GTFS was developed in 2005 by TriMet, the transit provider in Portland, OR, to enable the incorporation of Portland’s transit schedules into Google Maps.⁹ To develop the GTFS, TriMet published their transit schedule as a collection of CSV files to their website. TriMet then worked with Google Maps employees to ensure that those files were formatted in a way that would allow Google Maps to incorporate and visualize the data into its mapping platform. By 2006, six other cities had adopted the GTFS. By 2008, about a dozen transit operators were using the standard, and, by 2013, the number had climbed into the hundreds.¹⁰ Today, GTFS has become ubiquitous, employed by more than a thousand transit operators worldwide.¹¹

Though originally developed with Google in mind, the GTFS was never proprietary. The GTFS was intentionally designed around a simple database tool (a collection of CSV files) to ensure that any agency, large or small, would be able to easily adopt it.¹² This open-access strategy may be largely responsible for the GTFS’s success.

An updated version of the GTFS, called GTFS Realtime, became available in 2012.¹³ GTFS Realtime allows maps to convey dynamic information about when transit is actually arriving and departing, rather than relying on static,

⁸ “GTFS Static Overview Static Transit,” Google Developers.

⁹ Levine, “A Brief History of GTFS.”

¹⁰ McHugh, “Pioneering Open Data Standards: The GTFS Story.”

¹¹ “General Transit Feed Specification,” TransitWiki.

¹² McHugh, “Pioneering Open Data Standards: The GTFS Story.”

¹³ Ibid.

preset schedules that state when transit is scheduled to arrive and depart.¹⁴ The tradeoff of this improved accuracy is that GTFS Realtime requires inputs called “protocol buffers,” which are more complex than CSV files. The system uses one of two methods: fetch, which allows Google, or other developers, to retrieve data from a web location; or push, which allows users to upload data programmatically.¹⁵ Several large transit operators (including TriMet as well as operators in the San Francisco Bay Area; Boston; Louisville, KY; Madison, WI; and Hartford, CT) have switched to GTFS Realtime. Smaller transit agencies have found it more difficult to transition. Accommodating the advanced inputs required by GTFS Realtime involves hiring additional technical expertise, which may be hard to do under a limited operating budget. Smaller transit agencies may face other hurdles as well, such as difficulty procuring adequate in-vehicle tracking technology necessary to implement GTFS Realtime.¹⁶

An alternative to GTFS Realtime is a **REST API/XML**. Representational State Transfer (REST) is an architectural style for **A**pplication **P**rogramming **I**nterfaces (APIs). There are many alternative data-exchange formats that are compatible with RESTful APIs, including JSON and HTML, but the most commonly used in transportation spaces is **eXtensible Markup Language** (XML).¹⁷ RESTful API/XML is a generalized format and is not necessarily transit-specific, though it can be adapted to fit transit-specific needs. While GTFS Realtime reflects the actual, GPS-identified location of every vehicle or device in operation in a given system (i.e., “where are all the buses in the system?”), a RESTful API makes real-time arrival and departure predictions based on when a transit unit arrives or departs from a specific stop (i.e., “what’s the next bus at this stop?”). RESTful APIs hence provide “piece-wise” access to real-time transit data, while GTFS Realtime provides “bulk” access.

Importantly, piece-wise RESTful API data expressed in XML can be derived from bulk data included in the GTFS Realtime standards. A RESTful API essentially trades network completeness for reduced data-management requirements and greater object specificity (i.e., a user can request information on a single stop without having to download data on the whole transit system). In practice, a RESTful API/XML may be more feasible for smaller municipalities (which generally have fewer technical resources), while GTFS Realtime may be better suited to larger municipalities.¹⁸ If a planner wishes to view an entire transit map, however, RESTful APIs require individual requests to fetch stop data one at a time and thus may require more data storage space than GTFS Realtime.¹⁹ The REST/XML approach also tends to be less open. In order for third parties to access some REST/XML APIs (e.g., in Oahu, HI; Pinellas County, CA; and Chicago, IL), registration for an API key is required.²⁰

Example 2: The General Bikeshare Feed Specification

The GTFS has inspired data standards in other subsets of the transportation space. One example is the **General Bikeshare Feed Specification (GBFS)**.²¹ The GBFS is a data-standardization project for bikeshare programs launched in 2015 by the North American Bikeshare Association (NABSA).²² While use of the GTFS grew organically, expanding GBFS use required NABSA involvement. Because NABSA isn’t a household name in the same way that Google is, additional “top-down” effort from NABSA was needed to encourage stakeholders—including cities, transportation agencies, and bikeshare companies—to adopt the GBFS. This effort ultimately succeeded. As of 2019, around 230 bikeshare programs were using the specification.²³

¹⁴ “GTFS Static Overview,” Google Developers.

¹⁵ “Launch Process: Choose a Data Submission Method,” Google Support.

¹⁶ Accuardi et al., “The Data Transit Riders Want.”

¹⁷ “Publicly-accessible Public Transportation Data,” TransitWiki. Note that while XML is the most frequently used language for RESTful APIs in transportation spaces, the majority of transit data feeds are communicated through GTFS Realtime.

¹⁸ This is not always borne out in practice. REST/XML is currently in use by transit operators in Portland, the San Francisco Bay Area, Seattle, Los Angeles, Chicago, and several other large municipalities. Some cities employ GTFS Realtime for certain transportation systems and REST/XML for others.

¹⁹ “Interest for a Lite GTFS-TripUpdates,” Google Transit GitHub.

²⁰ Ferris, “GTFS-Realtime vs. REST/XML”; “Publicly-accessible Public Transportation Data,” TransitWiki.

²¹ “General Bikeshare Feed Specification,” North American Bikeshare Association.

²² “About NABSA,” North American Bikeshare Association.

²³ “General Bikeshare Feed Specification,” NABSA.

Example 3: The Mobility Data Specification

The GTFS and the GBFS gave rise to the **Mobility Data Specification (MDS)**.²⁴ The MDS was developed in 2018 to track vehicles operated by mobility providers in Los Angeles. Shared scooters became the first application of the MDS, but it was designed to collect mobility data from bikeshare, TNCs, and even potentially other mobility services that may yet emerge. The MDS was created by the Los Angeles Department of Transportation (LADOT), and like the GTFS, it was never intended for use exclusive to the Los Angeles area. Soon after the MDS was developed, administration of the standard transitioned from the LADOT to the nonprofit Open Mobility Foundation, and they are supporting dozens of cities in use of MDS for their jurisdiction.²⁵ According to OMF, MDS facilitates real-time data exchange among mobility providers, municipalities, and regulatory agencies. Mobility providers input data on vehicle location (which can be used to inform municipal planning and development of regulations), while public agencies input data important to mobility-service operations (such as information on regions where scooters and similar mobility vehicles are and are not allowed to park).²⁶

This two-way data sharing improves real-time collaboration between parties. Scooter companies have touted the cost and time that the MDS saves by streamlining and standardizing communication.²⁷ However, everyone at the table is not a willing party. It remains the voluntary choice of each city to adopt the standard, but many of the cities (including Los Angeles) mandate use of MDS as a prerequisite to obtaining a scooter provider license. Therefore, while the standardization rests on a “bottom up” voluntary framework, in many cities MDS is more “top down” than GTFS or GBFS.

As of July 2019, over 50 cities are implementing the MDS to track scooters in real time.²⁸ Despite its meteoric rise, the MDS is not without critics. The granular data included in the MDS raises privacy concerns due to the difficulty of anonymizing GPS level trip data, given that travel patterns can emerge when individuals travel from identifiable locations, e.g. home to work. In the event of a municipal or commercial database breach, some stakeholders argue that disaggregated trip data collected through the MDS could theoretically be re-identified—exposing riders’ personal travel information.²⁹ This problem is exacerbated by the fact that the MDS platform can store historic trip data indefinitely, which can make it easier to identify patterns and link trips to individuals. Reforms, such as some level of trip aggregation, may be necessary in order for the MDS to continue delivering public- and private-sector value without compromising security.³⁰

“Bottom-up” or “top-down”

The GTFS, GBFS, and MDS are notable in that each were adopted on a largely voluntary basis. Voluntary adoption has merits. When stakeholders independently select a standard, they are generally more invested in its success than if the standard were externally imposed. Stakeholders are also less likely to choose to adopt a standard they find cumbersome—hence, the simple fact that a company or city decides to adopt a particular standard is a sign that the standard is well designed and likely to grow. Finally, voluntary standards can foster a sense of collaboration and partnership between adopters.

But voluntary standard adoption has downsides. With voluntary adoption, there is often no governing body to ensure that standards can be easily met by all stakeholders. GTFS Realtime, for instance, has proven difficult for smaller transit agencies to adopt. Voluntary adoption of data standards also takes time to become widespread and may never be universal. The open-source GTFS Realtime has encountered competition from proprietary data-management platforms. MDS as a standard that is adopted voluntarily by over 50 cities, and will remain updated by an independent organization, but while adoption is voluntary, it is a hybrid model for standardization,

²⁴ “Mobility Data Specification,” City of Los Angeles.

²⁵ Pyzyk, “15 Cities Join Open Mobility Foundation.”

²⁶ Ibid.

²⁷ Zipper, “Cities Can See Where You’re Taking That Scooter.”

²⁸ “Home,” Open Mobility Foundation.

²⁹ Lubarsky, “Re-Identification of ‘Anonymized’ Data.” Personally Identifiable Information (PII) data may be re-identified through pseudonym reversal, by combining datasets to pinpoint the same individual across multiple datasets, or in the event that the data was insufficiently anonymized.

³⁰ Zipper, “Cities Can See Where You’re Taking That Scooter.”

because many cities are mandating its use by mobility providers. This blending of the “bottom up” and “top down” approaches may represent a possible middle-ground approach. In Recommendation 2, the MDS blended standardization and regulatory approach is compared to the regulatory actions of several states and cities who take an exclusively “top down” approach for regulation of TNCs.

Table 1. Key data standards/specifications in use for new mobility

Data Standard/ Specification	Primary Developer	Date of Development	Purpose & Details
General Bikeshare Feed Specification (GBFS)	North American Bikeshare Association (NABSA)	2015	Provides real-time information about bikeshare services, including information on pick-up and drop-off locations, number of bikes and docks at each location, availability of bikes for rent, system hours and days of operation, operating regions, system pricing, and system alerts (e.g., if a particular dock station is out of order).
General Transit Feed Specification (GTFS)	Tri-County Metropolitan Transportation District of Oregon (TriMet), Google	2005	Provides static information about public transit, including routes, schedules, arrival and departure times, service dates, and fare information. The GTFS has evolved into a de facto industry standard—it is currently used by more than a thousand transit agencies worldwide.
GTFS Realtime	Life Transit Updates, transit developers, Google	2012	Provides live updates about public transit services—including live departure times, service alerts, and vehicle positions—to users and app developers. GTFS Realtime is in use by several major transit agencies (e.g., the San Francisco Bay Area Rapid Transit), but smaller agencies have been slower to adopt these new standards due to high barriers to entry (e.g., technical expertise).
Mobility Data Specification (MDS)	Los Angeles Department of Transportation	2018	Facilitates collaboration among mobility providers and regulating entities in the collection and exchange of information. It is an open-source, standard that includes real-time (within 5 seconds) reporting through an API to track individual mobility vehicles/devices using a unique ID. The MDS includes vehicle type, event type (e.g., “scooter reserved”), reason for trip, trip time, locations of pick-up and drop-off, and percent remaining of the vehicle/device’s battery. In addition to vehicle-tracking in real time, routing information for each specific trip route must be reported within 24-hours.

Data Standard/ Specification	Primary Developer	Date of Development	Purpose & Details
Representational State Transfer / eXtensible Markup Language (RESTful API/ XML)	Roy Fielding, Bray et al.	2000 (RESTful API), 1996 (XML)	Offers a structured response-and-request interaction between a set of real-time transit data and a program (typically mobile) in order to update and visualize transit data on a map or scheduling platform. The RESTful API (2000) is the integration tool, while XML (1996) is the format for data exchange. Data handled through the RESTful API XML are “piece-wise,” real-time transit data. Access may be restricted by the host agency, requiring an API key to view and edit the underlying data.

Recommendation 2. Develop clear data-sharing requirements for transportation network companies (TNCs) and other mobility providers. (Cities, Regions, States)

Establishing uniform standards for mobility data (Recommendation 1) will facilitate data sharing across the public and private sectors. There is a particular need for clear data-sharing requirements with respect to TNCs. To date, many TNC companies have been largely reluctant to share data with public agencies and other end-users for reasons discussed above.

This section evaluates approaches that several states and cities have taken to encourage or mandate TNC data sharing. An important question is whether it is the role of states or cities to set data-sharing protocols for TNCs. The state-led approach generally yields reduced costs due to economies of scale, improved consistency and comparability across a greater area, more resources available for oversight, and greater public-sector leverage for negotiating with TNCs. Some cities have been empowered to lead when their states have demonstrated reluctance to get involved, or where data from state collection efforts are not shared with cities. State leadership may also make sense given that challenges associated with data collection can be more acute for many smaller or lower-resourced cities, which may lack the legal and institutional capacities to negotiate with TNCs directly over data policy.

Nevertheless, some cities will likely continue to push for data sharing beyond what their states require because city-led efforts can be better tailored to local needs and capacities. This could result in a “patchwork” of data-sharing requirements, which could generate additional costs and challenges for stakeholders operating in multiple jurisdictions. Significant city-led efforts could also lead to pressure from federal and commercial interests that may prefer uniformity over local control. Theoretically this pressure could result in federal policy that preempts local data collection in favor of larger scale uniformity, but this is not a guaranteed outcome. A federal policy debate could also lead to policy that supports cities in differentiating their mobility data collection strategies. Regional planning authorities may offer a good “middle ground” between state and city leadership, but no significant regional efforts to coordinate TNC data sharing exist to date.

The specifics of state- region- and city-led data-sharing approaches will vary depending on the unique preferences and capacities of the different jurisdictions. This makes generalizing best practices for TNC data-collection approaches difficult. Nevertheless, case studies can provide useful examples of how TNC data collection can work in practice. Below, we discuss five such case studies, including two states (California and Massachusetts) and four cities (New York; Portland, OR; and Austin, TX).³¹ Where relevant, we also refer to the City of Los Angeles’ data-collection strategy for scooters. The goal of including this scooter data collection strategy is to highlight the merits of an API-based approach—the MDS—that is currently in use in Los Angeles and dozens of other cities to regulate scooters. The MDS may be expanded in the future to include TNC data.³²

³¹ Appendix A provides a table comparing the specific data-collection strategies used by each of the case-study jurisdictions.

³² Nordrum, “Los Angeles to Require Scooter Companies to Share Data.”

California and **Massachusetts** have been at the forefront of state-led requirements for TNC data sharing. Both states expanded the missions of pre-existing regulatory bodies (the California Public Utilities Commission (CPUC) and the Massachusetts Department of Public Utilities (MDPU)) to include TNC oversight. Massachusetts created a new TNC Division within the MDPU to supervise TNCs operating in the state. Similarly, New York City (NYC) and Portland, OR, both expanded the regulatory missions of existing agencies (the **NYC Taxi and Limousine Commission (NY-TLC)** and the **Portland Bureau of Transportation (PBOT)**) to include TNCs operating within city limits. Austin, TX, requires TNCs to report directly to city administrators.³³ TNC data more detailed than the city requires is also made available to the public from RideAustin. RideAustin is a nonprofit rideshare program that has served the greater Austin area since 2016, when Uber and Lyft (temporarily) halted service to the region in response to tightening TNC regulations. The program was developed with contributions from the tech community as well as cash and in-kind donations from local partners. RideAustin's unique nonprofit status notably enables the program to circumvent proprietary concerns—commonly voiced by for-profit TNCs—related to data sharing.³⁴ The LADOT does not require any data sharing for TNCs, but requires that scooter companies operating in Los Angeles develop an API compatible with the MDS.

In all six of these jurisdictions, a governing authority or a group of governing agencies shaped and implemented a data-sharing mandate. While each jurisdiction solicited stakeholder feedback, the emphasis was on adding value for the public sector. Only in Los Angeles was the mandate intended to provide benefits for participating companies as well. This government-driven approach for TNC data sharing stands in stark contrast to the largely “bottom-up” emergence of standards for other mobility services, described under Recommendation 1. The fact that a widely accepted, voluntary data-sharing system has not emerged for TNCs could be partially linked to the highly competitive nature of emerging mobility markets. TNC providers may be less likely than transit agencies or bikeshare companies to share data that could inform competitors. TNC providers may also be particularly concerned that costs of collecting and sharing data could cut into profits. Mandatory TNC data-sharing requirements may therefore be necessary, at least in part, to ensure that transportation planners and other public officials have access to the mobility data they need. But these requirements will be most successful if designed with the perspectives of private-sector companies in mind.

Data-sharing requirements that are clearly aligned with specific planning purposes and public objectives will generally be more agreeable to all parties. However, the more data that TNCs are required to report, the harder (and more expensive) it will be for TNCs to comply. Compliance costs could then be passed onto consumers. In developing TNC data-sharing requirements, then, regulators and other public officials should carefully consider the following.

- **Temporal representation:** At what time interval (e.g., annually, quarterly, monthly, daily, or in real time) should data be reported?
- **Trip data and levels of aggregation:** Should data be aggregated at the zip-code or census tract level? Should each trip include latitude or longitude of all origin-destination pairs? Should trip data include all time spent in the vehicle, including dwell time, time spent picking up a passenger, and time transporting the passenger? Should TNCs report the number of passengers per trip and the type of vehicle used (e.g., EV, hybrid)?
- **Public availability of data:** Should data be made available for public use? Should any limitations be placed on who is granted access? Which third party (or parties) can host the data and regulate access?
- **Driver data:** Should TNCs report information on drivers (e.g., experience, earnings)?
- **Accessibility data:** Should TNCs report whether or not a rider has requested a wheelchair-accessible vehicle (WAV)? Should TNCs report if and when those rides are accepted by drivers?

³³ “Ordinance NO. 20160623-066.” City of Austin, Texas. Austin, TX, 2019.

³⁴ “Frequently Asked Questions.” RideAustin.

- **Emissions data:** Should TNCs report emissions and fuel-consumption information? If so, should this information be reported on a per-vehicle or fleet-wide basis?

The following sections explain how each of the aforementioned jurisdictions have addressed these questions.

Temporal representation

As of June 2019, both California and Massachusetts require TNCs to report data on an annual basis. In February 2019, the CPUC proposed increasing reporting frequency to a quarterly basis.³⁵ This proposal has not yet been approved.³⁶ NYC, Portland, and Austin require TNCs to report data on a monthly basis.³⁷ Los Angeles collects scooter vehicle-location data in real time,³⁸ although the specific routing of each trip can be reported at any time within the 24-hour period after the trip has been completed.³⁹ This difference in temporal representation may reflect a difference between state and city needs. States generally focus on longer-range planning, so annual reporting may be sufficient. Cities have a greater need for dynamic, shorter-term transportation-management decisions, and so may require more frequent reporting. If states wish to implement data-reporting requirements that successfully serve cities as well, more frequent reporting (e.g., monthly or quarterly) may be necessary. It is not clear when or if even more frequent reporting (e.g., weekly or daily) is necessary. If states intend data to be used primarily for long-range planning purposes, then annual reporting may be acceptable.

Trip data and levels of aggregation

The jurisdictions discussed in this report require different levels of trip aggregation for TNC reporting, but the rationales for these varying standards are unclear. California requires TNCs to aggregate trip origins and destinations at the zip-code level.⁴⁰ Beginning in 2023, California will require TNCs to report trip data to the California Air Resources Board (CARB) on a per-vehicle level due to emissions regulations passed in early 2019.⁴¹ Massachusetts requires TNCs to report data on the city or town where each trip begins and ends, as well as the trip's route (in miles) and length (in minutes).⁴² This information is aggregated and reported as totals entering and exiting each city. NYC requires TNCs to aggregate trip origins and destinations by "taxi zone."⁴³ Portland requires full disaggregation—TNCs must report the origin and destination latitude and longitude of each individual trip.⁴⁴ Austin does not require specific TNC trip data, only the number of trips completed.⁴⁵ However, RideAustin provides origin and destination latitudes and longitudes for individual trips.⁴⁶ Los Angeles uses the MDS API to collect GPS coordinates for the origin and destination of each scooter trip taken within its jurisdictional boundaries.

Public availability of data

Experts generally agree that only aggregated and anonymized TNC trip data should be widely available for public consumption.⁴⁷ Trip-level data and other potentially personally identifiable information (PII) should be kept confidential except under certain circumstances (e.g., for legal reasons) to minimize privacy concerns and security risks.⁴⁸ California, Massachusetts, Portland, and Los Angeles generally follow these practices. Each of these jurisdictions requires some degree of trip-level reporting to the overseeing agency. In California and Massachusetts, this data is publicly available only in aggregated reports. In Portland, TNC data is not available to

³⁵ "Required Reports TNCs Must Provide the CPUC," California Public Utilities Commission.

³⁶ "Rulemaking 12-12-011: Assigned Commissioner's Ruling Seeking Comments on Proposed Data Reporting Requirements," California Public Utilities Commission.

³⁷ "About TLC Trip Record Data," NYC Taxi & Limousine Commission; "Ordinance NO. 20160623-066," City of Austin, Texas; "City Code 16.40.20: Private For-Hire Transportation Regulations" City of Portland, Oregon.

³⁸ "Mobility Data Specification," City of Los Angeles.

³⁹ Zipper, "Cities Can See Where You're Taking That Scooter."

⁴⁰ "Required Reports TNCs Must Provide the CPUC," California Public Utilities Commission.

⁴¹ "Article 7. Transportation Network Companies [5430-5450]," California Public Utilities Commission.

⁴² "New State Website Shares Local Ride Hailing Data," Massachusetts Municipal Association.

⁴³ "TLC Trip Record Data," NYC Taxi and Limousine Commission.

⁴⁴ "City Code 16.40.20: Private For-Hire Transportation Regulations" City of Portland, Oregon.

⁴⁵ "Ordinance NO. 20160623-066," City of Austin, Texas.

⁴⁶ "RideAustin's Datasets," Data.World.

⁴⁷ "Guide to Identifying Personally Identifiable Information (PII)," University of Pittsburgh, Information Technology

⁴⁸ "Departmental Privacy Risk Management Policy," U.S. Department of Transportation.

the public but is accessible by approved individuals and agencies. In Los Angeles, access to scooter data and the MDS platform is restricted to mobility providers and City staff. NYC aggregates data by taxi zone, but this practice still provides some indication as to where TNC users are located geographically. RideAustin makes latitudes and longitudes of TNC trip origins and destinations publicly available.⁴⁹ This has raised concerns over the privacy afforded to RideAustin's drivers and riders. Moving forward, jurisdictions must address how much new mobility data to make public and in what manner.

Driver data

California requires TNCs to report that drivers have completed a TNC training course (mandatory in California), the number of driver violations and suspensions, and total hours logged annually by each driver.⁵⁰ Massachusetts requires TNCs to report if a driver has completed the background check required to drive a TNC in the state. Massachusetts also requires annual reporting of all driver and rider complaints (both content and frequency) and suspensions. State jurisdiction reigns over driver licensing, but it is common for taxi regulators (which include some states, counties, and cities) to impose additional requirements on commercial drivers. It is therefore unsurprising that some cities have additional requirements related to TNC drivers. Portland simply requires TNCs to report whether drivers have any criminal history.⁵¹ Austin requires fingerprint-based background checks.⁵² NYC does not require TNCs to report any special driver data, but New York State requires background checks for TNC drivers, in addition to reporting on accidents and driving license activity (e.g., license suspensions).⁵³

Accessibility data

Requiring TNCs to report data on wheelchair-accessible vehicles (WAVs) in their fleets is likely to be more helpful to cities than to states. Cities can use WAV data to identify and address local accessibility obstacles. For instance, an unusually high number of WAV ride requests in a particular neighborhood may mean that wheelchair-accessible transit options in that neighborhood need to be expanded. States are unlikely to respond dynamically to specific neighborhood needs, but states could support cities by providing funding opportunities for areas with high WAV riders. WAV-specific reporting requirements for TNCs are relatively limited. California collects accessibility information, but this data is held privately by the CPUC and is not currently shared with municipalities. Beginning in 2020, the CPUC will require TNCs to report the number of WAV rides requested and fulfilled, the wait times between WAV ride acceptance and pick-up, and any educational outreach efforts conducted by TNCs to reach disabled communities.⁵⁴ Massachusetts does not collect WAV information, but the City of Boston has a WAV pilot program that includes some mandatory reporting.⁵⁵ NYC and Portland require TNCs to report the number of people with wheelchairs requesting trips. Both cities ask for detailed information on all WAV requests, including dates, times, and locations of ride pick-ups and drop-offs. The City of Austin does not require data on WAV requests or usage. Though RideAustin does not offer WAV trips, it does partner with a third party to provide rides to people who require wheelchairs.⁵⁶ This WAV trip data is not publicly accessible.

Emissions data

None of the jurisdictions considered in this report currently requires TNCs to report emissions data. This is changing in California. California will soon require TNCs to annually report emissions-related statistics to the CARB. These statistics include total miles driven per vehicle annually as well as the percent share of miles completed by qualified zero-emissions vehicles. In 2019, CARB plans to issue a baseline for greenhouse gas emissions output by TNCs. In 2021, the agency will set a timeline for reducing emissions outputs relative to baseline. The timeline

⁴⁹ "RideAustin's Datasets." Data.World.

⁵⁰ "Required Reports TNCs Must Provide the CPUC," California Public Utilities Commission.

⁵¹ "16.40.240 TNC Operating Responsibilities and Prohibitions. Chapter 16.40 Private For-Hire Transportation Regulations." City of Portland, Oregon.

⁵² "Frequently Asked Questions." RideAustin.

⁵³ "FAQ for TNC Passengers," NY Department of Motor Vehicles; "2018 TLC Fact Book," NYC Office of Policy and External Affairs.

⁵⁴ "Required Reports TNCs Must Provide the CPUC," California Public Utilities Commission.

⁵⁵ "MassDOT and the MBTA to Partner with TNCs." Massachusetts Bay Transportation Authority.

⁵⁶ Gallaga, "In Their Own Words."

will take effect in 2023.⁵⁷

The key takeaway is that no single set of data-sharing requirements for TNCs has emerged yet. The “top down” approach employed in all the TNC mandate examples reflect the unique needs of each jurisdiction, and are not linked to broader industry-based standards. States and cities will need to work diligently to weigh the pros and cons of various policy strategies, keeping in mind the preferences of stakeholders within their jurisdiction. At the same time, states and cities should keep in mind the benefits of consistent reporting requirements across jurisdictional boundaries. Identifying approaches that best strike a balance between these two objectives will require considerable further analysis.

Recommendation 3. Establish publicly held big-data repositories, managed by third parties, to securely hold mobility data and provide structured access to states, cities, and researchers. (States, Federal)

Although the term “big data” has skyrocketed in popularity in recent years, there is not yet consensus around what exactly “big data” means.⁵⁸ In 2013, Cornell computer scientists proposed three criteria for determining whether a dataset qualifies as “big data”: (1) size of the dataset, (2) complexity of the behavior of the permutations, and (3) the technologies used to analyze the dataset.⁵⁹ The Method for an Integrated Knowledge Environment (MIKE2.0) project argues that dataset complexity, not dataset size, should be the key factor used to define big data, while Intel deems dataset size of greatest import.⁶⁰ Microsoft focuses on technology, considering “big data” to be that which requires “serious computing power” associated with “machine learning and artificial intelligence” for analysis.⁶¹ In the mobility world, “big data” generally refers to data collected passively through cell phones, vehicles, or other over-air or mobile technology. This type of data stands in contrast to data that is actively collected through surveys or other means of direct reporting.⁶²

Given the scale of mobility data and the costs of secure storage, the federal government is well placed to lead the development of a publicly held repository for large datasets that can help measure the impacts of new mobility. While states and municipalities may be capable of developing such repositories, there is value in avoiding duplication of efforts. A centralized mobility-data repository will provide stakeholders access to not only their own mobility data, but data from other regions as well. Data from multiple regions is often key to identifying patterns and unlocking insights. Expanding on existing centralized federal repositories, such as the Secure Data Commons⁶³ or the Transportation Secure Data Center⁶⁴ would reduce both upfront and ongoing costs. In order to make a centralized federal mobility data repository a reality, first a clear federal data policy framework is necessary. This may need to include outlining the types of data appropriate for storage in the federal mobility data repositories and the types of users and uses that will be allowed access to the data.

It is possible some states and local authorities will still see a need to develop their own data repositories. This could be justifiable if federal repositories do not emerge, or were they may be insufficient in meeting the needs of region-specific statutes or policy objectives.⁶⁵ To the extent possible, though, avoiding overlapping or duplicate efforts should be a priority. An overview of existing, relevant transportation data repositories and data-collection efforts is provided below and in Appendix A.

Regardless of the level at which data repositories are established, consideration should be given to having these

⁵⁷ “Clean Miles Standard.” California Air Resources Board.

⁵⁸ “The Big Data Conundrum,” Emerging Technology from the arXiv.

⁵⁹ Ward and Barker, “Undefined by Data.”

⁶⁰ “The Big Data Conundrum,” Emerging Technology from the arXiv.

⁶¹ Ibid.

⁶² Tosi, “Cell Phone Big Data.” The lines between active and passive data collection are blurring. For instance, active data collection can occur via cell phones using in-app polls or A/B tests.

⁶³ “About the Secure Data Commons Portal,” U.S. Department of Transportation, Secure Data Commons.

⁶⁴ “About the Transportation Secure Data Center.” U.S. Department of Energy, NREL..

⁶⁵ “Data.” Panel at 3 Revolutions Policy Conference, University of California, Davis.

repositories managed by a third party or parties, especially in light of the privacy challenges that the Freedom of Information Act (FOIA) and other transparency laws pose. Though FOIA exempts the disclosure of information (i) pertaining to trade secrets and (ii) that would otherwise “invade another individual’s personal privacy,”⁶⁶ it is unclear how federal, state, and local agencies in charge of a mobility-data repository would ensure that commercial and personal information is protected from open-records requests.⁶⁷ For instance, the Environmental Protection Agency came under fire in 2018 for exposing over 80 Social Security Numbers and other personally identifiable information on a FOIA request portal.⁶⁸ Engaging a third party for repository management can reduce the risk of breaches of privacy because the entity would establish clear protocols for addressing FOIA requests. An appropriate third party would have the resources and expertise to securely manage such a repository while also being subject to some public accountability. Examples of potentially appropriate third parties include universities and federally funded research and development centers such as National Laboratories. These entities are familiar with management of PII datasets and already have established institutional knowledge and capacity for ensuring the data is secure and not subject to direct FOIA requests.

Mobility-related big data in the public sector

Several agencies of the U.S. Department of Transportation (USDOT) are already coordinating and securing large data reserves. The **USDOT’s Secure Data Commons (SDC)** is an exemplary model of how to create a third-party platform for storing and analyzing sensitive data. The SDC engages two main types of users: data providers and data analysts. Data providers voluntarily add data however they deem appropriate. Providers can contribute data in real time, schedule regular batch uploads of historic data, or upload data on an occasional, ad-hoc basis. The SDC grants different levels of access to different types of data analysts based on criteria defined by data providers. Providers also set parameters for whether datasets (or data analyses conducted on the SDC platform) can be exported.⁶⁹

Data analysts must file an application and sign a data-sharing agreement to be granted access to the SDC. If approved, analysts will have access to a set of datasets tailored specifically to the level of access they are granted, as well as to statistical tools such as R, Python, and SQL. The SDC only includes two current mobility programs (each with several datasets): the **Waze Connected Citizens Program (CCP)** and the **Connected Vehicle Pilot Deployment Program (CVPDP)**.⁷⁰

The goal of the Waze CCP is to analyze Maryland traffic data provided by Waze (a subsidiary of Google), alongside crash data provided by the State of Maryland, in order to evaluate when and where traffic accidents are most likely to occur.⁷¹ ⁷² While the goal of this project is admirable, it is still relatively narrow in scope. There are undoubtedly myriad other lessons that could be learned from Waze datasets. But because SDC data providers are able to restrict data use to specific purposes, researchers have limited freedom to pursue new lines of inquiry.

The Connected Vehicle Pilot Deployment Program (CVPDP) houses data for connected-vehicle pilots in New York City, Wyoming, and Tampa, FL.⁷³ ⁷⁴ Like the Waze CCP, the CVPDP includes a very specific objective: “to support the independent evaluation of CV Pilot deployments.”⁷⁵ Independent evaluators who can receive access to the data include Noblis, the Texas A&M Transportation Institute, and Volpe.⁷⁶ Select datasets involved in these pilots

⁶⁶ “Frequently Asked Questions,” FOIA.gov.

⁶⁷ “State Freedom of Information Laws,” National Freedom of Information Coalition.

⁶⁸ Kopan, “Exclusive: Government Transparency Site Revealed Social Security Numbers, Other Personal Info.”

⁶⁹ “Data Provider,” U.S. Department of Transportation, Secure Data Commons.

⁷⁰ “Data Analyst,” U.S. Department of Transportation, Secure Data Commons.

⁷¹ “About the Secure Data Commons Portal,” U.S. Department of Transportation, Secure Data Commons.

⁷² “Safety Data Initiative Pilot Projects,” U.S. Department of Transportation.

⁷³ “Data Analyst,” U.S. Department of Transportation, Secure Data Commons.

⁷⁴ “20 Questions About Connected Vehicles,” U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology. The USDOT defines connected vehicles as those that “enable safe, interoperable networked wireless communications among vehicles, the infrastructure, and passengers’ personal communications devices.” Ideally, connected vehicles will allow cars and other vehicles (e.g., buses, trains) to “talk” to one another to prevent injuries, reduce travel times, and cut emissions.

⁷⁵ Ibid.

⁷⁶ “Connected Vehicle Pilot Deployment Program, Session 3: Evaluating Performance and Long-Term Sustainment,” U.S. Department of Transportation, ITS

(and a few other pilots including Virginia and San Diego) are also available for public access on a different USDOT platform, the **Intelligent Transportation Systems (ITS) Data Hub**.⁷⁷ The publicly available data includes safety message visualization data,⁷⁸ road weather demonstration data,⁷⁹ and several other datasets determined not to include personally or commercially sensitive information.⁸⁰

Determining which data should be made publicly available and which requires access restrictions to protect individual privacy and proprietary interests is at the core of the dialogue around mobility data sharing. Another example of a USDOT program that walks this line carefully and effectively is the **Transportation Secure Data Center (TSDC)**, maintained by the National Renewable Energy Laboratory (NREL) through a partnership between the USDOT and the U.S. Department of Energy. The TSDC aggregates data from travel surveys and studies—including household travel surveys and data collected from GPS—into a single, publicly available repository. NREL converts the data into an anonymized and consistent format prior to publication. NREL also grants researchers and other stakeholders access to the datasets for specific purposes, such as research. This limited-access model has enabled TSDC data to study transportation-system emissions, examine travel patterns, and plan alternative fuel stations.⁸¹

The USDOT is also developing another data portal, the **Data for Automated Vehicle Integration (DAVI)**, as a “multimodal initiative to identify, prioritize, monitor, and – where necessary – address data exchange needs for automated vehicles (AV) integration across the modes of transportation.” The DAVI website emphasizes that for automated vehicles, data will enable safer and more efficient “integration of AVs into the transportation system.”⁸² A recent example of where USDOT is expanding the use of data repositories is the spring 2019 release of a notice for funding opportunity for Automated Driving Assistance (ADS) demonstrations. Applicants are required to share demonstration data via an existing USDOT data repository or via a “third-party system where USDOT analysts can conduct their work.”⁸³

The extent to which the listed public platforms will grow or evolve is not yet clear. Whether one or more will emerge as the right fit for mobility data remains an open question.

Mobility-related big data in the private sector

Streetlight, INRIX, Amazon Web Services, and Replica (part of Google’s Sidewalk Labs and related subsidiary Waze) are a few of the well-recognized holders of proprietary “big data” with applications for mobility planning. These big mobility-data vendors carefully secure their datasets and/or package data into aggregated formats for use by planners and practitioners. Purchasing such data from these purveyors can be expensive. For smaller state agencies and cities, data costs represent a significant obstacle to data accessibility.

Alphabet’s **Sidewalk Labs** has a rich supply of movement data from cell phones, which can inform transportation and land-use planning.⁸⁴ One example of the Sidewalk capabilities is Sidewalk Toronto, a mixed-use data project for Toronto’s Eastern Waterfront. Sidewalk Toronto began in March 2017 as a collaboration between the City of Toronto and Sidewalk Labs. The project aims to integrate data into everyday city life in order to promote responsible urban growth.⁸⁵ Currently, Sidewalk Toronto manages two projects for mobility data sharing—**OldTO and Toronto Transit Explorer (ToTX)**. OldTO geotags historic photos of Toronto using Sidewalk Labs’ open-source

Joint Program Office.

⁷⁷ “Explore Our Data.” U.S. Department of Transportation, ITS DataHub.

⁷⁸ Cronin, “Vehicle Based Data and Availability.” Basic Safety Message (BSM) data includes vehicle size, position, speed, acceleration, brake system status, and other information regarding a vehicle’s expected/predicted path.

⁷⁹ “Road Weather Demonstration Data,” U.S. Department of Transportation. Road Weather Demonstration (RWD) data includes atmospheric pressure, dewpoint temperature, relative humidity, air temperature, surface temperature, and other vehicle-specific data (e.g., wiper status, speed).

⁸⁰ Ibid.

⁸¹ “About the Transportation Secure Data Center,” U.S. Department of Energy, NREL.

⁸² “Data for Automated Vehicle Integration (DAVI),” U.S. Department of Transportation.

⁸³ “NOFO: Automated Driving System (ADS) Demonstration Grants Department of Transportation DOT Federal Highway Administration,” U.S. Department of Transportation Highway Research and Development Program.

⁸⁴ “Home,” Sidewalk Labs.

⁸⁵ Ibid.

map to better understand urban development since the 1850s.⁸⁶ Toronto Transit Explorer (ToTX) is an open-source navigation tool that helps users optimize their travel around Toronto. Methods of travel included are public transit, bicycle, bikeshare, wheelchair, and walking. Data delivered to the user includes recommended route(s), route length, and estimated travel time based on public transit-data feeds. ToTX also enables to compare the efficiency and accessibility of different transportation methods along given routes. A shortcoming of this program is that transit data is based on weekday scheduled service rather than real-time service.

Another source of privately held data is **Streetlight Data**. Streetlight Data has been processing more than 60 billion new location records each month since 2014.⁸⁷ The company's data comes from smartphone apps, navigation programs, and census, parcel, and road network reports. Streetlight Data measures mobility of pedestrians, bikes, and vehicles across U.S. roads and "zones" using proprietary Streetlight Insight Metrics. Specifically, Streetlight Data features origin-destination pairs that track mobility through different geographic zones of interest (e.g., major thoroughfares, commercial districts). Mobility is evaluated based on statistics like trip volume, average trip time, annual average daily traffic, trip purpose, and more.⁸⁸ Streetlight Insight Metrics can also be customized by data-collection period, time of day, type of day (e.g., weekends), and trip type (e.g., personal, commercial). Streetlight Data's target audience is transportation and urban planners. In the past, Streetlight Data has worked with Napa Valley, CA, to assist with the development of their Countywide Transportation Plan and the Virginia Department of Transportation to identify high-demand routes in Northern Virginia.^{89, 90}

Other datasets are held privately but shared among several companies. The private sector has long shared data among industry networks to address issues such as cybersecurity, and safety. An official call for this type of structured data sharing was introduced in 1998 with Presidential Decision Directive 63 (PDD-63), which encourages Information Sharing and Analysis Centers (ISACs) to be "set up by the private sector in cooperation with the Federal government and modeled on the Centers for Disease Control and Prevention."⁹¹ The **Auto-ISAC** was established in 2015 to share information about vehicle-related cybersecurity risks on a secure data platform.⁹²

Other Business-to-Business (B2B) data-sharing platforms include the **ITS America Data Exchange**, a pilot project providing a platform for sharing data from Internet of Things (IoT) providers. Data on this exchange is available to all ITS-America member organizations.⁹³ The **Project for the Establishment of Generally Accepted quality criteria, tools and methods as well as Scenarios and Situations (PEGASUS)** is a second B2B data-sharing platform. PEGASUS was launched with support from the German Federal Ministry for Economic Affairs and Energy (BMWi) and focuses on highly automated vehicles (HAVs). For instance, the platform enables sharing of HAV test scenario designs.⁹⁴

Expanding access to privately held data will require policies to ensure that personally and commercially sensitive information is protected. One option is to narrowly restrict data access, e.g., to police officers or investigators with a warrant. A more transparent option—and one that would facilitate new insights—would be to also provide limited or full access to raw or aggregated datasets for research purposes. It is unlikely that any raw data will be made available to the broader public. Tough questions remain regarding the appropriate levels of aggregation and on what grounds other stakeholders (e.g., analytics companies, journalists) may be granted access to such data.

⁸⁶ "OldTO," Sidewalk Labs.

⁸⁷ "Home," Streetlight Data.

⁸⁸ "StreetLight InSight Metrics," StreetLight Data.

⁸⁹ "Transportation Demand Management in Northern Virginia," StreetLight Data.

⁹⁰ "Napa Valley Transportation Plan," StreetLight Data.

⁹¹ "Fact Sheet: Protecting America's Critical Infrastructures: PDD 63," The White House.

⁹² "Home," Auto-ISAC.

⁹³ "ITS America Data Exchange," The Intelligent Transportation Society of America.

⁹⁴ "PEGASUS Method: An Overview," PEGASUS.

Recommendation 4. Leverage innovative transportation-planning tools. (Cities, Regions)

In addition to streamlining mobility data sharing, there are opportunities to improve mobility data evaluation. The easier it is for planners and other practitioners to incorporate transportation data into decision making, the more informed and responsive those decisions will be.

There are both proprietary and open-source platforms for mapping and visualizing mobility data. Proprietary platforms typically have greater technical and analytical capabilities, but cost money to use. Relying on private platforms to store and/or visualize public data may also set a troubling precedent. Open-source platforms provide more transparency and accessibility, but they can be difficult and expensive for public agencies and other not-for-profit entities to maintain.

There are several private companies developing such platforms. **Populus** operates a multi-city mapping platform designed to assist cities in tracking shared mobility services such as bikes, scooters, and vehicles.⁹⁵ Platform data is location-based and time sensitive. Populus data is compiled using the **Populus Mobility Manager** and is available in real time, which allows cities to adjust planning decisions in response to observable trends. Populus' customers and partners include King County Metro Transit in Seattle, the Puget Sound Regional Council, the Seattle Department of Transportation, the U.S. Department of Energy, the District Department of Transportation in Washington, D.C., and the City of Arlington, Virginia.⁹⁶

Services that offer similar types of fleet-specific data include **Swiftly** and **Remix**. Swiftly is a private mobility data collection company that claims to automatically collect vehicle data to create live predictions of trip-travel times for transit fleets.⁹⁷ Importantly, Swiftly does not offer data on single-occupancy vehicles or new mobility service providers. For such data, one might turn to Remix, which offers an integrated suite of products that contain public transit, street-specific, and new mobility data, often in real time.^{98, 99} Both companies market themselves towards transit operators and schedulers as well as city and regional planners.¹⁰⁰

Coord is a private firm that offers a multi-city platform for visualization and analysis of publicly available curb-asset data. Data is collected through Coord's Surveyor mobile application (either by Coord or by others using the Surveyor tool app)¹⁰¹ and is standardized according to the Open Curb Assets Data Specification guidelines.^{102 103} The data includes locations and descriptions of curb assets (such as bollards, bike racks, bus stops, crosswalks, curb cuts, curb extensions, curb paint, fencing, fire hydrants, lane markings, parking meters, and parking signs), as well as information on the date and time each asset was surveyed. Coord's target audiences include local government agencies, mobility and logistics partners, and the architecture, engineering, and construction industries. Coord data can be found in the Open Curbs database, which currently includes data from Santa Monica, Denver, Paris, Milan, Los Angeles, and San Francisco.¹⁰⁴

An emerging leader in the open-source planning arena is **SharedStreets**,¹⁰⁵ a nonprofit that emerged from the World Bank's **Open Transport Partnership**.¹⁰⁶ SharedStreets offers a protocol for transportation data standardization and a repository for storage and sharing of open-source, non-proprietary geospatial data. The SharedStreets protocol employs a linear referencing system that is more flexible than traditional GIS positioning.¹⁰⁷ Rather than describing a reference point as a single dot on a map, the SharedStreets protocol describes a reference point in relation

⁹⁵ "Home," Populus.

⁹⁶ Ibid.

⁹⁷ "Home," Swiftly.

⁹⁸ "Home," Remix.

⁹⁹ "Remix: Our Customers," Remix.

¹⁰⁰ See Appendix A for specific implementation details.

¹⁰¹ "Home," Coord.

¹⁰² "Surveyor," Coord.

¹⁰³ "Open Curb Assets Data Specification," Coord.

¹⁰⁴ "Open Curb Data," Coord.

¹⁰⁵ "Home," SharedStreets.

¹⁰⁶ Muller, "The World Bank Launches New Open Transport Partnership."

¹⁰⁷ "SharedStreets Referencing System," Shared Streets.

to other street activities and landmarks. This supports clear and robust data linkage across different basemaps and GIS data layers. SharedStreets works by allowing users to maintain their own internal GIS basemaps, which lets users share individual segment references (rather than full renderings) to protect proprietary data and other intellectual property. SharedStreets is focused on encouraging collaboration between public and private entities in order to improve collection of transport data. The company has several relevant pilot projects, including a Mobility Metrics project that allows mapping of micromobility data. SharedStreets also has a **Taxi and TNC Activity** project operating in Washington D.C. and Toronto, which hosts “high resolution anonymized and aggregated data on for-hire vehicle pick-up and drop-off.”^{108, 109}

Mapping and analytical tools such as those described in this section offer promise for decision makers across the mobility landscape. These tools may be able to aid in finding new solutions to reduce barriers for data analysis and visualization, which can help non-technical audiences achieve informed and responsive policy solutions.

Conclusion

Mobility data can offer much for city, regional, and state planners. But there remain significant obstacles to collecting and sharing mobility data among public and private partners. Valuable data is held by the private sector, which may be reticent to share data with public actors for legitimate reasons. Meanwhile, many public officials without access to key data struggle to make informed decisions. Oversharing and undersharing mobility data are both problematic. An intermediate approach is needed.

This report recommends four principles that policymakers and other stakeholders should consider to expand mobility data sharing while addressing privacy and proprietary concerns. These are:

- (1) Foster voluntary agreement among mobility providers for a set of standardized data specifications. (Cities, Regions)
- (2) Develop clear data-sharing requirements designed for transportation network companies (TNCs) and other mobility providers. (Cities, Regions, States)
- (3) Establish publicly held big-data repositories, managed by third parties, to securely hold mobility data and provide structured access by states, cities, and researchers. (States, Federal)
- (4) Leverage innovative land-use and transportation-planning tools. (Cities, Regions)

The Mobility Data Standard (MDS) may represent a useful jumping-off point for Recommendation 1. The MDS was inspired by open and flexible transportation-data standards that came before it (such as the GTFS and the GBFS), but it goes a step further by providing a two-way street for data sharing among mobility providers and public-sector partners. The MDS was originally developed for scooters but is flexible enough that it could be readily extended to include other types of emerging mobility services. However, the MDS engenders some security concerns that may need to be addressed in order for the standard to achieve widespread adoption.

There is not yet a broadly accepted, replicable policy template for TNC data-sharing requirements (Recommendation 2). Nevertheless, the case studies presented in this report highlight the need for states and cities to link their data requests to specific public objectives. Public officials should also strive for consistent reporting requirements across jurisdictional boundaries. This will reduce administrative burdens on TNCs that could discourage new market entrants and will reduce compliance costs that could be passed onto consumers.

New data-collection efforts call for new, secure data repositories (Recommendation 3). Many such repositories already exist. One that may be particularly relevant to mobility data is the USDOT’s Secure Data Commons (SDC). The SDC includes protocols to protect personally and commercially sensitive information while preserving access

¹⁰⁸ “Mobility Metrics,” SharedStreets.

¹⁰⁹ “Taxi and TNC Activity,” SharedStreets.

for research. However, these protocols may sometimes be so strict as to make it difficult for researchers to pursue new lines of inquiry. We argue that federal agencies are best positioned to establish data repositories, though we acknowledge that there may be value for states, regions, and cities in establishing separate repositories for specific policy needs. Regardless of the level of governance at which a repository is established, assigning a publicly accountable third party to repository management can be an effective way to alleviate security concerns.

Data is not useful unless it is presented in a way that enables non-technical audiences, including decision makers, to glean useful insights. This report presents several innovative transportation-planning tools that could be leveraged to help integrate and analyze mobility data across networks and modes (Recommendation 4). Such tools can also help address privacy concerns by incorporating moderate aggregation. Among these, SharedStreets stands out as a way to empower governments to combine and present mobility data comprehensively.

Mobility data sharing can enable communities to make dynamic and responsive planning decisions. This paper presents recommendations to make data sharing simpler, cheaper, and more secure. We caveat, however, that these recommendations are preliminary. Considerable further review and analysis is needed to assess costs, barriers, and the extent to which they address data-sharing challenges presented herein.

Appendices

A. Examples of Existing Repositories for New Mobility Data

Program	Primary Developer	Repository Type	Publicly Accessible?	Free?	Purpose	Implemented By/In
Coord	Coord	Curbs data	Partially	No	Inform policy, regulation, and compliance regarding curb assets	-Santa Monica -Denver -Paris -Milan -Los Angeles -San Francisco
Data description	Locations, descriptions, and timestamps of curb assets					
Populus	Populus	Curbs data	Yes	No	Assist cities in tracking shared mobility services such as bikes, scooters, and vehicles	-King County Metro -Puget Sound Regional Council -Seattle Department of Transportation -U.S. Department of Energy -District Department of Transportation in Washington, D.C. -City of Arlington, Virginia
Data description	Aggregated location-based real-time data tracking shared mobility scenarios (e.g., bikes, scooters, vehicles). Distributions of available shared vehicles, bicycles, and scooters, measured in aggregate by location (e.g., census tract, district), time of day, and over time.					
Shared Streets	Open Transport Partnership	Curbs data	Yes	Yes	Standardize data to improve the collection of transport data	TBA
Data description	Reference points are described in relation to other street activity and landmarks.					
California Public Utilities Commission	CPUC	TNC data	Yes, in aggregate	Yes	Regulate TNCs operating in California	California
Data description	Includes (1) geographic data [zip codes where trips begin and end, per TNC vehicle]; (2) driver information [driver completion of a mandatory TNC training course, driver violations and suspensions, total hours logged annually by driver]; (3) accessibility information [WAV rides requested and fulfilled, wait times between WAV ride acceptance and pick-up (beginning in 2020)]; and (4) emissions data [vehicle-miles driven per vehicle type, percent vehicle type in fleet, percent share of miles completed by qualified zero-emissions vehicles].					
New York City Taxi and Limousine Commission	City of New York	TNC Data	Yes, by taxi zone	Yes	Regulate TNCs operating in New York City	New York City
Data description	Includes: (1) geographic data [origin and destination of each trip by “taxi zone”]; (2) driver information [driver completion of a mandatory background check, driver accidents or license activities]; (3) accessibility information [number of people with wheelchairs requesting trips]. Does not include emissions data.					

Program	Primary Developer	Repository Type	Publicly Accessible?	Free?	Purpose	Implemented By/In
Portland Bureau of Transportation	City of Portland, Oregon	TNC data	No	N/A	Regulate TNCs operating in Portland	Portland, Oregon
Data description	Includes: (1) geographic data [origin and destination of each trip reported by latitude and longitude]; (2) driver information [criminal activity of drivers]; and (3) accessibility information [number of people with wheelchairs requesting trips]. Does not include emissions data.					
RideAustin	Ride Share Austin, a nonprofit 501(c)(3) rideshare program	TNC data	Yes	Yes	Transparency regarding TNCs operating in Austin; inform transportation research	Austin, Texas
Data description	Includes: (1) geographic data [origin and destination of each trip reported by latitude and longitude]; and (2) driver information [driver completion of a mandatory fingerprint-based background check]. Does not include accessibility information or emissions data.					
Transportation Network Company Division		TNC data	Yes, in aggregate	Yes	Regulate TNCs operating in Massachusetts	Massachusetts
Data description	Includes: (1) geographic data [origin and destination of each trip reported by municipality]; and (2) driver information [driver completion of mandatory background check, driver and rider complaints and suspensions]. Does not include accessibility information or emissions data.					
Connected Vehicle Pilot Deployment Program (CV Pilot)	US Department of	Trip travel data	Partially	Yes	Encourage the use of connected vehicle applications by private and public stakeholders	-New York City DOT -Tampa-Hillsborough Expressway Authority, Florida -Wyoming DOT
Data description	Includes data from: (1) New York City [data on vehicle locations, path histories, and current directions; pedestrian use of crosswalks; traffic signals]; (2) Tampa, FL [data on vehicle speeds, wrong-way drivers, pedestrian use of crosswalks, public transit signals and routes, streetcar track conflicts, real-time traffic conditions]; and (3) Wyoming [data on roadside units, fleet vehicle movement, weather, and traffic].					
Data for Automated Vehicle Integration (DAVI)	US Department of	Trip travel data	N/A	N/A	Safe integration of AVs into existing transportation networks	Submission-dependent
Data description	Various data on automated vehicles.					
Fleet DNA	National Renewable Energy Laboratory	Trip travel data	Yes	Yes	Help manufacturers and fleet managers optimize service	Submission-dependent
Data description	Aggregated data on commercial fleet operations.					

Program	Primary Developer	Repository Type	Publicly Accessible?	Free?	Purpose	Implemented By/In
Replica	Sidewalk Labs	Trip travel data	Yes	No	Generate synthetic projections of travel activity to be used by urban planners	-Kansas City -Chicago, IL
Data description	Third-party mobile app data.					
Sidewalk Toronto	City of Toronto and Sidewalk Labs	Trip travel data	Yes	Yes	Integrate data into everyday city life to promote responsible urban growth	Toronto, Ontario
Data description	Geo-tagged historic photos of curb-asset data; optimized trip-route data.					
Streetlight Data	StreetLight Data, Inc.	Trip travel data	Yes	No	Measure mobility of pedestrians, bikes, and vehicles across U.S. roads and zones	-Napa Valley, California -Virginia
Data description	Origin-destination pairs through geographic zones of interest.					
Transportation Secure Data Center	National Renewable Energy Laboratory	Trip travel data	Yes, in aggregate	Yes	Relieve need for individual public agencies to collect transportation data	Study/survey dependent
Data description	GPS travel data from travel surveys and studies.					
Waze Connected Citizens Program (CCP)	U.S. Department of Transportation's Volpe National Transportation Systems Center.	Trip travel data	No	N/A	Improve police-reported crash predictions	Maryland
Data description	Crowdsourced crash data from Waze.					
Automotive Information Sharing and Analysis Center (Auto-ISAC)	Automakers	Vehicle	No	N/A	Share information about vehicle-related cybersecurity risks among industry partners	Membership includes BMW, Ford, General Motors, Hitachi, Honda, Hyundai, Kia, Mazda, Mercedes-Benz, Mitsubishi, Nissan, Subaru, Toyota, Volkswagen, and Volvo
Data description	Data to help protect critical U.S. infrastructure.					

References

- “16.40.240 TNC Operating Responsibilities and Prohibitions. Chapter 16.40 Private For-Hire Transportation Regulations.” City of Portland, OR. Accessed June 28, 2019. <https://www.portlandoregon.gov/citycode/article/689632>.
- “20 Questions About Connected Vehicles.” U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology. Accessed July 24, 2019. https://www.its.dot.gov/cv_basics/cv_basics_20qs.htm.
- “2018 TLC Fact Book.” NYC Office of Policy and External Affairs. Accessed June 18, 2019. https://www1.nyc.gov/assets/tlc/downloads/pdf/2018_tlc_factbook.pdf.
- “About.” Green Button Alliance. Accessed May 24, 2019. <http://www.greenbuttondata.org/>.
- “About NABSA.” North American Bikeshare Association. Accessed May 24, 2019. <https://nabsa.net/about/>.
- “About the Secure Data Commons Portal.” U.S. Department of Transportation, Federal Highway Administration. Accessed May 28, 2019. <https://its.dot.gov/data/secure/about.html>.
- “About the Transportation Secure Data Center.” U.S. Department of Energy, National Renewable Energy Laboratory (NREL). Accessed June 11, 2019. <https://www.nrel.gov/transportation/secure-transportation-data/tsdc-about.html>.
- “About TLC Trip Record Data.” NYC Taxi & Limousine Commission. Accessed June 6, 2019. <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>.
- Accuardi, Zak, Chris Pangilinan, Aaron Antrim, and Greg Rucks. “The Data Transit Riders Want.” New York, NY: Transit Center, 2018. http://transitcenter.org/wp-content/uploads/2018/12/The_Data_Transit_Riders_Want.pdf.
- Alexander, Lauren, Shan Jiang, Mikel Murga, and Marta C. González. “Origin–Destination Trips by Purpose and Time of Day Inferred from Mobile Phone Data.” *Transportation Research Part C: Emerging Technologies* 58 (2015): 240–250.
- “Application Programming Interface.” Wikipedia. Accessed June 14, 2019. https://en.wikipedia.org/wiki/Application_programming_interface.
- “Article 7. Transportation Network Companies [5430-5450].” California Public Utilities Commission. Accessed June 18, 2019. http://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=PUC&division=2.&title=&part=&chapter=8.&article=7.
- “About PEGASUS.” PEGASUS. Accessed June 30, 2019. <https://www.pegasusprojekt.de/en/about-PEGASUS>.
- “Back to Tranquility: Swiftly Restores Transit Calm on the Oregon Coast.” Swiftly. Last modified May 15, 2018. <https://blog.goswift.ly/back-to-tranquility-swiftly-restores-transit-calm-on-the-oregon-coast-fc98b1f5cca7>.
- “Baltimore Develops a Golden Method for Transit Signal Priority.” Swiftly. Last modified October 26, 2018. <https://blog.goswift.ly/baltimore-develops-a-golden-method-for-transit-signal-priority-f1b63381cb8f>.
- “CARTA Jump-Starts Customer Delight with Swiftly.” Swiftly. Last modified May 15, 2018. <https://blog.goswift.ly/carta-jump-starts-customer-delight-with-swiftly-6971c43568e3>.
- Chopra, Aneesh. “Modeling a Green Energy Challenge after a Blue Button.” The U.S. National Archives and Records Administration. Accessed June 17, 2019. <https://obamawhitehouse.archives.gov/blog/2011/09/15/modeling-green-energy-challenge-after-blue-button>.

“City Code 16.40.20: Private For-Hire Transportation Regulations – Chapter Applies to All Companies, Drivers, and Vehicles.” City of Portland, Oregon. June 29, 2018. https://www.portlandoregon.gov/citycode/28593#cid_689620.

“Clean Miles Standard.” California Air Resources Board. Accessed May 14, 2019. <https://ww2.arb.ca.gov/our-work/programs/clean-miles-standard>.

“Cleansed Data from Transportation Studies and Surveys.” The National Renewable Energy Laboratory. Accessed June 30, 2019. <https://www.nrel.gov/transportation/secure-transportation-data/tsdc-cleansed-data.html>.

“Connected Vehicle Pilot Deployment Program, Session 3: Evaluating Performance and Long-Term Sustainment.” U.S. Department of Transportation, ITS Joint Program Office. Accessed July 24, 2019. https://www.its.dot.gov/presentations/world_congress2017/CVP%20ITSWC17%20Session%203_Compressed_FINAL.PDF.

Cronin, Brian. “Vehicle Based Data and Availability.” U.S. Department of Transportation, Intelligent Transportation Systems Joint Program Office. Accessed July 24, 2019. https://www.its.dot.gov/itspac/october2012/PDF/data_availability.pdf.

“Curbside Data Management.” SharedStreets. Accessed June 30, 2019. <https://sharedstreets.io/curbside-management/>.

“CV Pilots News and Events.” U.S. Department of Transportation. Accessed June 18, 2019. <https://www.its.dot.gov/pilots/index.htm>.

“Data.” Panel at 3 Revolutions Policy Conference, University of California, Davis, March 2019.

“Data Analyst.” U.S. Department of Transportation, Secure Data Commons. Accessed May 14, 2019. <https://its.dot.gov/data/secure/data-analyst.html>.

“Data for Automated Vehicle Integration (DAVI).” U.S. Department of Transportation. Last modified May 10, 2019. <https://www.transportation.gov/av/data>.

“Data Provider.” U.S. Department of Transportation, Secure Data Commons. Accessed May 14, 2019. <https://its.dot.gov/data/secure/data-provider.html>.

“DCTA Revamps its Customer Service in the Heart of Cowboy Country.” Swiftly. Last modified September 10, 2018. <https://blog.goswift.ly/dcta-revamps-its-customer-service-in-the-heart-of-cowboy-country-e1ca18c7e1cb>.

De Montjoye, Yves-Alexandre, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. “Unique in the Crowd: The Privacy Bounds of Human Mobility.” *Scientific Reports* 3 (2013): 1376.

“Departmental Privacy Risk Management Policy, DOT Order 1351.18 Privacy Risk Management.” U.S. Department of Transportation. Washington, D.C., 2014. <https://www.transportation.gov/sites/dot.gov/files/docs/DOT%20Order%201351.18%20Privacy%20Risk%20Management%20Policy.pdf>.

“Experts Identified Areas for Operational Improvements Without Implementing Organizational Changes.” Government Accountability Office. Washington, DC, DC: U.S. Department of Transportation, March 13, 2017. <https://www.gao.gov/products/GAO-17-478>.

“Explore Our Data.” US. Department of Transportation, ITS DataHub. Accessed May 14, 2019. <https://www.its.dot.gov/data/>.

“Extensible Markup Language (XML) 1.0 (Fifth Edition).” W3C. Accessed June 18, 2019. <https://www.w3.org/TR/REC-xml/>.

“Fact Sheet: Protecting America’s Critical Infrastructures: PDD 63.” The White House. (May 22, 1998). <https://fas.org/irp/offdocs/pdd-63.htm>.

“FAQ for TNC Passengers.” New York Department of Motor Vehicles. Last modified May 14, 2018. <https://dmv.ny.gov/more-info/faq-tnc-passengers>.

“Feed Entities.” Google Developers. Accessed June 18, 2019. <https://developers.google.com/transit/gtfs-realtime/guides/feed-entities>.

Ferris, Brian. “GTFS-Realtime vs. REST/XML.” Google Groups (blog). n.d. <https://groups.google.com/forum/#!topic/gtfs-realtime/oCX7BxLE1A8>.

Fielding, Roy. “Architectural Styles and the Design of Network-based Software Architectures.” University of California, Irvine. Accessed June 18, 2019. <https://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>.

“Frequently Asked Questions.” Automotive Information Sharing and Analysis Center (Auto-ISAC). Accessed June 18, 2019. <https://www.automotiveisac.com/faqs/>.

“Frequently Asked Questions.” FOIA.gov. Accessed June 30, 2019. <https://www.foia.gov/faq.html>.

“Frequently Asked Questions.” Office of Information Policy (OIP), United States Department of Justice. Last modified June 2019. <https://www.foia.gov/faq.html>.

“Frequently Asked Questions.” RideAustin. Accessed June 17, 2019. <http://www.rideaustin.com/faq>.

Gallaga, Omar L. “In Their Own Words: All of Austin’s Ride-hail Companies Make Their Pitch for Your Business.” Austin American-Statesman. September 24, 2018. <https://www.statesman.com/news/20170606/in-their-own-words-all-of-austins-ride-hail-companies-make-their-pitch-for-your-business>.

“General Bikeshare Feed Specification” North American Bikeshare Association. Accessed May 24, 2019. <https://github.com/NABSA/gbfs/blob/master/systems.csv>.

“General Transit Feed Specification.” TransitWiki. Last modified December 12, 2017. https://www.transitwiki.org/TransitWiki/index.php/General_Transit_Feed_Specification.

“General Transit Feed Specification References.” The MobilityData Project. Accessed June 18, 2019. <https://gtfs.org/reference/static>.

“Green Button Connect My Data (CMD) Standard.” Green Button Alliance. Last modified August 2018. <https://www.greenbuttonalliance.org/assets/docs/Collateral/2018-08%20Green%20Button%20CMD%20and%20Testing%20Data%20Sheet.pdf>.

“Green Button for My Home.” Green Button Alliance. Accessed May 24, 2019. <http://www.greenbuttondata.org/residential.html>.

“GTFS Static Overview” Google Developers. Last modified July 26, 2016. <https://developers.google.com/transit/gtfs/>.

“GTFX-FLEX,” Mobility Data. Accessed May 24, 2019. <https://github.com/MobilityData/gtfs-flex>.

“Guide to Identifying Personally Identifiable Information (PII).” University of Pittsburgh, Information Technology. <https://www.technology.pitt.edu/help-desk/how-to-documents/guide-identifying-personally-identifiable-information-pii>.

“Home.” Auto-ISAC. Accessed May 28, 2019. <https://www.automotiveisac.com/>.

“Home.” Coord. Accessed April 22, 2019. <https://www.coord.co/>.

“Home.” Populus. Accessed May 28, 2019. <https://www.populus.ai/>.

“Home.” Remix. Accessed June 30, 2019. <https://www.remix.com/>.

“Home.” SharedStreets. Accessed May 28, 2019. <http://sharedstreets.io/>.

“Home.” Sidewalk Labs. Accessed May 28, 2019. <https://www.sidewalklabs.com/>.

“Home.” Sidewalk Toronto. Accessed May 28, 2019. <https://sidewalktoronto.ca/>.

“Home.” Streetlight Data. Accessed June 17, 2019. <https://www.streetlightdata.com/>.

“Home.” Swiftly. Accessed June 20, 2019. <https://www.goswift.ly/>.

“Home.” Open Mobility Foundation. Accessed July 21, 2019. <https://www.openmobilityfoundation.org/>.

“How It Works.” Wyoming Department of Transportation Connected Vehicle Pilot. Accessed June 18, 2019. <https://wydotcwp.wyroad.info/>.

“How Juneau Is Wooing Cruise Ship Passengers with a Transit Rebrand.” Swiftly. Last modified October 9, 2018. <https://blog.goswift.ly/how-juneau-is-wooning-cruise-ship-passengers-with-a-transit-rebrand-e5f36d3b9eee>.

“Introducing Replica.” Sidewalk Labs. Accessed June 30, 2019. <https://www.sidewalklabs.com/blog/introducing-replica-a-next-generation-urban-planning-tool/#>.

“ITS America Data Exchange.” The Intelligent Transportation Society of America. Accessed May 28, 2019. <https://www.itsa.org/itsa-data-exchange/>.

“ITS America Data Exchange.” The Intelligent Transportation Society of America. Accessed June 30, 2019. <https://itsa.terbine.io/#>.

“ITSA Data Exchange Cliff Notes.” The Intelligent Transportation Society of America. Accessed June 30, 2019. https://static1.squarespace.com/static/596fb16003596e0fa70a232f/t/5b1ff2568a922d10c3ef98e3/1528820310898/Q%26A_+ITSA+Data+Exchange.pdf.

Jenn, Alan, Katalin Springel, and Anand R. Gopal. “Effectiveness of Electric Vehicle Incentives in the United States.” *Energy Policy* 119 (2018): 349–356.

Kopan, Tal. “Exclusive: Government Transparency Site Revealed Social Security Numbers, Other Personal Info.” CNN. Accessed July 24, 2019. <https://www.cnn.com/2018/09/03/politics/foia-revealed-social-security-numbers/index.html>.

“Launch Process: Choose a Data Submission Method.” Google Support. Accessed May 24, 2019. https://support.google.com/transitpartners/answer/2529132?hl=en&ref_topic=2527461.

Levine, Kendra K. "A Brief History of GTFS." UC Berkeley: Institute of Transportation Studies Library. October 14, 2014. <https://library.its.berkeley.edu/node/265>.

Lubarsky, Boris. "Re-Identification of 'Anonymized' Data." Georgetown Law Technology Review 202 (2017).

"MassDOT and the MBTA to Partner with Transportation Network Companies to Support More Wheelchair-Accessible Vehicles." Massachusetts Bay Transportation Authority. Accessed April 1, 2019. <https://www.mbta.com/news/2019-04-01/massdot-and-the-mbta-partner-transportation-network-companies-support-more>.

McHugh, Bibiana. "Pioneering Open Data Standards: The GTFS Story," in Beyond Transparency: Open Data and the Future of Civic Innovation," edited by Brett Goldstein, 125–137. San Francisco: Code for America, 2013.

"MDS-DEV." City of Los Angeles. Accessed May 24, 2019. <https://github.com/CityOfLosAngeles/mds-dev/tree/master/processing>.

"Mobility Data Specification." City of Los Angeles. Accessed May 24, 2019. <https://github.com/CityOfLosAngeles/mobility-data-specification>.

"Mobility Metrics." Shared Streets. Accessed May 28, 2019. <http://sharedstreets.io/mobility-metrics/>.

Muller, Xavier. "The World Bank Launches New Open Transport Partnership to Improve Transportation through Open Data." The World Bank. Last modified December 19, 2016. <http://www.worldbank.org/en/news/press-release/2016/12/19/the-world-bank-launches-new-open-transport-partnership-to-improve-transportation-through-open-data>.

"Napa Valley Transportation Plan." StreetLight Data. Last modified March 26, 2018. <https://www.streetlightdata.com/napa-valley-transportation-plan>.

Narayanan, Arvind, and Vitaly Shmatikov. "Robust De-Anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset)." University of Texas at Austin (2008).

"New State Website Shares Local Ride Hailing Data." Massachusetts Municipal Association. Last modified May 2, 2018. <https://www.mma.org/new-state-website-shares-local-ride-hailing-data/>.

"NOFO: Automated Driving System (ADS) Demonstration Grants Department of Transportation DOT Federal Highway Administration." U.S. Department of Transportation Highway Research and Development Program, 2019. <https://www.grants.gov/web/grants/view-opportunity.html?oppld=310839>.

Nordrum, Amy. "Los Angeles to Require Scooter Companies to Share Data." IEEE Spectrum: Technology, Engineering, and Science News. Last modified February 19, 2019. <https://spectrum.ieee.org/telecom/wireless/mwc-barcelona-2019-los-angeles-to-require-scooter-companies-to-share-data>.<https://spectrum.ieee.org/telecom/wireless/mwc-barcelona-2019-los-angeles-to-require-scooter-companies-to-share-data>.

"OldTO." Sidewalk Labs. Accessed May 28, 2019. <https://oldto.sidewalklabs.com/>.

"Open Curb Assets Data Specification." Coord. Last modified March 11, 2019. https://docs.wixstatic.com/ugd/8b1277_8e32c9463b3743b7833b9c1e82f0b558.pdf.

"Open Curb Data." Coord. Accessed April 22, 2019. <https://www.coord.co/open-curb-data>.

"Ordinance NO. 20160623-066." City of Austin, Texas. Austin, TX, 2019. https://library.municode.com/tx/austin/ordinances/code_of_ordinances?nodeId=778633.

“Our Members.” The Intelligent Transportation Society of America. Accessed June 30, 2019. <https://www.itsa.org/our-members>.

“PEGASUS Method: An Overview.” PEGASUS. Published 2019, <https://www.pegasusprojekt.de/files/tmpl/Pegasus-Abschlussveranstaltung/PEGASUS-Gesamtmethode.pdf>.

“Project Scope.” New York City Department of Transportation. Accessed June 18, 2019. <https://www.cvp.nyc/project-scope>.

“Publicly-accessible Public Transportation Data.” TransitWiki. May 14, 2019. Accessed June 14, 2019. https://www.transitwiki.org/TransitWiki/index.php/Publicly-accessible_public_transportation_data#List_of_publicly-accessible_public_transportation_data_feeds:_static_data_in_the_GTFS_format.

Pyzyk, Katie. “15 Cities Join Open Mobility Foundation.” Smart Cities Dive. Last modified June 26, 2019. <https://www.smartcitiesdive.com/news/15-cities-join-open-mobility-foundation/557565/>.

“Remix for New Mobility.” Remix. Accessed June 30, 2019. <https://www.remix.com/new-mobility>.

“Remix: Our Customers.” Remix. Accessed June 30, 2019. <https://www.remix.com/customers>.

“Remix Public Transit.” Remix. Accessed June 30, 2019. <https://www.remix.com/transit>.

“Remix Streets.” Remix. Accessed June 30, 2019. <https://www.remix.com/streets>.

“Required Reports TNCs Must Provide the CPUC.” California Public Utilities Commission. Accessed June 6, 2019. <https://www.cpuc.ca.gov/General.aspx?id=3989>.

“Ride Austin’s Datasets.” Data.World. Last modified 2017. <https://data.world/ride-austin>.

“Road Weather Demonstration Data,” U.S. Department of Transportation. Accessed July 24, 2019.

“Rulemaking 12-12-011: Assigned Commissioner’s Ruling Seeking Comments on Proposed Data Reporting Requirements.” California Public Utilities Commission. Last modified February 8, 2019. <http://docs.cpuc.ca.gov/PublishedDocs/Efile/G000/M263/K645/263645339.PDF>.

“Rules and Policies – Protecting PII – Privacy Act.” U.S. General Services Administration. Accessed June 18, 2019. <https://www.gsa.gov/reference/gsa-privacy-program/rules-and-policies-protecting-pii-privacy-act>.

“Safety Data Initiative Pilot Projects.” U.S. Department of Transportation. Last modified April 8, 2019. <https://www.transportation.gov/SafetyDataInitiative/Pilots#waze>.

San Francisco County Transportation Authority (SFCTA), and Joe Castiglione. “Reply Comments of San Francisco County Transportation Authority to Assigned Commissioner’s Ruling Seeking Comments on Proposed Data Reporting Requirements of Rulemaking 12-12-011.” San Francisco, CA: Public Utilities Commission of The State of California, 2019.

Shaheen, Susan, Elliot Martin, Mikaela Hoffman-Stapleton, et al. “Understanding How Cities Can Link Smart Mobility Priorities Through Data.” UC Berkeley: Transportation Sustainability Research Center, (April 2018). Retrieved from <https://escholarship.org/uc/item/7303t6sw>.

“SharedStreets Referencing System.” Shared Streets. Last modified January 14, 2018. <https://github.com/sharedstreets/sharedstreets-ref-system>.

“Smart City Challenge: Lessons for Building Cities of the Future.” U.S. Department of Transportation, 2016. <https://www.transportation.gov/sites/dot.gov/files/docs/Smart%20City%20Challenge%20Lessons%20Learned.pdf>.

“Speeding Up Transit (and Project Timelines) at Santa Clara VTA.” Swiftly. Last modified October 19, 2018. <https://blog.goswift.ly/speeding-up-transit-and-project-timelines-at-santa-clara-vta-24b2f997576d>.

“State Freedom of Information Laws.” National Freedom of Information Coalition. Accessed June 30, 2019. <https://www.nfoic.org/coalitions/state-foi-resources/state-freedom-of-information-laws>.

“StreetLight InSight Metrics.” StreetLight Data. Accessed May 28, 2019. <https://www.streetlightdata.com/transportation-metrics/>.

“Surveyor.” Coord. Accessed April 22, 2019. <https://www.coord.co/surveyor>.

Sweeney, Latanya. “K-Anonymity: A Model for Protecting Privacy.” International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10, no. 05 (2002): 557–570.

“Tahoe Transportation District Gets Real-Time Arrival Information Powered by Swiftly.” Swiftly. Last modified September 7, 2016. <https://blog.goswift.ly/tahoe-transportation-district-gets-real-time-arrival-information-powered-by-swiftly-55e902c91cea>.

“Taxi and TNC Activity.” Shared Streets. Accessed May 28, 2019. <http://sharedstreets.io/taxi-tnc-activity/>.

“The Big Data Conundrum: How to Define It?” Emerging Technology from the arXiv. MIT Technology Review, October 3, 2013. <https://www.technologyreview.com/s/519851/the-big-data-conundrum-how-to-define-it/>.

“TLC Trip Record Data.” NYC Taxi and Limousine Commission. Accessed May 14, 2019. <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>.

Tosi, Davide. “Cell Phone Big Data to Compute Mobility Scenarios for Future Smart Cities.” International Journal of Data Science and Analytics, no. 4 (Spring 2017): 265. <https://doi.org/10.1007/s41060-017-0061-2>.

“Transportation Demand Management in Northern Virginia.” StreetLight Data. Last modified March 16, 2018. <https://www.streetlightdata.com/transportation-demand-management-in-northern-virginia>.

“TNC Permits Issued.” California Public Utilities Commission. Accessed June 26, 2019.

“Transportation Network Companies.” California Public Utilities Commission. Accessed June 26, 2019. <https://www.cpuc.ca.gov/tncinfo/>.

Ward, Jonathan S. and Adam Barker. “Undefined by Data: A Survey of Big Data Definitions.” arXiv, (September 2013). <https://arxiv.org/pdf/1309.5821.pdf>.

“What We’re Doing.” Tampa Hillsborough Expressway Authority. Accessed June 18, 2019. <https://www.tampacvpilot.com/what-were-doing/>.

Yee, Jane Y., Shivana Sidhar, and Lisa P. Tse. “Reply Comments of Uber Technologies, Inc. On Phase III.b. Scoping Memo and Ruling of Assigned Commissioner Track III (TNC Data).” Public Utilities Commission of The State of California, 2019.

Zipper, David. “Cities Can See Where You’re Taking That Scooter.” Slate, April 2, 2019. <https://slate.com/business/2019/04/scooter-data-cities-mds-uber-lyft-los-angeles.html>.