Shared Mobility Data Sharing

Opportunities for Public-Private Partnerships

April 23, 2019

TransLink New Mobility Lab

Prepared by:
Rainer Lempert, UBC Sustainability Scholar
Prepared for:
TransLink

TRANSLINK

Table of Contents

1. Executive Summary

A growing share of trips in Metro Vancouver are being taken through shared mobility providers such as bikesharing and carsharing. This is anticipated to grow with the near-term introduction of ridehailing and other sharing services. As these services become ever more present in our cities, and accessed by customers with mobile phones, the travel data generated has been flagged by regulators and transportation authorities as potentially valuable.

To date, the data generated from these services has been sporadic and inconsistent, reducing the capacity for agencies like TransLink to have the insights necessary to develop robust, thoughtful policies.

Without a standardized way to obtain data from shared mobility providers, TransLink will be unable to create and enforce necessary policies. This report summarizes challenges with sharing data and examines several emerging technologies that could be part of a standardized sharing solution.

Why are providers hesitant to share their data?

- *Data is currency* – Data is a valuable commodity. The act of sharing data requires companies to invest resources. Companies want something in return for access to their data. Sharing data also risks sharing proprietary information with competitors, a further disincentive.
- *Privacy concerns* – Human travel traces are easily identifiable. With access to granular trip data, a bad actor could easily exploit or harm the public.

What technologies could be part of a standardized solution to data sharing?

- *Data standards* – The established standard, General Bikeshare Feed Specification (GBFS), describes the realtime location of micromobility vehicles. An emerging standard, Mobility Data Specification (MDS), also describes historic trip data and has the technical capacity to include other modes of transportation, like ride hailing.
- *Data aggregators and repositories* – A solution to business and personal privacy risks is for a neutral organization, perhaps a nonprofit, to act as an intermediary. This host could aggregate the data, removing individual trip information, before sharing with regulators. SharedStreets is emerging as a viable data repository.

What is TransLink's path forward?

- *Determine Metro Vancouver's data standard* – Determine which data standard (GBFS or MDS) makes the most sense to adopt and enforce. This will largely depend on TransLink's and area municipalities in-house capacity to store, process, and analyze data.
- *Continue to assess emerging data hosts* – SharedStreets' technical capacities could be a part of a standardized data sharing solution. However, it is unclear if SharedStreets can navigate their model's legal and business issues at scale. Continued communication with insiders is crucial to assessing SharedStreets long term viability.

2. Introduction

Mobility data is critical to monitoring and forecasting transportation behavior. Regulatory authorities, like TransLink, rely on robust data to inform transportation planning and policy creation. The rise of shared mobility services, such as ride hailing, car sharing, bike sharing, and e-scooter sharing, has changed the transportation landscape. In order to continue to appropriately inform decision making, regulatory agencies need access to the data generated be these services.

This report summarizes the problems surrounding standardized data sharing and analyzes several emerging technologies that could be part of the solution. This is an evolving, fast-changing field, and as a result the conclusions and recommendations provided by this report should be treated as an initial “roadmap” for TransLink and municipalities in the Metro-Vancouver region to develop a data standard requirement for mobility providers.

2.1 Problem statement

Without consistent trip data, regulatory agencies like TransLink do not have the materials necessary to develop robust, thoughtful policies that optimize transportation efficiencies. There are already systems in place for data acquisition of certain types of trips. Public transit data can be accessed directly through the Compass system. Information on walking, biking, and private vehicles are estimated by regular surveys such as the Regional Trip Diary. Other types of transportation services, however, do not have systems for regulatory data collection. This is particularly true for shared mobility services.

A growing share of trips in Metro Vancouver are being absorbed by shared mobility, such as carsharing and bikesharing. This percentage will only grow with the introduction of ridehailing and e-scooter sharing services. Without a standardized way to obtain data from shared mobility companies, TransLink will be unable to create and enforce policies surrounding these services.

A well-constructed standardized agreement should make it simple, seamless, and safe for companies to share their data. Shared mobility companies are often reluctant to share their data with regulators. There are worries about compromising business intelligence and user privacy, as well as the internal expense of individually wrangling and sharing data for different municipalities with different needs and requests. Technical, legal, and business skills are required on the part of the regulator to standardize data sharing.

2.2 Methods and report structure

Research for this report was done in stages. An online literature scan was undertaken to find policies, techniques, and technologies used to facilitate data sharing between shared use providers and transportation regulatory agencies such as cities and transportation authorities. Relevant resources, including data aggregator and data standard repositories, such as Github, were analyzed and comprise the report's foundational information.

Due to the transient, fast developing nature of this space, these initial findings were not sufficient to adequately address the problem statement. Instead, these findings informed questions directed at knowledgeable transportation professionals at relevant companies and government agencies. These interviews, in turn, inform many of the opinions and recommendations found in this report. Interviews were arranged to solicit opinions from a diverse group of transportation professionals. In industry, employees at the local car sharing coop Modo, e-scooter sharing company Lime, and ride hailing company Uber were interviewed. In government, employees from the City of Vancouver, TransLink, City of Boston, District of Columbia, and United States Department of Transportation Volpe Center were interviewed.

Conversations and educational material from additional expert parties, including Mark Hallenbeck, the Director of the University of Washington's Transportation Data Collaborative, transportation technologies regulation lawyer at Best Best & Krieger Greg Rodriguez, and members of the transportation data aggregator SharedStreets, were incorporated into this report. The recommendations found in this report represent an attempt at synthesizing the wide-ranging insight, knowledge, and perspective of these expert interviews.

3. Background: challenges with data sharing

New technologies allow for cheaper, better data collection. Smartphones have revolutionized personal data collection through the development of location-based services, application-based tracking, and navigation apps. Methods for vehicle location data collection have also improved through advances in telemetry, WiFi and Bluetooth sniffers, and connected vehicles. While there is widely acknowledged need to access this transportation data sharing between providers and regulators challenges in sharing this data persists due to the value of data and privacy concerns.

3.1 Data is currency

The digitization of information has revolutionized economic practices across industries. The ever-growing wealth of data can be used to unlock new sources of economic value by accomplishing things previously impossible: spotting business trends, understanding political and social climates, combatting crime and disease, and so on. According to Craig Mundie, Senior Advisor to the CEO at Microsoft, "data are becoming the raw material of business."[1]

Take-away points

1. Data is a valuable commodity.
2. The act of sharing data requires companies to invest resources.
3. Companies want something in return for access to their data.
4. Willingness to share varies across type of service provider (e.g. ride hailing vs. bike sharing).

The value of data applies to the transportation sector as well. The meteoric rise of Silicon Valley based Transportation Network Companies (TNCs) such as Uber and Lyft, whose core competencies include data management and analysis, has further entrenched data as a valuable commodity within the transportation space. Other shared mobility companies, such car sharing and bike sharing organizations, operate with similar views and assumptions about the value of data. These companies use all vast quantities of data to operate, manage, and optimize their services.

3.1.1 Differing corporate perspectives

The inherent value of data disincentivizes shared mobility organizations from willfully sharing what they've collected with regulators unless they believe they are receiving something in return, whether that be a policy that promotes usage of their service, a reciprocal data exchange, or something else of mutual benefit. While all transportation companies would like to see a return on the data they share, their operating policies are

[1] The Economist, *"Data, data everywhere."* February 25, 2010.

not monolithic. Shared mobility organizations' views and practices regarding data sharing are as diverse as the services they offer. The Technology Innovation and Policy team at the US DOT Volpe Center, noted that, in their experience, "getting data from bike sharing typically isn't too tough, getting data from ride sharing companies is not so easy."

The US DOT's general explanation for this difference in practice is that companies with a less established product in the marketplace are more likely to agree to share data with governing bodies. The act of sharing data holds value for scooter and bike sharing companies, since it can illustrate bikeshare usage, exemplify transparency, and support municipalities efforts to build cycling infrastructure. Ride hailing companies are a more established business, boasting widespread market saturation. As a result of this they often have more leverage when negotiating data sharing agreements and are less likely share large, holistically representative datasets.

Interviews with Uber, including Data Science and Public Policy teams, largely echoed these views. They noted that an unsuccessful early attempt at data sharing with the City of Boston in January 2015. This attempt did not succeed because the City of Boston did not clarify how they wanted to use the data for analysis. Not knowing how the City would use the data, Uber was not incentivized to give full access to its information. Instead, the company shared highly aggregated data at the zip-code level, which was not useful for planning or regulatory purposes.

Scooter sharing companies have shared a slightly different perspective. E-scooter and bike sharing companies are more reliant on acceptance by regulatory agencies than ride hailing companies because they store their vehicles on the public right of way. As ride hailing companies don't require vehicle storage, Uber and Lyft can launch without a city's knowledge, which factors into their hesitation to share data. One scooter sharing company noted that: "[Ride hailing organizations] are afraid of public agencies having an accurate picture of their overall effects, including additions to vehicle miles traveled and congestion." E-scooter and bike sharing companies, on the other hand, "don't have anything to hide."

3.1.2 Labor costs of data sharing

In addition to the potential opportunity cost of not receiving a return on shared data, there is another, more tangible cost shared mobility providers need to factor. The process of sharing data with a municipality, which involves drafting and finalizing a data sharing agreement and developing an appropriate application programming interface (API), requires time, resources, and labor across multiple teams that the mobility provider could be putting towards other projects. These organizational labor costs increase with each additional municipality that asks for a data sharing agreement, especially if the terms and data parameters of these agreements differ. Standardized required data can mitigate these costs to shared mobility providers.

3.2 Privacy concerns

Data sharing is further complicated by inherent privacy concerns and risks. This is true across many industries, but is particularly important in transportation, where company and government data can track individual human mobility. Certain data sets, while useful for informing municipal policy decisions, hold risks if accessible by the public. This could occur through malicious intent if the data is not securely stored, or through Freedom of Information Law (FOIL) requests. For example, loose local FOIL regulations have impeded at least one data sharing agreement, between the City of Seattle and Uber, according to Uber employees.

Take-away points

1. Human travel traces are easily identifiable. A bad actor could exploit granular data towards public harm.
2. Sharing data risks sharing proprietary information with competitors.
3. These two factors further disincentivized providers from sharing granular trip data.

According to Mark Hallenbeck, the Director of the University of Washington's Transportation Data Collaborative, privacy concerns risks fall under two broad categories, *personal risks* and *business risks*.

3.2.1 Personal privacy risks

Research done by Montjoya et al. showed that human travel traces are easily and inherently identifiable.[2] This study analyzed mobility data for 1.5 million people in a European country over 15 months. With four randomly chosen spatiotemporal points it is possible to accurately a unique identify an individual 95% of the time. With only two points an individual can be identified 50% of the time. Even with a lower temporal and spatial resolution identifying unique travel traces is still a relatively easy endeavor (Figure 1).

---

[2] Y.-A. De Montjoye, C. A. Hidalgo, M. Verleysen, & V. D. Blondel, Unique in the Crowd: The Privacy Bounds of Human Mobility, Scientific Reports 3: 1376 (2013).

Given the ease that individual traces can be determined, shared mobility companies worry that giving away too much granular data can result in members privacy being compromised. Hashing relevant identifiers, a common practice that involves algorithmically mapping data to a set of values, has limited benefit. Anonymized wifi, transit farecard, location based service, and GPS data can often be reidentified. In 2014, an independent researcher was able to do this just this using a dataset of New York City

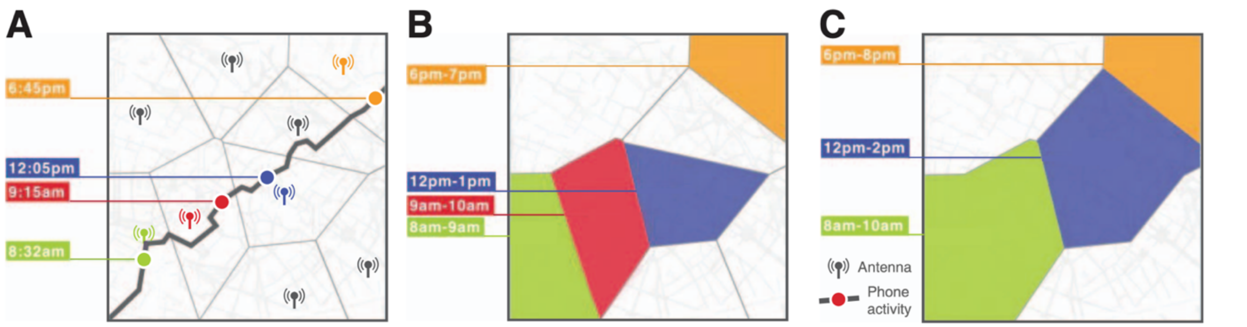

Figure 1 (A) Trace of an anonymized mobile phone user during a day. (B) The same user's trace as recorded in a mobility database. (C) The same user's trace with lower temporal and spatial resolution. *Source: Montjoya et al. (2013)*

taxicab trips.[3] The dataset was obtained legally by a Freedom of Information Law request. The researcher was able to un-hash taxi license and medallion numbers. Using the no longer anonymous dataset as well as several easily obtained photographs, he successfully identified several celebrities' trips as well as their tipping habits (Figure 2).

Figure 2 Individual trip details in the NYC taxicab dataset for celebrities Bradly Cooper and Jessica Alba. *Source: Neustar Research (2014)*

[3] Neustar Research, "*Riding With The Stars: Passenger Privacy in the NYC Taxicab Dataset.*" September 15, 2014.

The researcher was also able to find home locations of patrons to an adult entertainment venue. By querying the dataset to find pick taxicab pick-ups that occurred outside the club between the hours of midnight and 6am, the researcher found corresponding drop-off coordinates (Figure 3). The precision of the data set was such that drop-offs could be mapped to specific addresses in New York City.

The Neustar Research study, while ultimately benign in outcome, shows the risk inherent in the availability of granular, individual trip data. Shared mobility providers are acutely aware of these risks and as such are often hesitant before sharing trip data. Members of both Spin and Uber cited this as a concern when entering sharing data agreements with

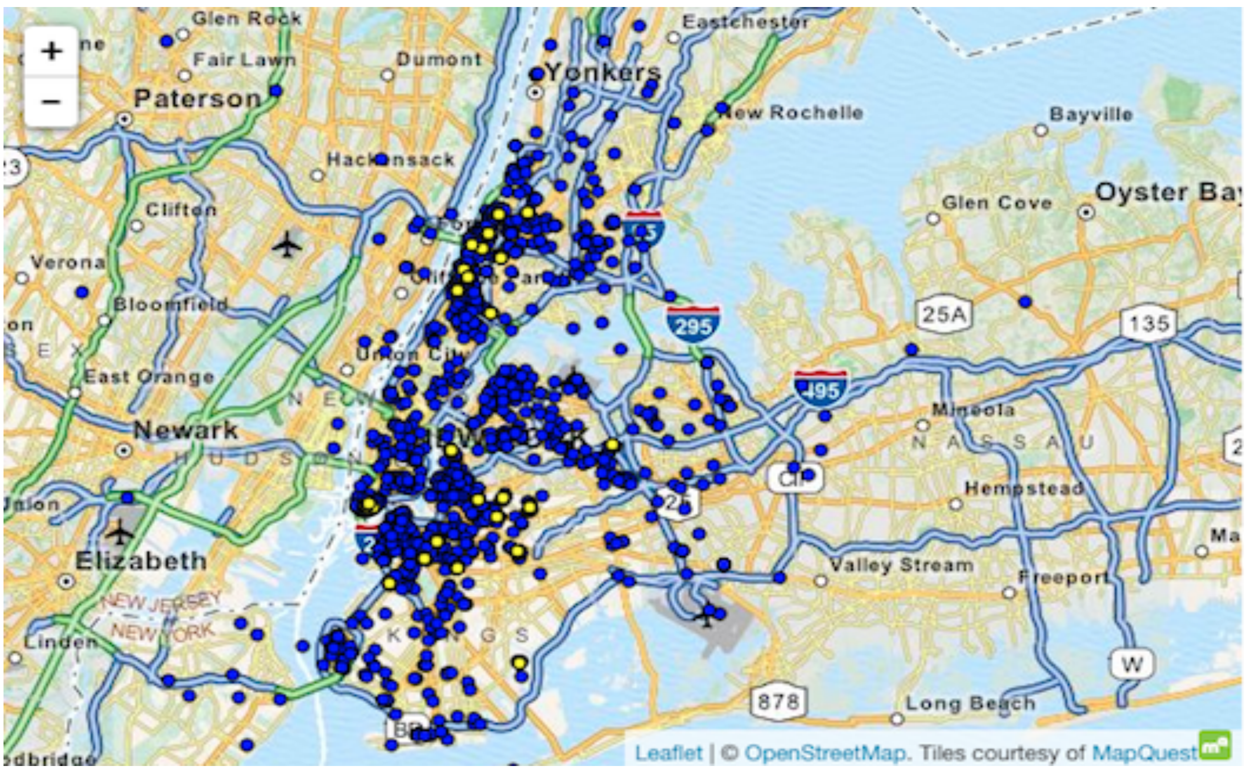

Figure 3 Addresses of adult entertainment patrons. Yellow dots represent points that are frequently clustered, implying a returning customer *Source: Neustar Research (2014)*

municipalities. Providers consider member privacy even when deciding to share data sets that do not contain individually hashed trip. For examples, a local Metro Vancouver Carshare provider considers risks inherent in sharing two “safe” data sets that, when combined, could reveal private information on members.

3.2.2 Business privacy risks

In addition to personal privacy risks, shared mobility providers worry about the business risks inherent in sharing data. Data sharing agreements could result in a reveal of proprietary information to both regulators and competitors. Information on the market the company serves, including its size, timing and location, could be revealed. Price points to consumers could also be revealed. In some cases, it could even be possible to recreate a company's algorithm for assigning rides. This would all serve to both devalue the data and undermine the potential profitability of the company. Companies that share previously private business information run the risk of increased regulatory burden, a decrease in profitability, and an increase in liability.

4. Data Standards: an evolution

Take-away points

1. GTFS is a structured list of real-time transit stop locations.
2. GBFS is targeted towards helping the customer make transportation decisions. It shows the real time or semi-real time locations of micromobility vehicles.
3. MDS directly builds on GBFS to help regulators understand local trends in micromobility. In addition to a required GBFS feed, it includes historical data and other personally identifiable information.

In 2005, the City of Portland and Google collaborated on data standard for public transit informationl, General Transit Feed Specification (GTFS). GTFS is a simple, structured list of real-time transit stops locations. After a successful pilot in Portland, GTFS was met with widespread adoption, today being used by more than 1,000 transit agencies worldwide.

Inspired by the success of GTFS and motivated by the rise in bikesharing, in 2015 the North American Bikeshare Association (NABSA) created General Bikeshare Feed Specification (GBFS). GBFS standardizes sharing real time or semi-real time micromobility data. There are currently hundreds of active GBFS feeds in over thirty countries.[4] In 2018, the City of Los Angeles created Mobility Data Specification (MDS), a standard built directly off of GBFS, as a way to help manage and regulate the influx of additional mobility services, including micromobility option such as e-scooters.

This section will go over the key differences between GBFS and MDS and contain recommendations for how to choose between these standards. It will also provide an overview of the benefits and challenges presented by MDS, including lessons learned from other cities that have or have not chosen to adopt this developing standard.

4.1. GBFS basics

GBFS describes the current state of a micromobility system at a single point in time. It does not cover any historical data, such as vehicle trips, and it does not contain other personally identifiable types of information. As a standard for micromobiity, GBFS does not describe the current state of other shared-use mobility services, such as ride-hailing and carsharing. Much like GTFS, GBFS data is publicly available, allowing a customer,

[4] A complete list of cities using GBFS can be found here: https://github.com/NABSA/gbfs/blob/master/systems.csv

regulator, or member of the public to access the data. Access is not contingent on an API key or any other form of authorization.

Every GBFS feed consists of between one and ten files (reproduced in full in the Appendix). Different files are required depending on the type of micromobility. For example, certain files required for station-based bikeshare do not apply to dockless services. Files are all delivered through an API.

GBFS was created to support station-based bikeshare. It is currently being updated to support dockless mobility. A field enabling distinction between micromobility modes is currently being added.

4.2. MDS basics

MDS builds on GBFS by also including historical information that regulators/transportation authorities can use for monitoring and data analysis purposes. MDS currently only supports micromobility (bike sharing and e-scooter sharing) but has the developmental potential to also support other modes of shared-use transportation data, such as ride hailing and car sharing. MDS is open source and can be freely implemented by any organization, assuming they have the requisite technological savvy. MDS consists of two components, a provider API and an agency API.

4.2.1 Provider API

The provider API is intended to be implemented by micromobility providers. It is meant to be queried by a municipality for data exchange containing a fleet of micromobility devices' operational information. This API also presents a historical view of all operations. The MDS provider API requires exposing a public GBFS feed as well for realtime micromobility data, enabling MDS operators to know the locations and statuses of all vehicles in the system.

In addition to the GBFS requirements (see section 4.1.) the MDS provider API requires two files, *trips.json* and *status_changes.json* (see Appendix for more details). The trips file details journeys taken by micromobility customers with a geo-tagged start and stop point. The trips file is set up to allow for ease of querying historical data. s*tatus_change.json* details the availability of all vehicles for customer use. Parameters include the availability battery life of all micromobility vehicles. Other parameters, such as time of use and vehicle type, are also included to allow for querying historical data.

4.2.2 Agency API

The agency API, which at the time of writing this report is not yet active, is to be employed by municipalities and other regulators in order to query the shared mobility providers data. The agency API is designed to help support micromobility operations by allowing the regulator to contain information on municipal matters that may affect operations. Much of the information in the agency API is similar to that in the provider API. However, providers are only allowed to query data related to vehicles in their own fleet. For example, an employee at Lime operating the MDS agency API would not be able to see data on Spin vehicles.

What is Mobility Data Specification?

MDS cheat sheet

Mobility Data Specification (MDS) is a standard that facilitates two-way data transmission between regulators and mobility as a service providers.

MDS currently only supports micromobility (bike sharing and e-scooter sharing) but has the developmental potential to also support other modes of shared-use transportation data.

MDS consists of two separate application programming interfaces (APIs):

Provider API

(in use)

- Implemented by mobility as a service providers
- Queried by regulators for analysis and planning
- Contains historical trip data

+ GBFS feed

Agency API

(in development)

- Implemented by regulators
- Queried by mobility as a service providers for operational support
- Contains historical trip data and information on public infrastructure

4.3. Benefits of MDS

4.3.1 Standardized two-way information sharing

Through its two APIs, MDS enables information sharing between shared mobility providers and regulators. The provider API allows regulators to pull data from the provider. This data can be used to inform customers of realtime vehicle locations through the required GBFS feed, and to help regulators manage, permit, and analyze micromobility services through the historical trip data. The providers benefit by pulling information from the agency API. This API will allow providers to directly and seamlessly communicate with regulators. Updates on road closures and parking restrictions, for example, will be communicated directly to the providers, who can move shared-use mobility vehicles and infrastructure (scooters, docking stations, dedicated parking spaces, etc.) as well as disseminate this information to their customers.

4.3.2 Developmental flexibility

While MDS and GBFS have been products of the City of Los Angeles' Department of Transportation and the NABSA, respectively, both standards have benefited from an iterative, collaborative software development infrastructure called GitHub, which allows a community of code contributors to collaborate on open source code projects. The GBFS Github repository lists 28 unique contributors, while the MDS Github repository lists 37 contributors. MDS contributors are representatives of both cities and vendors, including

MDS Benefits	MDS Challenges
1. *Two-way data sharing* – providers and regulators can both benefit from standardized data sharing.	1. *Barrier to entry* – storing and analyzing MDS data requires in-house knowledge and/or resources.
2. *Developmental flexibility* – the ongoing, collaborative software development process will enable MDS to keep up with future transportation innovations.	2. *Developmental disagreements* – contributing stakeholders from different sectors may disagree on standard features and methodologies for data analysis.
3. *Third-party compatibility* – there is a growing list of third-party companies that can provide MDS data storage and analysis support to agencies.	3. *Privacy risks* – MDS data presents significant privacy risks. Improper data usage could endanger public safety.

the Cities of San Francisco, Seattle, Austin, Santa Monica as well as shared-use industry providers Spin and Lime. This collaborative nature encourages skilled employees from public and private organizations to raise issues and make collaborative changes towards the global improvement of both standards. Additional cities that decide to adopt either of these standards, in particular MDS, are encouraged to contribute to the Github repositories.

To contribute to either GBFS or MDS, an organization needs fork the repository and submit a pull request on Github. A repository administrator will have the ability approve the changes. On the Github repositories it is easy to see who the most frequent and important collaborators are.[5]

MDS currently only supports data gathering on shared-use bicycles and scooters, but the long-term vision extends beyond that. The contributors of MDS are laying the foundation for the data standard to have capabilities to record historical trips by taxis and rideshare, although there are still significant hurdles in convincing these industries to upload their data to an MDS provider API (see Section 3). The collaborative, flexible MDS development environment will also enable data from future forms of transportation, whether that be autonomous cars, drones, or transportation modes we haven't conceived of yet, to be added to the standard.

4.3.3 Growing compatibility with third party sources

Access, management, and analysis of MDS data can be done entirely in-house by a regulator. However, not all agencies have the technical resources and know-how to access and process the massive data sets (see section 4.4.1.).

Competition between several third-party transportation analysis providers, including Remix, Ride Report, Populus, and SharedStreets, is sparking further innovation in this space. These companies are developing MDS-compatible plugins. They are offering a growing mix of data management and data analysis tools ranging from general solutions to plugins tailored to an agency's needs. For instance, in Los Angeles, Lime and Spin partnered with Remix, sharing their data with the Remix platform which is then shared with the City of Los Angeles. Remix was hired to provide an intermediary between the City and the businesses, as well as to provide dashboard data analytics and visualization. Figure 4 shows a screenshot of the Remix dashboard in Santa Monica.

[5] The most influential MDS contributors can be found here, while the most influential GBFS contributions can be found here.

Third party analyzers will provide a growing list of analyses, and also can tailor analyses to suit the agency's needs. Figure 5 shows three distinct data visualizations using the same MDS trips data done by SharedStreets. Each of these three SharedStreets analyses, for example, may be useful to support different planning and regulatory inquiries and decision-making.

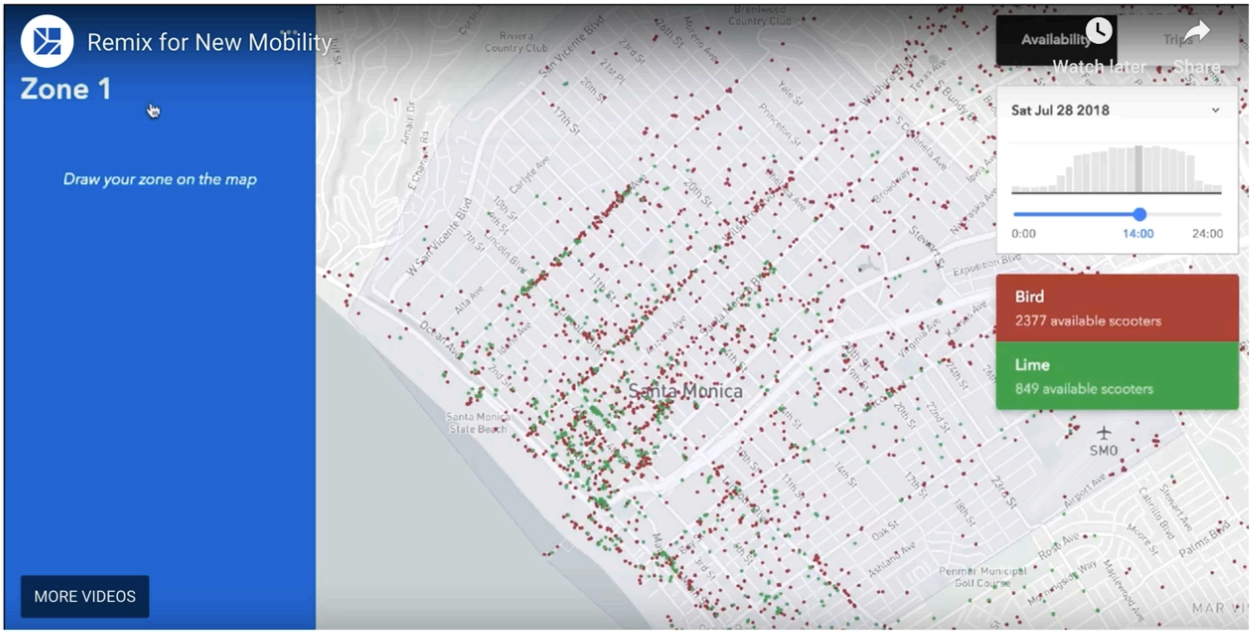

Figure 4 Screenshot of Remix dashboard showcasing MDS data in Santa Monica.

Figure 5 The same MDS data produces three distinct visualization analyses, all of which may be useful for different planning purposes. *Source: SharedStreets (2019)*

4.4 Challenges adopting MDS

4.4.1 Technical barrier to entry

Storing and analyzing MDS data requires resources on the part of a regulator through either in-house data capabilities or a third-party company. MDS requires systems that ingest, clean, store, and process the data via APIs. Without a dedicated team for this effort, or the funds to contract an outside company, it is difficult to take full advantage of the vast realtime datasets and insights that and MDS feed affords regulators.

Case study: Washington DC opts out of MDS

After determining that the costs of adding internal MDS functionality would outweigh the benefits of the platform, The Washington D.C. District Department of Transportation (DDOT) determined that, for the time being, will not mandate shared-use service providers to adhere to an MDS API through its dockless micromobility permitting process.

Instead, the DDOT requires micromobility providers to share regular static reports, in addition to a standard GBFS feed. These reports are retrospective rather than current but remain useful to for analysis and regulatory decision making. DDOT also reserves the right to make an MDS provider API mandatory in the future, if barriers to entry decrease (whether by continuing improvements to the MDS ecosystem or if DDOT gains additional technical resources and capabilities).

Case study: Portland's previous technological investments pay dividends

As an original collaborator on GBFS, the Portland Bureau of Transportation (PBOT) has historically had a strong institutional technological aptitude for shared-use data processing. With strong resources and knowhow at their disposal, PBOT decided to pilot MDS from July through November, 2018, PBOT built automated tools that enabled them to better understand their streets and enforce various micromobility regulations. For example, . PBOT was able to build useful visualizations and gain actionable insights, such as learning which areas should be priorities for bike lanes, through MDS data.

However, there were many technical challenges that arose from this investment. At first, providers did not understand the difference between MDS and GBFS and provided GBFS feeds as MDS APIs. Technical challenges related to resolving this disconnect required much of the PBOT data team's labor. Additionally, storing and analyzing the data was a large investment, both in labor and in resources..

4.4.2 Developmental disagreements

The wide variety of MDS contributors is consider a strength of MDS as many people have contributed to reviewing and ensuring the technical rigor and sophistication of the standard. However, different contributors may have different agendas for what they would like to see included and prioritized in the standard. This is especially relevant given some contributors are private organizations that might not always be keen to openly share all of their data.

While all parties may agree in the future on what features to include in the standard, harnessing the full power of MDS data will likely continue to present challenges. For example, differences in analytical methodology will need to be resolved and agreed to between parties in order to ensure that insights are accurate and appropriate, reducing conflicting interpretations of data. Currently there is no standard methodologies for calculating or validating even the most basic metrics, such as the number of devices on the streets and extent of vehicle utilization. Complications are further amplified through, for example, defining various municipal boundary coordinates, in turn leading to different operators, providers, and third-party analysis organizations attaining different analytical results for these and other parameters.

As a solution to this problem, third-party organizations such as Ride Report and SharedStreets, are working on an open source effort to create transparent methods for best-practice analytics methodologies, allowing both providers and regulators to reach the same conclusions with MDS data. This project is in its infancy and should be carefully followed and assessed as it continues to develop.

4.4.3 Privacy risks

MDS data, while valuable for regulatory decision making, could expose personal privacy risks (see section 3.2.1.). A letter written by the non-profit Center for Democracy and Technology outlines why MDS raises greater privacy and security concerns than GBFS.[6] MDS contains granular historical and real-time data location data, making it possible to compromise users location information. In order to protect this data from potentially malevolent parties, it is important to adopt clear and robust privacy and security policies around MDS data collection and use. This includes limiting access to use of MDS data, creating policy for data retention and deletion and building systems for secure data storage to protect against leaks.

[6] https://cdt.org/insight/comments-to-ladot-on-privacy-security-concerns-for-data-sharing-for-dockless-mobility/

5. Data Aggregators and Repositories

As covered in section 3, transportation providers are often uncomfortable or unwilling to share data with regulators, due to both business and personal privacy concerns. Willingness to share varies across different types of service providers. A standard like MDS is only useful if companies are willing to upload their trip data to the provider API. Some dockless micromobility providers have shown a willingness to openly share with regulators, and have been complying with early MDS pilots in cities like Los Angeles and Portland. TNC providers, on the other hand, have not been willing to share granular data, in any format, with regulators (see section 3).

Take-away points

1. As currently structured, MDS may never be a viable option for sharing TNC data.
2. A trusted data repository, managed by a neutral third-party host, may be a solution.
3. SharedStreets, a data aggregator from NACTO and the World Bank, is currently the most viable third-party. However, they need to invest significantly in order to scale

5.1 Third-Party hosts

A proposed solution to incentivize industry sharing is to have a third-party serve as a liaison between providers and the regulators. One host can, in theory, serve as a standard, trusted data repository across many municipalities. The host needs a strong, technical transportation data expertise to enable sharing and analysis. The host also needs a deep understanding of the legal system and the ability to coordinate building policies and protocols for data ownership, access and use in the interest of regulators, providers, and the general public. The host also needs to have the ability and willingness to build privacy-preserving algorithms to protect the public from the disclosure of unique traces (see section 3.2).

5.1.1 Private hosts

There are several different third-parties startups vying for market share in this space. Remix, Populus, and Ride Report are the most visible of these companies today, all having secured venture seed money. All three provide similar services and operate with a similar business model. Municipal governments and regulators contract their services to help manage transportation data. Each service has a dashboard that can showcase and analyze shared mobility data. All three organizations have growing experience working and facilitating use of MDS APIs (see section 4.3.3).

Some micromobility providers have been amenable to sharing their data with these private hosts. In Los Angeles both Lime and Spin partnered with Remix, allowing the company to host and control trip data as part of their government contract with the City. However, ridehailing companies like Uber and Lyft are less likely to share their data with another private organization.

5.1.2 Neutral hosts

Ride hailing providers worry about business risks of giving another company access to their data. A more attractive solution is a trusted, neutral host, and non-profit host. In the United States, Universities appear to be the best suited hosts, as exemplified by the University of Washington's Transportation Data Collaborative. Nonprofits, such as NACTO's SharedStreets, also offer promise as neutral hosts. Both of these organizations have had some success partnering with shared transportation companies. The Transportation Data Collaborative has served as a repository for scooter and bike sharing data in the Seattle metropolitan area, while SharedStreets has announced a partnership with Uber to run pilot projects. Of these two projects, SharedStreets, which is also backed by the World Bank's Open Traffic Project, is the more viable model to work at scale. It should be noted that the University of Washington has yet to receive and host ridehailing data.

5.2. Opportunities with SharedStreets

SharedStreets has recently received a lot of publicity, being featured in numerous publications and transportation blogs. This is in part due to its burgeoning partnerships with industry and municipal governments. Uber donated $250,000 to SharedStreets, and SharedStreets has active pilots with Seattle, Detroit, and Washington D.C. SharedStreets is marketing itself as a scalable panacea to problems involving transportation data sharing. It offers three primary services, *a referencing system, a data aggregator*, and a *data visualization and analysis service*.

What is SharedStreets?

1. *A referencing system* – new technology to identify street space and connect datasets.
2. *A data aggregator* – vehicle GPS points are tracked and aggregated, removing individual trip information.
3. *A data visualization and analysis service*

SharedStreets has growing functionality with GBFS and MDS. Cities can download plugins to pair SharedStreets with these APIs. SharedStreets can also be used with static data sets, although that would diminish some of the real-time benefits of the service. Figure 6 details a pipeline from data input to visualization and analysis using SharedStreets.

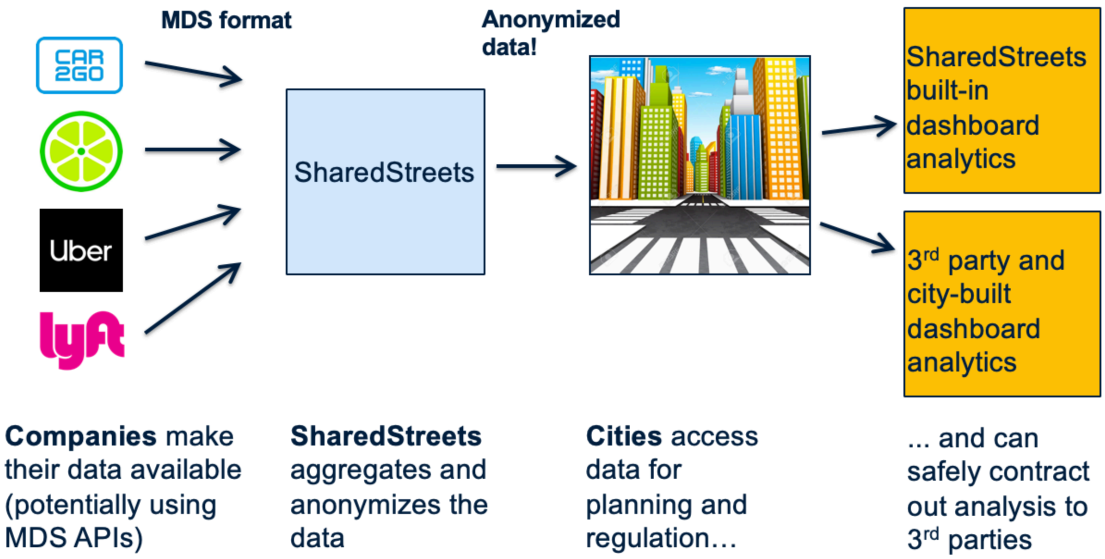

Figure 6 Potential SharedStreets pipeline. Data is collected, aggregated, and made available for analysis and visualization.

5.2.1 A referencing system

The initial innovation behind SharedStreets is its powerful referencing system. The referencing system redefines the street, linking previously unreconcilable data sets and allowing for easier communication and collaboration between different organizations (Figure 7). SharedStreets Referencing System provides universal IDs for different facets of the street and is meant be used as a complement to OpenStreetMap (OSM), allowing users to use OSM data without worrying about how the data is encoded.

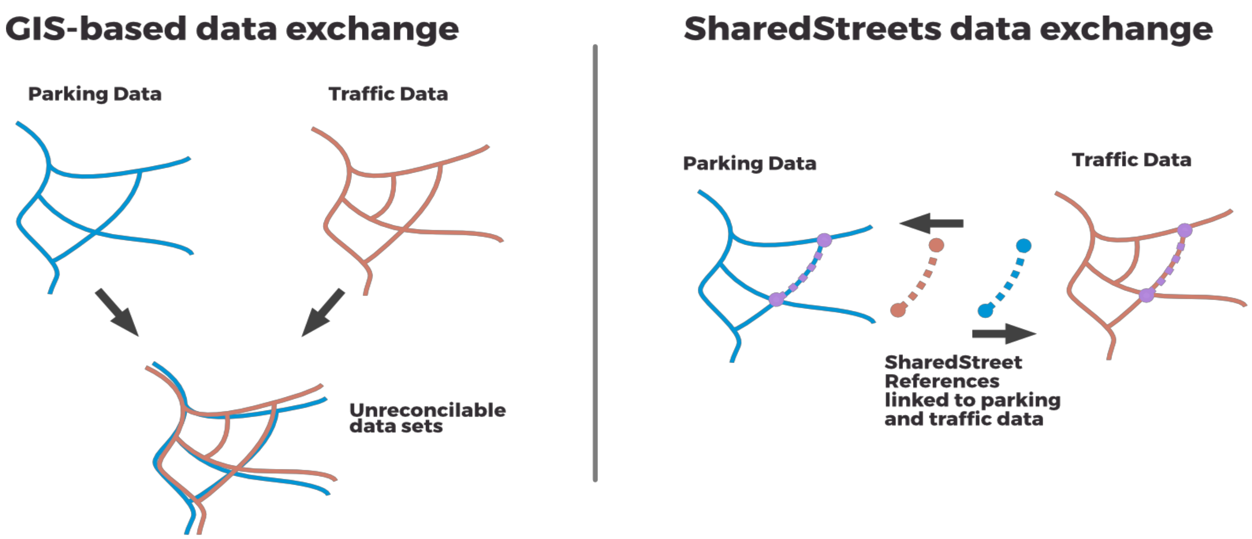

Figure 7 Streets become segments, each with their unique references, IDs, and attributes. This allows separate data sets to be easily reconciled. *Source: SharedStreets (2019)*

SharedStreets Referencing System is built by four layers of data. The first two levels are the street intersections and street references (Figure 8). Intersections and references make up, respectively, nodes and edges in a street system. The intersections describe road junctions, while the references describe street segments. Combined, these two layers can describe a simplified grid of any road network. On top of these two layers are

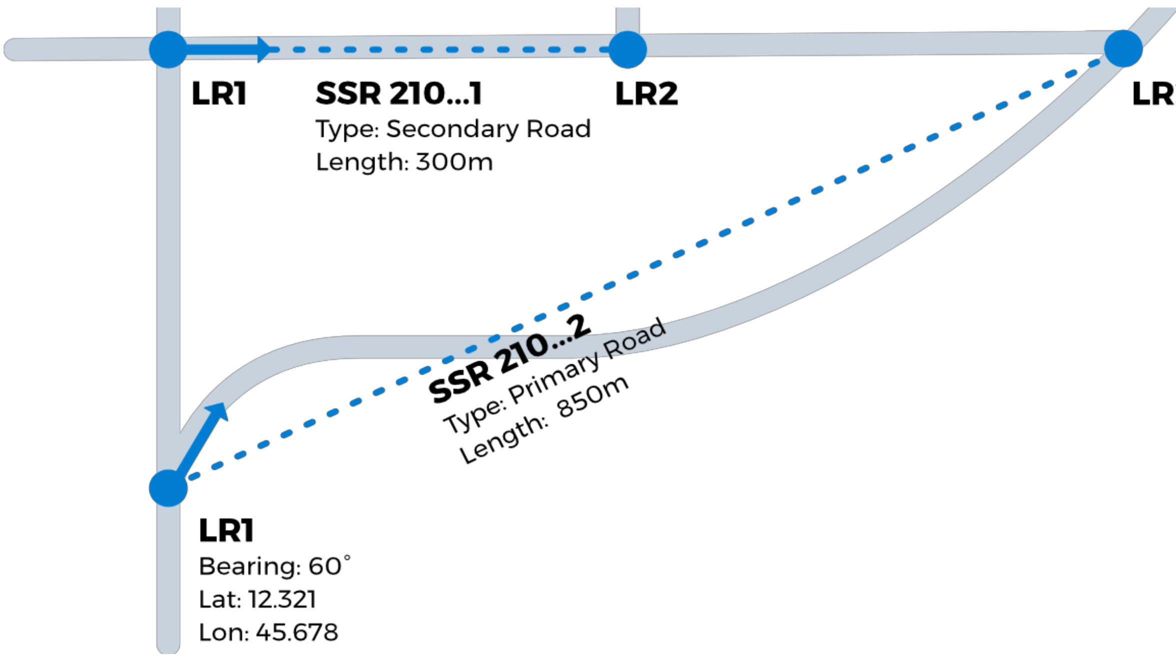

Figure 8 Schematic showing intersections and references. *Source: SharedStreets (2019)*

SharedStreets geometries, which describe three-dimensional shapes and eccentricities streets embody. Finally, the last layer is OSM metadata, which provides a framework to track changes in the basemap data that defines the other three levels.

5.2.2 A data aggregator

There are numerous personal privacy and business issues related to sharing granular, individual trip data (see Section 3). SharedStreets addresses this problem with a data aggregation strategy. Before aggregating data, a proprietary algorithm matches individual GPS origin, travel, and destination points generated by vehicles to a location on a roadway network (Figure 9). By finding likely routes, individua GPS points, that are often ambiguous or imprecise, are snapped to most probable road segments. In addition to location, the algorithm also assigns other parameters, such as speed and direction of travel. The data inputs for this algorithm are flexible, and could be provided by GBFS, MDS, or static data sets.

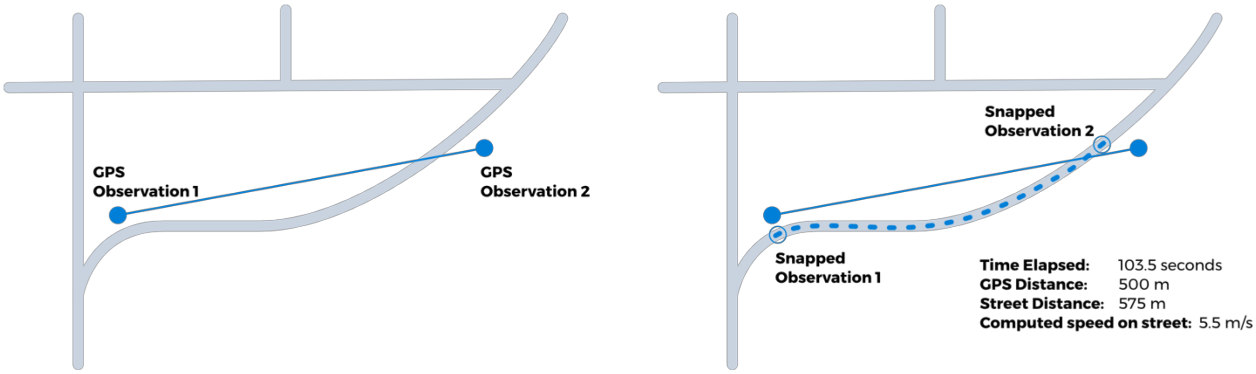

Figure 9 SharedStreets matching algorithm. Individual GPS points are tracked and a model find the most likely placement, route, and direction of travel for each point. *Source: SharedStreets (2019)*

SharedStreets does not store the individual outputs of its matching algorithm. Instead, a variety of temporal and geospatial aggregation strategies are used to ensure personal privacy and protect business interests. These strategies are based on the SharedStreets referencing system (Section 5.2.1.). Trip data are aggregated using a binned linear street reference. Streets are by default divided into 10 meter bins, although this parameter can easily be altered depending on analysis. Data is then further aggregated hourly and daily, into 168 (24 hours * 7 days) bins. After aggregation is complete, each 10 meter section of bin contains information on total number of trips, type of trips, and mean trip speed for each hour of the week.

There are various validation strategies SharedStreets uses to ensure that its aggregated data are accurate and high quality. The first level of validation is ensuring that there are no failures when GPS points are matched to the basemap. Errors at this stage could imply an improperly coded basemap. To check for this, SharedStreets aggregates matching failure rates by street edge. Identifying areas of highest match failure allows SharedStreets operators to update and improve basemaps. SharedStreets also has built-in features to validate trip speeds. GPS data is compared to vehicle odometer data, to ensure that both vehicle kilometers travelled, and speed data is accurate.

5.2.3 A data visualization and analysis service

In addition to their sophisticated referencing and trip matching technology, SharedStreets is building data analysis dashboards to compete with companies like Remix and Populus. While not yet operational, SharedStreets is running several pilots where they hope to showcase their platform's visualization and analysis potential. Results of these pilots have yet to be published at the time of this report's writing and it is difficult to know how viable SharedStreets visualization services will be in the long term.

Figure 10 A dashboard used for Taxi/TNC analysis. This analysis and visualization is set to be piloted in Washington DC. *Source: SharedStreets (2019)*

5.3 Challenges scaling SharedStreets

In order to unlock efficiencies inherent to their technical model, SharedStreets needs to scale across North America. If SharedStreets are able to standardize data aggregation the service could make it easy and worthwhile for collaboration and data sharing between transportation companies and regulators. This is especially true given SharedStreets growing compatibility with MDS. Combining and standardizing these two services could shift transportation companies willingness to share their data with regulators.

SharedStreets' core competency is their industry-leading capacity to ingest, store, process, and report/export transportation data. However, there are still several hurdles for SharedStreets to overcome before being considered a North American industry-standard data repository. While the SharedStreets team has multiple funding sources, including NACTO, the World Bank, and Uber, they currently have a very small team of four full-time employees. SharedStreets must expand its legal and business-facing capabilities if it hopes to scale and harness its technical potential.

5.3.1 A complex legal environment

According to Mark Hallenbeck, the Director of the University of Washington's Transportation Data Collaborative, the single biggest issue SharedStreets has to contend with is maneuvering a complex legal environment. SharedStreets is preparing to be a data aggregator and repository for major metropolitan areas in the United States and Canada. However, privacy laws at the state/provincial and municipal level differ greatly throughout North America. Due to this, SharedStreets rules and procedures will invariably not satisfy local laws and provisions in select municipalities.

A potential solution to this problem could be achieved through SharedStreets defining its procedures for different municipalities to adhere to local statutes. SharedStreets could also work with local law and policy makers to adapt current laws to the its platform's infrastructure. Unfortunately, the SharedStreets team does not currently have a dedicated legal team needed to navigate these complicated issues.

5.3.2 Unproven business model

In order for SharedStreets to scale, it needs to secure further investment and eventually generate revenue. SharedStreets is a new and mostly untested platform. Results from their pilots with Washington D.C. and Seattle remain unpublished and unproven. Unfortunately, it's too early to judge the viability of the business model.

SharedStreets Benefits

1. *Trusted repository* – due to its status as a non-profit, SharedStreets is viewed as a relatively safe organization for providers to share with (from a business intelligence perspective).
2. *Innovative technology* – with their referencing and matching systems, SharedStreets has developed a sophisticated framework that has the ability to accommodate multiple types of shared mobility data.

SharedStreets Challenges

1. *Complex legal environment* – unclear how SharedStreets will maneuver complicated municipal legal systems.
2. *Unproven business model* – unproven how SharedStreets will secure funding at scale.
3. *Focus on visualization tools* – SharedStreet's focus on data visualization/analysis tools may distract from developing and scaling its other offerings.

6. Paths forward

The shared mobility space is new, complex and quickly evolving. The past two years have brought new innovations to North America including a surge of shared micromobility options. While it is impossible to precisely map the future shared mobility landscape, it is clear that new innovations and technologies will continue to emerge.

Regulators need to stay nimble to keep pace with transportation innovations. Data standards and data aggregators could be a part of the solution. Standards like MDS and aggregators like SharedStreets are flexible technical tools that enable rigorous shared mobility data collection and analysis, allowing regulators to map mobility trends for planning and regulatory purposes. Combined with thoughtful policy, these tools could herald a standardized digital infrastructure that makes data sharing between provider and regulator easy, safe, and mutually beneficial. The complementary functionality of MDS and SharedStreets increases the potential viability of both tools.

Unfortunately, MDS and SharedStreets are currently in their infancy and remain largely untested. Unresolved issues threaten the standardized adoption of both tools. Requiring MDS puts a large resource burden on the regulator, with a high initial investment and potentially associated fixed costs. Not all regulatory agencies have the resources to make investing in MDS worthwhile. MDS also faces unresolved existential issues, the most notable being privacy. The MDS strategy of collecting and storing individual trip data could threaten both personal and business privacy.

Issues surrounding SharedStreets revolve primarily around its unproven model. It is unclear how SharedStreets, which is currently less tested than MDS, can operate at scale. To become the leading transportation data aggregator and live up to its technical prowess, SharedStreets needs to navigate complicated legal and business terrain. It is simply too early to know how the SharedStreets team will respond to these challenges.

Given these unresolved issues, there is no one best path forward for TransLink to holistically standardize its data sharing practices and recommend a standard for shared mobility operators in the Metro Vancouver Region to abide by. Uncertainty regarding future innovations in the data sharing space further complicates the matter. However, these two high-level steps should be taken:

1. Continue assessing viability of different data standards for the Metro Vancouver region, including liaisons with local shared use mobility providers (see page 28 for more details).
2. Closely follow SharedStreets over the next months and years, paying attention to their growth and evolving pilot programs (see page 29 for more details).

Path Forward: Data Standards

GBFS is the present. An MDS-like standard is the future.
How should TransLink navigate this evolving landscape?

1. Require one GBFS feed for all of Metro Vancouver

As new municipalities adopt micromobility services, require one aggregate GBFS feed for all of Metro Vancouver rather than individual municipal feeds. This will also set the precedent for future data sharing activities. While not technically challenging, this mandate will require municipal cooperation. However, this feed does not support ridehailing

2. Gauge technical feasibility/cost of MDS adoption

Review TransLink's in-house Information Technology and Analytics capabilities. Determine if there are sufficient resources/expertise and cooperation with area municipalities to create an aggregate Metro Vancouver MDS feed. Explore the options of using third-party transportation data companies (such as Remix, Populus, or Ride Report) to determine cost of using their platforms for any analysis needed.

Based on the outcome of these findings, there are three primary options:

a. Adopt MDS in-house (Portland model)

- Create internal systems for MDS data storage and analysis.
- Contribute to the MDS Github to ensure TL's requirements and interests, such as standards for ridehailing, are met
- *Ambitious project*; likely will require labor of at least one skilled employee full-time

b. Adopt MDS with third-party (LA model)

- Use third-party's systems for MDS data storage and analysis.
- Contribute to the MDS Github to ensure TL's requirements are met
- *Large project*; likely will require at least 50% of one skilled employee's time

c. Wait to adopt MDS (DC model)

- Wait for technology to improve and ease of adoption to increase.
- Require GBFS feed and regular static reports
- Determine if a third-party's dashboard/tools is viable for static report analysis

Path Forward: Data Aggregators

SharedStreets is a potential panacea but is largely unproven.
What is the best way to vet this developing organization?

1. Develop and maintain connections with knowledgeable leaders

To get an understanding of SharedStreets' continued growth and development, it is necessary to communicate with professionals directly involved withShared Streets. Expert testimony is essential to navigate this complex pathway in order to shed light on the state of pilot programs and updates to SharedSteets internal infrastructure (for example, if they are hiring a legal team). For a list contacts, please consult with the TransLink New Mobility Lab and this report's author.

2. Pay attention to upcoming webinars

Organizations like NACTO and NABSA have frequent webinars showcasing the latest technological advances. Members of SharedStreets and partners often present. Following these webinars is an easy way to stay updated.

3. Follow other third-party hosts

Part of SharedStreets' viability is the attractiveness of its non-profit model to shared mobility providers. However, other profit-based hosts, such as Remix, Populus and Ride Report, offer a growing array of products that could compete with SharedStreets. While these companies currently provide primarily data analysis and visualization, they are expanding their services and may soon offer data aggregation capabilities.