|  |
| --- |
|  |
| Passive Data Collection Guidebook |
| National Center for Applied Transit Technology |

|  |
| --- |
| AECOM |

Table of Contents

[Executive Summary 1](#_Toc93652511)

[Introduction 3](#_Toc93652512)

[Overview of Passive Data Collection Technologies 4](#_Toc93652513)

[Rider Data 4](#_Toc93652514)

[Passenger Counting 4](#_Toc93652515)

[Market Preferences 5](#_Toc93652516)

[Revenue Tracking 5](#_Toc93652517)

[Potential for Passenger Origin-Destination Data 5](#_Toc93652518)

[System Data 5](#_Toc93652519)

[Bus Route/System Performance 6](#_Toc93652520)

[Maintenance/Monitoring Data 6](#_Toc93652521)

[Driver Safety Tracking 6](#_Toc93652522)

[Vehicle Health Monitoring 6](#_Toc93652523)

[Automated Passenger Counters (APCs) and Automatic Vehicle Location (AVL) Systems 7](#_Toc93652524)

[Door Sensor-Based Passenger Counters 8](#_Toc93652525)

[Video-Based Passenger Counters 8](#_Toc93652526)

[Mobile “Smartphone” Enabled Systems 8](#_Toc93652527)

[Mobile Fare Payment 8](#_Toc93652528)

[Bus Tracking/Crowding Applications 8](#_Toc93652529)

[Aggregated Smartphone Movements Data 9](#_Toc93652530)

[Smart Card-Based Travel Data 9](#_Toc93652531)

[Agency Website Usage 9](#_Toc93652532)

[Driver Safety Tracking Systems 9](#_Toc93652533)

[Vehicle Health Monitoring/Performance Tracking Systems 9](#_Toc93652534)

[Passive Data Collection Vendor Types 11](#_Toc93652535)

[”Big Data” Vendors (Example: StreetLight Data) 11](#_Toc93652536)

[Trip Planning/Discovery Application Vendors (Example: Transit App) 13](#_Toc93652537)

[On-Site Collection Tools (Example: Avail Technologies) 14](#_Toc93652538)

[Conclusions 15](#_Toc93652539)

[Case Studies 17](#_Toc93652540)

[Introduction 17](#_Toc93652541)

[Current Practices/Applications of Technology 19](#_Toc93652542)

[Staff Capacity Requirements and Needs 20](#_Toc93652543)

[Benefits and Costs of Technologies 20](#_Toc93652544)

[Data Sharing 21](#_Toc93652545)

[Challenges 22](#_Toc93652546)

[Conclusion: Goals and Next Steps for Passive Data Collection Technology 23](#_Toc93652547)

[Best Practices and Recommendations 24](#_Toc93652548)

[Glossary 31](#_Toc93652549)

[APPENDIX A – Vendor Interview Notes 32](#_Toc93652550)

[StreetLight Data 32](#_Toc93652551)

[General Questions 32](#_Toc93652552)

[Technical Questions 33](#_Toc93652553)

[Maintenance Questions 33](#_Toc93652554)

[Client Questions 34](#_Toc93652555)

[Cost Questions 34](#_Toc93652556)

[Data Sharing 35](#_Toc93652557)

[Additional 35](#_Toc93652558)

[Transit App 36](#_Toc93652559)

[General Questions 36](#_Toc93652560)

[Technical Questions 37](#_Toc93652561)

[Client Questions 38](#_Toc93652562)

[Cost Questions 38](#_Toc93652563)

[Data Sharing 38](#_Toc93652564)

[Additional 39](#_Toc93652565)

[Avail Technologies 40](#_Toc93652566)

[General Questions 40](#_Toc93652567)

[Technical Questions 41](#_Toc93652568)

[Maintenance Questions 41](#_Toc93652569)

[Client Questions 41](#_Toc93652570)

[Cost Questions 42](#_Toc93652571)

[Data Sharing 42](#_Toc93652572)

[Additional 42](#_Toc93652573)

[APPENDIX B – Case Study Interview Notes 43](#_Toc93652574)

[CORTRAN 43](#_Toc93652575)

[General Questions 43](#_Toc93652576)

[Technology Questions 44](#_Toc93652577)

[Capacity Questions 45](#_Toc93652578)

[Cost Questions 45](#_Toc93652579)

[Data Sharing 46](#_Toc93652580)

[Additional 46](#_Toc93652581)

[Lane Transit District 47](#_Toc93652582)

[General Questions 47](#_Toc93652583)

[Technology Questions 48](#_Toc93652584)

[Capacity Questions 48](#_Toc93652585)

[Cost Questions 49](#_Toc93652586)

[Data Sharing 49](#_Toc93652587)

[Additional 49](#_Toc93652588)

[LANta 51](#_Toc93652589)

[General Questions 51](#_Toc93652590)

[Technology Questions 52](#_Toc93652591)

[Capacity Questions 53](#_Toc93652592)

[Cost Questions 54](#_Toc93652593)

[Data Sharing 54](#_Toc93652594)

[Seneca Transit System 56](#_Toc93652595)

[General Questions 56](#_Toc93652596)

[Technology Questions 57](#_Toc93652597)

[Capacity Questions 57](#_Toc93652598)

[Cost Questions 58](#_Toc93652599)

[Data Sharing 58](#_Toc93652600)

[Additional 58](#_Toc93652601)

[Victor Valley Transit Authority 59](#_Toc93652602)

[General Questions 59](#_Toc93652603)

[Technology Questions 60](#_Toc93652604)

[Capacity Questions 60](#_Toc93652605)

[Cost Questions 61](#_Toc93652606)

[Data Sharing 61](#_Toc93652607)

# Executive Summary

This *Passive Data Collection Guidebook* serves to provide an overview of the various passive data collection technologies available that can be utilized by transit agencies throughout the nation.

While new technologies can ease the burden of data collection, they can increase the complexity of maintaining and operating an agency’s assets. New technologies may interact with legacy systems in unexpected ways and managing the use of an agency’s resources to operate them – and derive the full benefit from them – can be challenging.

In this era of “big data”, more information is being collected automatically than ever before. These technologies hold tremendous potential for providers of public transportation to increase the efficiency and reliability of data previously collected “actively” (or manually), while also expanding the kinds of information that they can gather. However, these opportunities should be weighed against the ongoing needs of fully utilizing these technologies, especially for smaller agencies.

Several technologies are reviewed in this *Passive Data Collection Guidebook*. Some of the “key takeaways” and recommendations for each are summarized in **Table 1**:

**Table 1 – Recommendations and Key Takeaways**

|  |  |
| --- | --- |
| **Technology** | **Key Recommendation / “Takeaway”** |
| Automatic Vehicle LocationIcon  Description automatically generated | Best when used on the entire fleet.Operations/on-road supervisory staff should be trained to “trust the data” and respond when conditions or issues arise that require interventions to get the best value from an AVL system and maintain on-time performance.  |
| Driver Safety TrackingA picture containing text  Description automatically generated | Address any potential implications regarding the collective bargaining agreement prior to purchasing driver safety tracking systems.  |
| Automated Passenger Counters | The quantity and type of data generated by automated passenger counters require robust data analysis capabilities and an active planning paradigm in the transit agency in order to get the most value from such systems.  |
| Vehicle Health MonitoringText  Description automatically generated | Analyze what new data a vehicle health system will provide versus what data it will provide that is already available from drivers and maintenance crews to help determine if this type of system is worth the cost.If an agency can only afford this type of system for some buses (but not all) in the system, prioritize electric vehicles, given their relatively new technological needs. |
| Aggregated Mobile Phone Data Icon  Description automatically generated | Transit agencies should consider “piggybacking” or “partnering” with other organizations (such as a Metropolitan Planning Organization) when purchasing this type of data technology, as other partners are just as likely to be using it.This type of data can be helpful when conducting a system-wide study or redrawing routes for a system. Similar to automated passenger counters, it should be kept in mind that this technology generates a large quantity of data that will need to be analyzed and will require robust data analysis capabilities and an active planning paradigm in the transit agency. |
| Real Time Data Applications | Although these are typically a benefit of having automated passenger counters and automatic vehicle location systems, it is important that an agency keep its input data – particularly GTFS data – as up to date as possible, as they form the foundation of these systems. |
| Mobile Fare PaymentA picture containing text, clipart  Description automatically generated | Agencies should first address the equity and Title VI issues associated with mobile fare payment before adopting any particular system or vendor – with the proper preparations, these systems can become an additional option that do not place any undue burden on the “unbanked” or “underbanked”.  |

# Introduction

The Community Transportation Association of America (CTAA), through its National Center for Applied Transit Technology (N-CATT), has prepared this *Passive Data Collection Guidebook* as a reference guide outlining the technologies that can allow transit agencies and operators to collect various types of data in a “passive” manner. Passive data collection efforts are those that do not require the active collection of data either by surveyors, bus operators or other agency/operator personnel. They typically utilize electronic and/or digitized devices to collect and store the data, while the examination, compilation and analysis of the data requires the use of skilled personnel.

The *Passive Data Collection Guidebook* provides an overview of some of the uses of passive data collection technology, and then presents interviews with some vendors of passive data collection systems. These are then followed by case studies of some transit agencies throughout the nation, which illustrate some “real world” uses of the passive data collection technologies, examples of how they were adopted, and limitations on how the technology was deployed.

The intent of the *Passive Data Collection Guidebook* is to disseminate information to transit agencies and operators throughout the nation (particularly to small and medium size systems – meaning those systems with up to about 50 vehicles in maximum service in any given transit mode), and is specifically aimed at assisting transit managers (and Operations Managers, particularly) with understanding the use cases of different passive data collection systems, as this field is evolving quickly and presents a “moving target.”

Finally, the *Passive Data Collection Guidebook* presents a series of findings in terms of the relative utility and benefit of the various data collection technologies that are available.

**Overview of Passive Data Collection**

* Generally falls into categories of *RIDER DATA, SYSTEM DATA* and *MAINTENANCE/MONITORING DATA*
* Allows for the passive collection of data without utilizing “active” resources such as random surveying
* Provides a large, comprehensive and continuously collected amount of data
* Requires some robust level of data analysis capability at an agency
* Can be used as a foundation for improved real time customer information systems

# Overview of Passive Data Collection Technologies

Transit operators have a multitude of reasons to collect certain types of data, ranging from the overall management and administration of the transit system, to planning needs, to regulatory reporting and funding requirements. The ability to collect some data using “passive” data collection systems – as opposed to those “active” methods that require the collection of data either by surveyors, bus operators or other agency/operator personnel – allows for a significant amount of data to be collected, should that be desired.

Some of the data needs that passive systems are utilized to gather are described here. They can generally be categorized into three overarching categories: *Rider Data*, *System Data*, and *Maintenance/Monitoring Data*. They include the following:

## Rider Data

These are elements related to riders, such as ridership. They may include:

### Passenger Counting



Passenger boardings need to be accurately counted so that transit agencies can determine the level of use of their services, both for planning reasons as well as for regulatory reporting and funding requirements. For example, the National Transit Database (NTD) requires that transit agencies record ridership.

**Passive data collection systems that accurately count riders are able to provide ridership reporting data for the purposes of the National Transit Database**

In addition, some passenger counting systems that record boardings and alightings can also track the level of ridership on a vehicle, thus allowing for the potential to track crowding on board a vehicle both for planning purposes as well as in real time, which has proven useful during the COVID-19 pandemic because it allows riders to gauge the level of crowding on a vehicle.

### Market Preferences

Related to ridership tracking, some data collection systems can allow the transit operator to determine the relative popularity of certain services and times of day when services are needed (thus allowing for inferences regarding trip purpose). This data can be useful for planning purposes.

### Revenue Tracking



Depending on the fare payment systems utilized (see the *New Fare Systems and Payment Technology Guidebook*), some transit systems can track the sources of revenue (fares paid) by route, by customer type, by payment type and by various other factors. Similar to some of the other data collected by passive systems, this data can be useful for planning purposes, particularly in terms of fare policy.

**The *New Fare Systems and Payment Technlogy Guidebook* has more detailed information on fare collection technologies and their uses**

### Potential for Passenger Origin-Destination Data



Some advanced passive data collection technology allows for the collection of rider origins and destination locations, as opposed to boarding/alighting location. These can be useful for system planning purposes, as well as some potential regulatory reporting requirements.

## System Data

These are elements related to the performance of the transit system itself, and may include:

### Bus Route/System Performance



There are several systems that can track the actual position of a bus relative to the location where it is supposed to be and can thus develop a comprehensive dataset of on-time performance and other reliability metrics. These systems can be used both for planning purposes as well as in real time, informing riders of the arrival status of a specific bus.

## Maintenance/Monitoring Data

These elements relate to the monitoring of the vehicle fleet and (depending on the technology selected) the operators. They may include:

### Driver Safety Tracking



Some systems can track the driving characteristics of a bus operator, such as acceleration rates, sudden brake applications, et cetera, so that remedial training can be assigned.

### Vehicle Health Monitoring



There are systems available that can monitor certain data points on board a bus – for example, engine temperature, fluid levels, rate of fuel and/or electric power charge consumption, or specific fault indicators – and either record or transmit the data to an operator’s maintenance staff. This allows the staff to more closely monitor a vehicle’s performance and maintenance needs and keeps an operator’s fleet in a better *state of good repair*.

Over the past few years, various types of technologies have been developed to help collect these various types of data in a passive manner. This not only includes ongoing improvements to relatively older technologies like Automated Passenger Counters (APCs) and Automatic Vehicle Location (AVL) systems, but also newer technology driven by the proliferation of mobile smartphones that allow for data to more easily gathered.

All of the various passive data collection technologies can use multiple methods to transmit the data they collect – the data can be transmitted wirelessly to a “base station” server when a vehicle is in its storage facility, or it could be uploaded to a server in real time, or the data may need to be downloaded through an actual “hardwired” physical connection.

Some of the various types of passive data collection system technologies are described here. They include:

## Automated Passenger Counters (APCs) and Automatic Vehicle Location (AVL) Systems

These devices are most typically utilized in conjunction with each other. AVL systems utilize a global positioning system (GPS) to determine vehicle location.

APCs allow for the accurate recording of boarding and alighting data, and – when used in conjunction with AVLs – allow that on-and-off data to be matched and linked to a geographic location, most typically a bus stop or transit station. APCs must be calibrated and verified so that they can be used for National Transit Database (NTD) reporting and allow for the replacement of “active” methods such as surveying or utilizing bus operators to record ridership. They allow for the continuous comparison of ridership data with fare collection data, and also allow for the provision of real time information regarding the level of crowding on a vehicle. It should be noted that it is typically easier to have a system’s entire bus fleet equipped with AVLs and APCs, as that eliminates the need to “cycle” the buses equipped with APCs and AVLs through all the bus routes (and trips) and through all the differing schedule days (such as weekdays, Saturdays, Sundays, holidays, and so on) that a system operates so that a sufficient “sample” of ridership levels is developed. By having the entire fleet equipped with AVLs and APCs, it greatly simplifies the systemwide scheduling and vehicle assignment process.

Some types of APC and AVL systems also integrate the capability to measure other elements of transit system use using sensors, such as (for example) deployment of the wheelchair ramp or use of the bicycle rack. While these types of additional data points may not always be required for NTD reporting, they can serve a useful planning function as they can indicate (for example) where to prioritize accessibility improvements at bus stops and sidewalks that meet the Americans with Disabilities Act (ADA) or where to locate bicycle racks along the transit system.

Taken together, APC and AVL systems provide transit systems with a large amount of data (which must still be analyzed) regarding the use of the transit system and its routes on a temporal (i.e., time-based) basis, a geographic basis and a directional basis.

There are two technologies that are most typically used by APC systems:

### Door Sensor-Based Passenger Counters

These electronic devices use invisible infrared beams located at the doorways of the transit vehicle. Depending on the number of times and the order these beams are broken, the number of riders and whether they are boarding or alighting can be determined.

### Video-Based Passenger Counters

These devices use cameras and associated software programs to count riders. While usually utilized at stations or bus tops, these can also be used on board a transit vehicle.

**APC systems must be regularly calibrated and validated in order to be able to be used for National Transit Database reporting**

As mentioned previously, calibration of APC systems is needed in order to use them for NTD reporting, and to create a method to account for certain anomalies in the data (for example, children in arms, et cetera).

## Mobile “Smartphone” Enabled Systems

There are several ways in which mobile smartphones can be used for passive data collection. These include:

### Mobile Fare Payment

Smartphones can be used to pay fares through a variety of software types; depending on the software and the settings selected by the smartphone user, a wealth of data can then be gathered including the fare paid, where it was paid, what payment method was used, and so on.

### Bus Tracking/Crowding Applications

As previously mentioned, with the use of AVL and APC systems transit agencies can provide real time tracking information to riders so that they can view estimates of bus arrival times, a specific vehicle’s location, the level of crowding estimated on board, and other factors.

### Aggregated Smartphone Movements Data

Several vendors aggregate data from mobile smartphones (again, depending on the software and settings selected by the smartphone user) to provide a variety of data points, such as trip origins and destinations, travel mode shares, travel movements along corridors by direction and time of day, and so on. This data is typically anonymized so that the induvial user’s privacy is maintained.

## Smart Card-Based Travel Data

Some transit agencies use “smart cards”; these are fare payment cards that can be filled at fare vending locations or linked to credit and/or debit accounts. The data from the use of these cards provide certain types of information including the fare paid, where it was paid, what payment method was used, and so on.

## Agency Website Usage

Depending on the transit system, some agency websites can be utilized to gather data regarding various subjects, such as trip planner queries and other elements that can be useful for planning purposes.

## Driver Safety Tracking Systems

As previously mentioned, some software systems can track the driving characteristics of a bus operator, such as acceleration rates, sudden brake applications, et cetera, so that remedial training can be assigned.

## Vehicle Health Monitoring/Performance Tracking Systems

As previously mentioned, these systems allow for the tracking of specific elements on a vehicle so as to help improve vehicle maintenance and keep the fleet in a *state of good repair*. These systems may include items such as fuel/electricity use sensors, speed sensors, and so on.

The main theoretical advantage to passive data collection systems is that – as they are not “active” – the automated data collection system frees up labor hours for tasks that cannot be easily or as effectively automated, such as customer service kiosks, bus operations, and safety and security functions. Furthermore, because they are automated, these systems can provide a more continuous stream of reporting than systems that rely on random sampling.

These methods of collecting data can be used for a variety of purposes. As was previously mentioned, APC units can (for example) provide important information on the highest ridership routes or stops for service planning and marketing departments, especially when combined with demographic information from the Census Bureau and can assist with NTD reporting once the systems have been validated for those purposes. Data from mobile fare payment applications can support Title VI fare equity analyses by showing the different pass usage rates on various routes.

However, as will be described in other portions of the *Passive Data Collection Guidebook*, while the automated and passive collection of data allows for less labor to be used in *gathering* the data, it does not obviate the need for skilled staff to nonetheless *analyze* the data, particularly given the raw level of data that passive data collection systems can produce and the need to be able to “scrub” through the data to decipher certain anomalies and produce useful information.

**Data may be gathered passively – nonetheless, the analysis of any data collected must be done actively by skilled staff for all the data to be truly useful**

# Passive Data Collection Vendor Types

As described previously, there are various types of passive data collection technologies, with each providing differing use cases for the transit agencies that utilize them, as well as differing mthods of how these data are collected. These can range from utilizing “big data” vendors (who use the aforementioned “Smartphone” enabled systems to gather data) to vendors of hardware systems that are provided “on-site”.

In October and November of 2021, interviews were held with vendors of passive data gathering systems – StreetLight Data, Transit App, and Avail – to gather information and examples regarding the capabilities of their technologies and how passive data is being gathered and utilized in transit. The companies provide products that serve different business functions, but all three generate and/or utilize passive data to provide solutions to transit agencies and riders.

This section of the Guidebook summarizes the information obtained from the interviews to show the overarching capabilities of some of the current technologies as well as some challenges that have been observed in the transition from active to passive data collection.

## ”Big Data” Vendors (Example: StreetLight Data)

***Introduction***

Some companies mine and amalgamate data sourced through passive mobile phone location data and blended it with spatial or statistical data. This data is then refined and presented in a manner that allows practitioners to understand and utilize the data. This process is referred to as “Passive Data-Derived Analytics.”

Data analytics companies serve a variety of business functions, but three common transportation uses include vehicle tracking, last mile studies, and origin-destination data generation. Vehicle tracking is able to obtain speed, demographic, and location data from vehicles, which can be valuable when developing transit network designs and corridor studies as it provides insights to the mobility needs in an area beyond that of average annual daily traffic (AADT) or other more traditional methods of traffic analysis. For example, San Mateo County Transportation Authority utilized vehicle tracking to determine that low bus ridership might have been a result of transit schedules not matching commuting travel time patterns, as opposed to a lack of intrinsic demand for transit service. After adjusting the schedules to better align with travel time patterns, ridership grew by 30 percent.

Their data can also support last mile studies; for example, data from mobile phones can help planners understand demand at park-and-ride lots both in terms of volume, times of use, and mode of access to and from a park-and-ride lot.

These companies can also provide detailed origin-destination information that includes locations and mode selection. For one company to develop software that can identify the mode in which people are traveling, contractors were hired to record their mode of travel (by keeping a “diary” of their travel); this data was subsequently used to develop a training set, which teaches the machine learning software to correctly classify a person’s mode of travel.

Agencies should ensure that all data used by these vendors is anonymized and, when evaluating bus networks, it doesn’t track individual bus operators; rather, data is provided at the system or route level.

***Typical Client***

Clients of these companies vary in type and in size; however, traditionally, clients consist of large metropolitan planning organizations (MPOs) or State agencies, and in many cases includes planning efforts for a transit system. However, the number of small and mid-sized organizations utilizing data vendors is becoming increasingly common as big data becomes more engrained in transportation planning processes. This type of data can serve both urban and rural populations – companies have found that their data provides rural practitioners value as passive monitoring is cheaper than in-person data collection methods.

***Overall Compatibility and Data Sharing***

Look for vendors that have a policy that encourages sharing data. There is flexibility built into the data sharing process, but the general approach for the vendor interviewed is to restrict clients’ ability to aggregate data (whose methods may be viewed as proprietary for a vendor) but to share results. The vendor interviewed, for example, contributes to Open Streets, meaning it shares anonymized data with the public; however, project-specific data can be limited on a case-by-case basis based on the preference of the client.

***Benefits and Challenges***

This type of business model is built on minimizing agency requirements and does not require client agencies to have a background in data analytics. However, if the user does have data analytics capabilities, the product’s overall utility might be increased. The product is meant to decrease the burden of “big data analysis” on agencies, given it is expensive and time consuming for firms and agencies to collect, store, and process detailed data. If the vendor maintains the data collection software itself, there is no need for localized maintenance efforts by agencies. Some parties also have ethical concerns about government data management practices; in some cases, these can be assuaged when the data is managed and stored by a third party vendor such as StreetLight Data.

As transit agencies tend to gather data from various sources, the lack of standardized data is a significant challenge. Given the lack of standardization, the data vendors’ products work best when paired with OpenStreetMap data. Some vendors are working on developing an industry standard that addresses how to best centralize data.

Finally, the data amalgamation provided by vendors generally focuses on individuals with mobile smartphones, meaning data collected may not fully reflect a given community. The data would be most likely to omit demographic groups that are less likely to have mobile smartphones, such as seniors. Though this means data is not entirely comprehensive for a given study area, most vendors feel the risk of not including at-risk populations and non-smartphone users is limited because traditional methods like in-person surveying also tend to have these biases. Additionally, mobile phone ownership is growing, and the present context is different from previous decades.

## Trip Planning/Discovery Application Vendors (Example: Transit App)

***Introduction***

The goal of trip planning and discovery applications (or “apps”) are to help people get around their cities, regardless of size, through a display of nearby mobility options in a mobile smartphone application. The app utilizes open source data, including data that agencies publish on vehicle routes, bike shares, scooter locations, and other types of travel modes. The app can also work with data submitted by transit agencies to help their users understand and utilize the various transportation options available. The data is displayed as real-time information to the app user to allow for multi-modal trip planning as well as fare collection.

Apps have deployed directional features as well, such as “GO” in Transit App. These are travel companions in the app that provide the user with audio directions and guidance at different stages of their trip. The app can track when and where the app is opened, when users view nearby routes, trips that users plan within the app, and when and where tickets are purchased. Agencies can also pay to have additional data analytics performed on their data.

Pioneer Valley Transit Authority in Springfield, Massachusetts and the Big Blue Bus in Santa Monica, California are examples of agencies that are incorporating ticketing into these apps to simplify the ticketing and payment system. All data used and collected through the app is anonymized for privacy purposes.

***Typical Client***

Client agencies are typically smaller urban or rural agencies. Nonetheless, such apps consider agencies with approximately 150 buses their primary client size because they are large enough to benefit from the app and need the assistance.

***Overall Compatibility and Data Sharing***

Apps only require the presence of cellular data or Wi-Fi for users to access the real-time data and utilize the fare collection component of the app. Some are also compatible with kiosks, which could mitigate accessibility concerns as some demographic groups are less likely to have mobile smartphones for use of the app.

There are agreements that client agencies must sign with a consultant before they share data, as there are rules around data aggregation, anonymity, and privatization which must be complied with. The apps offer ways to provide data to a client regarding route reliability and ridership levels.

***Benefits and Challenges***

Trip discovery and planning apps combine real-time route information, multimodal trip planning, fare collection, and audio directions into an app that can be directly used by transit riders. The transit agency does not need any specific settings or staffing requirements to utilize the app; however, having an in-house data analyst is helpful. For the apps to function properly, the transit agency must keep schedules, such as General Transit Feed Specification (GTFS) data, and real-time information up-to-date, as this is what is communicated through the app. It was noted in an interview with Transit App that it is best for agencies to rely on firms with digital expertise to build their schedule files rather than attempting to do this task themselves.

## On-Site Collection Tools (Example: Avail Technologies)

***Introduction***

Certain vendors provide end-to-end packages for agencies that include hardware, software, and various data collection and analysis services to help an agency passively collect data on vehicle health, passenger counting, and fare collection. The vendor interviewed – Avail Technologies – also provides a public-facing app for smartphones. Most contracts with agencies are full-service bids where Avail goes onsite to conduct a needs assessment, study current practices, provide Standard Operating Procedures, and maintain products in use.

Vehicle health monitoring has become an increasingly important part of some services as the expanded deployment of electric vehicles introduces “range anxiety” and other areas of concern associated with the relatively new technology being utilized. Butler Transit Authority is an example of a transit agency that initially needed a public-facing informational app but then expanded into a much larger scale of data gathering and distribution. This was an example of “growing out” the pieces and capabilities of on-site services and seeing the value in automating various processes.

***Overall Compatibility and Data Sharing***

These systems import an agency’s scheduling data rather than developing an agency’s schedules themselves. There are many sources for scheduling packages that are meeting the needs of agencies, so products are designed to be as compatible with those packages as possible.Some vendors, such as Avail, generate and utilize numerous variations of data including some protected data, such as payroll information, that cannot be shared. However, all data generated with Avail software or hardware belongs to the agencies and can be used at the agencies’ discretion. All data is stored in the Cloud.

***Benefits and Challenges***

These systems are able to provide stop-level detail, which can provide critical information when making decisions about changing stops or rescaling a system. By knowing who is riding, when they are riding, and which stops they are utilizing, agencies can make informed decisions that provide benefits to the most people.

Products can also provide detailed data on buses that can indicate if drivers are accelerating or braking too quickly and reducing the fuel or energy efficiency of the vehicles.

All data is provided in real time via cellular networks. While cellular networks are the only requirement, Avail stated in an interview that it is important that agencies also keep their General Transit Feed Specification (GTFS) data as up-to-date and accurate as possible. Issues have arisen that are tied to an underlying schedule problem, which can have downstream impacts and lead to confusion with the data. Additionally, once the services are rolled out, there is a maintenance plan to which agencies need to adhere. If the equipment is not routinely calibrated, it will get “out of sync” and provide inaccurate data.

While some small agencies have done very well with the rapid increase in data, on-site vendors’ service model is designed to fill the gaps so agencies can function seamlessly regardless of size and resources available. However, a current challenge with passive data is establishing trust in the data. The data is not perfect, but the systems are automated and run continuously, meaning the same data is being provided consistently, which can reduce the noise that is included in manually collected data. Building acceptance of automation is currently a challenge and will likely be one of the top issues agencies face when incorporating passive data into their systems and decision-making processes. Many small agencies outsource their information technology (IT) efforts, which leads to the critical need to have a “champion,” ideally in operations, that advocates for data collection and the employment of the tools that are available.

## Conclusions

The availability of passive data collection in transit will continue to grow as technology and a focus on the ability to utilize “big data” progress. The vendors interviewed fill different market needs, but they all focus on utilizing data collected passively to provide a product that allows agencies to make more data-driven decisions and for passengers to have accurate “real time” information (and thus have better and more reliable access to transit systems of various sizes and modes). A key shift that is occurring with the rise of passive data collection is the speed in which this data is communicated. More data is being communicated in real-time to both agencies (such as Avail providing vehicle health data), and passengers (such as Transit App providing arrival times and bus locations).

As with all data-driven tools, these vendors rely on accurate and standardized data, making it extremely important to keep underlying data inputs – such as a system’s General Transit Feed Specification (GTFS) information – up to date. This sets the stage for potential future initiatives that would further standardize data reporting processes for agencies on a larger scale.

**It is vitally important for transit agencies to keep underlying data inputs that various vendors use – such as the GTFS information – up to date**

While the availability of data continues to increase and vendors continue to offer innovative passive data collection capabilities and solutions, there is still a need to cultivate trust in the data, particularly among operations staff at transit agencies. Several of the passive data collection technology vendors interviewed mentioned that it is important for agencies to have a “champion” that advocates for the adoption and utilization of the passive data collection technologies that are available and becoming ever more present.

# Case Studies

## Introduction

Five transit agencies - that range in geography, size, and services offered - were interviewed to provide case studies on how agencies are using passive data collection systems.

The interviewed agencies included Victor Valley Transit Authority (VVTA) in California, CORTRAN in Roanoke, Virginia, the Lehigh and Northampton Transportation Authority (LANta) in the Lehigh Valley in Pennsylvania, Lane Transit District (LTD) in Eugene, Oregon, and Seneca Transit System in western New York.

As shown in **Table 2**, the National Transit Database (NTD) information has been provided for each transit agency participating in the case study interviews. The ridership (in terms of annual unlinked trips), vehicle revenue miles (VRM), vehicle revenue hours (VRH), vehicles operated in maximum service (VOMS), and operating funds expended have been defined. For VOMS, both the number of buses and the demand response (DR) total have been provided for reference, along with other modes if applicable.

**Table 2 – Case Study National Transit Database Information (2020)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Transit Agency** | **Ridership** | **VRM** | **VRH** | **VOMS** | **Operating Funds Expended** |
| **CORTRAN** | 13,840 | 173,958 | 13,580 | 4 (DR) | $790,278 |
| **Lane Transit District** | 8,533,030 | 7,067,504 | 437,933 | 134 (DR)86 (bus/BRT)16 (vanpool) | $54,346,157 |
| **LANta** | 4,036,833 | 4,964,631 | 358,144 | 82 (DR)78 (bus) | $37,725,778 |
| **Seneca Transit** | 12,463 | 203,699 | 5,682 | 2 (bus) | $558,494 |
| **Victor Valley Transit Authority** | 1,971,684 | 9,146,033 | 333,612 | 42 (DR)53 (bus)224 (vanpool) | $24,148,955 |

These agencies and their locations are shown in **Figure 1**.

**Figure 1 – Transit Agencies Interviewed**



These agencies provide services ranging from demand-response service utilizing contracted vehicles to fixed route systems. They also currently collect and utilize varying degrees of passive data and have clientele ranging from only seniors and those with disabilities, as is case with CORTRAN, to largely students and “essential” workers, as is the case with LTD and LANta.

The following sections will outline the current “use case” practices, applications of available technologies, tradeoffs, data sharing concerns, challenges, and some next steps associated with passive data for the agencies.

## Current Practices/Applications of Technology

All the interviewed agencies utilize passive data; however, there is significant variety in the quantity and level of detail being collected. VVTA, LTD, and LANta have computer-aided dispatch/automatic vehicle locator (CAD/AVL) systems and automatic passenger counters (APCs) on all buses, though the APCs are not certified for use in gathering data for the National Transit Database (NTD) at VVTA. The combination of CAD/AVL systems and APCs gather boarding data, passenger load data, as well as stop-level data. The data is utilized to analyze performance and ridership and then develop data-driven solutions to provide the most benefit to the largest number of riders. The AVL systems are also able to track sudden changes in speed that indicate extreme acceleration or braking, which can provide data in cases where there is an incident or even if a vehicle’s fuel consumption is outside of optimal parameters.

While Seneca Transit does not have APCs on their vehicles, the agency does use the “Passio GO!“ AVL and rider mobile app system that records passenger load and shows vehicle location through an online portal; this data is available in real-time as well as for later reference in cases of an incident. While the AVL records vehicle location, Seneca Transit utilizes active data collection to obtain passenger boarding information by having the driver push a button each time someone boards. This data is sent automatically to the agency. They use Passio GO! to eliminate the use of paper ridership data collection.

CORTRAN partners with Via to provide demand response transit. While Via is the vehicle operator, the vehicles are leased by a third party. Technology on the vehicles is limited to drivers using their own personal devices and the only data provided to the agency is from a report generated by Via.

The availability of data is continuing to increase, and some agencies are including vehicle health monitoring systems (sometimes referred to as “telematic” systems) on their vehicles. These are particularly relevant for electric vehicles. LTD indicated that they collect data from their electric buses on how quickly they are charging and the intensity of braking and accelerating. They currently have 10 electric buses but will soon have 30, dramatically increasing the available data from their fleet. Battery electric buses at VVTA also utilize a telematic system that passively collects data; however, this is not included on the compressed natural gas (CNG) buses. As an alternative to telematic systems, LANta has fuel pump and fluid software to capture how much fuel and fluids a bus is using at the point of delivery (on the service line). This is used to flag when a bus is outside of the target parameters, which triggers their maintenance department to inspect the bus.

Additional data can also be collected using apps that facilitate mobile ticketing. LANta is able to collect data from the fare payment app they use (Token Transit), which passengers can use to buy bus passes with their mobile phone, that provides data on when and where people are activating passes. This provides additional insights regarding demand for the transit system.

## Staff Capacity Requirements and Needs

While the vendors that were interviewed indicated that their products are designed in a way that does not require additional staffing or data analytics backgrounds, the agencies overwhelmingly cited the heightened staffing needs associated with passive data systems. VVTA summarized this by stating “to accurately collect and use passive data, you need active monitoring and analysis.”

**“…to accurately collect and use passive data, you need active monitoring and analysis…”**

Anomalies in the data are not infrequent and require attention to gain beneficial insight from the data. LANta indicated that they do have the ability to handle technology support and data analysis internally, but it does require a lot of staffing resources. At LANta, the planning staff handle the majority of the data analysis, but they also have a committee that reviews route-level data to identify where issues exist and distinguish problems from anomalies. For example, at the end of a route a driver may move the bus to a certain location to take a layover, and the APC and CAD/AVL systems will record that the trip has begun early because the vehicle is moving when it should not be on a route.

The interviews with the transit agencies drove home the point that while trained data scientists or analysts might not be needed, there is a need for tech-savvy, data-oriented employees to monitor and review the data. As was indicated in the interviews with the vendors, it is important to have a “champion” for the technology who is trained and can take the lead on diagnosing data anomalies and repairing the hardware. LANta stated that they have access to a strong labor market, but it has been difficult to hire people who are intellectually curious enough to find and diagnose anomalies in the data. This was echoed by LTD, as they take pride in the sophistication of their passive data collection but continue to struggle with staffing. This has been magnified by labor shortages facing most of the country since the beginning of the pandemic. CORTRAN noted a similar situation, though at a smaller scale, stating that while their staff would like to have additional capacity in managing and analyzing the provided data, it isn’t possible at this time. Generally, it appears that agencies collect data to varying degrees and while they stay within their staffing capabilities, they typically find themselves pushed near the limit of what their resources allow in terms of being able to analyze and prepare data.

**It is vitally important to have a “champion” for the technology who is trained and can take the lead on diagnosing data anomalies**

## Benefits and Costs of Technologies

Systems that enable passive data collection can drastically increase the amount of data available to transit agencies and can contribute to data-driven solutions regarding route changes, stop frequency, trouble-shooting problem areas, and many other applications. However, these systems represent an expense for the agencies. Given the inevitability of cost constraints, all agencies will have to assess the tradeoffs between the perceived costs and the potential benefits of passive data collection. LTD found that the implemented passive data collection technologies converted costs from labor to capital expenditures. For example, they are able to save some resources by doing on-board counts with APCs rather than hand-counts, but now instead needed to purchase, calibrate and maintain the APCs.

Both LANta and VVTA reported exploring the potential of comprehensive vehicle health monitoring/telematic systems, but felt the cost was too high. VVTA considered installing telematic systems on the entirety of the bus fleet but ultimately found it was not a feasible option due to costs. They have had some issues in the past with drivers not following procedures, which has led to mechanical issues (e.g., overheating engines), so they do see the benefit of all buses having telematic systems, but the perceived benefits do not outweigh the potential costs. While LANta did not feel they should fund the installation of vehicle health monitoring systems themselves, PennDOT is paying to upgrade all buses on fixed routes within Pennsylvania using Avail. This will allow PennDOT to collect the same data from all transit systems.

While some passive data collections systems were deemed infeasible due to cost, the interviewed agencies emphasized that the focus of these systems should be on improved service and reliability instead of any potential cost savings. Rather than assessing the benefit of the passive data collection using a traditional benefit-cost analysis in which the benefits are monetized, the agencies shared cohesive feedback that it is not about saving money but about delivering a better product to its customers. The data provided and the speed in which that data is provided allows agencies to make adjustments more dynamically. Seneca Transit noted that the motivation for employing the Passio system was improved accuracy of data. LANta noted that while cost savings may have been an initial motivation, they realized this is not actually an outcome. For example, with the installation of data collection systems, they thought they would not need supervisors in the field. However, this was not the case as the systems do not have the ability to interact with upset customers, change parking/layover accommodations, and many other needs that regularly require the presence of a trained professional. They iterated that the systems do help with the reliability and performance of the systems and, therefore, helps in delivering their mission – but they did not necessarily save resources. There is now less focus on using passive data collection systems to save money than using it to make sure the best service possible is provided.

**The focus of using these technologies should be on improved service and reliability for riders instead of any potential cost savings**

## Data Sharing

With an increase in available data, agencies must determine the degree to which they want and are able to share data. This varies based on the agency, the contracts they hold, and policies in their respective geographies. For example, although the contract that CORTRAN holds with Via states that no ridership data can be shared with researchers or the public, other agencies such as VVTA and Seneca Transit have faced no issues with sharing data. VVTA indicated that since the data is their operating data, it is theirs to use and distribute as they please; however, they did indicate that there has been some push-back with the battery electric buses over concern that information could be gathered without the VVTA’s approval.

Data privacy is a key concern for the transit agencies. To protect sensitive data CORTRAN reported that all data is anonymized, and LANta does not share any data from their paratransit system due to HIPPA protections. While LTD noted that data privacy is a key concern, they also indicated that Oregon has very strong open records laws, so there is not a lot of flexibility to keep data secret unless it would cause a specific privacy issue. LANta also discussed being subject to right-to-know laws. If they have a system that is collecting data and automatically generates a report, they would have to share this data if someone submitted a right-to-know request. For this reason, they do not upload all camera footage automatically; instead, they manually get footage if there is an incident. With this practice, they are able to cite that they do not normally store this footage, so they are not required to get the footage solely for a random request.

## Challenges

While passive data collection can allow agencies to provide a more efficient and reliable service, it does not come without challenges even outside those of cost and staffing capabilities. One difficulty that was noted with APCs is that they must regularly be manually calibrated or reset, which introduces room for error. This then introduces, for example, the question of whether failing to reset the APCs at the start of a route is worth disciplining drivers. Issues were reported by VVTA regarding compatibility and support with the APCs since they switched to their new CAD/AVL system. They are currently in the process of upupdating all the APCs due to this lack of compatibility and recommend that agencies always procure APC and CAD/AVL as a joint purchase. While there is significantly more data when using passive data collection systems, there are still shortcomings that limit the usage. CORTRAN noted that while data collection is easier with their contract with Via, their portal only provides data from recent months, meaning older data is still not easily accessible. LTD reported that while the data can be helpful for planning purposes, the data they obtained from StreetLight Data has margins of error that are large enough to make it challenging at times to use the data in the context of transit planning.

Seneca Transit reported challenges understanding how the data is obtained, transmitted, and reported. The Passio system does not provide any information other than ridership to the agency. They also utilize their contractor (First Transit) to track preventive maintenance activities, but they are not sure what methodology is used for this tracking.

As was indicated in the interviews with the vendors, the agencies have also faced challenges in establishing “buy-in” for the uses and types of data they collect. LANta reported that dispatchers frequently see issues reported by the data collection systems but do not report them because it happens on a regular basis. They tend to assume the system is wrong, so they do not react to the data in front of them. This presents a challenge as those in operations are the people who can use the information to impact the overall quality of the service. CORTRAN also reported instances in which passengers disagree with the routes that are recommended by the mapping technology, leaving drivers to determine if they should trust the generated route or take the route requested by the riders.

**Collecting data “is hard, but worth it…a lot of value…even though it’s really challenging…”**

While there are numerous challenges associated with data collection, the agencies do largely report that it allows them to provide a better service to their clients. LTD stated that “for small- and medium-sized agencies, collecting this data is hard but worth it. They get a lot of value for their money even though it’s really challenging.”

## Conclusion: Goals and Next Steps for Passive Data Collection Technology

As the agencies continue to expand their competency in data collection, management, and utilization, they are also looking at other areas that could benefit from more modern technology and data collection. One area that agencies would like to improve is the efficiency of [fare collection and ticketing](https://n-catt.org/guidebooks/new-fare-payment-systems-and-payment-technology/). This includes taking steps to encourage online ticketing rather than having riders use cash, as collecting cash on the buses is onerous. Seneca Transit had been exploring touch free/mobile ticketing but suspended further investigation due to COVID. However, the pandemic has made them question whether the cost of collecting fares is worth the administrative effort of collecting and counting cash, in addition to the operational delays due to people counting out change at the farebox.

Improvements to staffing capabilities is also something agencies would like to improve. Several of the agencies expressed issues regarding finding and retaining highly qualified staff. LTD is considering offering options such as allowing people to work remotely as a strategy to recruit talent.

The availability of data will continue to increase as technology progresses and electric vehicles become more commonplace in transit fleets. Agencies will be tasked with finding a balance between utilizing the available data to improve their systems and staffing and funding limitations. There will inevitably be redundancy in what the various technologies and data collection systems can offer and what drivers and maintenance employees already provide to agencies. Agencies will need to determine what types of data collection provide new and useful information as well as where data collection provides necessary redundancy that improves operations and maintenance capabilities. While case-studies and lessons-learned from other agencies can inform these decisions, it will ultimately be unique to each agency.

# Best Practices and Recommendations

In this portion of the *Passive Data Collection Guidebook*, the “state of the practice” is laid out for each of the various passive data collection technologies in a series of tabular matrices (in **Tables 3 through 9**), which allows the user to compare different elements for each technology, including the uses and shortcomings of each technology, the organizational/staffing capacity needed for implementation and operation, relative costs (when available), and other elements.

These matrices are intended to aid small-to-medium sized transit agencies in thinking about the relative cost and complexity of each technology, the benefits that these technologies bring, and other considerations before undertaking a more extensive pursuit of any new passive data collection technology.

Finally, the *Passive Data Collection Guidebook* provides a summary of “Recommendations” in terms of the key takeaways for each technology in **Table 10**.

**Table 3 – Automatic Vehicle Location Summary**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Technology** | **“Use Case” Function** | **Agency’s Staffing Needs** | **Agency’s Maintenance Needs** | **Overall Complexity** | **Value Added** | **Shortcomings** | **APPROXIMATE Cost****(*If Available*)** | **Estimated Vendor Diversity** |
| Automatic Vehicle Location (AVL) | * Tracks vehicles along routes
* Identify common “problem areas”
* Monitors on-time performance
 | * Manual corrections to logs can be required due to inaccurate readings
* Requires staff to review anomalies to obtain useable data
 | * Routine calibration
 | * Relatively low complexity to use
* Some degree of complexity can arise when analyzing the data and identifying issues and anomalies
 | * Increased awareness of movement along routes
* Continuously pinpoints “problem spots” in the system impacting on-time performance
* Most value added if installed on all buses in fleet
 | * Can sometimes incorrectly classify bus movements (for example, revenue vs. deadhead moves)
 | Unit cost:~ $500/AVL unit~ $150 software license/unit“Hardwire” installation:~ $2,000Ongoing support fees:~ $65 per month/unit, which includes configuration updates, reporting and data storage | * High
 |

**Table 4 – Driver Safety Tracking Summary**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Technology** | **“Use Case” Function** | **Agency’s Staffing Needs** | **Agency’s Maintenance Needs** | **Overall Complexity** | **Value Added** | **Shortcomings** | **APPROXIMATE Cost****(*If Available*)** | **Estimated Vendor Diversity** |
| Driver Safety Tracking | * Helps monitor sudden changes in speed
 | * Requires staff to review anomalies to obtain useable data
 | * Routine calibration
 | * Relatively low complexity to use
* Some degree of complexity can arise when analyzing the data and identifying issues and anomalies
 | * Can identify driving patterns outside of optimal parameters
* Most value added if installed on all buses in fleet
 | * Anomalies in the data could lead to inaccurate assumptions if not properly diagnosed
 | Varies on specifics of system purchased and extent to which included with original bus equipment | * High/Moderate
 |

**Table 5 – Automated Passenger Counters Summary**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Technology** | **“Use Case” Function** | **Agency’s Staffing Needs** | **Agency’s Maintenance Needs** | **Overall Complexity** | **Value Added** | **Shortcomings** | **APPROXIMATE Cost****(*If Available*)** | **Estimated Vendor Diversity** |
| Automated Passenger Counters (APC) | * Continuously counts passengers boarding and alighting at all times, as well as some other potential data points (such as wheelchair ramp deployments)
 | * Must regularly be manually calibrated or reset
* Requires staff to identify anomalies to determine is APC needs to be recalibrated

Robust data analytics capabilities are needed to gain full benefits of the data | * Routine calibration
 | * Relatively low complexity to use, but does place additional responsibility on drivers to correctly record trip start/end points
 | * Less noise in passenger count data

Comprehensive ridership data for all trips at all times* Eliminates need for manual on-board counts

Can also provide data on other use parameters, such as bicycle rack use | * Inaccurate data if APCs are not properly maintained and calibrated
 | * Unit cost:
* ~ $5,000/APC unit

~$900 software license/unit~ $700 Mobile Data Terminal/unit~ $850 hardware installation/unit* Ongoing support fees:
* ~$30 per month/unit which includes configuration updates, reporting and data storage
 | * High
 |

**Table 6 – Vehicle Health Monitoring Systems Summary**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Technology** | **“Use Case” Function** | **Agency’s Staffing Needs** | **Agency’s Maintenance Needs** | **Overall Complexity** | **Value Added** | **Shortcomings** | **APPROXIMATE Cost****(*If Available*)** | **Estimated Vendor Diversity** |
| Vehicle Health Monitoring Systems | * Tracks various metrics of vehicle performance to identify potential issues in real time
 | * Generates large amounts of data that must be analyzed and processed
* Requires operations and maintenance staff to monitor data and respond if an issue is detected
 | * Requires routines maintenance and calibration
 | * Low initial complexity as the vendor and/or original equipment manufacturer will install hardware and software
* A much higher degree of complexity can arise from the robust generation of large amounts of data
 | * Highly beneficial to electric vehicles as it can quickly identify when a bus is outside of its target parameters which could reduce range
* Providing real-time feedback on vehicle health could help avoid more costly maintenance and repairs that would be required if the issue was not detected
 | * Data is not always trusted by staff
* Can provide a high degree of redundant data
 | * Unit cost:
* ~ $1,000/bus (parts only)
 | * Moderate
 |

**Table 7 – Aggregated Mobile Phone Data Summary**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Technology** | **“Use Case” Function** | **Agency’s Staffing Needs** | **Agency’s Maintenance Needs** | **Overall Complexity** | **Value Added** | **Shortcomings** | **APPROXIMATE Cost****(*If Available*)** | **Estimated Vendor Diversity** |
| * Aggregated Mobile Phone Data
 | * Tracks all roadway users using mobile phone data
* Provides detailed origin-destination and mode choice information
 | * Robust data analytics capabilities are needed to gain full benefits of the data
 | * None
 | * Moderate due to degree of data that must be analyzed to reach meaningful conclusions
 | * More detailed level of detail regarding origin-destination data and mode choice than traditional modes of data collection
 | * Potential for large margins of error on the data
* Relies on mobile phone data so there might be gaps in the data due to demographics of users
 | * Varies on specifics of contract with agency and/or MPO
 | * Moderate/Low
 |

**Table 8 – Real Time Data Applications Summary**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Technology** | **“Use Case” Function** | **Agency’s Staffing Needs** | **Agency’s Maintenance Needs** | **Overall Complexity** | **Value Added** | **Shortcomings** | **APPROXIMATE Cost****(*If Available*)** | **Estimated Vendor Diversity** |
| Real Time Data Applications | * Provide real-time information about bus arrivals and locations
 | * Relatively minimal staffing needs
 | * None
 | * Relatively low complexity
 | * Gives passengers more detailed and accurate information than what may otherwise be available
 | * Accuracy of the apps are tied to the accuracy of the data provided by the agency, thus requiring the agency to robustly maintain its data (e.g., GTFS data)
 | * Typically contained within the cost of the APC and AVL systems
 | * High
 |

**Table 9 – Mobile Fare Payment Summary**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Technology** | **“Use Case” Function** | **Agency’s Staffing Needs** | **Agency’s Maintenance Needs** | **Overall Complexity** | **Value Added** | **Shortcomings** | **APPROXIMATE Cost****(*If Available*)** | **Estimated Vendor Diversity** |
| Mobile Fare Payment | * Allow riders to purchase tickets within an application rather than on the bus
 | * Relatively minimal staffing needs
 | * None
 | * Relatively low complexity
 | * If ticketing is available within the app, this can reduce the amount of cash that needs to be collected and processed
 | * Typically requires riders to have a smartphone and bank account/credit
 | * Typically, approximately 10% of each fare goes to the app vendor
 | * High/Moderate
 |

**Table 10 – Recommendations and Key Takeaways**

|  |  |
| --- | --- |
| **Technology** | **Key Recommendation / “Takeaway”** |
| Automatic Vehicle LocationIcon  Description automatically generated | Best when used on the entire fleet.Operations/on-road supervisory staff should be trained to “trust the data” and respond when conditions or issues arise that require interventions to get the best value from an AVL system and maintain on-time performance.  |
| Driver Safety TrackingA picture containing text  Description automatically generated | Address any potential implications regarding the collective bargaining agreement prior to purchasing driver safety tracking systems.  |
| Automated Passenger Counters | The quantity and type of data generated by automated passenger counters require robust data analysis capabilities and an active planning paradigm in the transit agency in order to get the most value from such systems.  |
| Vehicle Health MonitoringText  Description automatically generated | Analyze what new data a vehicle health system will provide versus what data it will provide that is already available from drivers and maintenance crews to help determine if this type of system is worth the cost.If an agency can only afford this type of system for some buses (but not all) in the system, prioritize electric vehicles, given their relatively new technological needs. |
| Aggregated Mobile Phone Data Icon  Description automatically generated | Transit agencies should consider “piggybacking” or “partnering” with other organizations (such as a Metropolitan Planning Organization) when purchasing this type of data technology, as other partners are just as likely to be using it.This type of data can be helpful when conducting a system-wide study or redrawing routes for a system. Similar to automated passenger counters, it should be kept in mind that this technology generates a large quantity of data that will need to be analyzed and will require robust data analysis capabilities and an active planning paradigm in the transit agency. |
| Real Time Data Applications | Although these are typically a benefit of having automated passenger counters and automatic vehicle location systems, it is important that an agency keep its input data – particularly GTFS data – as up to date as possible, as they form the foundation of these systems. |
| Mobile Fare PaymentA picture containing text, clipart  Description automatically generated | Agencies should first address the equity and Title VI issues associated with mobile fare payment before adopting any particular system or vendor – with the proper preparations, these systems can become an additional option that do not place any undue burden on the “unbanked” and “underbanked”.  |

**END OF GUIDEBOOK**

# Glossary

APC – Automated Passenger Counters

AVL – Automatic Vehicle Location

CBA – Collective Bargaining Agreement

Deadhead – Non-revenue service (can be miles, hours, et cetera)

DR – Demand Response

FR – Fixed Route

GPS – Global Positioning System

GTFS – General Transit Feed Specification

MPO – Metropolitan Planning Organization

# APPENDIX A – Vendor Interview Notes

## StreetLight Data

Vendor: [Streetlight Data](https://www.streetlightdata.com/)

Vendor Contact: Laura Schewel

Product Type: Data Mining and Amalgamation

Date of Interview: 11/4/21 @ 3:30 PM

General Questions

1. Can you describe your product in one or two sentences?

StreetLight Data sources passive cellphone location data and blends it with spatial or statistical data and presents refined data in a manner that allows practitioners (clients) an easy way to sort through and understand data. They refer to this action as “Passive Data-Derived Analytics”.

Their services provide three benefits:

1. It is expensive and time consuming for other firms to collect and process detailed data
2. It is also expensive to store data.
3. There are also ethical concerns about government data management practices that can be reduced with 3rd party (i.e., Streetlight Data) storage.
4. What business functions does your product serve?

Business functions vary but three common transportation cases include the following:

1. Car Tracking
	1. StreetLight Data will work with transit authorities to track car movements (a competitor to busses or trains) in a community
	2. This is valuable for things like network designs or corridor studies
	3. This involves more than just cars “pinging” a location, StreetLight Data can track speeds, demographics, etc.
		1. Laura said she could share methodology document\*
	4. A sample client was the San Mateo County Transportation Authority. StreetLight Data on car patterns revealed that low bus ridership might have been a result of schedules not matching travel patterns (informed by vehicle data)
		1. After adjusting schedules to accommodate these times, ridership grew by 30%
2. Last Mile Studies
	1. Their services can help planners understand demand at park & rides both in terms of volume, times, and how one travels to/from a Park & Ride
3. Transit Authorities
	1. They can provide detailed origin-destination information. This includes locations, mode selection (bus, bike, etc.)
		1. To determine a user’s mode, they have a “Nielsen Diary” model where they hire contractors to record how they travel, building a database which machine learning can then use to identify other people’s modes

Technical Questions

1. What are the minimum agency requirements in order for your product to work? (e.g., 4G service, GTFS data, etc.)

StreetLight Data’s Business model is built on minimizing agency requirements. The platform only requires agencies to provide an email and OpenStreetMap (OSM) information. If GTFS information is available, it is encouraged, but not required.

Data amalgamation focuses on individuals with smartphones meaning any data collected would not totally reflect a given community.

1. Is your product compatible with other typical transit technologies?
The product works best when paired with OSM data. This noted, there is an industry-wide data collection and processing challenge: unlike GTFS, collection and presentation standards do not exist, meaning different firms and agencies cannot easily blend data. Future solutions could involve standardizing data presentation, even if it just involves establishing spreadsheet design standards
2. [if public-facing] What technology requirements does the public have in order to use your service?

While this is a passive-data collection service, it still requires its data sources to have smartphones. Though this means data is not entirely comprehensive for a given study area, StreetLight Data feels the risk of not including at-risk populations and non-smartphone users is limited because traditional methods like in-person surveying also tend to have these biases. Additionally, mobile phone ownership is growing, and the present context is different from previous decades.

Maintenance Questions

1. What level of maintenance do clients typically perform themselves?
Streetlight Data does all maintenance, client involvement is minimal-to-non-existent.
2. What is the expected lifecycle of your product?

StreetLight Data operates on an annual-subscription basis.

1. What level of maintenance support is provided by your company?
StreetLight Data does all maintenance of data meaning there is no need for support.
2. What on-the-ground resources are needed in order for the product to run smoothly? (e.g., on-site IT staff, in-house data analyst, etc.)

Users need to have basic computer skills and a willingness to learn how to use the software (15-minute training video). There is no need for a user to have a background in data analytics. The platform is designed for someone who isn’t a data scientist. This noted, if a user has a background in data analytics, it exponentially increases the product’s utility.

Client Questions

1. What is the typical size of your clients?

Clients vary in type and in size. Traditionally, clients consist of large MPOs or State agencies. This noted, the business model is changing, and small/mid-size organizations are becoming more common. Currently, Streetlight Data has more than 700 clients ranging from Lancaster, PA to New York City.
2. Do you have any examples of small urban or rural clients effectively using your product?
Streetlight Data serves both urban and rural populations. The company is unique to its industry in that it serves rural areas, and it finds that it provides rural practitioners, especially utilities, value as passive monitoring is cheaper than in-person data collection methods.
3. Do you have any best practices or things to avoid for clients?

As stated above, the biggest challenge involves standardizing data. This is challenging when a transit authority is mixing data sources. StreetLight Data is working on developing an industry standard that addresses how to best centralize data.

Cost Questions

1. Is it less expensive to implement your product for smaller agencies?

StreetLight Data structures fees in a way that is context specific. For a long-term project this may involve the population size served by an MPO. For a smaller, specific project, they use a zone-based model that acts like geofencing. Clients can purchase data from a certain number of geographic zones covering a project area.
2. What is the typical capital cost for implementation?

Capital costs are very context specific.
3. Is there an ongoing fee for using the service? If so, what is it?

There is an annual subscription fee that is dependent on the size and scale of a project.

Data Sharing

1. Are you able to share the data you collect with researchers?

StreetLight Data has a policy that encourages sharing data. It can be flexible, but the general approach is to restrict the aggregation of data but share results.
2. Are there any limits to data sharing with the public?

StreetLight Data contributes to OpenStreets meaning it shares anatomized data with the public. This noted, project specific data can be limited on a case-by-case basis.
3. Do you have any problems with other vendors sharing data with you?

The issue comes back to a lack of data standards. In an ideal world, spreadsheets of data all have the same column design and structure, regardless of the data source. This is not the current case.

Additional

1. Do you have any additional comments or questions?

All StreetLight Data is anonymized. Additionally, when evaluating bus networks, it doesn’t track individual drivers, just the system or routes as a whole.

## Transit App

Vendor: Transit App

Vendor Contact: David Block-Schachter

Product Type: App (for “smart” phones – also has a desktop version)

Date of Interview: October 27, 2021

General Questions

1. **Can you describe your product in one or two sentences?**

The goal of the app is to help people get around their cities (of all sizes) through a display of nearby mobility options in the app.

The app pulls open data into its system. This data includes, but is not limited to, data that agencies publish on vehicle routes, including bike shares, scooter locations, et cetera. The data is displayed as real-time information if applicable to the data set.

Client agencies are located in approximately 300 cities across North America.

The app can work with data submitted by transit agencies to help their users get around. Through inclined partnerships, the app can publicize an agency’s data set and customize the data features. Agencies will typically pay the app to perform additional data analysis tasks for them.

There are concerns with privacy so everything is anonymized in the app. When a user first opens the Transit App it asks for the user’s location so it can show them nearby mobility options. The app has a general idea where users are located rather than specific addresses.

They can work with the agency to show specific mobility options. Through the agency contracts, the app provides the data back to the agency to use for their own purposes.

Data Collection: The app tracks the following data:

General location every time the app is opened

Records whenever the user clicks on nearby routes to view them

Records planned trips (origins and destinations) and which routes the user interacts with. It is important to note that there are notable differences between planned trips and actual trips performed.

Multimodal trip planning is a feature that is included and viewable in the app.

“GO” is a travel companion in the app that provides the user with audio directions and guidance at different stages of their trip.

The app also shares the user’s location so it can provide accurate information to other users who want to use a similar mobility mode or route in the future.

The data that is shared comes from the users. The data that is used in the app is mediated through the agency, bikeshare company, et cetera.

1. **What business functions does your product serve?**

The app integrates fare collection services into it.

Technical Questions

1. **What are the minimum agency requirements in order for your product to work? (e.g., 4G service, GTFS data, etc.)**The agency does not need any specific settings or requirements to participate in the app.
2. **Is your product compatible with other typical transit technologies?**It is compatible with kiosks.
3. **[if public-facing] What technology requirements does the public have in order to use your service?**

Cellular data or Wi-Fi is needed to access the real-time data in the app.

Survey work has helped the app (on-board surveys) understand the demographics of its users. (Note: slightly oversamples students, but other demographics are similar to the general transit user).

Maintenance Questions

1. **What level of maintenance do clients typically perform themselves?**
Clients must keep schedules (e.g., GTFS data) and real-time information up to date.
2. **What is the expected lifecycle of your product**?
There is no “lifecycle” that clients or users need concern themselves with – they constantly update a new version of the app approximately every three weeks. Agencies receive new features first.
3. **What level of maintenance support is provided by your company?**
There is no need for “maintenance support” on the client side. Nonetheless, the app offers ways to give the data back to a client in a way that they may not have seen before (e.g., route reliability and ridership levels).
4. **What on-the-ground resources are needed in order for the product to run smoothly? (e.g., on-site IT staff, in-house data analyst, etc.)**

An in-house data analyst is helpful, but not necessary.

Client Questions

1. **What is the typical size of your clients?**
Clients are typically small urban or rural agencies. The app considers agencies with approximately 150 buses their primary client size because they are just large enough to work with the app and need the assistance.
2. **Do you have any examples of small urban or rural clients effectively using your product?**
In Massachusetts they are integrating ticketing into the app for Pioneer Valley. The Big Blue Bus in Santa Monica is another example.
3. **Do you have any best practices or things to avoid for clients?**

Clients should stop trying to build the schedule files (i.e., GTFS data) themselves – it is typically more accurate if they can farm it out to a firm that has more digital expertise than always trying to do it in-house.

Cost Questions

1. **Is it less expensive to implement your product for smaller agencies?**
Costs vary depending on the level of features selected and the agency size. They review the number of boardings, the number of vehicles, and the number of routes to determine what an appropriate fee should be for each client agency.
2. **What is the typical capital cost for implementation?**
The app costs approximately $500 for a small agency.
3. **Is there an ongoing fee for using the service? If so, what is it?**

“Royale” is a subscription service provided to the user for an additional fee to access additional features in the app. Agencies can pay for this if they want their users to have access to it. There are monthly and annual subscription options.

Data Sharing

1. **Are you able to share the data you collect with researchers?**
There is a form of agreement that the agency must sign with a consultant company before they share data with them. There are rules around data aggregation, anonymity and privatization that must be complied with.
2. **Are there any limits to data sharing with the public?**
All data is anonymized, but the agency may decide what to share in public reports.
3. **Do you have any problems with other vendors sharing data with you?**

No.

Additional

1. **Do you have any additional comments or questions?**

The app also works with IKE kiosks.

## Avail Technologies

Vendor: Avail Technologies

Vendor Contact: Rick Spangler, Ryan Harshbarger

Product Type: hardware and software data collection and support services

Date of Interview: 11/16/2021

General Questions

1. **Can you describe your product in one or two sentences?**

Avail provides an end-to-end software package for agencies that includes hardware, software, and various data collection and analysis services to help an agency make the leap from active to passive data collection. This includes driver behavior data such as when a driver is being heavy on the gas or regularly slamming on the brakes, which is relevant given many agencies want to improve the range on their new electric buses. There is a “scorecard” that allows training opportunities to be identified.

Avail also provides public-facing apps for mobile phones that agencies can use to show their riders where buses are, how close to schedule they are operating, et cetera. Avail imports scheduling data rather than developing schedules themselves. There are other excellent scheduling packages meeting the needs of agencies, so Avail’s products are designed to be as integrated with those packages as possible.

The first software package Avail developed analyzed reporting on farebox data because agencies were beginning to buy electric fareboxes. The genesis of many of Avail’s offerings came from the desire to automate reporting for the National Transit Database (NTD) requirements. The product was a way to visualize the large amounts of data to present the information in a digestible way.

1. **What business functions does your product serve?**

Most contracts with agencies are full-service bids where Avail goes onsite to conduct a needs assessment, study current practices, provide Standard Operating Procedures (when and where to use technologies), maintain products, etc.

Vehicle health monitoring is an increasingly important part of Avail’s offerings. While individual drivers have range anxiety with electric vehicles, there are tangible dollars that can be assigned to every roadcall that is both costly and inconvenient for agencies. Vehicle health monitoring has thus become more crucial as it provides more clear and consistent information about the bus condition and whether a roadcall is needed. With electric buses there is now the question of whether the maintenance/repair needs to be completed out-of-house, which adds additional cost and time to the maintenance/repair. Analysis of the vehicle’s whole health has gotten more important given repairs can be more complicated. This is ultimately a cost minimization exercise.

Technical Questions

1. **What are the minimum agency requirements in order for your product to work? (e.g., 4G service, GTFS data, etc.)**

Data communications is a backbone of the services. Cellular coverage is now cheap and widespread, and people want data in real time. All data is provided in real time via cellular networks.

Maintenance Questions

1. **What level of maintenance do clients typically perform themselves?**

Once the services are rolled out, there is a maintenance plan to which agencies need to adhere. If the equipment is not routinely calibrated, it will get “out-of-sync” and provide inaccurate data.

1. **What on-the-ground resources are needed in order for the product to run smoothly? (e.g., on-site IT staff, in-house data analyst, etc.)**

It was discussed that small agencies have the same transit challenges as large transit agencies; although they are on a smaller scale (and given that they are smaller agencies with fewer resources), some challenges are actually more significant for a smaller agency, as there is less ability to recover from incidents or irregular operations. Some small agencies have done very well with the data, but Avail’s service model is to fill the gaps so agencies can function seamlessly regardless of size and resources available. A lot of small agencies outsource their IT, which leads to the critical need to have a “champion” for data collection technology advocating to use the tools that are available – this does not need to be an IT person (in fact, it could be more useful if the “champions” are operations personnel, as they benefit most directly from data collection automation).

Client Questions

1. **Do you have any examples of small urban or rural clients effectively using your product?**

Two examples of small rural systems that has benefitted from Avail’s services are Butler Transit Authority and Crawford Area Transportation Authority (both in western Pennsylvania). In smaller municipalities it is life changing to have an app with bus times, and there are people at small agencies that take this very seriously. Butler initially needed such an app, but this expanded into a much larger scale of data gathering and data distribution for various aspects of the agency. This was an example of growing out the pieces of the system and seeing the value in automating various processes.

1. **Do you have any best practices or things to avoid for clients?**

Avail concurred that it is important that agencies maintain their General Transit Feed Specification (GTFS) data as up-to-date and accurate as possible. For example, some common issues are tied to an underlying schedule problem, which can have downstream impacts. It is critical to pay attention to stops in the right order or the data will not always make sense.

Cost Questions

1. **Is it less expensive to implement your product for smaller agencies?**

Pricing is tied to the number of vehicles they are deploying hardware on. The pricing for the maintenance package is standard. They do also keep an eye on the count of riders and overall agency size when considering some pricing decisions.

Data Sharing

1. **Are you able to share the data you collect with researchers?**

Avail has a lot of variations of data, which does include some protected data, such as payroll information. However, the agency data belongs to the agencies, so they can do with it what they want. Data is stored in the cloud.

Additional

1. **Do you have any additional comments or questions?**

The risk for missing key information points is bigger for smaller agencies, given their (general) lower frequency of service. For example, when you are making decisions about cutting stops or scaling back a system, it is critical that you know who is riding, when they are riding, and which stops they are utilizing – all items that can actually be more impactful when a system is smaller than in larger transit systems.

A current challenge with passive data is establishing trust in the data. The data is not perfect, but the systems are automated and run 24/7, meaning the same data is being provided consistently. This reduces the “noise” that is included in manually collected data. However, getting people to emotionally accept automation is still a challenge.

# APPENDIX B – Case Study Interview Notes

## CORTRAN

Agency: CORTRAN

Agency Contact: Paula Benke

Product Type: Demand-response transit services

Date of Interview: 11/5/21

General Questions

1. Could you describe your service in a few sentences? (e.g., deviated fixed route, demand response, etc.)

CORTRAN partners with Via to provide demand-response transit services to qualifying individuals in the greater Roanoke area. While Via is the “operator” the vehicles are leased by a third party, though they are CORTRAN-branded. In many ways, this setup mirrors a TNC in that drivers are independent contractors.
2. Could you describe your customer base? Seniors, students, low-income, etc.

To use the services, customers must satisfy the following:
	1. County residency
	2. Their age must be 70+ or they must have a disability
	3. Need transportation to and from destinations within the County of Roanoke, Town of Vinton, City of Salem, and City of Roanoke
3. Can you give a brief overview of the kinds of data collection activities that you have to undertake for asset management, NTD reporting, grant reporting, etc.?

They use data to apply for grants and when they report their funding. Funding reporting is not as strict for this service.

 Data collected includes the following:

* 1. User information (name, address, phone, etc.)
	2. Origin-destination
	3. Trip purposes
	4. Payment method
	5. No-shows

For reporting, Via provides some data, though there isn’t a formal reporting structure in place. Via provides data as requested by CORTRAN.

Technology Questions

1. What passive data collection technologies do you currently have on your vehicles?

Via leases the vehicles form a 3rd party (Buggy). Technology on the vehicles is limited to drivers using their own personal devices. This is similar to many TLC models where drivers accept riders on their personal devise. All vehicle use, including unlocking the car or accepting riders, takes place on the phone (via an app).
2. What technologies have you explored but not implemented?

Prior to Via, vendors were not as effective in providing the service.
3. What has been your biggest disappointment with technology in the past?

It is difficult to pull past data. The portal seems to only store data from recent months and data from prior years isn’t easily accessible.

Additionally, the following issues occasionally arise:

* 1. Some riders disagree with the routes that the mapping technology generates
	2. There seem to be app glitches that occur.
		1. Via is quick to service though it feels like Via typically feels the glitches are a “first-time” incident
		2. Related, Via is based overseas, meaning there sometimes is a delay in service due to time differences
		3. Lastly, there have been transition pains in terms of acquiring ADA-accessible vehicles
1. Do you go looking for technologies with passive data collection capabilities, or is it a side-benefit of other technologies that you procure for other reasons?

n/a

Capacity Questions

1. What staff capacity do you have to support technology at your agency?

Staff capacity at CORTRAN is small, with one individual (Paula) managing the operation.

1. Do you have in-house staff or contractors to serve your technology needs? Or do you rely entirely on vendor support?

While CORTRAN staff would like to have additional capacity in managing and analyzing the via-provided data, it isn’t possible at this time. In place, Via has a semi-user-friendly system that allows her to generate reports.

1. Are capacity limitations due to budget or lack of expertise in your geographic area?

Paula doesn’t have a transportation planning/data sciences background (Human Resources) though she finds the data effective. If she had additional staff and more time to focus on the project, she’d like to further explore and understand the data.

CORTRAN does utilize the county’s GIS (Geographical Information Systems) department for mapping and verification of address data when Via is not able to or cannot provide data in a timely manner. This happens rarely, but it has occurred within the last year.

Cost Questions

1. Do you find that passive data collection technologies save you money?

The monthly cost for Via is based on two components:
	1. Service hours (how long the call center for rides is open)
	2. Driver hours (the hours a driver is available to take rides)
		1. Drivers are paid continuously, even while deadheading
2. Do you find data collection and reporting requirements burdensome to the point that you are looking for opportunities to invest in automated capabilities?

CORTRAN staff wish they had more time to process the data available, though there hasn’t been a discussion on automation.

Data Sharing

1. Are you able to share the data you collect with researchers?

The contract with Via states that no ridership data can be shared with researchers. All data is anatomized (no birthdays stored, etc.).
2. Are there any limits to data sharing with the public?

Data cannot be shared with the public.
3. Are you permitted to share the data developed by your passive data collection technology in reports and other media? Do you regularly do so?

There are very tight restrictions on when data can be shared.

Additional

1. Do you have any additional comments or questions?

Service is generally effective, though it can be patchy in rural portions of the service area. The majority of rides take place in an urban context.

## Lane Transit District

Agency: Lane Transit District (LTD)

Agency Contact: Andrew Martin

Date of Interview: 10/29/21

General Questions

1. **Could you describe your service in a few sentences? (e.g., deviated fixed route, demand response, etc.)**
* LTD provides fixed route, BRT service, demand response, demand taxi – demand response service is all contracted out.
* Also runs the Rhody express in Florence. That is contracted service, though the vehicles and preventive maintenance are all provided by LTD.

1. **Could you describe your customer base? Seniors, students, low-income, etc.**
* Students are 30% – 40% of the ridership; university students are a lot of the ridership and they started a student pass program for younger students. All high school, middle school, and elementary school students get passes and can ride for free.
1. **Can you give a brief overview of the kinds of data collection activities that you have to undertake for asset management, NTD reporting, grant reporting, etc.?**
* LTD collects a lot of data. For a mid-sized agency, it is definitely on the high end of data collection. 2007 is when they equipped vehicles with CAD / AVL on their buses, which are also 100% equipped with APCs. They also do APC validation in-house.
* For asset management, they use a system called EAM in maintenance. They contract out maintenance for paratransit vehicles and do not have the paratransit fleet hooked into the EAM.
* The agency has 10 electric buses, soon to be 30. They are collecting a lot of data from the electric buses through the manufacturer. E-buses are a lot smarter; they can see how quickly they are charging, what drivers’ acceleration and braking habits are, etc.
* AVL also on the paratransit vans, and paratransit scheduling is through TripSpark Novis.

Technology Questions

1. **What passive data collection technologies do you currently have on your vehicles?**
* Fareboxes are really not smart. Drivers have to do a visual verification of exact change. But they did purchase touch cards and a mobile fare payment app so they know when cards tap or scan in. They are not collecting location-based information or other met-data. They are using something called “Fare Collection as a Service” or something similar from the mobile payment vendor.
* Every 4 years LTD does a Title VI analysis, and they have to do origin/destination surveys through on-board surveys. Now that there is fare card data they want to do an analysis similar to the MBTA ODX model.
1. **What technologies have you explored but not implemented?**
* None that the interviewee was aware of.
1. **Do you go looking for technologies with passive data collection capabilities, or is it a side-benefit of other technologies that you procure for other reasons?**
* It’s more that LTD gets new technology unrelated to data collection but gets to enjoy new data collection capabilities when it becomes available. Sometimes they are concerned with data collection, but usually that isn’t driving the procurement decision.

Capacity Questions

1. **What staff capacity do you have to support technology at your agency?**
* They don’t have people hyper-specialized in data at a mid-sized agency like LTD. But over the years there have been some highly effective data/tech-oriented employees. For instance, they have a really customized version of Transit Master (Trapeze) that scrubs data, etc. It is challenging for a small agency to maintain that kind of expertise. The IT director is retiring, and it’s going to take 2 – 3 people to replace his skillset.
* LTD has also benefitted immensely from having the back-end infrastructure set up. They are leaders in GTFS, NTD data collection – LTD is very sophisticated but staffing is the challenge.
1. **Are capacity limitations due to budget or lack of expertise in your geographic area?**
* LTD has trouble attracting planners. One of the things they are talking about is letting people work remote full time. For example, the database manager is local but works remotely and could conceivably move away. Specialized IT positions might be full-time remote and could be located anywhere. For example, they have a grant to develop a MaaS system and the project manager is going to be located in Pittsburgh.

Cost Questions

1. **Do you find that passive data collection technologies save you money?**
* It converts cost from labor to capital. They save a lot of money by doing on-board counts with APCs instead of hand-counts. But more importantly, the quality of the data is so much higher than anything you could get manually.
* This improved quality of data translates directly to improved quality of service and customer experience. They can adjust time points much more dynamically to improve OTP, for example.

Data Sharing

1. **Are you able to share the data you collect?**
* Not aware of any issues with data restrictions. Data privacy is a key concern. Oregon has very strong open records laws and so there is not a lot of flexibility to keep data secret unless it is a privacy issue.

Additional

1. **Do you have any additional comments or questions or thoughts?**
* The main performance report that is compiled for the board is ridership, financial, and other metrics. It’s an Excel file that has to be updated manually. Some of the reports are automated and newer software is providing more canned reports, but they don’t always meet agency needs.
* However, having so many custom reports and maintaining them through staff turnover makes the custom reports very difficult. One change to the system causes reports to break. One example is when a staff position left and then a finance code changed, and the custom report prevented them from paying invoices.
* In terms of Streetlight, the MPO recently updated the Regional Transportation Plan using the platform. However, the margins of error make it challenging to use in the transit context.
* Final quote: “For small- and medium-sized agencies, collecting this data is hard but worth it. They get a lot of value for their money even though it’s really challenging.”

## LANta

Agency: Lehigh and Northampton Transportation Authority

Agency Contact: Owen O’Neil

Product Type:

Date of Interview: 12/1/2021

General Questions

1. **Could you describe your service in a few sentences? (e.g., deviated fixed route, demand response, etc.)**

LANta provides a fixed-route bus system as well as paratransit services. The paratransit system is fairly large compared to the fixed route system as they have the ADA program as well as various human service transportation systems for seniors and those with medical needs (LANtaVan).

**Could you describe your customer base? Seniors, students, low-income, etc.**

For the fixed-route system, the average rider is in the 30-50 age group, disproportionately female, and generally an hourly wage earner ($15-20 an hour). The majority of riders are people commuting between their homes in urban areas and jobs in suburban areas. This is common for the area as many people live in urban areas and then need to travel to specific employers that are not in downtowns, and historically this was the case in the Lehigh Valley with large manufacturers (e.g., Mack Truck) and the Bethlehem Steel Mill being located away from the traditional downtown areas.

1. **Can you give a brief overview of the kinds of data collection activities that you have to undertake for asset management, NTD reporting, grant reporting, etc.?**

On the fixed-route system, all buses have APCs that track the number of people boarding the bus. The APCs can also provide a greater level of detail, down to stop-level data. They also use farebox software that gives ridership information but does not give stop-level data.

Last year LANta switched to Optibus for scheduling software.

LANta uses Token Transit to enable people to buy passes with their phone. This also provides data about when and where people are activating passes.

On the demand-response system it is easier to have specific data because they know exactly where people are picked up and dropped off.

The Planning Department leads the effort of digging into the data for processing needs. They lead a committee consisting of Planning, Operations and Customer Service staff that regularly reviews information. They use the data to look at performance/ridership and then make decisions from this information.

For more detailed data analysis, there is a committee consisting of Planning, Operations and Customer Service staff that regularly looks at route level data to identify where issues exist and works to separate issues from anomalies. For example, the layover activities at the end of line tend to create anomalies that require manual analysis (e.g., an operator pulls off to another spot to take layover, and the APC and AVL systems erroneously think the next trip has started). If certain days are not running smoothly, they are able to investigate what the driver is doing differently. Driver-specific information is available from the AVL system.

Technology Questions

1. **What passive data collection technologies do you currently have on your vehicles?**

LANta does not currently have vehicle health monitoring systems installed on buses, but they do have the fuel pump and fluid software to capture how much fuel/fluids and bus is using. This can be used to flag when bus is outside of the target parameters, which can trigger the maintenance department to say something is off.

For braking and acceleration information, the AVL system and camera systems track speed and sudden changes in speed. They can look at this if there is an incident.

1. **What technologies have you explored but not implemented?**

Vehicle health monitoring has been pitched before, but they did not think it was worth the cost to fund on their own. However, there is now a Fixed Route ITS program where PennDOT is paying to upgrade buses on fixed routes using AVAIL. This statewide initiative will include vehicle health monitoring and will allow PennDOT to have the same data from all transit systems.

LANta does already have some AVAIL equipment, so their equipment and software will be updated. Systems that do not currently have any AVAIL products will receive all new products.

LANta does maintenance in-house and has procedures and programs in place to monitor buses. Using AVAIL will provide real-time information such as if an engine is running hot or other real time issues that might need to be addressed. There will be some degree of redundancy with the products offered and what LANta already has in place.

1. **What has been your biggest disappointment with technology in the past?**

One difficulty with APCs is that they rely on manual calibration, which introduces room for error. This then introduces the question of whether failing to recalibrate the APCs is this really worth disciplining drivers.

They do have mobile data terminals so they can see what the APC is showing. It will beep at them is they are running early or late. The schedule data is uploaded from Optibus into AVAIL.

Capacity Questions

1. **What staff capacity do you have to support technology at your agency?**

LANta does have the ability to handle technology support and data analysis in-house, but it does take a lot of staff resources. For both AVAIL and the farebox program, there are always “exceptions” that occur where people must then go in and clarify the data. This takes dedicated staff and the smaller the agency, the fewer staff they have to track these types of issues.

They are basically part of the Philadelphia and New York City metro area, so they do have a good labor market, but they are still experiencing labor issues, like most places in the current market. However, it has always been hard to get people who are tech savvy and intellectually curious enough to troubleshoot the exceptions. It was noted that it is hard to get these people at the pay a rural transit system can offer.

The IT staff can make sure the systems are running but it is the Planning and Operations staff who need to figure out the anomalies. This usually falls on the planning staff, but dispatchers and operations supervisors need to be invested to really make the system work well. Frequently dispatchers notice issues but do not report them because it happens every day – they tend to assume the system is wrong, so they do not react to the data in front of them. This presents a challenge as these are the people that can use the information they receive to impact the overall quality of the service.

This supports the idea that passive data systems really drive what riders perceive as the reliability of the service. LANta feels passive data collection (and dissemination) systems are not “money savers”, per se, but rather services that help people provide better and more reliable service “on the street”.

LANta is now using TransitApp. They are currently pushing this more than the AVAIL app, which works but is not great. This app has a direct link with Token Transit, so users can pay fares from within the app. TransitApp can show crowding, but they do not have this turned on because it has not been an issue.

Cost Questions

1. **Do you find that passive data collection technologies save you money?**

The various technologies do not appear to save the system money. This was originally the thought and motivation, but, over time, they have realized (for example) that you still need supervisors out on the streets. However, the technology does help with the reliability and performance of the system and, therefore, helps in delivering their mission. There is now less focus on using it to save money than using it to make sure the best service possible is provided.

In the beginning they tried to not have supervisors on the streets but there are other needs for supervisors, such as altercations, parking requests, etc., that the technology can obviously not handle.

1. **Do you find data collection and reporting requirements burdensome to the point that you are looking for opportunities to invest in automated capabilities?**

Collecting cash on the buses is quite onerous. They are trying to get people to pre-purchase passes or use mobile ticketing. With cash, they need to take measures to ensure the cash is secure and have designated staff to count the cash. Armored car services are getting more expensive due to staffing issues so the cost of processing every dollar that comes in is costing more money.

Most cash comes in on the vehicles but they also have customer service windows that accept cash. These are more onerous because it is less cash, but you have to go through all the same steps to secure, count, and deposit cash.

LANta is just starting to roll out ticket vending machines, but this still involves cash. However, they can hire a service to handle bringing in the cash. LANta still must process this cash. Vending machines will also take credit cards. They would run into a Title VI issue if they decided to charge less for customers using credit cards given fewer low income-people have credit cards.

Data Sharing

1. **Are you able to share the data you collect with researchers?**

Especially on the paratransit side, LANta doesn’t share anything because of HIPPA protections.

LANta is subject to Right-to-Know laws. If they have a system that is automatically collecting data and would produce a report by just pressing a button, they would have to share this data if someone submitted a Right-to-Know request.

They do not upload all camera footage because if uploaded, it is subject to a Right-to-Know request. Instead, they manually get footage if there is an incident. Data is often requested by attorneys (or in worse cases, for example, someone who is stalking another person). With their practice, they can say they would not normally get this footage, so they are not going to get the footage just for the request.

## Seneca Transit System

Agency: Seneca Transit System (STS)

Agency Contact: Sharon Ray, Judy, Tammy

Date of Interview: 11/1/21

General Questions

1. **Could you describe your service in a few sentences? (e.g., deviated fixed route, demand response, etc.)**
* STS operates public transportation for the Seneca Nation across two territories, Allegheny and Cattaraugus. There is 35 miles between the two territories, with two buses running back and forth connecting them.
* Ridership since COVID has gone down and the agency is still not collecting fares. It receives FTA funding as well as NYSDOT to run their fixed route deviated service (up to a quarter of a mile).
* They run purchased transportation. First Transit is the service provider.
* They do not currently charge fares due to COVID but planning on charging fares in the future. They have looked at touch free/mobile ticketing but suspended further investigation due to COVID. The pandemic has really made them question whether the cost of collecting fares is worth the administrative effort of collecting and counting cash, in addition to the operational delays due to people counting out change at the farebox.

1. **Could you describe your customer base? Seniors, students, low-income, etc.**
* Biggest customer base is middle-aged, 20s through 50s. Cattaraugus residents travel to the Park and Shop in Gowanda NY, and a lot from Allegheny going out to Cattaraugus. They are also trying to get a circulator within the city of Salamanca.
1. **Can you give a brief overview of the kinds of data collection activities that you have to undertake for asset management, NTD reporting, grant reporting, etc.?**
* Most of the data collection and reporting is done by First Transit. Less than a year ago, they purchased a Passio Go! to get rid of paper ridership data collection. However, so far, the paper counts are used sometimes and then put into an Excel spreadsheet.
* When they use Passio, every time someone gets on, the driver pushes a button to record the boardings. Data is sent via MiFi data link.
* Passio doesn’t provide any information other than ridership to STS. There is also vehicle load information but it’s not clear how that is transmitted or reported.
* First Transit tracks preventive maintenance activities, but they are not sure what methodology they use to track it.

Technology Questions

1. **What passive data collection technologies do you currently have on your vehicles?**
* The agency has AVL systems which record vehicle location and show it through the Passio online portal. It also records the passenger load. They use the AVL to track the vehicles in real time and also to go back and see vehicle locations if there was an issue.
1. **What technologies have you explored but not implemented?**
* STS is looking to purchase new buses shortly and they are investigating what could make the technology set up better. They are interested in acquiring APCs and potentially new farebox technology.
1. **What has been your biggest disappointment with technology in the past?**
* There haven’t been any big disappointments with the technology, the challenges have been more with training and implementation. A lot of the drivers are reluctant to change established practices.
1. **Do you go looking for technologies with passive data collection capabilities, or is it a side-benefit of other technologies that you procure for other reasons?**
* New technology is not a major focus of their operations, it is more just an issue that they think about when they procure new vehicles, for example.

Capacity Questions

1. **What staff capacity do you have to support technology at your agency?**
* The tribe has an IT department, but they haven’t really assisted at all with data collection. The transit planner generally handles data collection and analysis.
1. **Do you have in-house staff or contractors to serve your technology needs? Or do you rely entirely on vendor support?**
* First transit provides the staff, Seneca Transit provides the vehicles and equipment.
1. **Are capacity limitations due to budget or lack of expertise in your geographic area?**
* The biggest issue is keeping bus drivers, there is a lot of turnover. If they could get staff consistency, it would be much easier to implement new practices or technology.

Cost Questions

1. **Do you find that passive data collection technologies save you money?**
* The motivation for the Passio technology was accuracy of data collection, not really savings.

Data Sharing

1. **Are you able to share the data you collect?**
* Never had any problems with sharing data.

Additional

1. **Do you have any additional comments or questions or thoughts?**
* Passenger counting has been a business case where they needed new technology but they tried to employ this new technology (Passio Go!) and it has been a rocky implementation – drivers don’t like the new technology, there is high driver turnover, and technology problems. They think that it might have been easier if it had been directly operated service, but not sure about that. If directly operated, at least they could control the pay scales a little bit better to reduce turnover. Furthermore, it is difficult in a rural area to procure services, First Transit is one of two vendors who would even bid on the job, so they are limited.

## Victor Valley Transit Authority

Agency: Victor Valley Transit

Agency Contact: Simon Herrera, Craig Barnes, Dustin Strandberg

Product Type:

Date of Interview: 12/7/2021

General Questions

1. **Could you describe your service in a few sentences? (e.g., deviated fixed route, demand response, etc.)**

VVTA offers fixed-route and flexible/on-demand transit.
2. **Could you describe your customer base? Seniors, students, low-income, etc.**

Prior to the pandemic there were a lot of students, both college and high school, that used the transit system on a regular basis. However, this has shifted since pandemic; now it is mostly low-income people and those without cars. They are unsure if this will revert to the pre-pandemic composition or if this will be a long-term shift.

Ridership is currently lower than it was a year ago. Prior to the pandemic approximately 48% of riders were students. Victor Valley College still has a lot of online classes, so those students are not fully back. Even the classes that are back in-person are allowing students to miss 75% of the semester in-person. For younger students, parents have gotten used to driving their kids everywhere. VVTA is campaigning for parents to get their kids back on the bus.

1. **Can you give a brief overview of the kinds of data collection activities that you have to undertake for asset management, NTD reporting, grant reporting, etc.?**

All buses have APCs, but they are not NTD certified. The APCs provide stop-level data. T hey are in the process of updating the APCs because they recently changed their AVL provider to Syncromatics (GMV).

The web-based app has passenger load data, but this is not shown on the mobile app. The public-facing apps also show the estimated arrival time of the buses.

Data from the paratransit system is pulled from the manifests.

Technology Questions

1. **What passive data collection technologies do you currently have on your vehicles?**

As mentioned, all buses have APCs. VVTA owns their own buses and Keolis does their maintenance in-house. Battery electric buses do have a telematic (i.e., vehicle health monitoring) system, which passively collects data, that is not included on the CNG buses. All buses have a tablet for pre-trip analysis. Only three buses are set up to collect data automatically. These can send e-mails if issues arise, but they would have to be programmed to do so.

1. **What technologies have you explored but not implemented?**

They have investigated installing telematic systems on the entirety of the bus fleet but do not see this as a feasible option currently. They have had some issues in the past with drivers not following procedures, which has led to mechanical issues (e.g., overheating engines). Given this, they do see a benefit of all buses having telematic systems, but the benefits do not outweigh the costs.

1. **What has been your biggest disappointment with technology in the past?**

They would have purchased new APCs with the CAD/AVL system. There has been a compatibility and support issue with the APCs since they switched to the new AVL. They are currently in the process up updating all the APCs due to this issue.

Capacity Questions

1. **What staff capacity do you have to support technology at your agency?**

In terms of hardware, the most critical person to have is a champion for the technology. This is someone that is trained and can take the lead on diagnosing and repairing the hardware. However, everyone should be provided training on how to program and replace AVLs.

From an administrative point of view, there is someone always monitoring and reviewing the data. There are key people for pulling ridership statistics and other data.

***To accurately collect passive data, you need active monitoring.***

There are currently so many systems that collect data they must decide what and how much data they want because there can be too much data.

1. **Are capacity limitations due to budget or lack of expertise in your geographic area?**

They have not encountered capacity issues. The people suited for these jobs need to be able use Excel and know what data to look for. This can be trained if you have the right person.

A mechanic is a mechanic, so given proper training, they should be able to work on anything on a bus – electric or not.

Cost Questions

1. **Do you find that passive data collection technologies save you money?**

VVTA does not look at passive data collection as a cost-saving tool, aside from maybe telematics. Instead, they look at what provides the most value to VVTA, which is good, accurate data that can be utilized.

They are charged annually on the ITS system (includes APCs) per vehicle.

1. **Do you find data collection and reporting requirements burdensome to the point that you are looking for opportunities to invest in automated capabilities?**

Pulling on-time performance data prior to earlier this year was a manual and tedious process. They have invested in a new program (see below) that has made it much easier to make service changes. They are now using Swiftly, which pulls historical data and allows them to make informed schedule adjustments when needed. This can show a heat map with actual times where the problems are most frequently located.

They would like to have an automated fuel system. There is currently a lot of room for error in entering all the needed information.

Data Sharing

1. **Are you able to share the data you collect with researchers?**

They have not had any issues with sharing data as it is their operating data with which can do what they want.

The biggest push-back was with the battery electric buses over concern that other people were going to gather information without the VVTA’s approval. As an example. It took multiple meetings to get TransTrack access to the data.