STATEWIDE DATA STANDARDS TO SUPPORT CURRENT AND FUTURE STRATEGIC PUBLIC TRANSIT INVESTMENT

Interim Report

Task #1: Perform a comprehensive review on the current state of the art and state of the practice of transit ridership data

PROJECT SPR 803

STATEWIDE DATA STANDARDS TO SUPPORT CURRENT AND FUTURE STRATEGIC PUBLIC TRANSIT INVESTMENT

Interim Report

Task #1: Perform a comprehensive review on the current state of the art and state of the practice of transit ridership data

PROJECT SPR 803

by

J. David Porter, PhD, Phillip Carleton, Sylvan Hoover

School of Mechanical, Industrial and Manufacturing Engineering
Oregon State University
Corvallis, Oregon 97331

for

Oregon Department of Transportation
Research Section
555 13th Street NE, Suite 1
Salem OR 97301

December 2016

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Statewide Data Standards to Support Current and Future Strategic Public Transit Investment		5. Report Date December 2016 6. Performing Organization Code
7. Author(s) J. David Porter, PhD, Phillip Carleton, and Sylvan Hoover School of Mechanical and Industrial Engineering 204 Rogers Hall Oregon State University, Corvallis, OR 97331		8. Performing Organization Report No.
9. Performing Organization Name and Address Oregon Department of Transportation Research Section 555 13th Street NE, Suite 1 Salem, OR 97301		10. Work Unit No. (TRAIS) 11. Contract or Grant No. SPR 803
12. Sponsoring Agency Name and Address Oregon Dept. of Transportation Research Section and Federal Highway Admin. 555 13th Street NE, Suite 1 400 Seventh Street, SW Salem, OR 97301 Washington, DC 20590-0003		13. Type of Report and Period Covered Interim Report 14. Sponsoring Agency Code
15. Supplementary Notes		
16. Abstract		
17. Key Words Ridership; data standard	18. Distribution Statement Copies available from NTIS, and online at http://www.oregon.gov/ODOT/TD/TP_RES/	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages XXX 22. Price

Technical Report Form DOT F 1700.7 (8-72) Reproduction of completed page authorized Printed on recycled paper

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	millimeters squared	mm^2
ft^2	square feet	0.093	meters squared	m^2
yd^2	square yards	0.836	meters squared	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	kilometers squared	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	ml
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	meters cubed	m^3
yd^3	cubic yards	0.765	meters cubed	m^3
	NOTE: Volumes greater than 1000 L shall be shown in m^3.			
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams	Mg
		TEMPERATURE (exact)		
°F	Fahrenheit	(F-32)/1.8	Celsius	°C

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	millimeters squared	0.0016	square inches	in^2
m^2	meters squared	10.764	square feet	ft^2
m^2	meters squared	1.196	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	kilometers squared	0.386	square miles	mi^2
		VOLUME		
ml	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	meters cubed	35.315	cubic feet	ft^3
m^3	meters cubed	1.308	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.205	pounds	lb
Mg	megagrams	1.102	short tons (2000 lb)	T
		TEMPERATURE (exact)		
°C	Celsius	1.8C+32	Fahrenheit	°F

*SI is the symbol for the International System of Measurement

ACKNOWLEDGEMENTS

[TBD]

DISCLAIMER

This document is disseminated under the sponsorship of the Oregon Department of Transportation and the United States Department of Transportation in the interest of information exchange. The State of Oregon and the United States Government assume no liability of its contents or use thereof.

The contents of this report reflect the view of the authors who are solely responsible for the facts and accuracy of the material presented. The contents do not necessarily reflect the official views of the Oregon Department of Transportation or the United States Department of Transportation.

The State of Oregon and the United States Government do not endorse products of manufacturers. Trademarks or manufacturers' names appear herein only because they are considered essential to the object of this document.

This report does not constitute a standard, specification, or regulation.

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

1.0 EXECUTIVE SUMMARY

As the overall population of the United States grows, the most rapid growth is seen in urban and suburban areas. Following this same trend, the demand for public transit services is increasing across the nation. Transit service providers, state transportation agencies, metropolitan planning organizations, technology vendors, and other stakeholders will be expected to create effective solutions to address the continuing challenges created by this growth trend. Furthermore, these solutions will have to be based on decision making driven by accurate, real-world data.

Despite advancements in the analysis and visualization of the supply side, large information gaps remain in the understanding of the demands imposed on the transit system. In particular, the ridership data needed to set policy, develop plans, and prioritize investments in the State of Oregon is often not easily available and/or not in a useful standard format. This is mainly due to the fact that transit agencies vary widely in their ability to provide sufficient quantities of high-quality ridership data, and what can be provided is too often of little use due to a high level of aggregation, broad scope, sparsity, errors, and lack of standardization.

The inconsistencies in the availability, format, and quality of ridership data make it very difficult (if not impossible) for ODOT and entities with an interest in multi-agency transit networks to make effective and informed decisions to address challenges presented by growing transit demands. Therefore, the objective of this project is to develop a comprehensive (yet flexible) ***public transit ridership data standard***. This standard will be designed to improve the processes of ridership data collection, management, reporting, and analysis. An additional objective of this project is to support the core functionality of the standard through the development of open-source, web-based software tools.

This interim report documents the findings of a comprehensive review performed on the current state of the art and state of the practice of transit ridership data. Understating the research on ridership data collection methods, opportunities, and limitations is an important first step in the process of using transit ridership data. While there are existing federal requirements for ridership data reporting, these requirements are relatively lenient with regards to the level of resolution, quality, and detail required. As it may be expected, this fact leads to the use of a wide range of data collection methods by transit agencies. The state of the practice review was conducted to gain an understanding of current practices for ridership data among Oregon transit agencies and a select number of transit agencies in the United States and abroad.

SUMMARY OF THE STATE OF THE ART REVIEW

The literature synthesized in the review of the state of the art reveals that different areas of ridership data collection, analysis, and accessibility are being researched. However, they do not follow a structured or concise approach.

The state of data collection in present literature seeks to evolve accurate means of collecting ridership data by leveraging combinations of proven methods with rapidly advancing sensing technologies and increasing computational power. These methods include numerous optical, physical, electromagnetic, and analytic options. While there exist promising developments in improving ridership data collection, no one method has yet proven to be the de facto solution or dominant strategy. Evidence also exists that the methods to collect ridership data are expanding, and with each new advancement, the structure of the data collected may change. Therefore, the methods to manage and analyze ridership data must also advance to guarantee access to a breadth of data in usable formats, thus facilitating their broad application.

Although the methods for distributing and making ridership data accessible are still new, development of a standard is key for proliferation. The growth experienced by GTFS is a good example to follow. Standardizing formats (e.g., CSV, XML, etc.) and content across formats (e.g., passenger counts, fare types, etc.) will support the development of advances in the collection, analysis, and distribution of ridership data.

SUMMARY OF THE STATE OF THE PRACTICE REVIEW

The main instrument to collect information to prepare the state of the practice review was an online questionnaire. The instrument is composed of a total of 33 questions, most of which are specifically designed to elucidate current practices related to ridership data. In total, 147 individuals identified as having a role with a transit agency/service were directly invited to respond to the online questionnaire. In the end, 53 responses were considered reasonably complete and valid giving a response rate of 36%. From the 53 responses, 47 were provided by transit agencies/services that operate in Oregon and six were provided by transit agencies/services that operate elsewhere in the United States or overseas.

State of Ridership Data Quality and Collection

Several of the responses that centered around ridership data quality reveal the current state of issues such as ridership data resolution and frequency. For Oregon transit agency/service respondents, the majority of ridership data is collected for Bus and Commuter bus modes at low resolution levels (i.e., route and trip levels), with a more moderate amount of higher resolution stop-level data collection. Also, route- and system-level ridership data was reported as mostly available on a daily basis, while fewer reported having stop-level data on a daily basis. A ridership data standard would likely benefit those currently collecting high-resolution data frequently, and possibly motivate and direct those agencies who would like to improve the quality of their collected ridership data.

In present practice, the lack of automated ridership data collection means was clear, especially for smaller, rural, and/or non-fixed route agencies/services. The majority of agencies and service providers still rely on manual ridership data collection and cite presently unaddressed hurdles which prevent them from adopting automated means. Again, a ridership data standard would help provide guidance and direction to agencies interested in advancing their data collection practices and methods or in updating their data collection technology.

State of Ridership Data Analysis

In present practice, the degrees of ridership data analysis are highly correlated with the size of the fleet and the availability of resources at the different transit agencies. The development of a ridership data standard would ease the creation of common analysis tools and processes which will benefit transit agencies of all sizes, making it easier to share ridership data with interested parties (much as occurred with GTFS). The development of shared analysis tools should reduce barriers for smaller agencies to make effective use of their own ridership data.

State of Ridership Data Availability

The state of data availability and management closely mirrors that of data analysis. There does not yet exist a widely used ridership data standard. Therefore, collection of responses from agencies was undertaken with the understanding that the distribution method of ridership data employed by each agency may be unique. Increasing the number of transit agencies releasing their data will benefit analysis and planning, and providing a data standard for distribution would aid employment of the data.

CONCLUSIONS AND OPPORTUNITIES FOR FUTURE WORK

The review of the current state of ridership data reveals clear opportunities for improvement and growth. The amount and resolution of ridership data collected across agencies varies, and many still collect data using infrequent and/or inaccurate means.

Providing smaller agencies with approachable, affordable, and useful means to collect, manage, and analyze ridership data would address one of the clearest deficits revealed by responses. At present, the varying types of services provided and the limited resources available to smaller agencies appear to have slowed the adoption of precision automated ridership data collection, making access to ridership data further impeded. The existing literature shows many opportunities for advancing ridership data, with promise from new technologies and methods and guidance for implementing proven processes.

A significant opportunity for future work is to develop a unified platform on which data are collected and analyzed to streamline progress in the field. Much as open approaches for GTFS have spurred the development of transit scheduling and rider planning, an open approach for ridership data may yield similar results. The evolution of GTFS informs that a similar evolution will be required for a ridership data standard. The framework for such an evolution and the continued support and maintenance of the standard will be vital to its success. The standard must be flexible and simple enough to be used across a wide spectrum of agencies, including those still relying on sampled manual collection as well as those with fully automated stop-level data collection means.

2.0 INTRODUCTION

As the overall population of the United States grows, the most rapid growth is seen in urban and suburban areas. Following this same trend, the demand for public transit services is increasing across the nation. According to the Bureau of Transportation Statistics (2016), transit demand (as measured through passenger miles traveled) increased by over 8% in the United States between 2010 and 2014. Transit service providers, state transportation agencies, metropolitan planning organizations, technology vendors, and other stakeholders will be expected to create effective solutions to address the continuing challenges created by this growth trend. Furthermore, these solutions will have to be based on decision making driven by accurate, real-world data.

The Oregon Department of Transportation (ODOT) is fully aware that the performance analysis and service planning of transit systems hinge on data availability and reliability and has been forward-thinking in developing new tools to simplify and standardize the understanding and analysis of the complex transit service supply. One example is the open-source, web-based *Transit Network Analysis Software Tool* which fuses the General Transit Feed Specification (GTFS) data feeds of over 60 Oregon fixed-route transit service providers with several data sources including census, Park & Ride, employment, and Title VI (Porter et al., 2016).

Despite advancements in the analysis and visualization of the supply side, large information gaps remain in the understanding of the demands imposed on the transit system. In particular, the ridership data needed to set policy, develop plans, and prioritize investments in the State of Oregon is often not easily available and/or not in a useful standard format. This is mainly due to the fact that transit agencies vary widely in their ability to provide sufficient quantities of high-quality ridership data, and what can be provided is too often of little use due to a high level of aggregation, broad scope, sparsity, errors, and lack of standardization.

2.1 PROJECT OBJECTIVE

The inconsistencies in the availability and quality of ridership data make it very difficult (if not impossible) for ODOT and other transit stakeholders to make effective and informed decisions to address challenges presented by growing transit demands. Therefore, the objective of this project is to develop a comprehensive (yet flexible) ***public transit ridership data standard***. This standard will be designed to improve the processes of ridership data collection, management, reporting, and analysis. An additional objective of this project is to support the core functionality of the standard through the development of open-source, web-based software tools.

2.2 REPORT ORGANIZATION

The remainder of this report is organized as follows. Chapter 3 presents a review of the state of the art with respect to transit ridership data. Chapter 4 discusses the state of the practice regarding the collection, storage, sharing, reporting, and analysis of ridership data currently utilized by a sample of Oregon and out-of-state transit agencies/services. Finally, Chapter 5 presents the conclusions, future opportunities, and implications for the development of a ***public transit ridership data standard***.

3.0 STATE OF THE ART REVIEW

Understating the research on ridership data collection methods, opportunities, and limitations is an important first step in the process of using transit ridership data. Also, federal funding to transit agencies in the United States is can be impacted by ridership; therefore, the vast majority of agencies must employ some form of ridership data collection. While there are existing federal requirements for ridership data reporting, these requirements are relatively lenient with regards to the level of resolution, quality, and detail required (Federal Transit Administration, 2016). As it may be expected, this fact leads to the use of a wide range of data collection methods by transit agencies. Many transit agencies have a strong desire to go beyond the basic reporting requirements and collect more detailed passenger counts, as demand estimation and forecasting is a significant component of their planning processes (Boyle, 1998).

The state of art review is organized into three sections. Section 3.1 examines current research in ridership data collection methods, whereas Section 3.2 examines current research in methods for employing ridership data to improve transit understanding and service. Section 3.3 discusses the current state of open availability of various types of public transportation data. Finally, Section 3.4 presents a summary of the main findings of the state of the art review.

3.1 DEVELOPMENTS IN RIDERSHIP DATA COLLECTION

The research in transit ridership data collection is fairly well established, as it spans several decades and has maintained its relevancy. Prior work has been reported in many different areas including a comprehensive review of already established automatic passenger counter (APC) practices (Hodges, 1985); an investigation of the use of off-the-shelf pressure sensor mats and software to test improved passenger counting and classification (Greneker et al., 1996); and more recently, the utilization of on-board security cameras to validate trip counts obtained through other methods (Kirby, 2016). While the specific technologies and methods employed may vary among transit agencies, ridership data is generally derived from passenger counts determined by some combination of manual counting, farebox data, and APCs (Boyle, 1998).

Strathman and Hopper (1991) noted that APCs hold significant advantages over manual counting in that this technology can produce greater quantities of more disaggregate data that can be more readily available, at lower cost, and with improved accuracy. Boyle (1998) also noted that manual counting is very labor intensive and time consuming, thus introducing random errors and a “burnout” factor that may affect the reliability of the data.

While APC systems can address the issues pertaining to manual data collection, inconsistencies in their operation often require larger amounts of data to be screened out, thus requiring larger samples sizes and standard plans for sampling and validation (Strathman & Hopper, 1991). For instance, Kimpel et al. (2003) found that up to 35% of APC data is rejected. These known APC issues contribute to some agencies being satisfied with their manual counting procedures and not seeing enough incentive to move away from manual counting (Boyle, 1998). Most of the literature, however, focuses on the automated methods. Therefore, the following two sections highlight seminal and recent literature on ridership collection methods based on APCs and fareboxes.

3.1.1 Automatic Passenger Counters

Advents in technology have allowed for a transition in transit from predominantly manual counting to a greatly increased adoption of APC technology (Boyle, 1998). APCs utilize a host of technologies to detect passenger boardings and alightings, including infrared (IR) light beam cells, laser scanners, IR cameras, piezoelectric mats, microwave radars, switching mats, video cameras, Bluetooth and Wi-Fi Sensors, radio frequency identification (RFID) "smart cards," and vehicle air suspension pressure sensors. Table 3-1 summarizes these technologies as they appear in the literature.

APCs that utilize IR beam technology and switching mats have long been studied (Greneker et al., 1996; Hodges, 1985), and several sources in the literature provide guidance to practitioners about their implementation and use (Boyle, 2008; Federal Transit Administration, 2016; Fihn and Finndahl, 2011). Although these two types of APCs are considered old technologies, they are relatively inexpensive (i.e., less than five hundred dollars) and can provide finer levels of detail on the data collected when compared to manual counts (Bauer et al., 2011). However, these types of APCs cannot resolve certain instances accurately (e.g., simultaneous boardings or a mother carrying a child) and require regular calibration and validation which can be a challenge for transit agencies with limited resources (Boyle, 2008; Federal Transit Administration, 2016). A validation case study performed by TriMet in Portland, OR showed that the data collected with IR camera-based APC counters requires post processing and validation to address over- and under-counting (Strathman et al., 2005).

Applications of APCs that utilize video technology for data collection are now taking advantage of recently developed video data processing techniques. For example, a project currently underway with the North Carolina Department of Transportation is investigating the use of new processing algorithms to count passenger trips collected with pre-existing transit vehicle security cameras (Kirby, 2016). Another study conducted in China used cameras mounted above the doors of buses to count crowds of passengers attempting to simultaneously enter a bus. Clustering algorithms designed to separate out individual passengers from a crowd by tracking feature trajectories were employed to analyze the video data and produced passenger counts with an accuracy of 96.5% (Yang et al., 2010). Yahiaoui et al. (2010) showed that using two video cameras at different angles to produce a three-dimensional image of bus entryways could yield high accuracy (i.e., 97%) and reliability in passenger counts. García-Bunster and Torres-Torriti (2008) suggested an alternative approach which uses a camera mounted at a bus stop coupled with a specialized density-based algorithm to count the passengers as they wait on the bus. The results show that passengers could be detected with an accuracy of 94.1% with the proposed method.

A variety of wireless technologies have also been used as the basis of operation of APCs. A case study conducted in Seattle demonstrated the feasibility of using Bluetooth-based and Wi-Fi-based APCs combined with GPS data to link boarding and alighting data to specific individuals (Dunlap et al., 2016). While this study showed that origin-destination estimation via this data collection approach is quite feasible, there are still limitations which prohibit its successful implementation. More specifically, since not all passengers are likely to have the required technology so that they can be detected, it would not be effective at providing complete and accurate counts. Kostakos et al. (2013) reached similar conclusions in their study, which also used Bluetooth-based APCs (i.e., only 12.8% of passengers had discoverable devices), and they also particularly noted the privacy concerns that arise with the use of wireless technologies.

Table 3-1: APC and AFC Technologies Reviewed in Literature

	TECHNOLOGY							
ARTICLE	**Pressure Sensitive Mats**	**IR Light Beam**	**IR Camera**	**Video Camera**	**Bluetooth**	**WiFi**	**RFID Smart Cards**	**Air Suspension Pressure**
Dunlap et al. (2016)					X	X		
Kirby (2016)				X				
Kostakos et al. (2013)					X		X	
Kotz et al. (2015)		X		X				X
Kutti (2013)	X	X	X	X				
Peterson et al. (2013)							X	
Bauer et al. (2011)	X	X						
Oberli et al. (2010)							X	
Yahiaoui et al. (2010)				X				
Yang et al. (2010)				X				
Bunster and Torriti (2008)				X				
Strathman (2005)			X					
Boyle (1998)	X	X						
Greneker et al. (1996)	X							
Strathman (1989)								
Hodges (1985)	X	X						

A novel approach to passenger counting, which monitors the air pressure of the ride suspension system of a transit vehicle, was recently proposed (Kotz et al., 2015). This study found that inferring the vehicle's mass through pressure changes and translating this result into a passenger load was feasible and produced passenger counts that were 97.6% accurate. The researchers also noted that this method provided opportunities to generate additional information such as passenger load distributions throughout the vehicle cabin. The method was, however, sensitive to deviations from the assumed average passenger mass and to bus kneeling events.

Newly emerging technologies and methods will likely play a role in in the advancement of APC-based passenger data collection approaches, even if they do not replace the tried and tested current methods (i.e., IR beam, IR camera, and pressure sensitive mats). In particular, the rapidly advancing research on autonomous vehicle technology is making a compelling case for automation in public transit (Polzin, 2016). In probable future scenarios where public transit vehicles become fully autonomous, APCs will need to play an increased roll in ridership data collection.

3.1.2 Farebox Counting

Passenger counts are often derived from farebox data. With the increasing adoption of automated fare collection (AFC) through the use of RFID enabled "smart cards," transit agencies have been able to collect richer data than APCs often allow. For instance, in addition to stop-level boardings and alightings, smart cards can hold passenger specific information, such as transfer status, rider ID, fare type, and rider category (Pelletier et al., 2011).

There have been many advantages of adopting AFC systems noted in the literature, including reduced boarding times, reduced driver workload, flexible and creative fare structures, and reduced costs (Pelletier et al., 2011). In a field test conducted by Peterson (2013), it was found that using RFID smart cards with a medium detection range allowed recording passenger boardings with an average accuracy of 88%. However, after a lifecycle cost analysis was performed, it was found that the economic advantages of such a system would be dependent on ridership levels and the fare structures in place for different rider types (i.e., free student or discounted fare riders vs. full-fare riders). Similarly, Oberli et al. (2010) found that an average accuracy of up to 91% is achievable with RFID smart cards, but that accurate detection was highly dependent on the specific location of the smart card, such as being held by the passenger in their hand versus being located in a wallet or a backpack.

3.2 INNOVATIVE USES AND ANALYSIS OF RIDERSHIP DATA

The goal of ridership data collection and dissemination is to be able to apply that data. Application examples include resource optimization, meeting reporting requirements, and providing advocacy resources. However, the depth of data available may limit the scope of the analyses and applications that are feasible. Most of the prior research discussed in this section focuses on extracting useful synthesizes from limited available data through prediction and extrapolation. Something to notice while reading through the discussed research is the somewhat sporadic areas of study. Current research consistently looks at datasets of local or regional transit agencies, not attracting input from transit agencies outside of the geographic focus of the study. The nature of the ridership data is unique to the study, so generalization is not often discussed nor is it easily applied between different data sets.

Predicting ridership data from external sources provides a means of estimating ridership in the absence of empirically collected data. Chu et al. (2004) developed mathematical constraints to improve the estimation capabilities of prior models for average weekday boardings. Their results showed that the newly developed constraints improved the accuracy of the estimates by 54%. Chu et al. (2006) further expanded their model by developing a framework released as the Transit Boardings Estimation and Simulation Tool (T-BEST). Dill et al. (2013) utilized transit service, land use, and socio-demographic characteristics to predict stop-level ridership using data from three Oregon transit agencies (i.e., TriMet, Lane Transit District, and Rogue Valley Transit District). Their regression model worked well in predicting stop-level ridership in more urban environments, with transit service characteristics (i.e., transfer stops, transit centers, proximity to other stops, Park & Ride services, service headways, and service coverage hours) being the category of independent variables having the most impact.

Development of models to predict behavior can compensate for a lack of data. With limited data at hand, the need to extrapolate models from the data becomes vital. With the most standard digitally deployed technologies being automated fare collection (AFC) and automated vehicle location (AVL), a desire exists to learn more about ridership from that data. Nassir et al. (2011) attempted to model origins and destinations of riders using AFC and AVL data. The model was validated against APC data, with 98% of model transactions matching APC recordings. By validating models, additional data can be derived from what is collected.

Attempts to extrapolate data are evolving rapidly as the means of data collection increase. Transit Cooperative Research Program (TCRP) Synthesis 66 provided a summary of ridership forecasting techniques as of 2006 (Boyle, 2006). While this TCRP synthesis report is now over a decade old, the main challenges identified are still unresolved: lack of data, inconsistent collection methods across agencies, and an evolving modeling base of knowledge. Barabino et al. (2014) proposed a framework to be applied to buses to allow transit agencies to better manage their data and use them beneficially. The proposed framework provides for ingesting APC data, cleaning and verifying the data, and presenting the data in a profile beneficial to transit agencies.

Ensuring validity is one of the most significant challenges when analyzing large volumes of stop-level ridership data. Therefore, quality assurance (QA) methods are key to ensuring that the data being ingested for reporting and modeling is valid. Saavedra (2010) applied a new approach to QA where 612,000 stop-level records were examined, and as many as 85,680 (i.e., 14%) failed the QA measures. The fault in the data was associated largely with poor passenger balancing algorithms (i.e., processes to correct APC counting errors) or inaccurate APC equipment. Poor quality schedule data contributed to weak AVL-APC data. A core concern related to ridership data was revealed in this study, i.e., ridership data collection methods cannot be solely relied upon as a data source. Reliable means of QA are also necessary to ensure proper reporting and planning.

Once ridership data has a high assurance of accuracy, applications of the data may benefit transit services. The gains seen from the analysis of ridership data include improved efficiency, more accurate reporting, and new understandings of ridership. Reddy et al. (2009) detail how New York City Transit (NYCT) developed analyses based on an AFC system to infer origin-destination data and to determine mile calculations using schedule-driven, shortest-path algorithms. NYCT successfully retained a Federal Transit Administration-approved sampling methodology for Section 15 reporting while improving understanding of system use.

Like NYCT, the Utah Transit Authority sought to resolve new transit data from existing collection systems. Fayyaz et al. (2016) developed a model to predict dwell time (i.e., the amount of time a bus is at a bus stop) by analyzing other variables such as time used by cash payments and prepaid passes. They developed a genetic algorithm and regression-based modeling approach and validated their model empirically on a single bus route. They also proposed that their model can be transferred to other systems equipped with APCs to improve service optimization and performance assessments.

To see what gains are possible with sound analysis of ridership data, application studies evolve perceptions of functioning transportation systems. Changes to transit service (or changes to the environment surrounding a service) can impact the utilization of public transportation for both existing and potential riders. To determine how ridership is gained or lost, Trepanier (2010) used smart card fare collection data from "Société de transport de l'Outaouais" with more than 80% of riders using their smart cards for fare payment. Trepanier (2009) used the same data source to show the effectiveness of using smart card data "to measure transit supply and demand indicators." These gains and losses can be used to justify or disprove the analysis and applications discussed above.

The prediction, analysis, and application of ridership data will continue to evolve. Key to this work will be valid and consistent data collected through increasingly automated means.

3.3 ADVANCEMENTS IN TRANSPORTATION DATA STANDARDS AND OPEN DATA

The movement for the public availability of transit data is occurring globally, but challenges faced in providing data to the public slow expansion. To evolve data availability, effectively published standards should include supporting communities of developers to further both the applicability and the adoption of open transit data standards.

The quintessential open transit data standard is the General Transit Feed Specification (GTFS). Originally established as a partnership between Google and TriMet, GTFS is a transit data specification with global adoption. Two key components to its success are its immediate utility through Google's Transit and its ongoing adaptation initially centered on an online Google Groups forum (moderated by Google) and later on GitHub. The access the public had to Google's Transit (Google Maps Transit, 2016) for planning their public transit trips incentivized transit agencies to provide data about their network in a format compliant with GTFS. The ongoing adaptation though publicly accessible forums has allowed transit agencies globally to contribute to the standard in ways that allow further adoption by diverse agencies.

An example of both the adaptation and adoption of GTFS is a project conducted in Mexico City (Eros et al., 2014). Mexico City's transit system included vehicle types and schedules not originally accommodated by GTFS. However, by engaging the online GTFS community and implementing the necessary adaptations for their network, they were able to apply GTFS.

Two transit agencies recently discussed their approaches to working with and providing open transit data. Chandesris and Remy (2016) investigated applying open transit data within the French transit system. Their primary challenges were associating the collected data with human activities, achieving benefits by providing and applying the data in real time, and interweaving what has been learned from working with the rail system to other transit modes. The Massachusetts Bay Transit Authority and the Massachusetts Department of Transportation are developing methods to better provide their data to the public (Paget-Seekins & Tribone, 2016). Their primary challenges were aggregating data to protect privacy, choosing the best protocols for data distribution, assuring data quality, and the specific logistical decisions associated with large datasets.

A comprehensive review of the current state of open data within transit agencies globally can be found in TCRP Synthesis 115 (Schweiger, 2015). In TCRP Synthesis 115, four key lessons were learned in the development of open transit data:

- Data quality and accuracy are critical to the success of an open data program.
- Open data are not free.
- Recognize that opening data will create changes within and external to the agency.
- Engagement and developing relationships with developers is key to success as well.

Applying the lessons of other open transit data standards will be key to the success of an open ridership data standard.

3.4 STATE OF ART SUMMARY

The literature synthesized in this chapter reveals that different areas of ridership data collection, analysis, and accessibility are being researched. However, they do not follow a structured or concise approach.

The state of data collection in present literature seeks to evolve accurate means of collecting ridership data by leveraging combinations of proven methods with rapidly advancing sensing technologies and increasing computational power. These methods include numerous optical, physical, electromagnetic, and analytic options. While there exist promising developments in improving ridership data collection, no one method has yet proven to be the de facto solution or dominant strategy. Evidence also exists that the methods to collect ridership data are expanding, and with each new advancement, the structure of the data collected may change. Therefore, the methods to manage and analyze ridership data must also advance to guarantee access to a breadth of data in usable formats, thus facilitating their broad application.

Although the methods for distributing and making ridership data accessible are still new, development of a standard is key for proliferation. The growth experienced by GTFS is a good example to follow. Standardizing formats (e.g., CSV, XML, etc.) and content across formats (e.g., passenger counts, fare types, etc.) will support the development of advances in the collection, analysis, and distribution of ridership data.

4.0 STATE OF THE PRACTICE REVIEW

A state of the practice review was conducted to gain an understanding of current practices for ridership data among Oregon transit agencies and a select number of transit agencies in the United States and abroad. The main instrument to collect information to prepare this review was an online questionnaire developed by the research team in consultation with the ODOT Technical Advisory Committee (TAC). The online questionnaire was designed, distributed, and analyzed through the web-based survey platform Qualtrics. After all the responses to the online questionnaire were received, additional data was gathered through direct follow up with questionnaire respondents and through general web searches. The main objective of conducting a state of the practice review was to inform and direct the development of the requirements, structures, and functions of the ***public transit ridership data standard***.

In this chapter, the structure, distribution, and results of 33 questions presented to transit agencies both in Oregon and out-of-state are presented and analyzed. First, the structure and synthesis of thought that informed the development of the questionnaire is presented. Second, the method used to identify target agencies, contact the agencies, and provide agencies access to the questionnaire is discussed. Next, the results of the questionnaire are analyzed to gain a better understanding of the practices for ridership data. Finally, a summary of the findings of the state of the practice review is presented.

4.1 ONLINE QUESTIONNAIRE

4.1.1 Structure of the Questionnaire

The TCRP Synthesis 77 report (Boyle, 2008) was used as the foundation for the questions included in the online questionnaire. Then, its scope and content were refined and tailored to the specific needs of this project through an interactive process with the TAC.

The final online questionnaire distributed to transit agencies is included in Appendix A. The instrument is composed of a total of 33 questions. The first question asks for the name of the transit agency and the service being reported on, while the last five questions collect contact information about the respondent to facilitate a follow up (if needed). The remaining 27 questions are specifically designed to elucidate current practices related to ridership data while balancing the desire for rich and complete information with the need to achieve clarity and brevity to accommodate the intended respondents.

As depicted in Figure 4-1, the online questionnaire centered on the main components of ridership data (i.e., collection, management, analysis, and use) with a focus on the specific tools and methods utilized in practice. The goal was to compose a complete picture of the lifecycle of ridership data as it exists in practice among the transit agencies contacted, which vary in size and availability of resources to dedicate to these tasks.

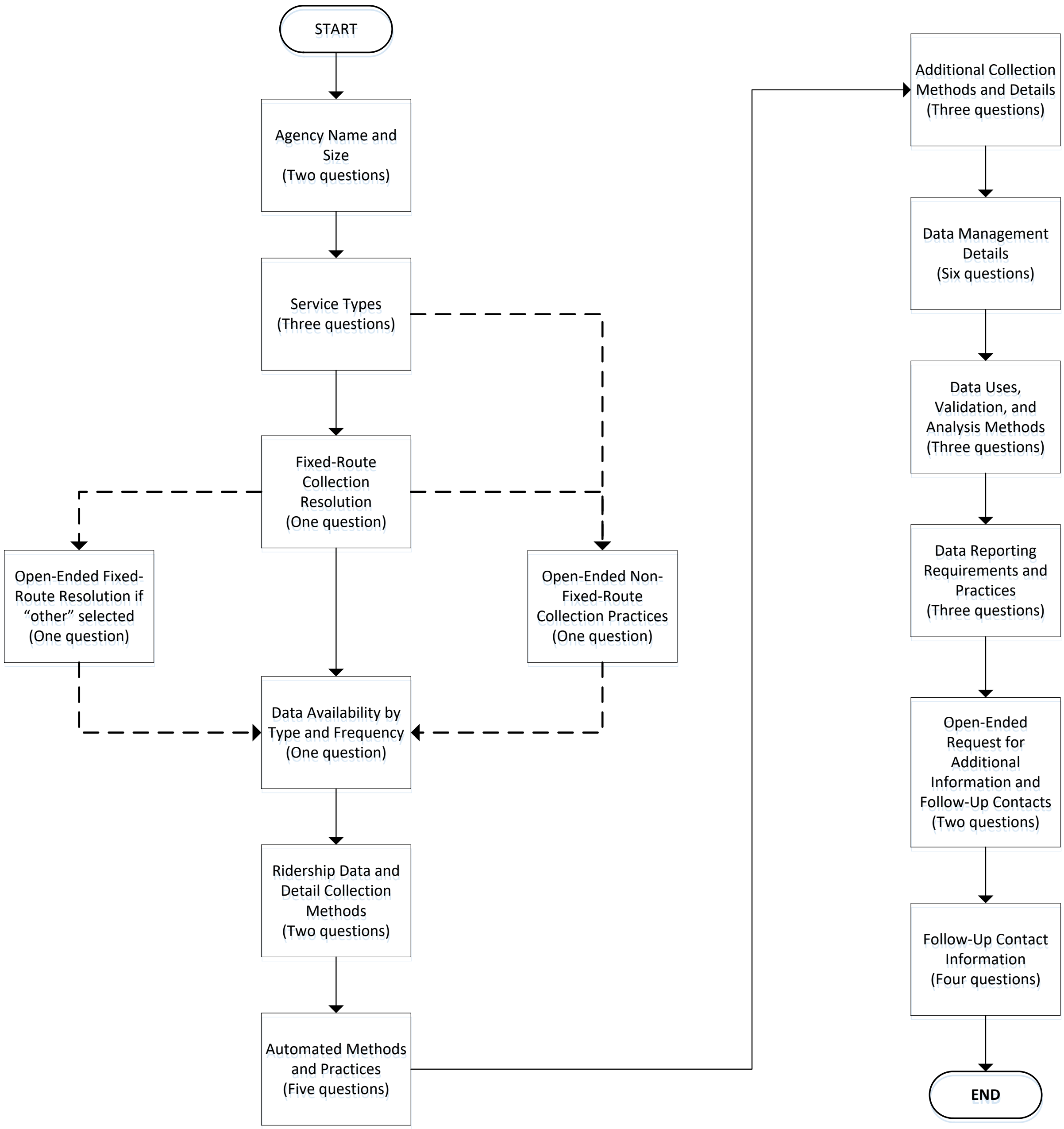

Figure 4-1: Flow and Logic Structure of the Questionnaire

4.1.2 Process to Identify Respondents

In coordination with the TAC, a total of 138 Oregon transit agencies/services were identified as potential respondents to the online questionnaire. An initial list of contacts from each of the six transit regions (i.e., Region 1, Region 2a, Region 2b, Region 3, Region 4, and Region 5) was generated from the Oregon transit service list spreadsheet. The initial list was expanded with additional contacts for each transit region provided by the Oregon regional transit coordinators (RTCs).

Nine out-of-state transit agencies were also identified as potential respondents. These out-of-state transit agencies were targeted because of their recognition of excellent performance as indicated by performance metrics calculated by a third party using data available through the National Transportation Database (NTD). Six of the out-of-state transit agencies operate in the United States, and three operate overseas.

Finally, a broad invitation was posted on Google-hosted Transit Developers forum. It is important to mention that there were no preconceived notions as to how many potential respondents could be reached through this outlet. However, this forum has many participants whose roles relate to public transit, so it was also targeted as a potential source for participants.

4.1.3 Inviting Respondents

In total, 147 individuals identified as having a role with a transit agency/service were directly invited to respond to the online questionnaire.

Potential respondents in each of the six Oregon transit regions were invited to participate via a personalized email message sent through the Qualtrics platform (see Appendix B). For the out-of-state transit agencies, a specific contact person was identified (if possible) and personalized email invitations were also sent to these individuals through the Qualtrics platform. Any updated out-of-state contacts were sent a personalized invitation via regular email.

Finally, an anonymous link alongside an explanatory post was submitted to the Google Group Transit Developers, an active forum for those involved with transit data and software development.

4.1.4 Final Number of Respondents

A total of 66 transit agencies/services responded to the online questionnaire with different levels of completion and validity. From these, 53 responses were considered reasonably complete and valid giving a response rate of 36%. Out the 53 responses, 47 were provided by the Oregon transit agencies/services listed in Appendix C. The geographical location of these 47 Oregon-based transit agencies/services is depicted in Figure 4-2.

The out-of-state transit agencies/services that responded to the online questionnaire are Blacksburg (Virginia) Transit, RTC of Southern Nevada, San Diego Metropolitan Transit System, King County Metro (Seattle, WA), Community Transit (Snohomish County, WA), and Transport of London. An additional 13 responses were received through an anonymous link which can be understood as stemming from the posting to the Google group forum. It is important to note that the responses received from users of the Google group forum were not accounted for in the calculation of the response rate of the online questionnaire.

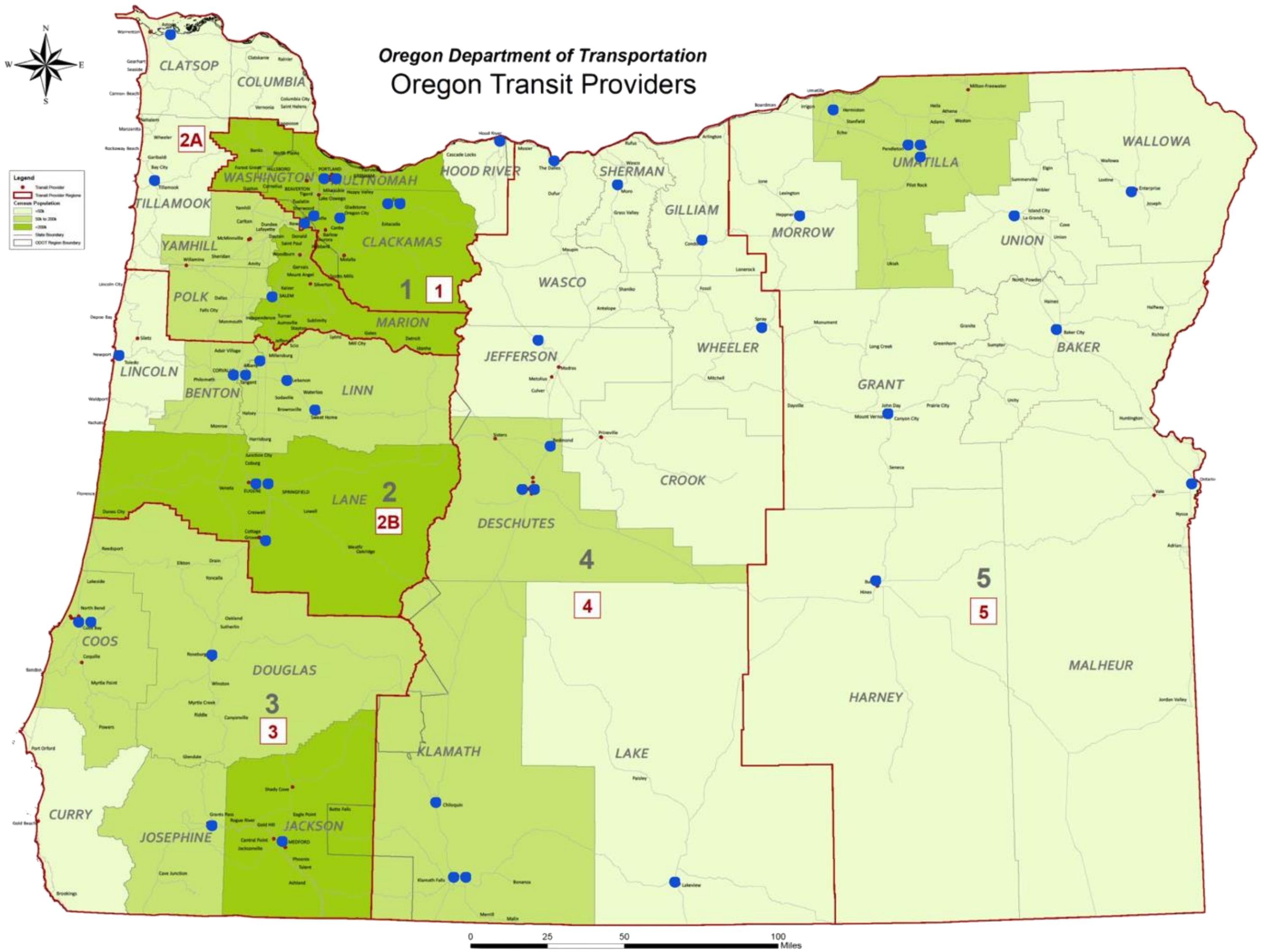

Figure 4-2: Map of Oregon Transit Agency Responses

4.2 RESULTS

When interpreting the results presented in this section, it is important to note that while there were 53 questionnaires that were considered reasonably complete and valid, many individual questions had a much smaller response count. Therefore, the corresponding response percentages reflect only those respondents that were considered as having answered the question appropriately. Furthermore, many text-based answers had high variability. Often these open-ended questions were not addressed as expected, and only the meaningful responses were included in the analysis. Occasionally (and only in very clear cases), certain responses were assumed if the context of other questions made the intended answer obvious.

4.2.1 Size and Composition of Fleet

A total of 47 transit agencies/services operating in Oregon reported on the size and composition of their fleets. Regarding fleet size, the smallest is one vehicle (i.e., Burns Paiute – Tribal Transit Services) and the largest is composed of 1,052 vehicles (i.e., TriMet).

As depicted in Figure 4-3, 39 transit agencies/services (i.e., 83%) own all the vehicles they operate. In contrast, four transit agencies/services (i.e., 8.5%) contract all the vehicles in their fleet. Four transit agencies/services (i.e., 8.5%) reported operating a mixed fleet.

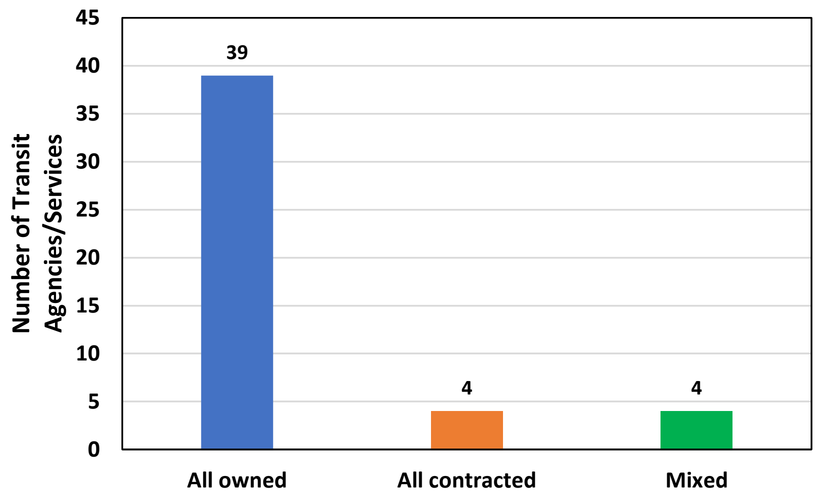

Figure 4-3: Composition of Fleet – Transit Agencies/Services in Oregon

The data depicted in Figure 4-4 shows that 45 transit agencies/services (i.e., 95.7%) have a fleet size between one and 80 vehicles. Only two transit agencies/services in Oregon (i.e., 4.3%) have a fleet size larger than 80 vehicles.

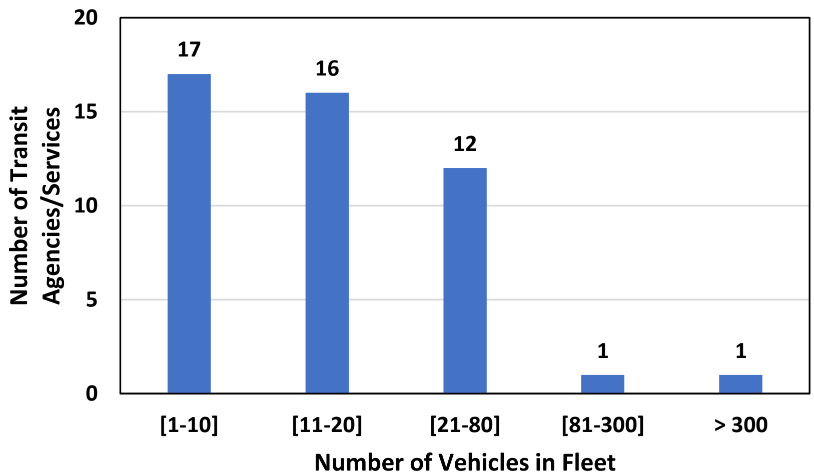

Figure 4-4: Distribution of Fleet Size – Transit Agencies/Services in Oregon

Six out-of-state transit agencies/services reported on the size and composition of their fleets. Their fleet sizes range from 66 to 9,400 vehicles. As depicted in Figure 4-5, four transit agencies/services (all operating in the United States) own 100% of the vehicles in their fleet. In contrast, Transport of London contracts 100% of its vehicles. Only one transit agency/service (also located in the United States) reported operating a mixed fleet.

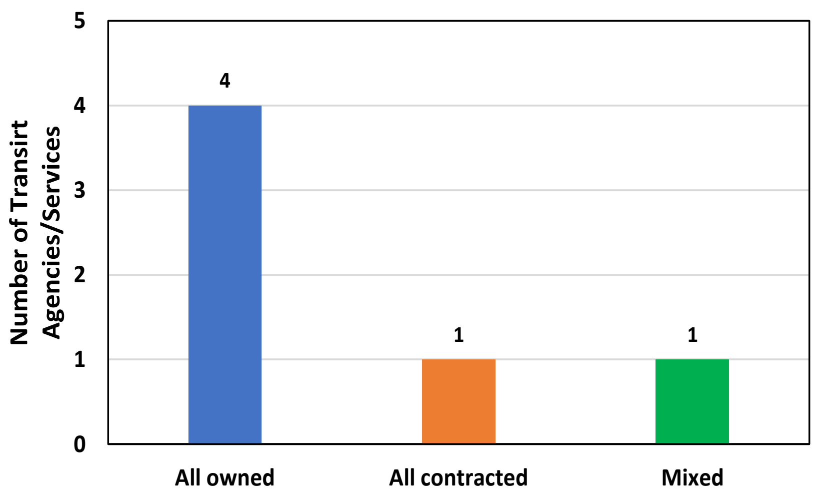

Figure 4-5: Composition of Fleet – Out-of-State Transit Agencies/Services

4.2.2 Type of Service Area

A total of 47 transit agencies/services operating in Oregon reported on the type of area they serve (i.e., rural, urban, or both). As depicted in Figure 4-6, 29 transit agencies/services (i.e., 61.7%) provide services only in rural areas, four (i.e., 8.5%) serve only urban areas, and 14 (i.e., 29.8%) provide services in both rural and urban areas.

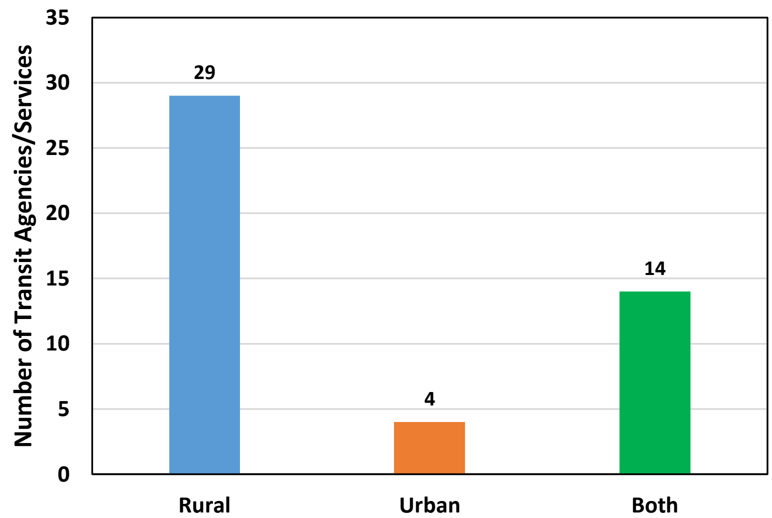

Figure 4-6: Type of Area Served – Transit Agencies/Services in Oregon

Six out-of-state transit agencies/services reported on the type of area they serve. As depicted in Figure 4-5, two of these transit agencies/services provide services only in urban areas (i.e., 33.3%) and four (i.e., 66.7%) provide services in both rural and urban areas.

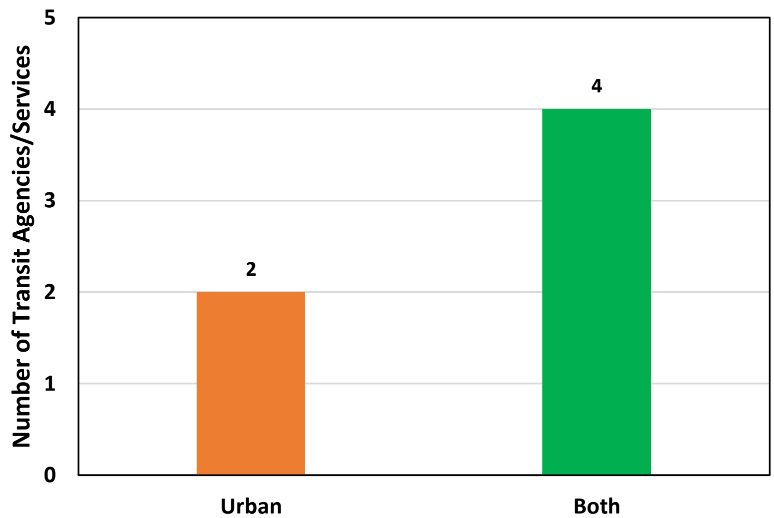

Figure 4-7: Type of Area Served – Out-of-State Transit Agencies/Services

4.2.3 Type of Service

A total of 47 transit agencies/services operating in Oregon reported on the type of service they provide (i.e., fixed route, non-fixed route, or both). As depicted in Figure 4-8, six transit agencies/services (i.e., 12.8%) only provide fixed route service and 13 (i.e., 27.7%) only provide non-fixed route service. In contrast, 28 (i.e., 59.6%) of transit agencies/services provide a combination of fixed route and non-fixed route services.

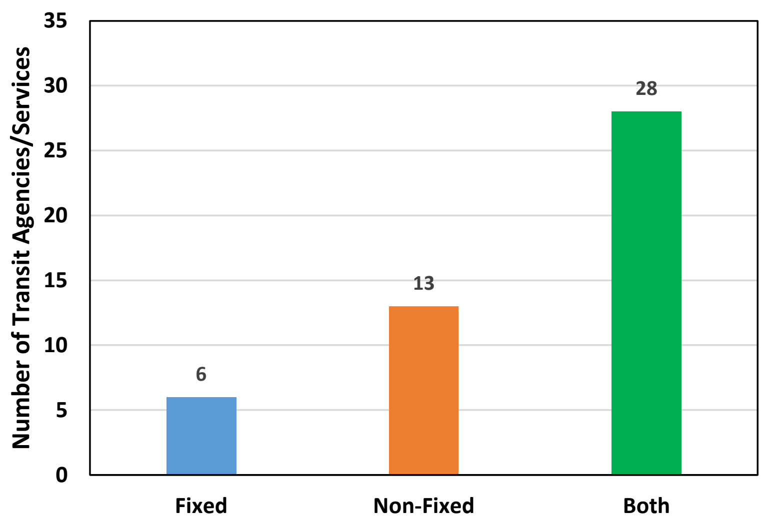

Figure 4-8: Type of Service – Transit Agencies/Services in Oregon

Six out-of-state transit agencies/services reported on the type of area they serve. As depicted in Figure 4-9, one of these transit agencies/services provides only fixed route service (i.e., 16.7%) and five (i.e., 66.7%) provide a combination of fixed route and non-fixed route services.

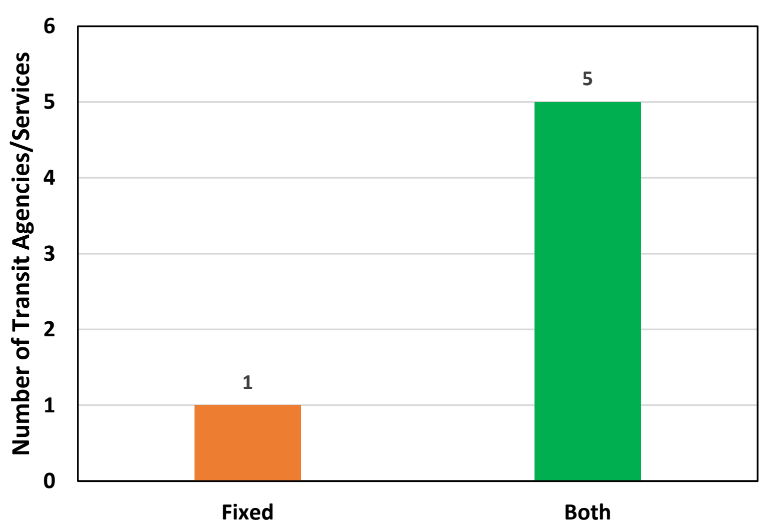

Figure 4-9: Type of Service – Out-of-State Transit Agencies/Services

4.2.4 Levels of Ridership Data Collection – Fixed Route Modes

A total of 31 transit agencies/services operating in Oregon reported on the levels of ridership data collection they perform. There were seven possible answers for the level of ridership data collection (i.e., “Stop,” “Segment,” “Trip,” “Route,” “System,” “Other,” and “N/A”) for the following 13 fixed route modes:

- Bus
- Commuter Bus
- Bus Rapid Transit
- Trolleybus
- Vanpool
- Ferryboat
- Commuter Rail
- Hybrid Rail
- Heavy Rail
- Light Rail
- Streetcar Rail
- Monorail/Automated Guideway
- Aerial Tramway

Table 4-1 shows the results for transit agencies/services located in Oregon. The fixed route modes for which ridership data are collected at more levels are “Bus” and “Commuter Bus.” Very little ridership data collection efforts were reported for the fixed route modes “Vanpool,” “Hybrid Rail,” “Light Rail,” “Streetcar Rail,” “Bus Rapid Transit,” and “Trolleybus,” whereas no ridership data collection efforts were reported for the fixed route modes “Ferryboat,” “Commuter Rail,” “Heavy Rail,” “Monorail/Automated Guideway,” and “Aerial Tramway.”

Figure 4-10 and Figure 4-11 depict more specific information about the levels at which transit agencies/services located in Oregon collect ridership data for the fixed route modes “Bus” and “Commuter Bus.” In both cases, the majority of transit agencies/services reported collecting fixed-route “Bus” service ridership data predominantly at the route level, followed in frequency (also in both cases) by trip and stop levels.

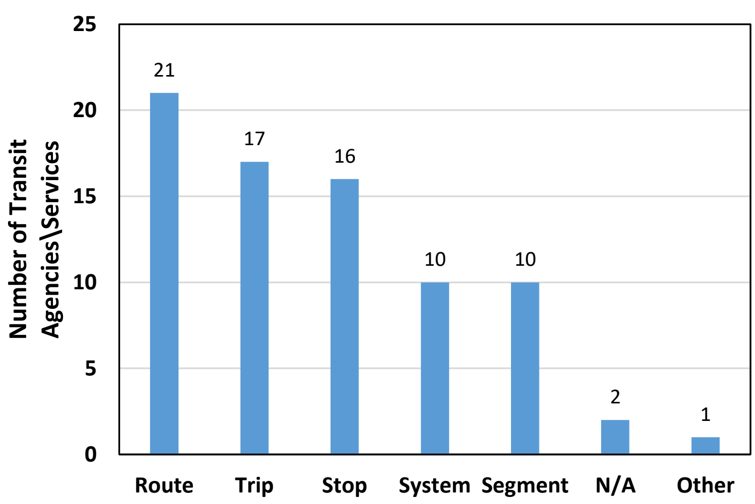

Figure 4-10: Most Frequent Levels of Ridership Data Collection for “Bus” Fixed Route Service – Transit Agencies/Services in Oregon

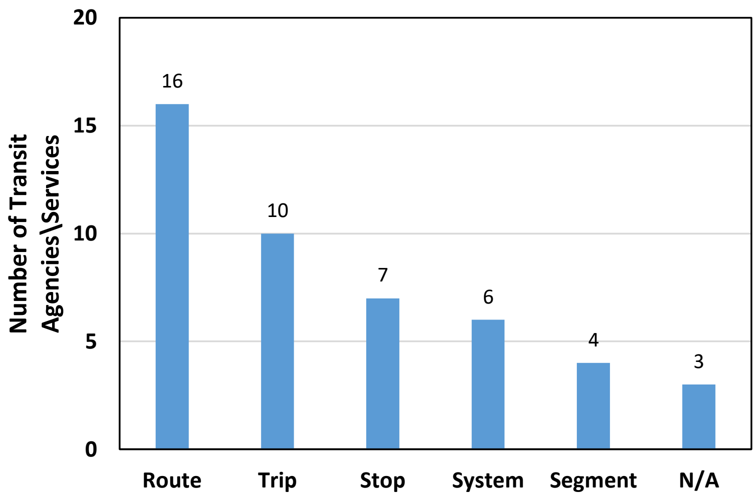

Figure 4-11: Most Frequent Levels of Ridership Data Collection for "Commuter Bus" Fixed Route Service – Transit Agencies/Services in Oregon

Table 4-2 shows the results for the six transit agencies/services located outside of Oregon. In this case, the fixed route modes for which ridership data are collected at more levels are "Bus," "Commuter Bus," and "Bus Rapid Transit." Very little ridership data collection efforts were reported for the fixed route modes "Vanpool," "Trolleybus," "Ferryboat," "Light Rail," "Streetcar Rail," and "Aerial Tramway," whereas no ridership data collection efforts were reported for the fixed route modes "Commuter Rail," "Hybrid Rail," "Heavy Rail," and "Monorail/Automated Guideway."

For illustration purposes, Figure 4-12 depicts more specific information about the levels at which transit agencies/services located outside of Oregon collect ridership data for the fixed route mode "Bus." In this case, the majority of transit agencies/services reported collecting fixed-route "Bus" service ridership data most commonly at the stop level, followed in frequency by trip, route, and system levels.

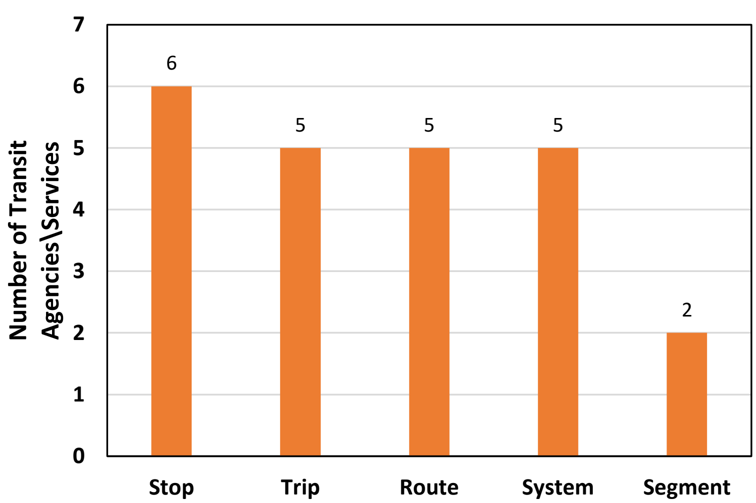

Figure 4-12: Most Frequent Levels of Ridership Data Collection for "Bus" Fixed Route Service – Out-of-State Transit Agencies/Services

Table 4-1: Levels of ridership data collected for each of the following fixed-route modes – transit agencies/services in Oregon

FIXED ROUTE MODES	Count	Stop	%	Segment	%	Trip	%	Route	%	System	%	Other	%	N/A	%
Bus	30	16	53.3%	10	33.3%	17	56.7%	21	70.0%	10	33.3%	1	3.3%	2	6.7%
Commuter Bus	22	7	31.8%	4	18.2%	10	45.5%	16	72.7%	6	27.3%	0	0.0%	3	13.6%
Vanpool	16	2	12.5%	2	12.5%	2	12.5%	3	18.8%	3	18.8%	0	0.0%	12	75.0%
Hybrid Rail	16	1	6.3%	1	6.3%	1	6.3%	1	6.3%	1	6.3%	0	0.0%	15	93.8%
Light Rail	16	1	6.3%	1	6.3%	1	6.3%	1	6.3%	1	6.3%	0	0.0%	15	93.8%
Streetcar Rail	16	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	1	6.3%	15	93.8%
Bus Rapid Transit	15	0	0.0%	0	0.0%	0	0.0%	1	6.7%	0	0.0%	0	0.0%	14	93.3%
Ferryboat	15	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	15	100.0%
Commuter Rail	15	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	15	100.0%
Heavy Rail	15	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	15	100.0%
Monorail/Automated Guideway	15	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	15	100.0%
Aerial Tramway	15	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	15	100.0%
Trolleybus	14	0	0.0%	0	0.0%	0	0.0%	2	14.3%	1	7.1%	1	7.1%	12	85.7%

Table 4-2: Levels of ridership data collected for each of the following fixed-route modes – out-of-state transit agencies/services

FIXED ROUTE MODES	Count	Stop	%	Segment	%	Trip	%	Route	%	System	%	Other	%	N/A	%
Bus	6	6	100.0%	2	33.3%	5	83.3%	5	83.3%	5	83.3%	0	0.0%	0	0.0%
Commuter Bus	6	5	83.3%	2	33.3%	4	66.7%	4	66.7%	4	66.7%	0	0.0%	1	16.7%
Bus Rapid Transit	6	3	50.0%	1	16.7%	3	50.0%	3	50.0%	3	50.0%	0	0.0%	3	50.0%
Vanpool	5	0	0.0%	0	0.0%	1	20.0%	0	0.0%	2	40.0%	0	0.0%	3	60.0%
Trolleybus	4	1	25.0%	0	0.0%	1	25.0%	1	25.0%	1	25.0%	0	0.0%	3	75.0%
Ferryboat	4	1	25.0%	0	0.0%	1	25.0%	1	25.0%	1	25.0%	0	0.0%	3	75.0%
Commuter Rail	4	1	25.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	4	100.0%
Hybrid Rail	4	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	4	100.0%
Heavy Rail	4	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	4	100.0%
Light Rail	4	2	50.0%	1	25.0%	1	25.0%	2	50.0%	2	50.0%	0	0.0%	2	50.0%
Streetcar Rail	4	2	50.0%	0	0.0%	1	25.0%	2	50.0%	1	25.0%	0	0.0%	2	50.0%
Monorail/Automated Guideway	4	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	4	100.0%
Aerial Tramway	4	0	0.0%	0	0.0%	0	0.0%	0	0.0%	1	25.0%	1	25.0%	3	75.0%

4.2.5 Levels of Availability of Ridership Data – Fixed Route Modes

A total of 44 transit agencies/services operating in Oregon reported on the availability of the ridership data they collect. There were seven possible answers for ridership data frequency (i.e., "Daily," "Weekly," "Monthly," "Quarterly," "Annually," "As Needed," and "N/A") associated with the following levels of ridership data:

- System ridership
- Route-level ridership
- Route segment ridership
- Stop-level boardings/alightings
- Performance measures
- Schedule adherence
- Running Times

Table 4-3 shows that all 44 transit agencies/services in Oregon have system-level ridership data available on some basis. Although having stop-level boardings/alightings data available at some level seems to be common among transit agencies/services in Oregon, 13 out of 38 respondents (i.e., 34.2%) indicated not having this level of ridership data available at all. Figure 4-13 graphically compares the levels of availability or ridership data reported by transit agencies/services operating in Oregon.

Table 4-4 shows that (with the exception of running times) all six out-of-state transit agencies/services have ridership data available more often at different levels. Figure 4-14 graphically compares the levels of availability or ridership data reported by transit agencies/services operating outside of Oregon.

Table 4-3: Frequency of availability for different levels of ridership data for fixed-route modes – transit agencies/services in Oregon

LEVEL OF RIDERSHIP DATA	Count	Daily	%	Weekly	%	Monthly	%	Quarterly	%	Annually	%	As Needed	%	N/A	%
System ridership	44	26	59.1%	11	25.0%	24	54.5%	15	34.1%	12	27.3%	11	25.0%	1	2.3%
Route-level ridership	41	20	48.8%	12	29.3%	15	36.6%	12	29.3%	9	22.0%	9	22.0%	7	17.1%
Performance measures	39	11	28.2%	9	23.1%	11	28.2%	9	23.1%	8	20.5%	14	35.9%	7	17.9%
Stop-level boardings/alightings	38	12	31.6%	6	15.8%	9	23.7%	6	15.8%	6	15.8%	12	31.6%	13	34.2%
Running Times	38	16	42.1%	9	23.7%	8	21.1%	7	18.4%	4	10.5%	14	36.8%	7	18.4%
Schedule adherence	37	14	37.8%	5	13.5%	9	24.3%	5	13.5%	4	10.8%	16	43.2%	8	21.6%
Route segment ridership	36	6	16.7%	4	11.1%	5	13.9%	3	8.3%	2	5.6%	10	27.8%	16	44.4%

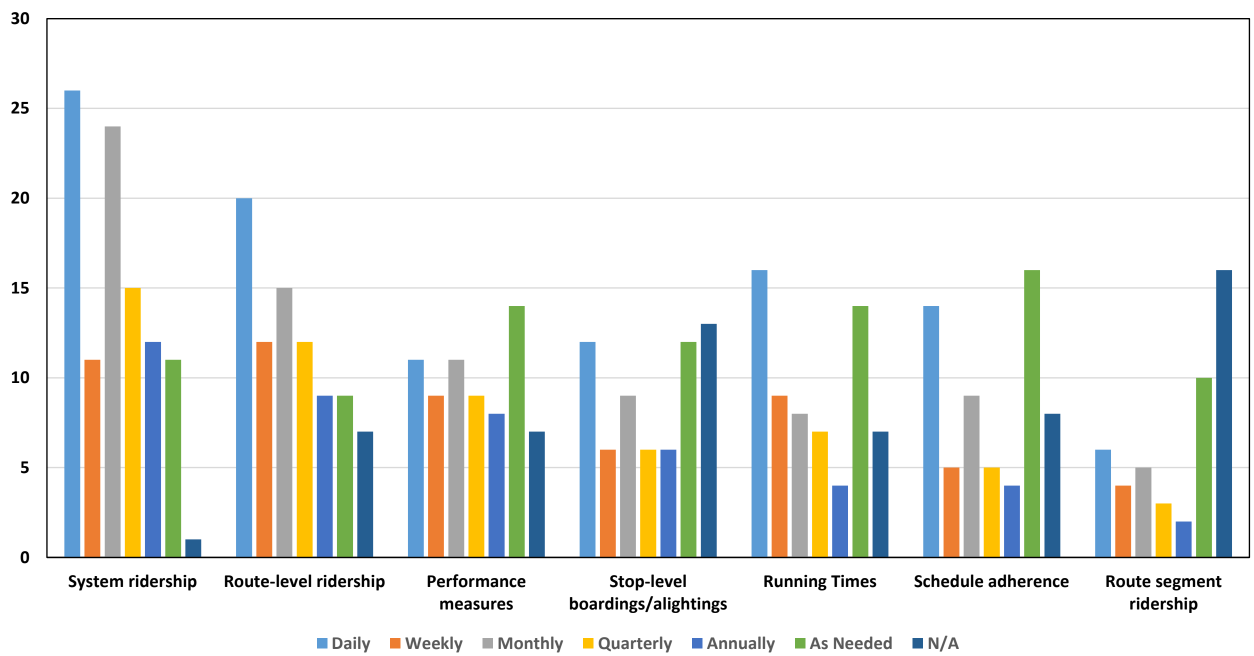

Figure 4-13: Frequency of Availability for Different Levels of Ridership Data for Fixed-Route Modes – Transit Agencies/Services in Oregon

Table 4-4: Frequency of availability for different levels of ridership data for fixed-route modes – out-of-state transit agencies/services

LEVEL OF RIDERSHIP DATA	Count	Daily	%	Weekly	%	Monthly	%	Quarterly	%	Annually	%	As Needed	%	N/A	%
System ridership	6	5	83.3%	3	50.0%	3	50.0%	4	66.7%	3	50.0%	2	33.3%	0	0.0%
Route-level ridership	6	5	83.3%	3	50.0%	4	66.7%	3	50.0%	3	50.0%	2	33.3%	0	0.0%
Stop-level boardings/alightings	6	4	66.7%	2	33.3%	2	33.3%	4	66.7%	3	50.0%	1	16.7%	0	0.0%
Performance measures	6	5	83.3%	3	50.0%	4	66.7%	3	50.0%	3	50.0%	2	33.3%	0	0.0%
Schedule adherence	6	4	66.7%	3	50.0%	4	66.7%	3	50.0%	3	50.0%	2	33.3%	0	0.0%
Running Times	6	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Route segment ridership	5	1	20.0%	1	20.0%	1	20.0%	2	40.0%	2	40.0%	2	40.0%	1	20.0%

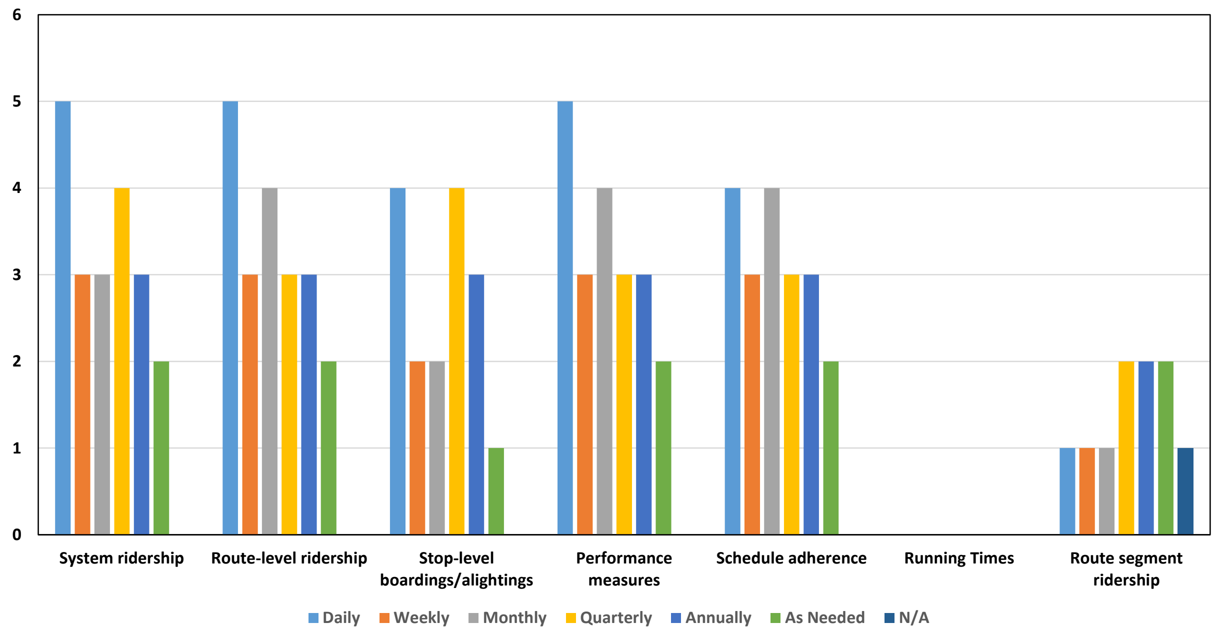

Figure 4-14: Frequency of Availability for Different Levels of Ridership Data for Fixed-Route Modes – Out-of-State Transit Agencies/Services

4.2.6 Methods of Ridership Data Collection

A total of 43 transit agencies/services operating in Oregon reported on the methods utilized to collect ridership data. In the online questionnaire, respondents were presented with four possible main answers (in the form of checkboxes) including "Automated passenger counters (APC)," "Registering fareboxes," "Handheld data collection units (e.g. Mobile Data Terminals - MDT)," and "Driver's trip log." A fifth option, in the form of an open-ended textbox, was also provided.

Figure 4-15 shows that a driver's trip log is the dominant ridership data collection method used by 33 (i.e., 76.7%) transit agencies/services in Oregon. In contrast, only eight respondents (i.e., 18.6%) reported using MDTs and five (i.e., 11.6%) reported using registering fareboxes or APCs. Other methods to collect ridership data were reported by a very small percentage of Oregon transit agencies/services.

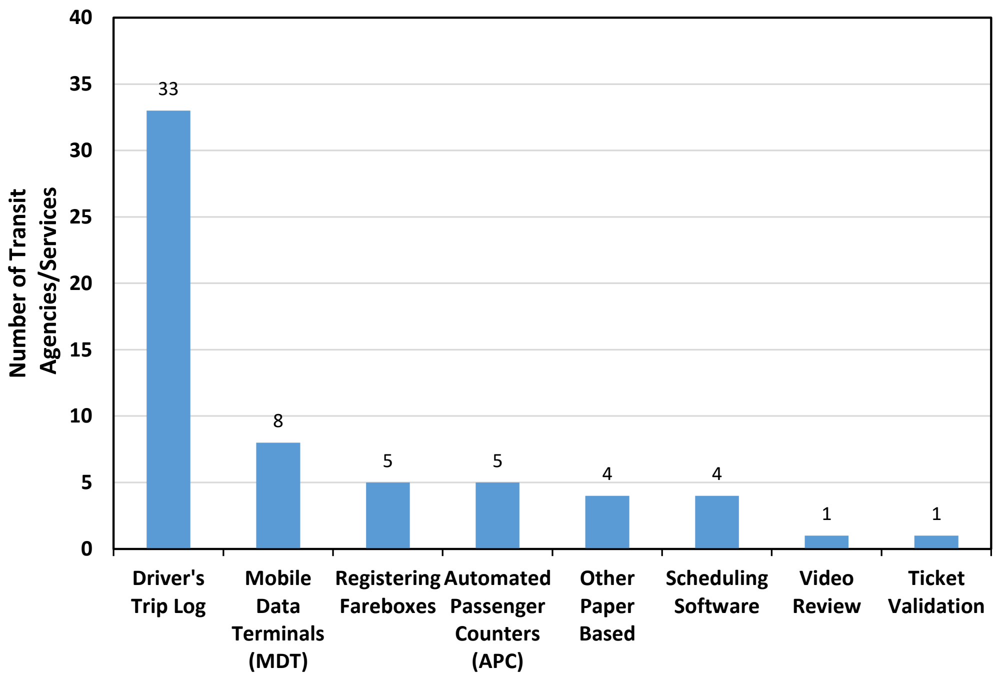

Figure 4-15: Methods of Data Collection – Transit Agencies/Services in Oregon

Figure 4-16 depicts the results for the out-of-state transit agencies/services. In this case, five out of six transit agencies/services (i.e., 83.3%) reported using APCs as the main method to collect ridership data. Also of note is that a larger portion of out-of-state respondents (i.e., 50%) reported using "Registering Fareboxes," which can also be thought of as an automated data collection method.

Figure 4-17 shows which data collection methods are used more frequently by different transit agencies/services (both in Oregon and out-of-state) relative to their fleet size. These results clearly show large and medium size transit agencies/services report using APCs and registering fareboxes much more frequently than smaller size transit agencies/services. In contrast, the use of driver's trip logs is more common in small transit agencies/services when compared to larger transit agencies/services.

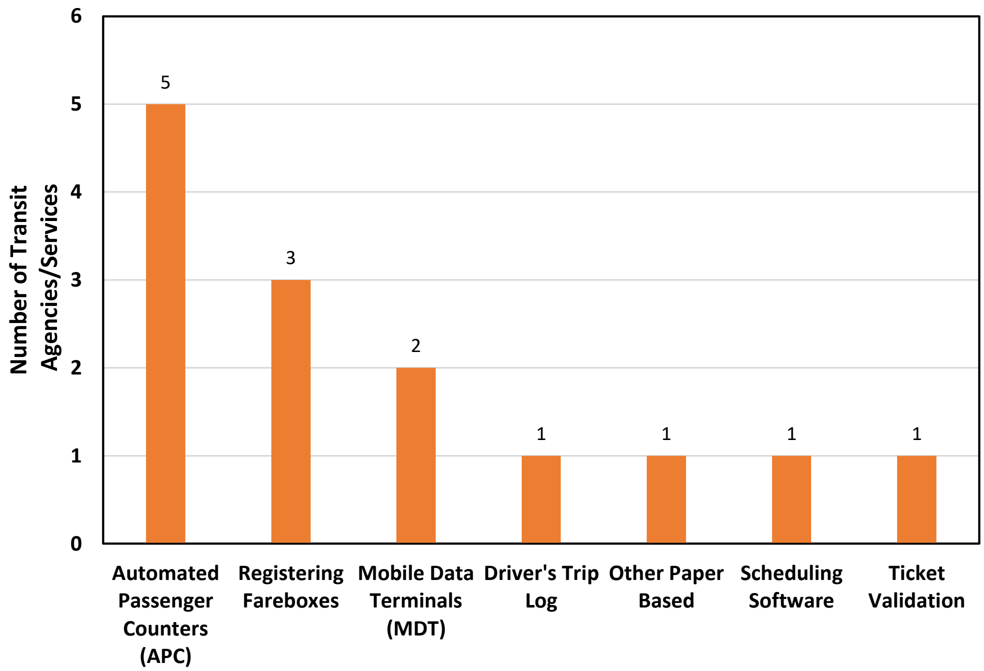

Figure 4-16: Methods of Data Collection – Out-of-State Transit Agencies/Services

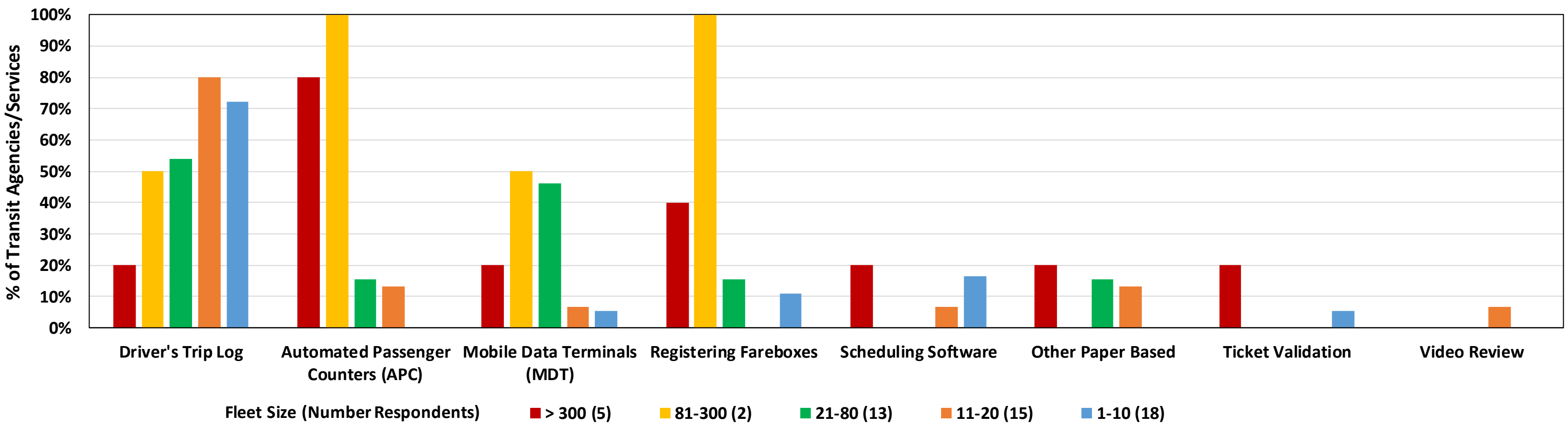

Figure 4-17: Methods of Data Collection by Fleet Size

4.2.7 Collection of Ridership Data Supplementary Details

A total of 43 transit agencies/services operating in Oregon provided more details about which types of supplementary details of ridership they collect. In the online questionnaire, respondents were presented with multiple checkbox selections representing potential supplementary details of ridership including "Wheelchairs," "Fare Types," "Special Rider Types," "GPS Coordinates," "Timestamps," "Bicycles," "Transfer Status," and "Other Medical or Mobility Devices." A final option, in the form of an open-ended textbox, was also provided.

Figure 4-18 presents the responses received from Oregon transit agencies/services. The results show that 29 respondents (i.e., 67.4%) reported collecting wheelchair-related data, while 25 (i.e., 58.1%) reported collecting fare types. Also, special rider types (i.e., 49.5%) and data on other types of medical or mobility devices (i.e., 37.2%) were reported as common types of supplementary ridership collected. Other interesting categories supplied by respondents included data on rider car sickness and needed caregivers or assistants. However, there was only one instance of each reported and (in context) these responses were in regards to demand response transit or paratransit services.

It is important to mention that six transit agencies/services in Oregon (i.e., 14%) do not collect supplementary ridership information.

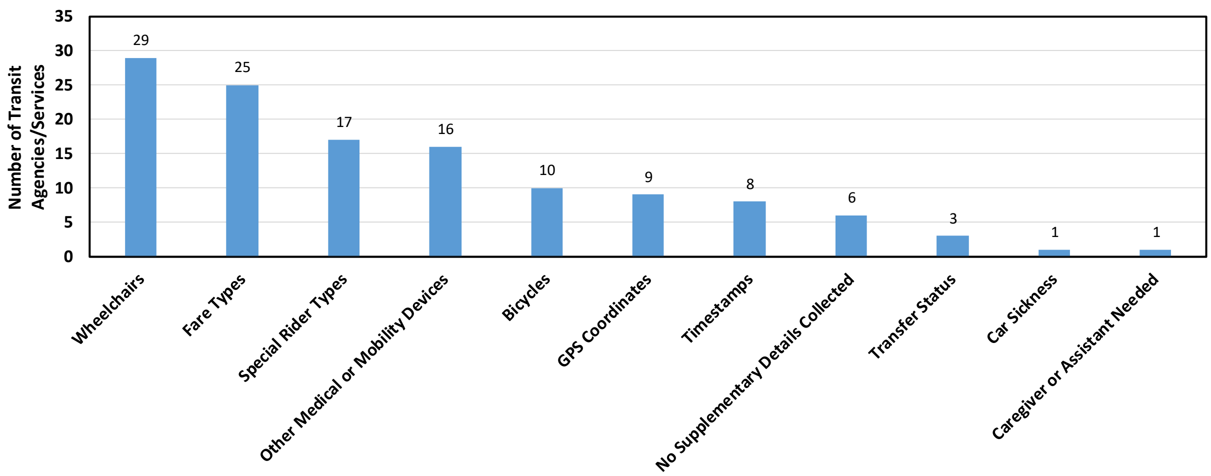

Figure 4-18: Ridership Supplementary Detail Data Types – Transit Agencies/Services in Oregon

Figure 4-19 presents the results for out-of-state transit agencies/services. In this case, all six respondents reported collecting GPS coordinates and timestamps. For comparison purposes, only 20.9% of Oregon transit agencies/services reported collecting GPS coordinates and 18.6% reported collecting timestamps.

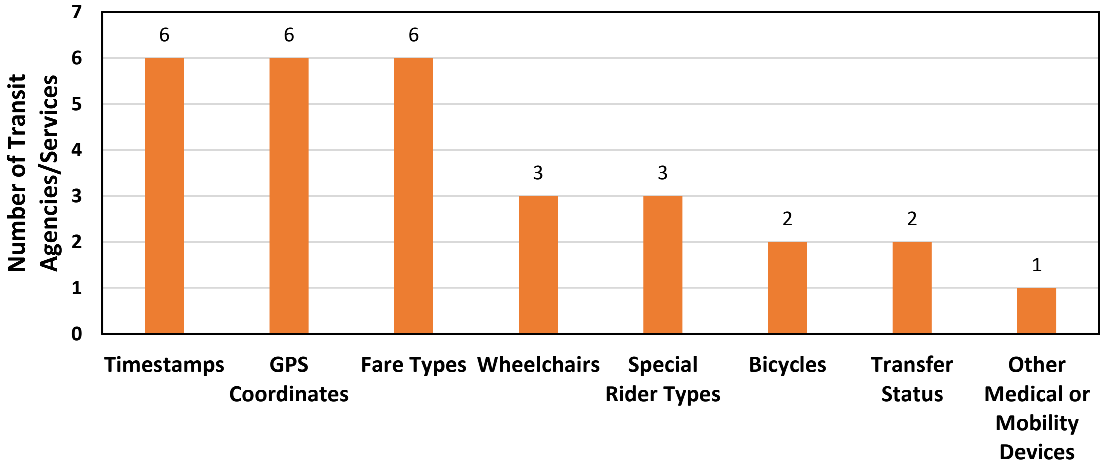

Figure 4-19: Ridership Supplementary Detail Data Types – Out-of-State Transit Agencies/Services

In this part of the online questionnaire, respondents were also asked to elaborate on how supplementary details of ridership are collected. The results for this question for Oregon transit agencies/services are depicted in Figure 4-20, while Figure 4-21 depicts the results for out-of-state transit agencies/services. Once again, "Driver Trip Log" was reported as the dominant collection method by Oregon transit agencies/services (i.e., 84.6%), while automated systems (i.e., "AVL or APC" and "Other Automated Systems") were the most prevalent methods utilized by out-of-state agencies (i.e., 66.7% for both). However, it should be noted that this particular open-ended sub-question had a relatively low response rate from both Oregon and out-of-state agencies (i.e., 29.5% and 50%, respectively).

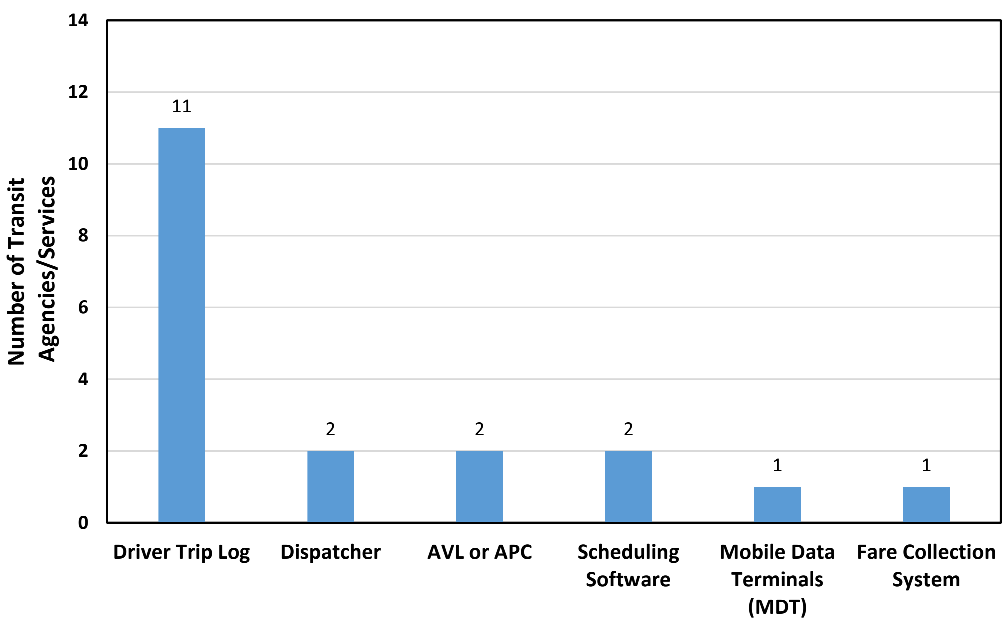

Figure 4-20: Ridership Supplementary Detail Collection Methods – Transit Agencies/Services in Oregon

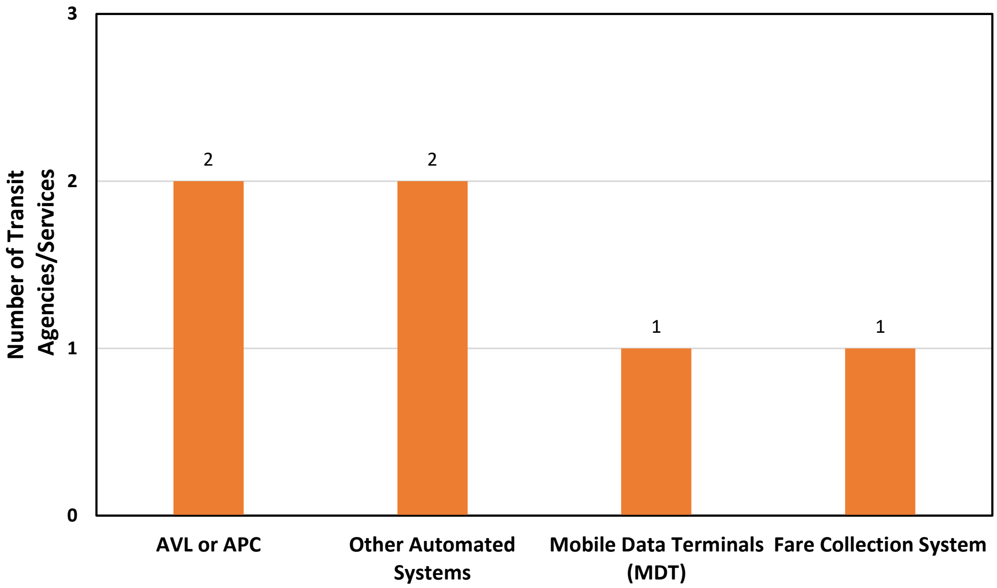

Figure 4-21: Ridership Supplementary Detail Collection Methods – Out-of-State Transit Agencies/Services

4.2.8 Automated Ridership Data Collection

The use of APCs by transit agencies/services was addressed through a series of questions aimed at gaining a deeper understanding of their adoption and implementation.

As previously illustrated in Figure 4-15 (see Section 4.2.6), only five out of 47 Oregon transit agencies (i.e., 10.6%) reported using APCs. These five transit agencies are TriMet, Salem Area Mass Transit District, Rogue Valley Transportation District (RVTD), Hood River County Transportation District, and Corvallis Transit System (CTS).

Respondents were then asked about the percentage of their fleet that is equipped with APCs. Again, the same five transit agencies in Oregon provided a non-zero or non-blank response. Only CTS reported having 100% of their fleet equipped with APCs. However, TriMet has 100% of its bus fleet and 50% of its light rail fleet equipped with APCs. It is interesting to note that the fleets of almost all the out-of-state respondents are equipped (to some extent) with APC technology. The sole exception is Transit for London, but this respondent has the largest fleet size at 9,400 vehicles.

A follow-up question was posed specifically for those agencies reporting some APC adoption (i.e., less than 100%) to learn how the APC equipped vehicles are chosen to service routes. Figure 4-22 shows the assignment methods employed by the seven transit agencies/services (in and out-of-state) that responded to this question. Besides the two respondents who have no need for a specific plan (mainly due to high enough proportion of APC equipped vehicles), all other respondents employed their own method. For example, one method mentioned was to maximize route coverage, another was to meet specific collection needs, and another was a simple random assignment.

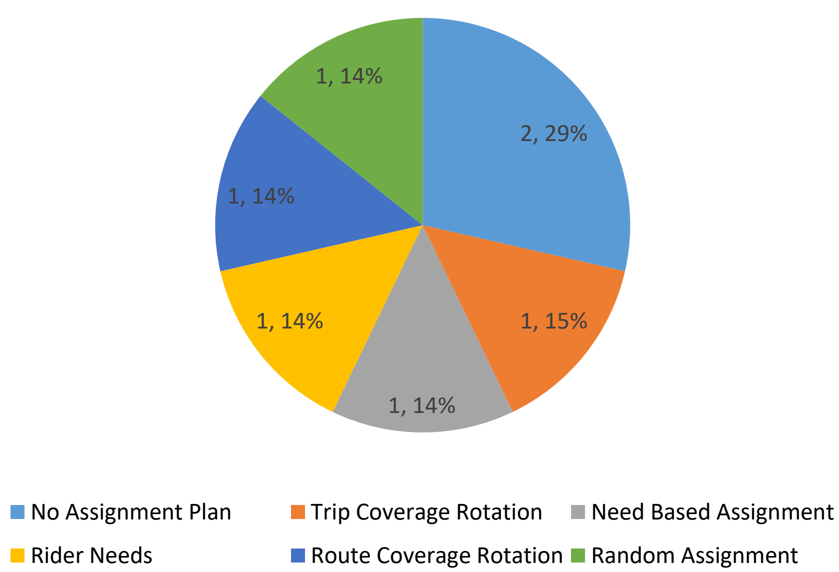

Figure 4-22: APC Equipped Vehicle Assignment Method

Anticipating a low adoption of APC technology in Oregon, especially among the smaller transit agencies/services, an open-ended question was posed to find which barriers might exist to a transit agency/service when implementing APC technology on their fleet. The results for this question are depicted in Figure 4-23 and show that from 34 Oregon agencies/services that responded to this question, 20 (i.e., 58.8%) cited financial resources as the main barrier. Others cited barriers included a perception of APCs being unnecessary (i.e., 20.6%) and concerns about APC accuracy (i.e., 14.7%). Several respondents cited agency-specific strategic or logistical constraints such as the timing of capital replenishment cycles or being unsure of demand response APC integration.

Figure 4-24 shows that three of the five (i.e., 60%) out-of-state transit agencies/services also cited financial resources as a barrier to the adoption of APC technology.

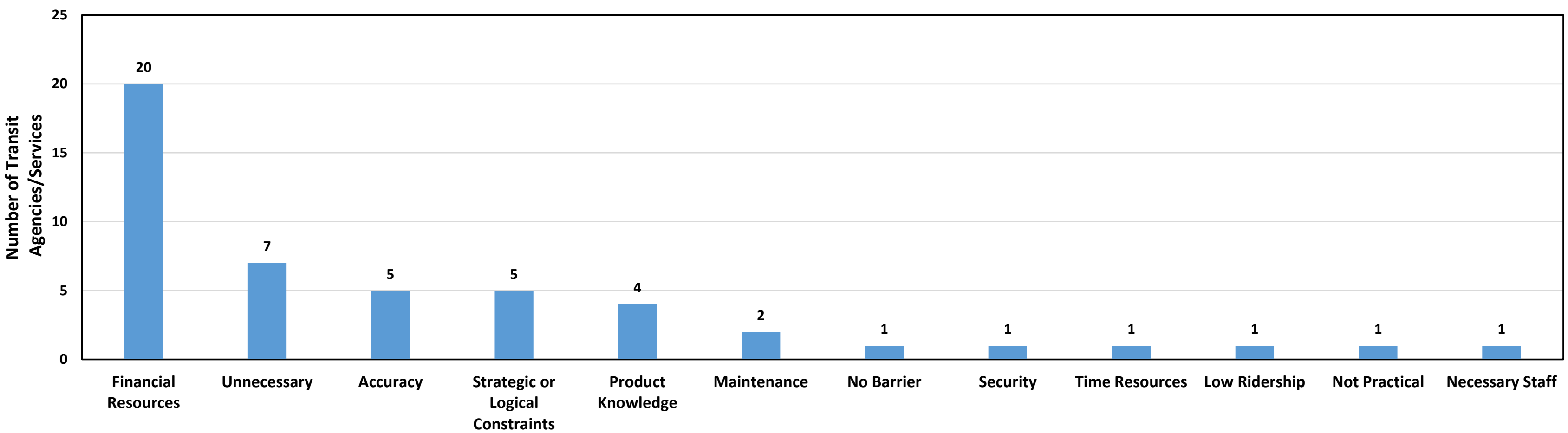

Figure 4-23: Barriers to APC Adoption – Transit Agencies/Services in Oregon

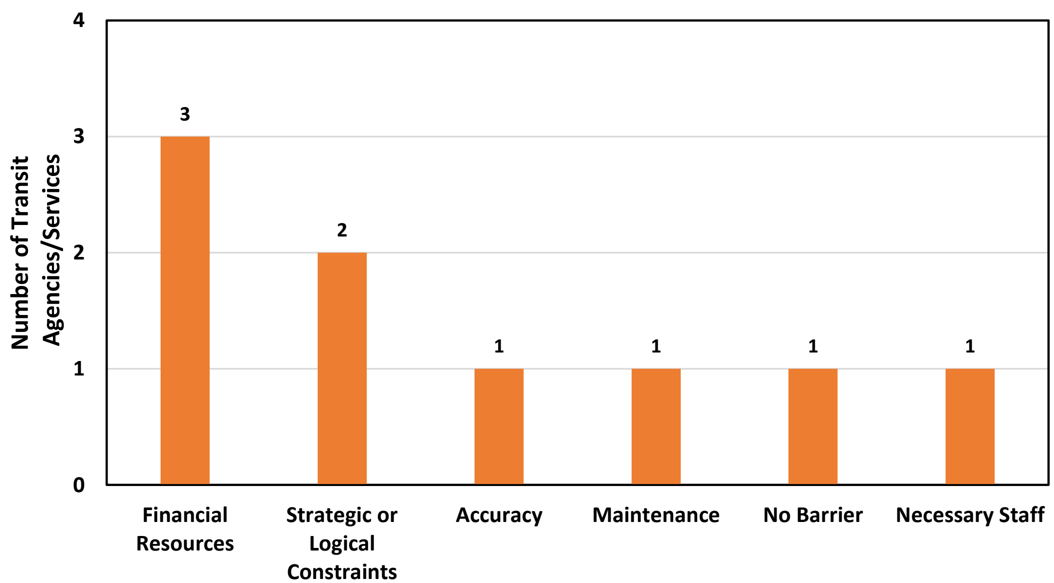

Figure 4-24: Barriers to APC Adoption – Out-of-State Transit Agencies/Services

Figure 4-25 and Figure 4-26 depict the results for a question asking what combinations of automated and manual methods are used to collect ridership data. This question had low response counts for both Oregon and out-of-state transit agencies/services (i.e., five and three, respectively). The common theme among these responses was that manual counts or surveys are often used in conjunction with automated methods (e.g., APC, AFC, and scheduling software) to validate the automated method.

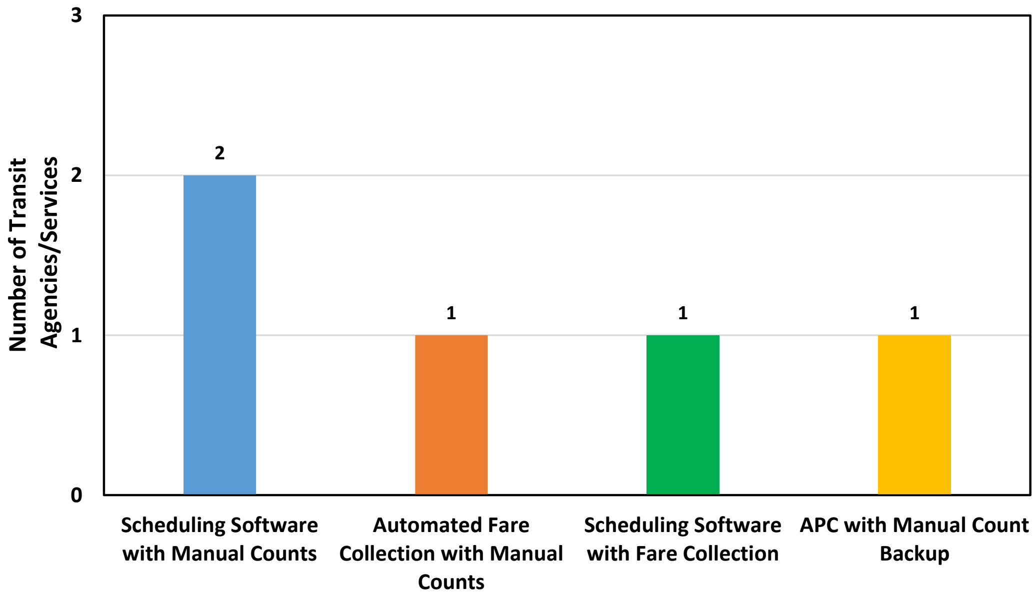

Figure 4-25: Automated and Manual Combination Methods – Transit Agencies/Services in Oregon

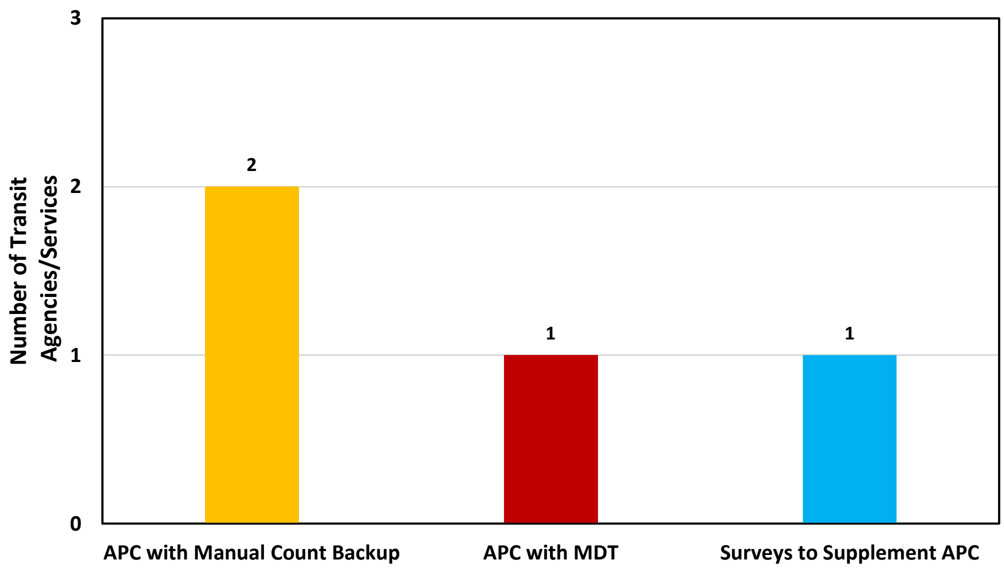

Figure 4-26: Automated and Manual Combination Methods – Out-of-State Transit Agencies/Services

Information was also collected regarding the specific manufacturers and models of the device(s) and/or software utilized to automatically collect ridership data. Most products were only mentioned once, but INIT's IRMA APC and Route Match software were each mentioned by three out of 12 (i.e., 25%) of the respondents (both in and out-of-state). Table 4-5 lists the various technologies mentioned in three different categories (i.e., APC/AFC, Mobile Data Terminals, and Software).

Table 4-5: Products commonly used to automatically collect ridership data

APC/AFC	Mobile Data Terminals (MDTs)	Software
• Dilax • Iris IRMA • INIT IRMA • IVT • Connexionz, Inc. APC Contractor • Genfare Odyssey Farebox	• DDS Wireless Vector9000 • DDS Wireless mSlate • Mentor Ranger • Samsung GT-N8913 • Unspecified Tablet	• UTA APC Software • Route Match • Ecolane

Additionally, respondents were asked to explain any differences in their ridership data collection practices that may have occurred in the past three years. While this question was aimed at further revealing any recent trends in the practice of ridership data collection, the response rate was very low (i.e., only five Oregon agencies responded). The items of note, as shown in Figure 4-27, are that three Oregon transit agencies/services have transitioned from paper-based collection methods to an increased used of MDTs and, as seen in Figure 4-28, two out-of-state agencies have increased their use of APCs.

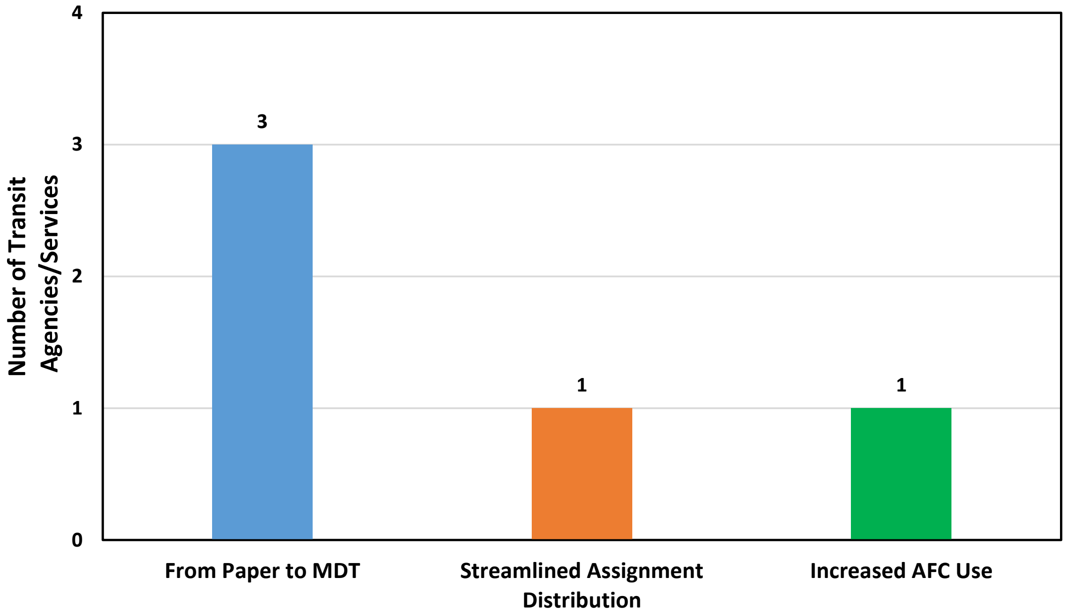

Figure 4-27: Recent Changes in Ridership Data Collection Methods – Transit Agencies/Services in Oregon

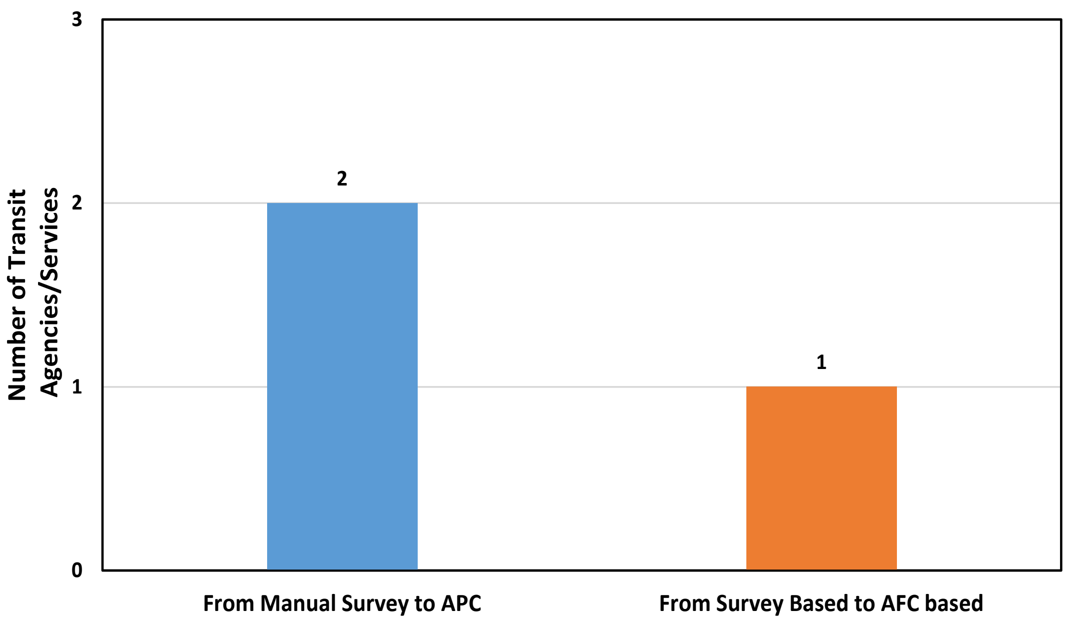

Figure 4-28: Recent Changes in Ridership Data Collection Methods – Out-of-State Transit Agencies/Services

4.2.9 Uses of Surveys and Sampling for Ridership Data

The final two questions regarding ridership data collection focused on the uses of alternate collection methods to infer ridership. More specifically, respondents were asked about their use of manual surveys and sampling methods to estimate ridership. These questions were both structured alike, each with four sub-questions relating to the frequency of use, sample size, proportion of trips studied, and estimation methods.

The results reveal that these questions were likely not applicable to many agencies as evidenced by only a small quantity of appropriate responses. With regards to the use of survey methods, five complete responses were collected from Oregon transit agencies/services and only one response was assessed as complete from an out-of-state respondent. A small number of additional responses to certain sub-questions were deemed appropriate and included in the analysis.

The results to the four sub-questions are depicted in Figure 4-29, Figure 4-30, Figure 4-31, and Figure 4-32. In Figure 4-32, it is worth noting that no response provided details on the use of surveys to estimate ridership. While two respondents (i.e., 40%) stated they estimate in general terms, the other respondents were using surveys for other purposes.

In regards to ridership sampling, only two appropriate responses were recorded. This seems to indicate that statistical sampling is only minimally performed among Oregon transit agencies/services. Many agencies commented on their daily 100% ridership counts with manual methods, and several other agencies repeated the same responses provided to the previous question regarding ridership surveys. Both cases were excluded and not considered here as statistical sampling. One agency, King County Metro in the Seattle metropolitan area, used daily random sampling for rider data collection in their large vanpool fleet. Northeast Oregon Public Transit used quarterly ride-along manual count sampling to note ridership trends, but not to estimate ridership counts.

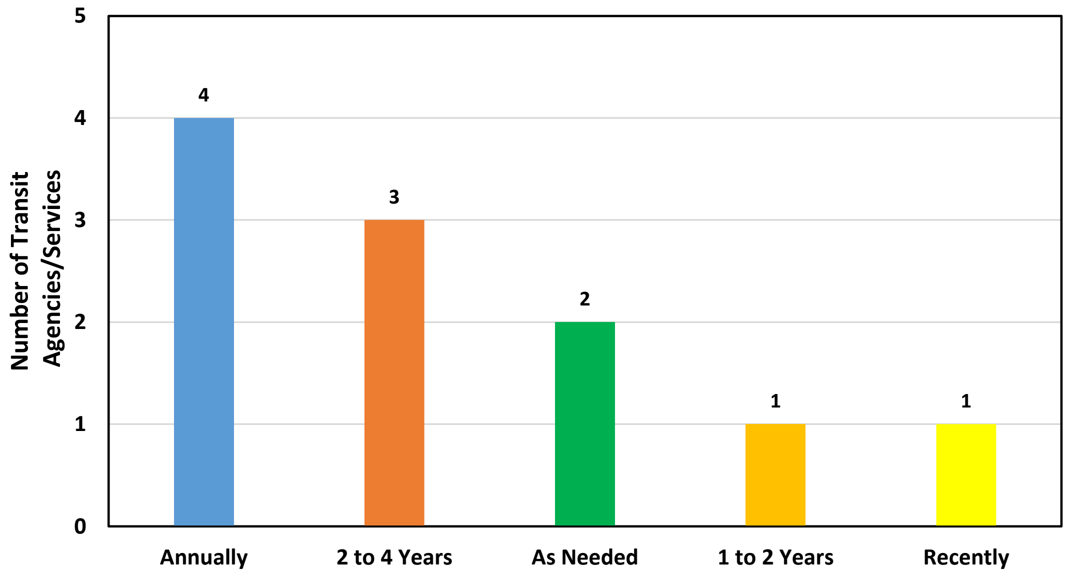

Figure 4-29: Survey Frequency – Transit Agencies/Services in Oregon

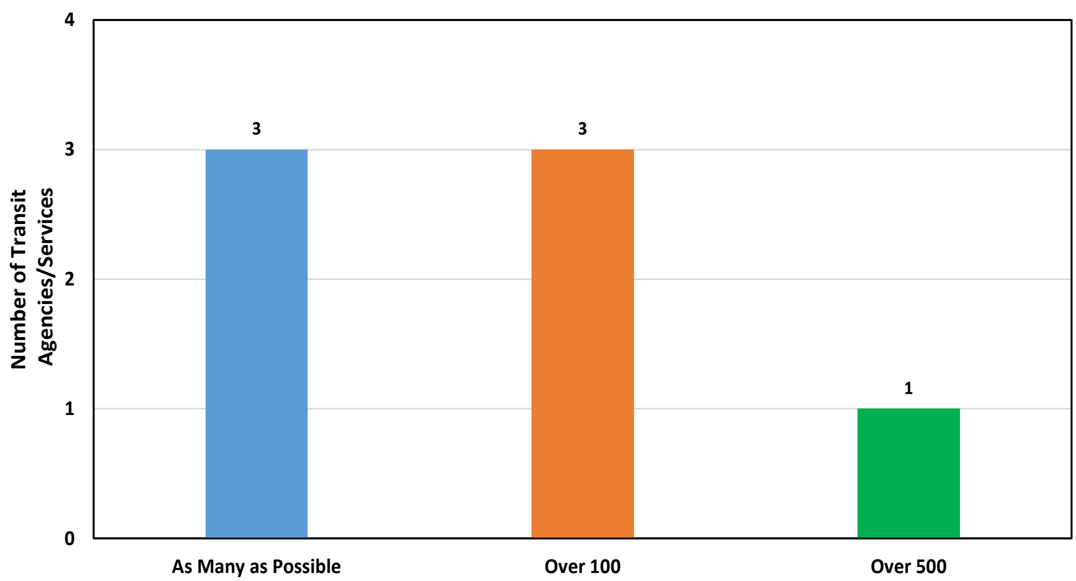

Figure 4-30: Distribution Size – Transit Agencies/Services in Oregon

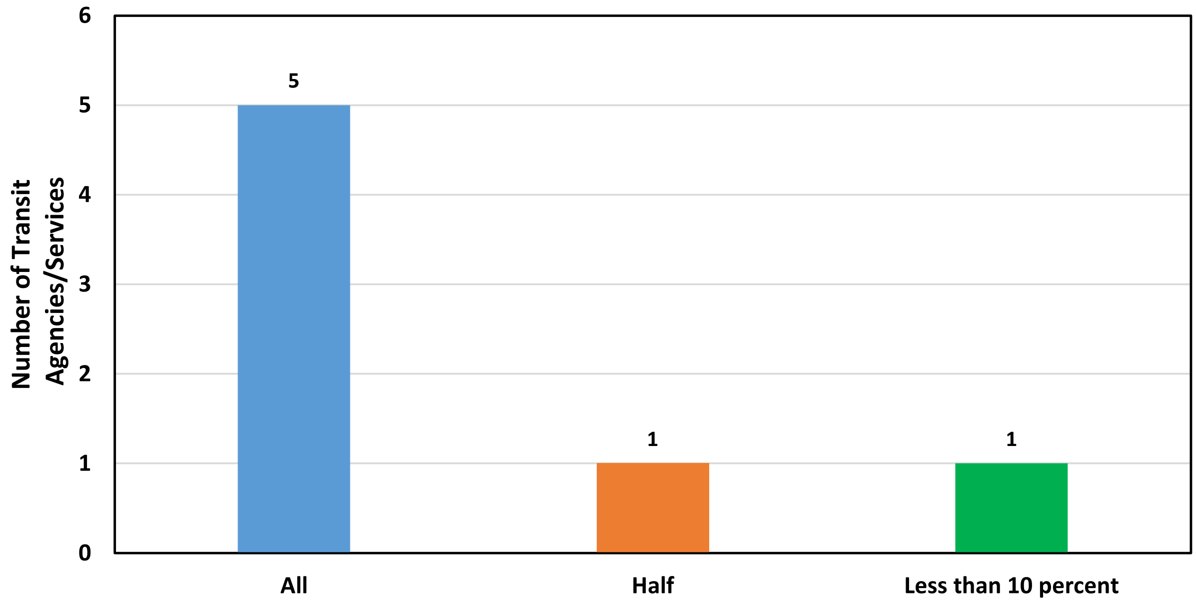

Figure 4-31: Proportion Surveyed – Transit Agencies/Services in Oregon

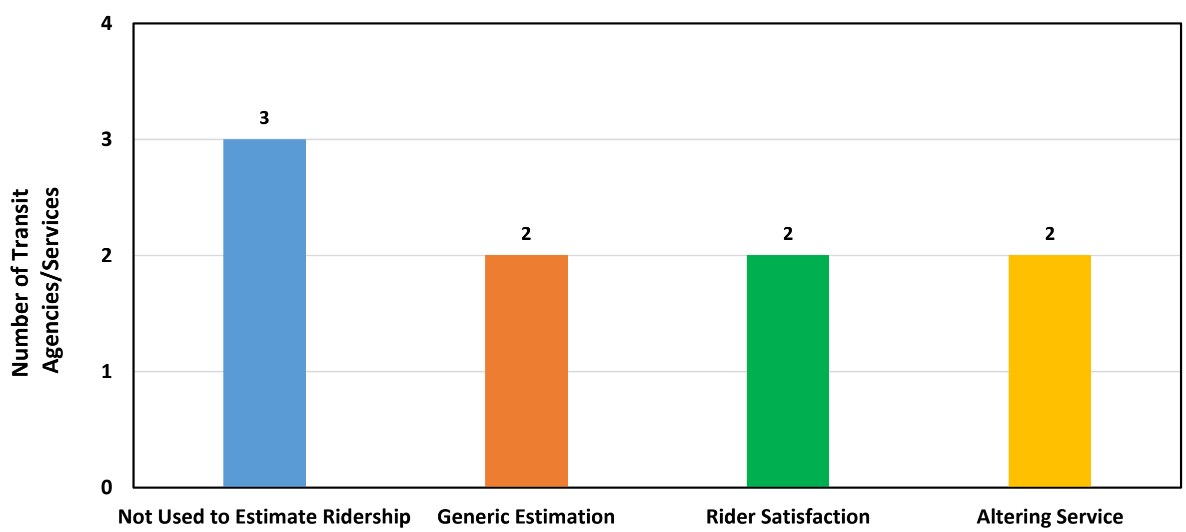

Figure 4-32: Survey Uses – Transit Agencies/Services in Oregon

4.2.10 Transit Agency Staff Resources for Ridership Data

An agency will require some dedication of staff resources to make ridership data available. The questionnaire inquired "[w]hat is the total employee FTE [full time equivalent] allocated to the collection and management of ridership data?"

Due to some clear outliers in the responses, a filter based on FTE-to-fleet-size was established and applied. Based on this filter, respondents with a FTE-to-fleet-size ratio greater than one standard deviation from the mean were eliminated. The result was a mean FTE-to-fleet size ratio of 0.092 (e.g., an agency with 100 vehicles in their fleet would have approximately 9 FTE allocated to the collection and management of ridership data). This values appears high, and if agencies with fewer than 100 vehicles in their fleet are excluded, the mean ratio drops to 0.004 (i.e., less than 1/20th the ratio after only the standard deviation filter was applied, and translating to a 100-vehicle fleet having .4 FTE allocated to the collection and management of ridership data).

The suggestion from these results is that there exists a baseline of resources required for ridership data management and analysis, and that such is disproportionately burdensome on smaller transit agencies.

4.2.11 Ridership Data Transfer Methods

A total of 27 transit agencies/services in Oregon reported having means of transferring ridership data collected digitally. As shown in Figure 4-33, the most common means was “Real-time dynamic or periodic remote retrieval” with eight transit agencies/services (i.e., 29.6%) currently using this approach. The prevalence of remote retrieval in Oregon is very similar to that seen with out-of-state agencies (i.e., 28.6%).

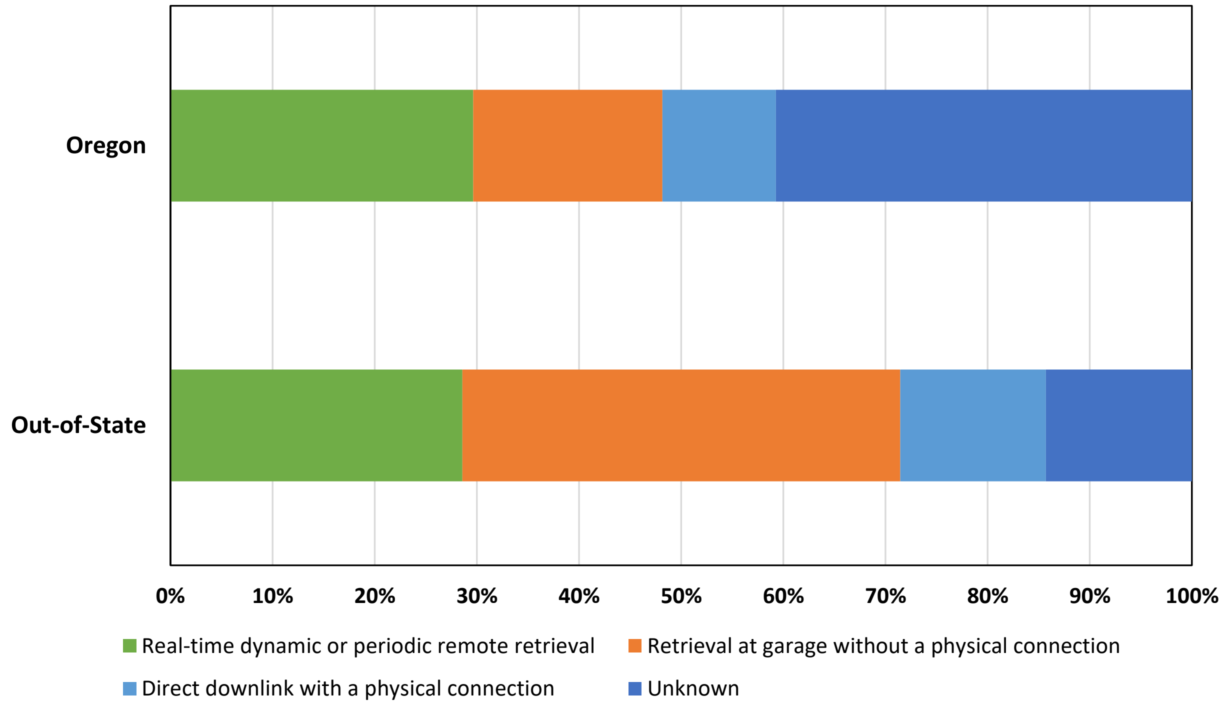

Figure 4-33: Ridership Data Retrieval Methods – Oregon vs Out-of-State

A plurality of transit agencies/services in Oregon (i.e., 11 out of 27 or 40.7%) were unsure as to the means of data transfer. For Oregon respondents, this does appear to correlate to smaller fleet sizes as depicted by Figure 4-34. The greatest proportion of unknown data retrieval methods is from agencies with 11-20 vehicles in their fleet followed by agencies whose fleets have 1-10 vehicles.

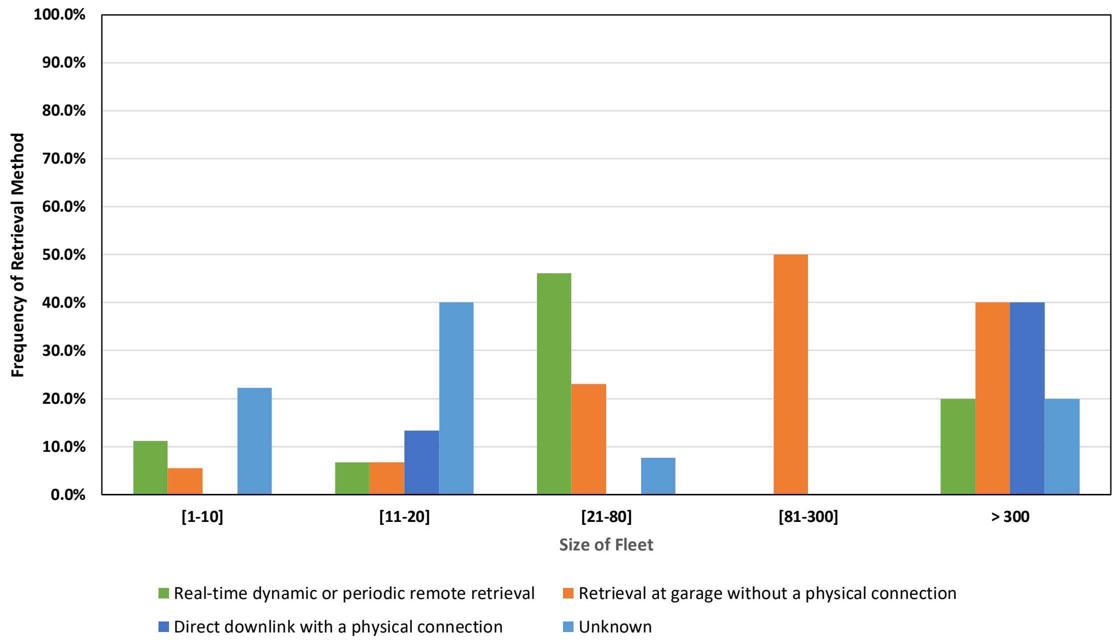

Figure 4-34: Ridership Data Retrieval Methods – Agency Fleet Size

4.2.12 Ridership Data Storage Location

The storage location of ridership data is influenced by the fleet size of the transit agency. As depicted in Figure 4-35, local hardcopy storage is preferred by 33.3% of respondents whose fleet includes 10 vehicles or less. This is the most common means for storage for transit agencies/services of that size alongside local network storage. Local hardcopy storage is still used by some transit agencies/services with 11 to 20 vehicles in their fleet, but it is no longer the most common method. No transit agencies/services with more than 21 vehicles employ local hardcopy storage.

Figure 4-35 also shows that the use of local network storage continually increases with the size of an agency's fleet, with a low of 33.3% for agencies with 10 or fewer vehicles to a high of 60% for agencies with more than 300 vehicles in their fleet.

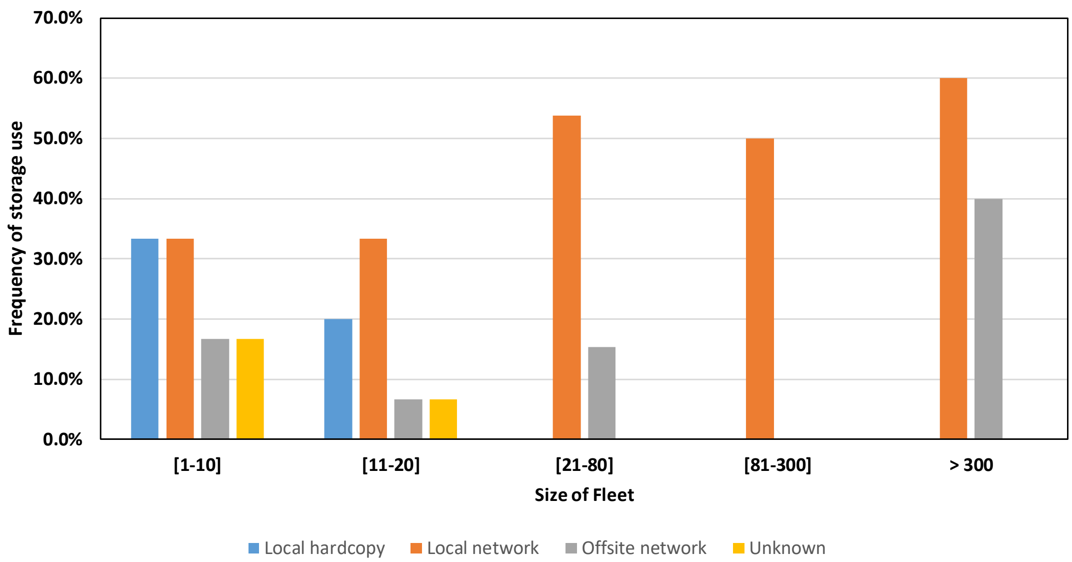

Figure 4-35: Ridership Data Storage Location – Agency Fleet Size

4.2.13 Ridership Data Storage Format

A total of 37 Oregon transit agencies/services reported on the format in which they store their ridership data. As depicted in Figure 4-36, Excel spreadsheets were the predominant data storage medium used by Oregon transit agencies/services (i.e., 62.2%) compared to 40.0% for out-of-state transit agencies/services. Out-of-state transit agencies/services preferred the use of relational databases much more (i.e., 100% vs. 21.6%) when compared to Oregon transit agencies/services.

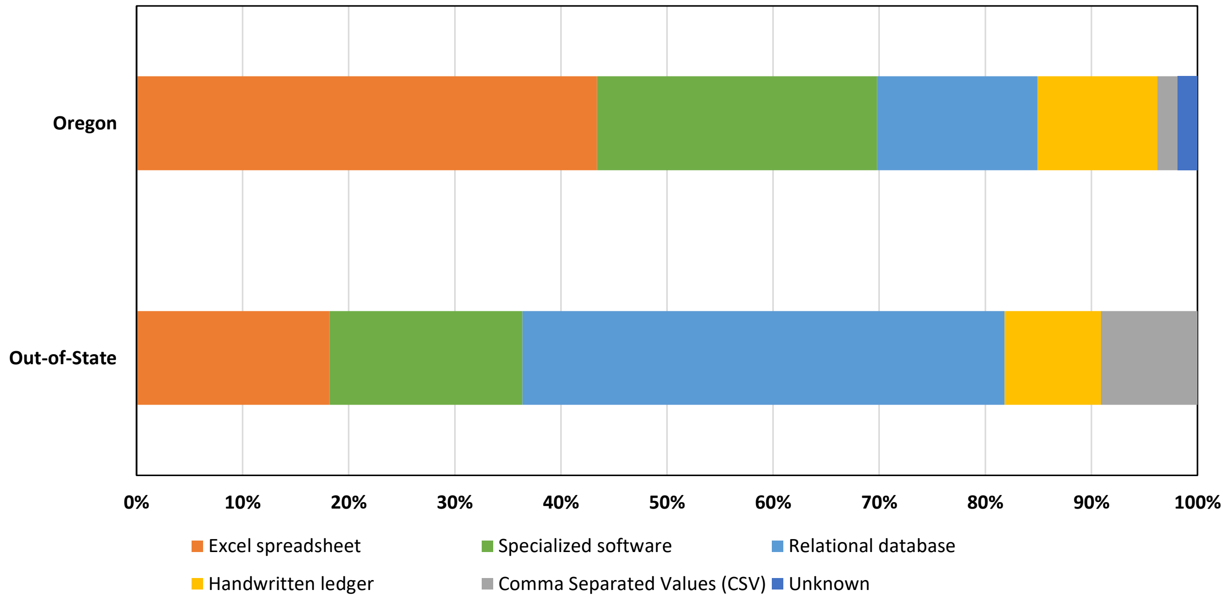

Figure 4-36: Ridership Data Storage Format – Oregon and Out-of-State Transit Agencies/Services

The difference in storage format does not appear as significant when examining fleet size. Transit agencies/services with a fleet size of fewer than 300 vehicles consistently reported the use of Excel spreadsheets and specialized software. It is only for transit agencies/services with fleet sizes of greater than 300 vehicles where Excel is present in fewer than 25% of responses, as illustrated in Figure 4-37.

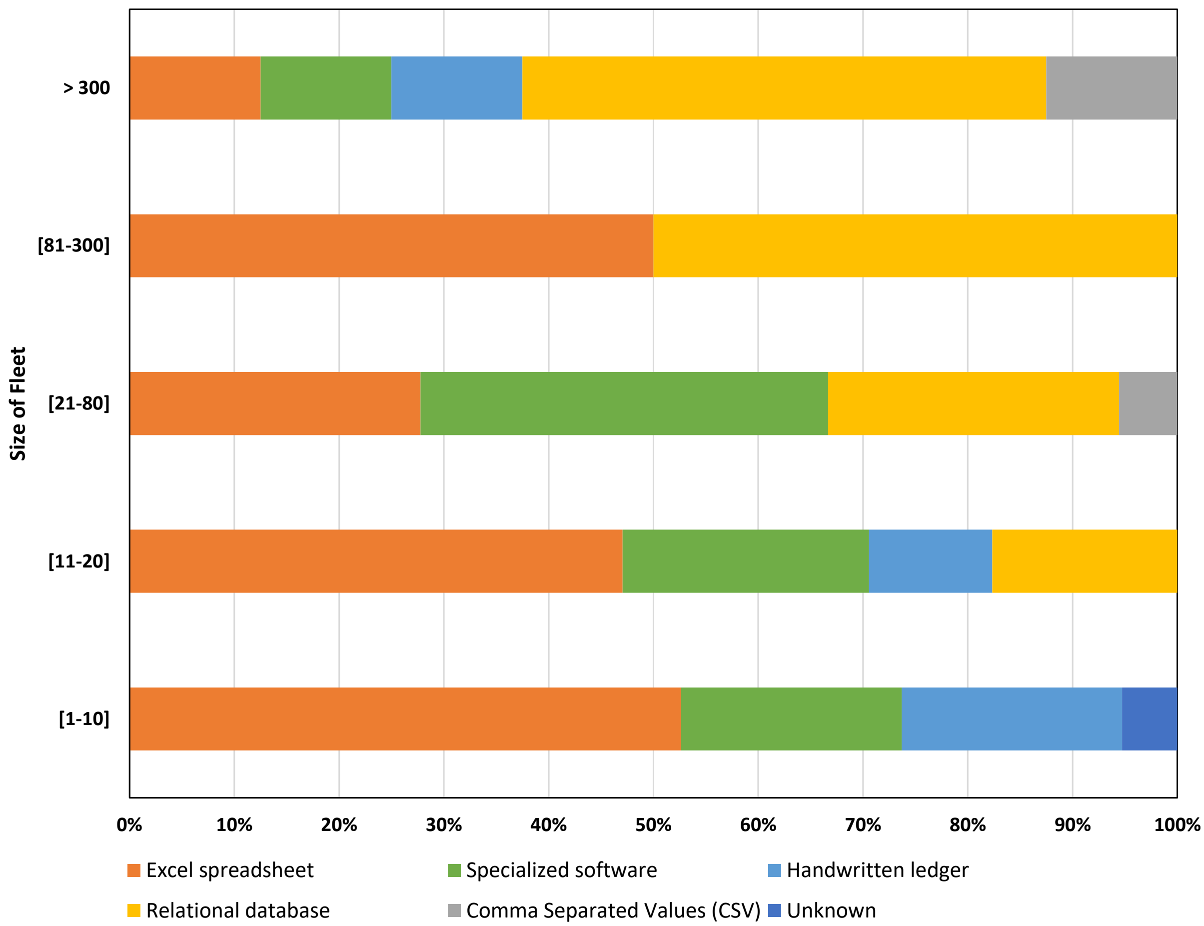

Figure 4-37: Ridership Data Storage Format – Agency Fleet Size

4.2.14 Ridership Data Access

A total of 34 Oregon transit agencies responded to the question of whether their ridership data is considered private or open. As seen in Figure 4-38, 11 Oregon transit agencies/services (i.e., 32%) indicated that their ridership data was considered open. This compares favorably to the results obtained from out-of-state transit agencies/services, depicted in Figure 4-39, which show that 20% consider their data open.

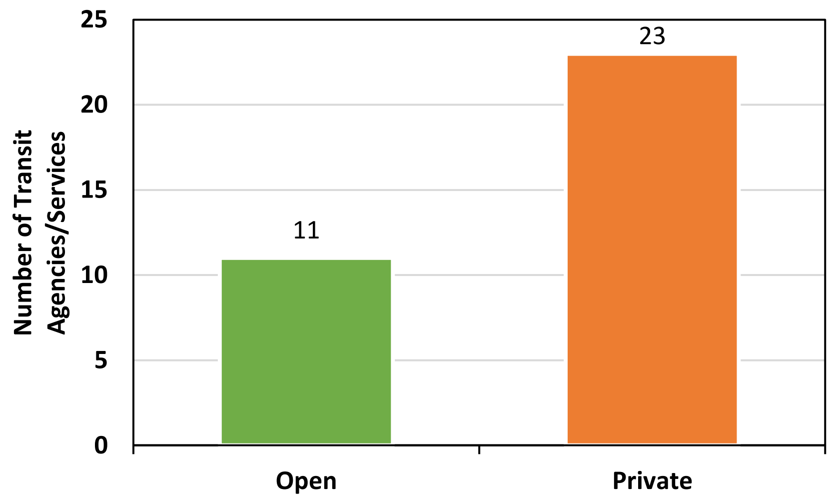

Figure 4-38: Ridership Data Accessibility – Oregon Transit Agencies

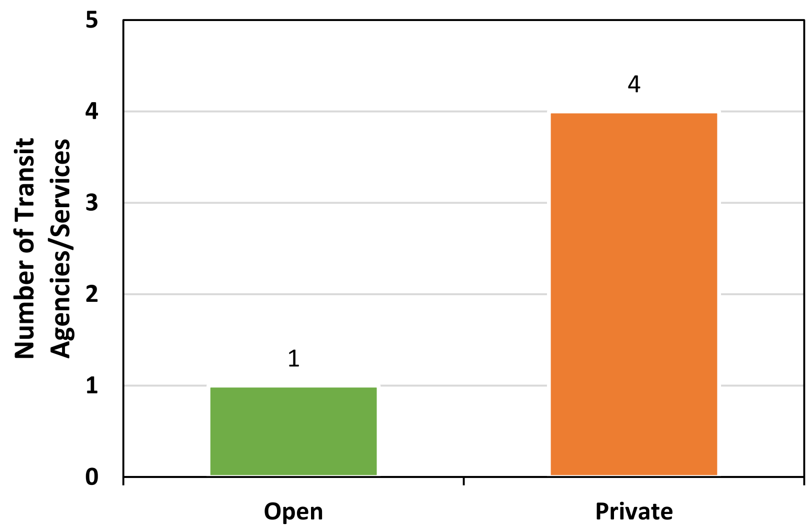

Figure 4-39: Ridership Data Accessibility – Out-of-State Transit Agencies

As seen in Figure 4-40, the fleet size of a transit agency/service was not shown to be a significant contributor of ridership data being considered open by a transit agency/service. Fleets with 81 to 300 vehicles only had two responses, both of which had private data, but with such a small sample size, it is difficult to draw conclusions.

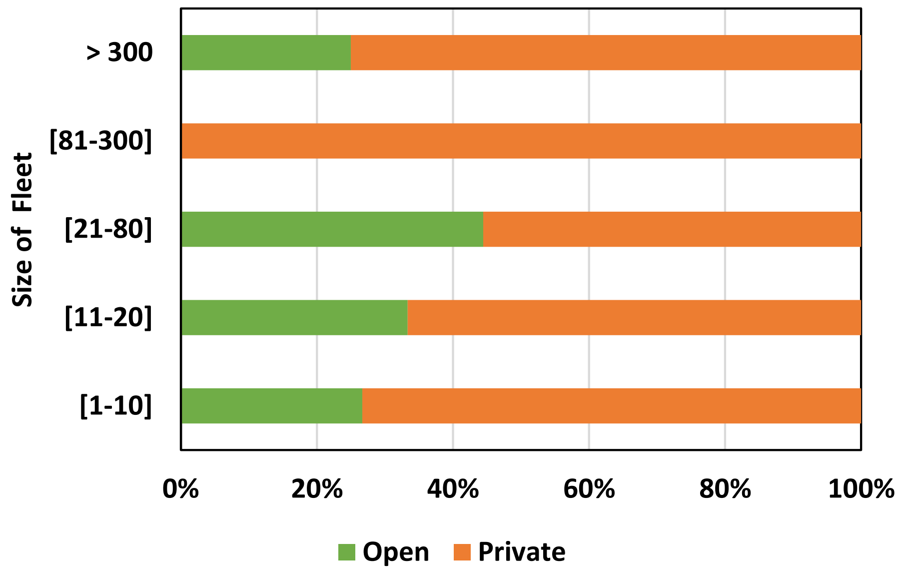

Figure 4-40: Ridership Data Accessibility – Agency Fleet Size

The online questionnaire also allowed a text response for respondents to elaborate about the "openness" of their ridership data. In categorizing responses, transit agencies/services that described their data as open with significant administrative action required to access it had their process classified as private. Furthermore, since the online questionnaire did not provide a specific definition of open data, some transit agencies took an existing process to release data as such data being open. In keeping with accepted intent of open standards and data, accessibility is a key consideration in determining status.

4.2.15 Ridership Data Uses and Employment

A total of 14 categorical uses of ridership data were recorded. Twelve of these were multiple choice options and two were developed from text responses. As depicted in Figure 4-41, in all but two uses (i.e., validate travel demand models and generic funding), out-of-state transit agencies/services used ridership data more frequently than Oregon transit agencies/services. No noteworthy difference was observed with ridership data use as agency size varied.

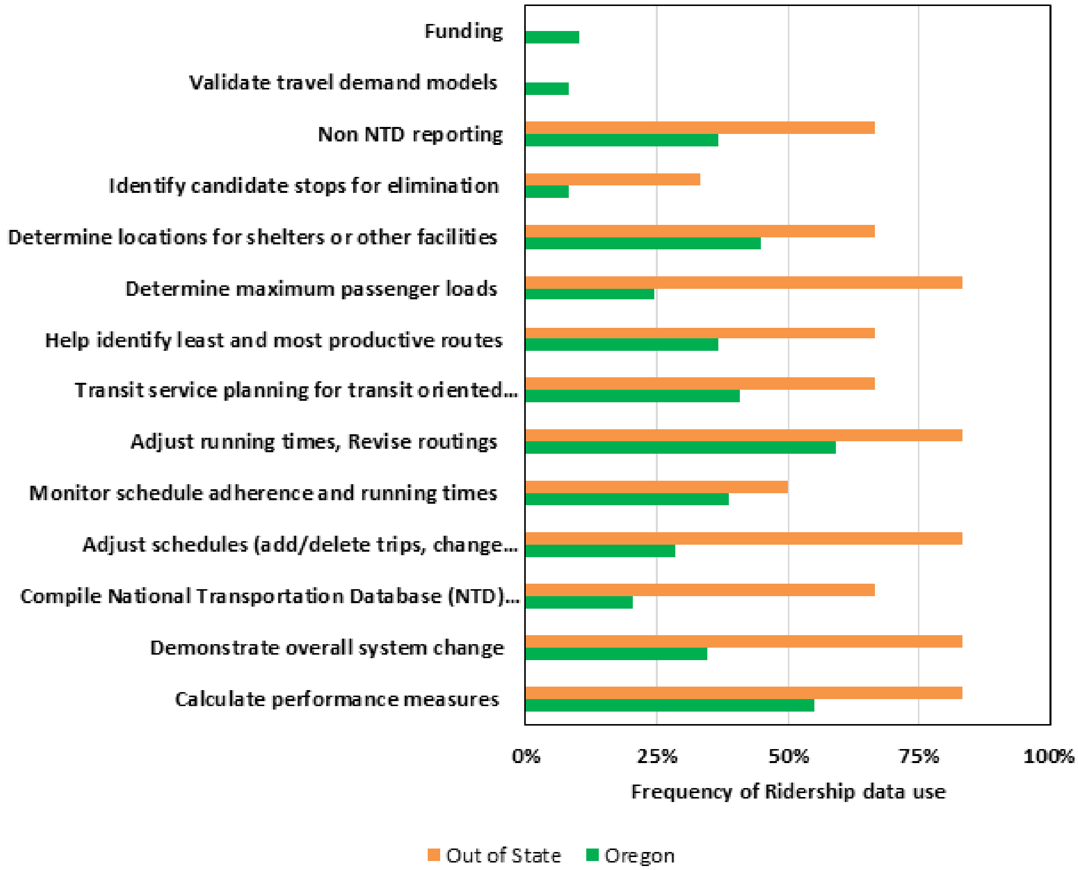

Figure 4-41: Ridership Data Uses – Oregon vs Out-of-State

4.2.16 Ridership Data Processing Steps

Thirty-three Oregon transit agencies/services provided 77 responses for steps taken to process their ridership data. The responses (i.e., “Compare with fare revenue,” “Look for unexplained variations across trips,” “Compare totals across days,” “Rely on the professional judgment of planner’s schedules,” “Use an automated program to analyze data,” “Compare boarding/alighting totals and adjust as needed,” and “Compare with manual counts”) were answered evenly throughout transit agencies of differing sizes. On average, 11 agencies responded to using each of the processing steps.

4.2.17 Ridership Data Analysis Tools

Oregon transit agencies/services provided 35 responses for analysis tools used. Of those responses, nine (i.e., 25.7%) use staff analysis and eight (i.e., 22.9%) use Excel. The remaining responses (i.e., “Unknown,” “Ecolane,” “RouteMatch,” “Transit Ace,” “NTD reporting,” “Access,” “ArcMap,” “SAS,” “SQL,” “Adept,” “TBEST,” and “Ride Express”) were marked by three or fewer agencies. No significant differentiation was observed between transit agencies/services with different fleet sizes.

4.2.18 Ridership Data Reporting Requirements

Oregon transit agencies/services provided 44 responses for reporting requirements of ridership data. Of those responses, 12 (i.e., 27.3%) report to NTD, 10 (i.e., 22.7%) to state government, and eight (i.e., 18.2%) for internal use. The remaining responses (i.e., “Partner organizations,” “City government,” “County government,” “Regional government,” and “Grant funding”) were used by four or fewer agencies. No significant differentiation was observed between transit agencies/services of different fleet sizes.

4.2.19 NTD Validation for Ridership Data

A total of 22 Oregon transit agencies/services responded regarding the steps taken with their ridership data for purposes of NTD validation. As depicted in Figure 4-42, 10 Oregon transit agencies/services (i.e., 45%) indicated that they were unsure about what steps were taken to validate their data for NTD reporting, and some indicated not knowing the purpose or process of NTD validation. This may be because not all respondents would necessarily be NTD reporters, but it indicates a disparity in knowledge regarding NTD procedures amongst those identified as transit agencies/services for the online questionnaire.

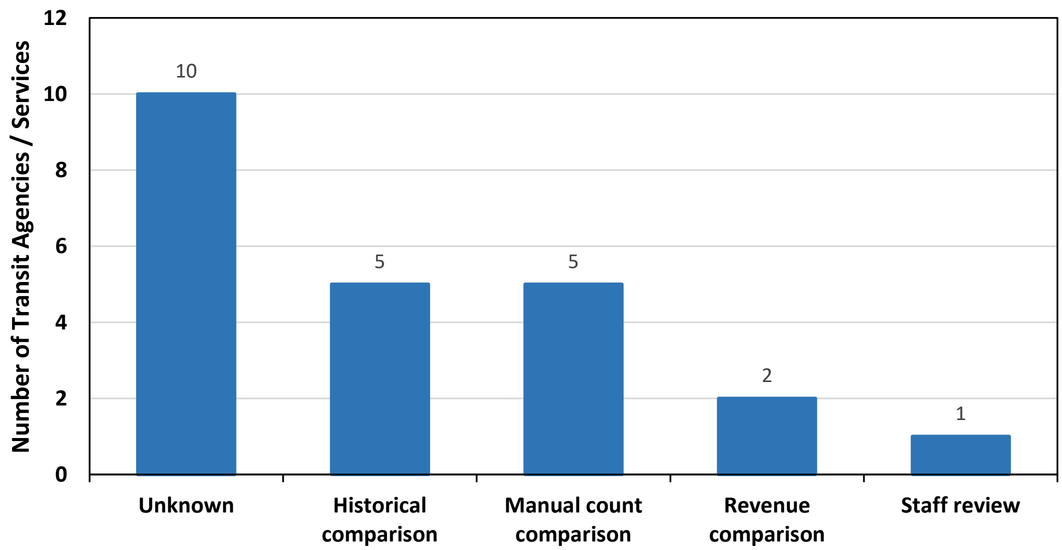

Figure 4-42: NTD Validation for Ridership Data – Oregon Transit Agencies

A large number of “Unknown” responses about methods of NTD data validation were tied to smaller Oregon transit agencies/services. As seen in Figure 4-43, no transit agencies/services with a fleet size larger than 81 vehicles responded as not knowing their NTD validation methods.

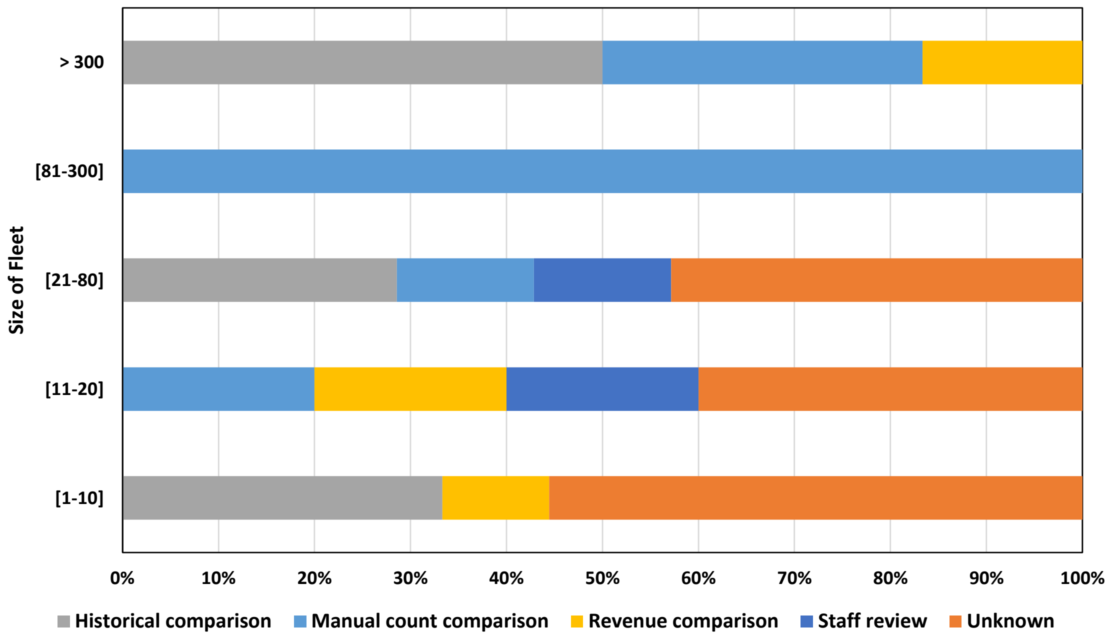

Figure 4-43: NTD Validation for Ridership Data – Agency Fleet Size

Out-of-state transit agencies/services rely mostly on historical comparison for their NTD validation methods, which is also the most common recorded method of active NTD comparison provided by Oregon transit agencies/services. As Figure 4-44 and Figure 4-45 show, once Oregon respondents with unknown validation methods are removed, Oregon and out-of-state transit agencies/services employ validation methods at similar rates.

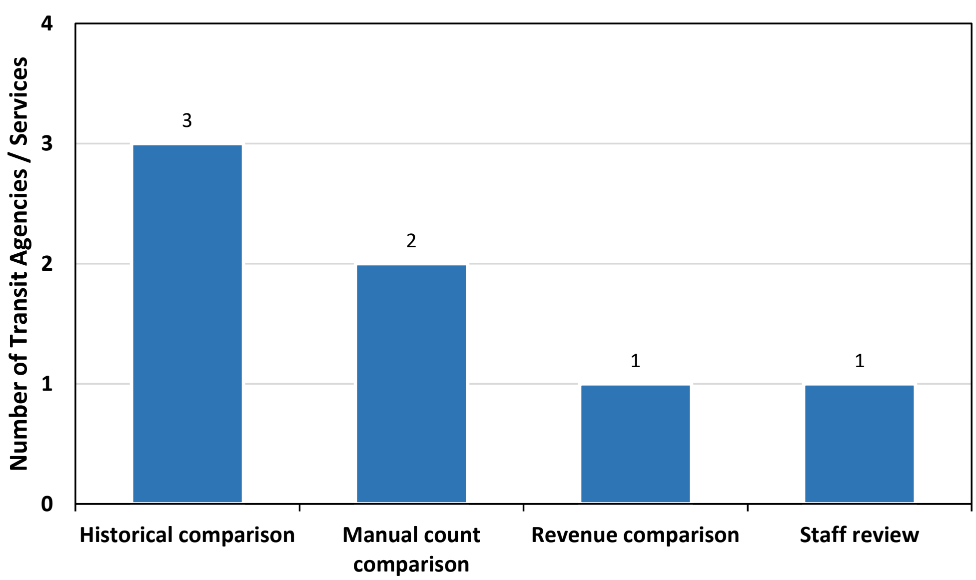

Figure 4-44: NTD Validation for Ridership Data – Out-of-State Transit Agencies

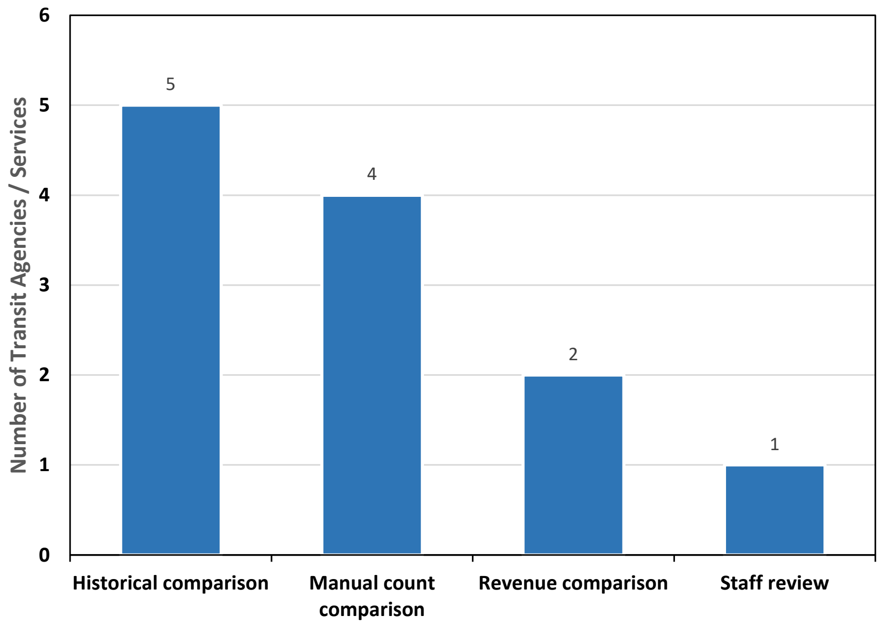

Figure 4-45: NTD Validation for Ridership Data – Oregon Transit Agencies with Only Known Methods

4.3 STATE OF PRACTICE SUMMARY

4.3.1 State of Ridership Data Quality and Collection

For Oregon transit agency/service respondents, the majority of ridership data is collected for Bus and Commuter bus modes at low resolution levels (i.e., route and trip levels), with a more moderate amount of higher resolution stop-level data collection. The generally larger out-of-state agency respondents report a higher proportion of stop-level ridership data collection, but also note relatively high proportions of route, trip, and system level ridership data.

For Oregon transit agencies/services, route- and system-level ridership data was reported as mostly available on a daily basis, while fewer reported having stop-level data on a daily basis. Conversely, out-of-state agencies show a higher proportion of daily stop-level boarding and alighting data availability than in-state agencies. A ridership data standard would likely benefit those currently collecting high-resolution data frequently, and possibly motivate and direct those agencies who would like to improve the quality of their collected ridership data.

As depicted in Figure 4-1(see section 4.1.1), 12 out of 33 questions of the questionnaire concentrate on developing a better understanding of issues associated with data collection and other related concerns. In present practice, the lack of automated ridership data collection means was clear, especially for smaller, rural, and/or non-fixed route agencies/services. The majority of agencies and service providers still rely on manual ridership data collection and cite presently unaddressed hurdles which prevent them from adopting automated means. While some smaller transit agencies/service did comment that automated data collection is not currently warranted or practical, several others expressed interest in exploring the option if it were more feasible under existing constraints. There was a trend noted in some agencies of increasing use of MDTs, AFC systems, and specialized transit software. Again, a ridership data standard would help provide guidance and direction to agencies interested in advancing their data collection practices and methods or in updating their data collection technology.

4.3.2 State of Ridership Data Analysis

In present practice, the degrees of ridership data analysis are highly correlated with the size of the fleet and the availability of resources at the different transit agencies. With regards to this ridership data category, the following key trends are noted:

- Agencies with the largest fleets have dedicated systems and resources for analysis and planning, while agencies of smaller scales often rely on staff experience and spreadsheets for planning decisions.
- Compared to larger agencies, smaller agencies report greater staff resources (relative to the size of the agency) being put toward ridership data analysis.
- As smaller agencies are dedicating more resources while still not adopting more advanced analysis, reducing the barriers for smaller agencies advancing their ridership data techniques should advance the state of their ridership data analysis significantly.

The development of a ridership data standard would support larger agencies by easing the sharing of ridership data with interested parties (much as occurred with GTFS) and reducing some barriers for smaller agencies by allowing common analysis services amongst such agencies.

4.3.3 State of Ridership Data Availability

The state of data availability and management closely mirrors that of data analysis. There does not yet exist a widely used ridership data standard. Therefore, collection of responses from agencies was undertaken with the understanding that the distribution method of ridership data employed by each agency may be unique. With regards to ridership data availability, the following key trends are noted:

- Oregon agencies have a greater rate than out-of-state respondents for open ridership data, but 68% of Oregon agencies still consider their ridership data private.
- There does not exist a shared understanding amongst Oregon transit agencies of the requirements of open ridership data.

Increasing the number of transit agencies releasing their data will benefit analysis and planning, and providing a data standard for distribution would simplify use of the data.

5.0 CONCLUSIONS AND OPPORTUNITES FOR FUTURE WORK

5.1 CONCLUSIONS

The review of the current state of ridership data reveals clear opportunities for improvement and growth. The amount and resolution of ridership data collected across agencies varies, and many still collect data using infrequent and/or inaccurate means. For efficient analysis and continued progress to occur, a low cost, accurate, simple, and unified means of data collection and management should be developed. The specific approaches will need to vary based on the mode and service type for which it is intended, but a common core of reliability, accuracy, and comparability should exist throughout.

Providing smaller agencies with approachable, affordable, and useful means to collect, manage, and analyze ridership data would address one of the clearest deficits revealed by responses. At present, the varying types of services provided and the limited resources available to smaller agencies appear to have slowed the adoption of precision automated ridership data collection, making access to ridership data further impeded. Furthermore, questionnaire results have shown that, stemming from the coverage goals of smaller agencies, high priority is not always placed on ridership data. This may point to a need for further incentives and communication of benefits for some agencies. The existing literature shows many opportunities for advancing ridership data, with promise from new technologies and methods and guidance for implementing proven processes.

5.2 OPPORTUNITIES FOR FUTURE WORK

The disjointed nature of current research highlights the fractious nature of ridership collection means and analysis. Therefore, a significant opportunity for future work is to develop a unified platform on which data are collected and analyzed to streamline progress in the field. Much as open approaches for GTFS have spurred the development of transit scheduling and rider planning, an open approach for ridership data may yield similar results.

The evolution of GTFS informs that a similar evolution will be required for a ridership data standard. The framework for such an evolution and the continued support and maintenance of the standard will be vital to its success. The standard must also be flexible in order to provide input means for both those still relying on sampled manual collection as well as those with fully automated stop-level data collection means.

6.0 REFERENCES

Barabino, B., Di Francesco, M., & Mozzoni, S. (2014). An offline framework for handling automatic passenger counting raw data. *IEEE Transactions on Intelligent Transportation Systems, 15*(6), 2443-2456.

Bauer, D., Ray, M., & Seer, S. (2011). Simple sensors used for measuring service times and counting pedestrians: Strengths and weaknesses. *Transportation Research Record: Journal of the Transportation Research Board,* (2214), 77-84.

Boyle, D. K. (2006). *TCRP synthesis 66: Fixed-route transit ridership forecasting and service planning methods*. Washington, D.C.: Transportation Research Board, National Research Council.

Boyle, D. K. (2008). *TCRP synthesis 77: Passenger counting systems*. Washington, D.C.: Transportation Research Board, National Research Council.

Boyle, D. K. (1998). *TCRP synthesis 29: Passenger counting technologies and procedures*. Washington, D.C.: Transportation Research Board, National Research Council.

Bureau of Transportation Statistics. (2016). *Passenger travel facts and figures 2016.* Washington, D.C.: U.S. Department of Transportation.

Chandesris, M., & Remy, A. (2016). New (big) data for mobility knowledge and management. some experiences and open challenges. Paper presented at the *Transit Data 2016: 2nd International Workshop on Automated Data Collection Systems*, Northeastern University, Boston, Massachusetts.

Chu, X. (2004). *Ridership models at the stop level.* (Final Report No. NCTR-473-04, BC137-31). Tampa, FL: National Center for Transit Research (NCTR).

Chu, X., Polzin, S. E., Pendyala, R. M., Siddiqui, N. A., & Ubaka, M. (2007). Framework of modeling and forecasting stop-level transit patronage. Paper presented at the *Transportation Research Board 86th Annual Meeting,* (07-1632)

Dill, J., Schlossberg, M., Ma, L., & Meyer, C. (2013). Predicting transit ridership at the stop level: The role of service and urban form. Paper presented at the *92nd Annual Meeting of the Transportation Research Board, Washington, DC,*

Dunlap, M., Li, Z., Henrickson, K., & Wang, Y. (2016). Estimation of origin and destination information from bluetooth and wi-fi sensing for transit. Paper presented at the *Transportation Research Board 95th Annual Meeting,* (16-6837)

Eros, E., Mehndiratta, S., Zegras, C., Webb, K., & Ochoa, M. (2014). Applying the general transit feed specification to the global south: Experiences in Mexico City, Mexico—and beyond. *Transportation Research Record: Journal of the Transportation Research Board,* (2442), 44-52.

Fayyaz, S., Kiavash, S., Liu, X. C., & Porter, R. J. (2016). A genetic-algorithm and regression-based model for analyzing fare payment structure and transit dwell time. Paper presented at the *Transportation Research Board 95th Annual Meeting,* (16-4815)

Federal Transit Administration, Office of Budget and Policy. (2016). *2016 NTD policy manual.* Washington, D.C.: U.S. Department of Transportation. Retrieved from https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/Report%20Year%202016%20Policy%20Manual_1.pdf

Fihn, J., & Finndahl, J. (2011). *A framework for how to make use of an automatic passenger counting system*. (Master's Thesis). Uppsala University, Uppsala, Sweden.

García-Bunster, G., & Torres-Torriti, M. (2008). Effective pedestrian detection and counting at bus stops. Paper presented at the *Robotic Symposium, 2008. LARS'08. IEEE Latin American,* 158-163.

Google Maps Transit. Accessed on 2 Dec. 2016. https://maps.google.com/landing/transit/index.html

Greneker, E., Murphy, K., Johnson, B., & Rausch, E. (1996). *Improved passenger counter and classification system for transit operations.* (Technical report). Washington DC: Transportation Research Board of The National Academies.

Hodges, C. C. (1985). *Automatic passenger counter systems: The state of the practice. final report.* (No. DOT-I-87-36). Washington, D.C: Office of Planning Assistance, Urban Mass Transportation Administration.

Kimpel, T., Strathman, J., Griffin, D., Callas, S., & Gerhart, R. (2003). Automatic passenger counter evaluation: Implications for national transit database reporting. *Transportation Research Record: Journal of the Transportation Research Board,* (1835), 93-100.

Kirby, J. (2016). *Using security cameras to count transit passengers*. (Project Abstract). North Carolina State University, Raleigh, NC. (FHWA/NC/2017-16). Retrieved from https://trid.trb.org/view.aspx?id=1424438

Kostakos, V., Camacho, T., & Mantero, C. (2013). Towards proximity-based passenger sensing on public transport buses. *Personal and Ubiquitous Computing, 17*(8), 1807-1816.

Kotz, A. J., Kittelson, D. B., & Northrop, W. F. (2015). Novel vehicle mass-based automated passenger counter for transit applications. *Transportation Research Record: Journal of the Transportation Research Board,* (2536), 37-43.

Kuutti, J. (2012). *A test setup for comparison of people flow sensors* (Licentiate's Thesis). Aalto University, Espoo, Finland.

Nassir, N., Khani, A., Lee, S., Noh, H., & Hickman, M. (2011). Transit stop-level origin-destination estimation through use of transit schedule and automated data collection system. *Transportation Research Record: Journal of the Transportation Research Board,* (2263), 140-150.

Oberli, C., Torres-Torriti, M., & Landau, D. (2010). Performance evaluation of UHF RFID technologies for real-time passenger recognition in intelligent public transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, *11*(3), 748-753.

Paget-Seekins, L., & Tribone, D. (2016). Releasing data to the public: APIs, dashboards, and data policy. Paper presented at the *Tranist Data 2016: 2nd International Workshop on Automated Data Collection Systems,* Northeastern University, Boston, Massachusetts.

Pelletier, M., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, *19*(4), 557-568.

Peterson, D. (2013). *Effortless passenger identification system: Final report. transit IDEA project 70*. Washington D.C.: Transportation Research Board, National Research Council.

Polzin, S. E. (2016). *Implications to public transportation of emerging technologies.* University of South Florida: National Center for Transit Research.

Porter, J. D., Kim, D. S., Ghanbartehrani, S., Mohseni, A. & Barahimi, P. (2016). *Transit network analysis software tool wiki*. Retrieved from https://tnasoftwaretool.engr.oregonstate.edu/TNAtoolAPI-Webapp/wiki/#!index.md

Reddy, A., Lu, A., Kumar, S., Bashmakov, V., & Rudenko, S. (2009). Entry-only automated fare-collection system data used to infer ridership, rider destinations, unlinked trips, and passenger miles. *Transportation Research Record: Journal of the Transportation Research Board,* (2110), 128-136.

Saavedra, M. R. (2010). *An automated quality assurance procedure for archived transit data from APC and AVL systems* (Master’s Thesis). University of Waterloo, Waterloo, Ontario, Canada.

Schweiger, C. L. (2015). *TCRP synthesis 115: Open data: Challenges and opportunities for transit agencies.* Transportation Research Board, National Research Council.

Strathman, J. G., & Hopper, J. R. (1991). Evaluation of automatic passenger counters: Validation, sampling, and statistical inference. *Transportation Research Record no. 1308, Public Transit Research: Rail, Bus, and New Technology*, (1308), 69-77.

Strathman, J., Kimpel, T., & Callas, S. (2005). Validation and sampling of automatic rail passenger counters for national transit database and internal reporting at TriMet. *Transportation Research Record: Journal of the Transportation Research Board*, (1927), 217-222.

Trépanier, M., & Morency, C. (2010). Assessing transit loyalty with smart card data. Paper presented at the *12th World Conference on Transport Research*. Lisbon, Portugal.

Trépanier, M., Morency, C., & Agard, B. (2009). Calculation of transit performance measures using smartcard data. *Journal of Public Transportation, 12*(1), 5.

Yahiaoui, T., Khoudour, L., & Meurie, C. (2010). Real-time passenger counting in buses using dense stereovision. *Journal of Electronic Imaging, 19*(3), 031202-031202-11.

Yang, T., Zhang, Y., Shao, D., & Li, Y. (2010). Clustering method for counting passengers getting in a bus with single camera. *Optical Engineering, 49*(3), 037203-037203-10.

APPENDIX A

ODOT SPR 803

This questionnaire will help the Oregon DOT and OSU researchers better understand your transit ridership data collection, storage, sharing, reporting, and analysis. Please answer each relevant question to the best of your knowledge. Please leave blank any question(s) that may not pertain to your transit agency or services. We thank you in advance for your support in answering this questionnaire.

1. Agency and Service(s) Name:

2. Number of vehicles in fleet:
   Owned (1)
   Contracted (2)

3. Type of Area Served (select all that apply):
- ❑ Urban (1)
- ❑ Rural (2)

4. What services are operated?
- ❍ Fixed and non-fixed route (1)
- ❍ Only fixed route (2)
- ❍ Only non-fixed route (3)

Answer If What services are operated? Only Non-fixed Route Is Not Selected

4a. What are the levels of ridership data collected for each of the following fixed-route modes?

	Stop (14)	Segment (15)	Trip (16)	Route (17)	System (18)	Other (19)	N/A (20)
Bus (124)	❑	❑	❑	❑	❑	❑	❑
Commuter Bus (125)	❑	❑	❑	❑	❑	❑	❑
Bus Rapid Transit (126)	❑	❑	❑	❑	❑	❑	❑
Trolleybus (127)	❑	❑	❑	❑	❑	❑	❑
Vanpool (128)	❑	❑	❑	❑	❑	❑	❑
Ferryboat (129)	❑	❑	❑	❑	❑	❑	❑
Commuter Rail (130)	❑	❑	❑	❑	❑	❑	❑
Hybrid Rail (131)	❑	❑	❑	❑	❑	❑	❑
Heavy Rail (132)	❑	❑	❑	❑	❑	❑	❑
Light Rail (133)	❑	❑	❑	❑	❑	❑	❑
Streetcar Rail (134)	❑	❑	❑	❑	❑	❑	❑
Monorail/Automated Guideway (135)	❑	❑	❑	❑	❑	❑	❑
Aerial Tramway (136)	❑	❑	❑	❑	❑	❑	❑

Answer If What services are operated? Only fixed route Is Not Selected

4b. How are ridership data collected on non-fixed route service(s) (e.g., Demand Response, Paratransit, etc.)?

Answer If What are the levels of ridership data collected for each of the following fixed-route modes? - Other Is Selected

4c. What other levels of ridership data are collected?

5. Please indicate the availability of data for each of the following items.

	Daily (1)	Weekly (2)	Monthly (3)	Quarterly (4)	Annually (5)	As Needed (6)	N/A (7)
System ridership (1)	❑	❑	❑	❑	❑	❑	❑
Route-level ridership (2)	❑	❑	❑	❑	❑	❑	❑
Route segment ridership (3)	❑	❑	❑	❑	❑	❑	❑
Stop-level boardings/alightings (4)	❑	❑	❑	❑	❑	❑	❑
Performance measures (5)	❑	❑	❑	❑	❑	❑	❑
Schedule adherence (6)	❑	❑	❑	❑	❑	❑	❑
Running Times (7)	❑	❑	❑	❑	❑	❑	❑

6. How are ridership data collected? Select all that apply.

❑ Automated passenger counters (APC) (1)
❑ Registering fareboxes (2)
❑ Handheld data collection units (e.g. Mobile Data Terminals - MDT) (3)
❑ Driver's trip log (4)
❑ Other: (please explain) (5) ____________________

7. Which supplementary details about ridership data are collected in addition to counts?

❑ Timestamps (2)
❑ GPS coordinates (3)
❑ Fare types (4)
❑ Transfer status (5)
❑ Special Rider Types (6)
❑ Bicycles (7)
❑ Wheelchairs (8)
❑ Other medical or mobility devices (9)
❑ Other details (please specify) (10) ____________________
❑ Please describe how the selected details are collected: (11) ____________________
❑ No supplementary details collected (1)

8. What percentage of the fleet (i.e., owned and subcontracted) is equipped with APCs?
______ Owned (1)
______ Subcontracted (2)

9. If not all vehicles in the fleet are equipped with APCs, how are the APC-equipped vehicles assigned to service runs?

10. What barriers, if any, does your organization see to implementing APC devices?

11. If combinations of automated and manual methods are used, please describe how each is used?

12. If automated methods are used, what are the manufacturers and models of the devices?

13. If manual surveys are conducted to collect ridership data, please answer the following questions:
How often are these surveys conducted? (1)
How many surveys are handed out? (2)
On what proportion of trips or routes do you apply these surveys? (3)
How are the survey data used to estimate ridership? (4)

14. If sampling methods are employed as an alternative to detailed ridership counts, please answer the following questions:
How often are these samples taken? (1)
What is the sample size? (2)
On what proportion of trips or routes do you perform sampling? (3)
How are the sample data used to estimate ridership? (4)

15. If the process(es) by which your agency collects ridership data have changed in the last three years, could you please explain the differences?

16. What is the total employee FTE allocated to the collection and management of ridership data?

17. How are the raw ridership data transferred from electronic collection devices to storage?
- ❑ Direct downlink with a physical connection (1)
- ❑ Retrieval at garage without a physical connection (2)
- ❑ Real-time dynamic or periodic remote retrieval (3)
- ❑ Removable storage medium (i.e., diskette, memory stick, memory card, etc.) (4)
- ❑ Other: (please explain) (5) ____________________

18. How are the raw and/or processed data physically stored (e.g., network drive, cloud storage, memory stick, files in filing cabinet, etc.)?

19. In what formats are ridership data stored?
- ❑ Comma Separated Values (CSV) (1)
- ❑ Excel spreadsheet (2)
- ❑ Relational database (3)
- ❑ Specialized software (4)
- ❑ Handwritten ledger (5)
- ❑ Other: (please specify) (6) ____________________

20. Are the ridership data considered open data or considered private to the agency?

21. If ridership data are made available to the public, please provide a URL.

22. What processing steps are needed to edit and validate collected ridership data?
- ❑ Compare with fare revenue (1)
- ❑ Look for unexplained variations across trips (2)
- ❑ Compare totals across days (3)
- ❑ Rely on the professional judgment of planner’s schedules (4)
- ❑ Use an automated program to analyze data (5)
- ❑ Compare boarding/alighting totals and adjust as needed (6)
- ❑ Compare with manual counts (7)
- ❑ Other: (please explain) (8) ____________________

23. What are the purposes for which ridership data are collected and used at your agency? Select all that apply.
- ❑ Demonstrate overall system change (1)
- ❑ Help identify least and most productive routes (2)
- ❑ Identify candidate stops for elimination (3)
- ❑ Determine maximum passenger loads (4)
- ❑ Monitor schedule adherence and running times (5)
- ❑ Calculate performance measures (6)
- ❑ Adjust schedules (add/delete trips, change headways) (7)
- ❑ Adjust running times, Revise routings (8)
- ❑ Determine locations for shelters or other facilities (9)
- ❑ Compile National Transportation Database (NTD) reports (10)
- ❑ Validate travel demand models (11)
- ❑ Transit service planning for transit oriented development (12)
- ❑ Other: (please explain) (13) ____________________

24. What methods and/or software tools are used to analyze collected ridership data?

25. What ridership data reporting requirements (internal and external) currently exist for your agency?

26. What are the resolution, frequency, and supplementary details required for reporting?

27. For NTD reporters, please describe any steps that are taken to validate ridership data for NTD reporting purposes.

28. Please provide any additional information that will aid in better understanding how your agency/service collects, stores, shares, reports, and analyzes ridership data.

29. Are you willing to provide contact information for possible follow-up questions?
- ❍ Yes (1)
- ❍ No (2)

Answer If Are you willing to provide contact information for possible follow-up questions? Yes Is Selected

30. Contact name:

Answer If Are you willing to provide contact information for possible follow-up questions? Yes Is Selected

31. Preferred method of contact:
- ❍ Telephone (1)
- ❍ E-mail (2)

Answer If Are you willing to provide contact information for possible follow-up questions? Yes Is Selected

32. Contact telephone:

Answer If Are you willing to provide contact information for possible follow-up questions? Yes Is Selected

33. Contact e-mail:

APPENDIX B

EMAIL INVITATION:

Dear ${m://FirstName} ${m://LastName},

The Rail & Transit Division of the Oregon Department of Transportation (ODOT) and Oregon State University (OSU) are collaborating on a project to develop a new public transit ridership data standard. More information about this project can be found by clicking here.

An online questionnaire has been developed to gain an accurate understanding of the current status of transit ridership data in Oregon. You are being invited to respond because you have been identified as a public/private transit service provider, at least 18 years of age, associated with collecting, storing, sharing, reporting, or analyzing ridership data. Completing the online questionnaire will take approximately 10 minutes. You will receive no financial compensation or other benefit for participating and there are no foreseeable risks to completing the online questionnaire.

It would be greatly appreciated if you could respond to the online questionnaire by 6pm Friday, October 7th to inform in the development of the standard. By completing and submitting the online questionnaire, you are giving us your informed consent. This means that this research study has been explained to you and that you agree to participate. We encourage you to print this page for your records.

TO ANSWER THE ONLINE QUESTIONNAIRE, ACCESS THE FOLLOWING URL:
${l://SurveyURL}

The regional transit coordinator (RTC) in your region, Karyn Criswell, is aware of this study and may be providing support with follow up emails in the upcoming weeks. For questions about the online questionnaire, contact Dr. J. David Porter at (541) 737-2446 or by email at david.porter@oregonstate.edu. For questions about your rights as a research participant, contact the Oregon State University Institutional Review Board (IRB) Human Protections Administrator at (541) 737-8008, by email at irb@oregonstate.edu, or by mail at B308 Kerr Administration Building, Corvallis, OR 97331-2140.

If you feel you are not the right person at your agency to answer this questionnaire, you may send the appropriate person's name, title, and contact in**formation to david.porter@oregonstate.edu.**

Thank you in advance for your participation.

Sincerely,

J. David Porter, Ph.D.
Oregon State University
(V) 541-737-2446
david.porter@orst.edu

APPENDIX C

LIST OF RESPONDENTS – OREGON TRANSIT AGENCIES/SERVICES

	AGENCY NAME
1	Basin Transit Service
2	Benton County Rural and Special Transportation
3	Blue Star Charters and Tours
4	Burns Paiute - Tribal Transit Services
5	Central Oregon Breeze
6	Central Oregon Intergovernmental Council
7	City of Albany, Albany Transit System
8	City of Hermiston Senior & Disabled Taxi Ticket Program
9	City of Lebanon Dial-a-Bus
10	City of Pendleton
11	City of Sandy, Transit
12	Clackamas County Social Services, Mt Hood Express
13	Community Action Program of East Central Oregon, Door to Door non-emergent medical transportation
14	Community Connection of Northeast Oregon
15	Community Connection of Wallowa County
16	Confederated Tribes of the Umatilla Indian Reservation/Kayak Public Transit
17	Coos County Area Transit Service District
18	Corvallis Transit System
19	Douglas County
20	Ecoshuttle
21	Gilliam County Transportation
22	Grant County Transportation District
23	Hood River County Transportation District
24	Josephine Community Transit, Josephine County - fixed route, commuter route, demand response and paratransit
25	Klamath Basin Senior Citizens' Center
26	Lake County Public Transit
27	Lane Transit District E&D Services
28	Lincoln County Transit
29	Malheur Council on Aging & Community Services
30	Mid-Columbia Council of Governments
31	MTR Western
32	Northeast Oregon Public Transit
33	Opportunity Foundation of Central Oregon
34	South Metro Area Regional Transit (SMART)
35	Rogue Valley Transit District (RVTD)
36	Salem Area Mass Transit District (Cherriots, CARTS, CherryLift, West Salem Connector, and RED Line)
37	Senior Citizens of Sweet Home, Inc. (Linn Shuttle) Linn Shuttle, Sweet Home Dial-A-Bus, Linn County DD transportation

	AGENCY NAME
38	Sherman County Community Transportation
39	South Lane Wheels
40	Sunset Empire Transportation
41	TAC Transportation, Inc.
42	The Klamath Tribes-Transportation
43	The Loop Morrow County Transportation
44	Tillamook County Transportation District
45	TriMet
46	Warm Springs Transit
47	Wheeler County Community Transportation

LIST OF RESPONDENTS – OUT-OF-STATE TRANSIT AGENCIES/SERVICES

	AGENCY NAME
1	Blacksburg Transit, a division of the Town of Blacksburg
2	Community Transit, Snohomish County, WA
3	King County Metro
4	RTC of Southern Nevada
5	San Diego Metropolitan Transit System
6	Transport For London