CAPABLE: Count All Pedestrian and Bicycle Locations Efficiently

State of the practice and recommendations for improving bike/ped data in TETC

Characterizes the state of the practice and makes recommendations for improving the availability, quality, and cost of bike and pedestrian data in The Eastern Transportation Coalition.

Final Report

September 2024

Prepared by:

Zachary Vander Laan
Senior Faculty Specialist
Center for Advanced Transportation Technology (CATT), University of Maryland

Stanley Young, PE, Ph.D.
Chief Data Officer
The Eastern Transportation Coalition

Executive Summary

This report presents findings from the CAPABLE (Count All Pedestrian and Bicycle Locations Efficiently) special cooperative project (SCOOP) – a cooperative project funded by states in The Eastern Transportation Coalition (TETC) that focuses on bike and pedestrian ("bike/ped") data. The project, which ran from December 2023 – September 2024, characterizes the state of bike/ped data and proposes steps the Coalition can take to advance its availability, quality, and cost.

Transportation agencies across The Eastern Transportation Coalition have shown interest in quantifying bicycling and pedestrian activity to support planning and safety initiatives – examples of which include determining where to plan and design bike/pedestrian infrastructure, assessing project impacts, contextualizing crash statistics, and identifying economic opportunities for local investment. Traditionally, measuring bike/ped activity has required deploying permanent or portable sensors to count pedestrians and cyclists at strategic locations. More recently, private industry has introduced novel data products that use Location Based Service (LBS) data from cell phones, and other sources to estimate bike and pedestrian activity at large spatial scales -- analogous to similar products that have been developed for estimating motorized traffic volumes. Several of these industry-sourced data sets, referred to as "estimated/crowdsourced bike/ped products", are currently sold in the Coalition's Transportation Data Marketplace (TDM) as ancillary products and are expected to become core data products in the next TDM procurement cycle.

A core emphasis of the CAPABLE project was to document the current state of bike/ped data – both in terms of Coalition agency count programs and emerging commercial data products. To this end, the project team gathered information from several sources, including (i) conversations with a steering committee composed of bike/ped data professionals from Coalition states and other external subject matter experts, (ii) a survey of state and local agencies within the Coalition, (iii) briefings by industry leaders, and (iv) a review of existing literature. The following takeaways emerged from this process:

- **The accuracy of estimated / crowdsourced bike/ped products depends on count data collected by governmental agencies, non-profit organizations, and agency support consultants.** Even with high-quality data sources, modeled estimation products require accurate, standardized counts to calibrate their models, and vendors have identified the lack of widely available calibration count data as a key impediment to improving accuracy.

- **Bike/ped count programs *do* exist in the Coalition and beyond, but data tends to be siloed across organizations and stored in different formats**. Data collection is usually executed by local agencies (cities and MPOs) to support diverse use cases, but the extent to which this information is shared with state DOTs or other entities varies state-by-state.

- **Agencies face institutional challenges and resource constraints**. In the absence of a federal mandate, many agencies do not prioritize bike/ped data collection, which often includes labor-intensive standardization efforts ("data wrangling"). As such, there is significant variation in the data collected across states and metro areas, depending on priorities, budgets, and presence of local champions within agencies.

- **Efforts to aggregate bike/ped data are in their infancy.** FHWA is trying to be a national aggregator of bike/ped data, but there are challenges to widespread adoption – in part

because submission is voluntary and can be technically demanding. Other multi-agency repositories exist, but either have limited adoption or are limited to single states.

- **Technology advancements are making computer vision-based data collection more attractive.** Technology, particularly video image processing technology enabled by AI/ML, is expected to improve the cost and quality of bike/ped data collection.

- **Better mapping and location referencing are needed for bike/ped infrastructure.** Industry is beginning to acknowledge that georeferencing and mapping for bicycle and pedestrian facilities needs to be more sophisticated. A current effort by the Bureau of Transportation Statistics, called NC-BPAID, is working towards developing a national bike/ped infrastructure data standard.

In light of these findings, the project team developed a framework for improving the quality, cost, and availability of bike/ped data within Coalition, and highlighted potential opportunities for Coalition involvement (see Section 4). Based on committee feedback and additional stakeholder conversations, the project team recommends the following next steps:

- **Establish an ongoing Coalition bike/ped committee** to ensure that data formats, specs, and future procurement of bike/ped data in the TDM align with Coalition member needs. The steering committee formed for the CAPABLE was highly engaged and provided invaluable feedback on the use cases, challenges, and insight into the nuances of bike/ped data collection within the Coalition. Re-forming – and perhaps expanding – the committee is a critical first step to any follow-up work. In addition to providing feedback on data formats and specifications, the committee will also be invited to inform/consult on Traffic Data Marketplace (TDM) data validation activities related to bike/ped data.
- **Create a consistent data schema for bike/ped count data**, to be developed in coordination with practitioners and enforced through the TETC Transportation Data Marketplace procurement. The *bike/ped count schema* would unambiguously describe bike and pedestrian counts through a prescribed data format, making it easier to exchange count information and populate databases in a standardized way.

  Future data collection services through the TDM would be expected to adhere to this format (via RFP requirements), but it would also serve as a common format that legacy data could be translated to, and that agencies could use when hiring consultants to assist with data collection. It could also be used to develop a calibration database that would enable TDM vendors to improve estimation accuracy.

- **Establish Coalition resources to support agencies with bike/ped data collection and analysis.** This recommendation addresses the fact that bike/ped data collection is often (though not always) undertaken at the local level, and many small agencies/municipalities do not have internal resources or prior experience “wrangling” the data to common formats or implementing QA/QC methods. In response, the plan should be developed to:
  - Identify how the Coalition can support agencies with “data wrangling” and management
  - Provide ongoing education on best practices for bike/ped data collection and analysis (e.g., emerging technologies, QA/QC methods, etc). This may include written materials, webinars, or a training program.

Acknowledgements

The project team would like to thank the following members of the CAPABLE steering committee for providing guidance and feedback throughout the project.

TETC Member State Representatives

Anna Bergeron (Connecticut DOT)

Melanie Zimyeski (Connecticut DOT)

Christopher Roberts (Connecticut DOT)

Sam Harris (Georgia DOT)

Ron Knezevich (Georgia DOT)

Bob Frey (Massachusetts DOT)

Martha Koch (Massachusetts DOT)

Sarah Searcy (North Carolina DOT)

Eric Weis (Rhode Island DOT)

Mason Perrone (Rhode Island Department of Administration)

Will Rogers (Tennessee DOT)

David Cook (Virginia DOT)

Eric Courchesne (Alabama Transportation Institute)

External Representatives

Shawn Turner (Texas Transportation Institute)

Liz Stolz (SRF Consulting)

Josh Roll (Oregon DOT)

Andy Duvall (National Renewable Energy Lab)

Table of Contents

1. Introduction

Transportation agencies across The Eastern Transportation Coalition have shown strong interest in quantifying bicycling and pedestrian ("bike/ped") activity to support planning and safety initiatives. Safety applications are particularly relevant at the moment, as non-vehicular (bike/ped and other micromobility mode) injuries and fatalities have significantly and disproportionately increased relative to their motor vehicle counterparts in recent years (increase of 44% vs 22% from 2011-2020, respectively[1]).

Traditionally, measuring bike/ped activity has required deploying permanent or portable sensors to count pedestrians, cyclists, and other micromobility users (e.g., scooters, skateboards, wheelchair users) at strategic locations. More recently, private industry has introduced novel data products that use Location Based Service (LBS) data from cell phones, GPS data, and other sources to estimate bike and pedestrian activity at large spatial scales -- analogous to similar products that have been developed for estimating motorized traffic volumes.

Several of these industry-sourced data sets, referred to as **estimated or crowdsourced bike/ped data**, are currently sold in Coalition's Transportation Data Marketplace (TDM) as ancillary products, and are expected to become core data products in the next TDM procurement cycle. Although feedback from agencies using these products has been generally positive, the products are still in early-stage maturity and agencies would like to see improvements in quality. Preliminary conversations with vendors highlighted that one of the greatest impediments to improving accuracy is that high-quality bike/ped counts, which are needed for model calibration, are not widely available or standardized.

Recognizing the need for better bike/ped count data, the CAPABLE (**C**ount **A**ll **P**edestrians **A**nd **B**icycle **L**ocations **E**fficiently) project – a quick-response cooperative project funded by state agencies within the Coalition, was developed. The intent of this project was to (1) summarize the state of practice for acquiring bike/ped count data, including both traditional count programs and emerging data sources, (2) document the current bike/ped data collection activities taking place across TETC member states and share feedback on successes, challenges, and lessons learned, and (3) propose and demonstrate a recommended methodology for Coalition members to plan and execute bike/ped data collection, with regional cooperation and industry in mind. A steering committee composed of both Coalition representatives and national subject matter experts was assembled to provide guidance in alignment with these objectives.

The remainer of this report is organized in the following manner. Section 2, titled *State of Bike/Ped Data*, includes an inventory of bike/ped count programs across the Coalition, an overview of commercial offerings (both counters and estimated / crowdsourced data) and summary of lessons learned. Section 3, titled, *Bike/Ped Data Schemas*, investigates several existing data schemas for standardizing bike/ped data into a common format, including FHWA's Traffic Monitoring Guide (TMG), Portland State's BikePed Portal, and TTI/TxDOT's Bike and Pedestrian Count Exchange (BP|CX). Finally, Section 4, titled *Opportunities and Next Steps*, presents a framework of the bike/ped data ecosystem and makes recommendations for high-impact opportunities that can improve the availability, quality, and cost of bike/ped data to TETC members.

[1] 2024 USDOT Notice of Proposed Rulemaking

2. State of Bike/Ped Data

Although traffic monitoring has typically focused on motorized vehicles, there is growing interest in also measuring pedestrians, bicyclists, and other micromobility modes. Many agencies want to encourage using these modes, as they improve congestion and reduce emissions via fewer car trips, are broadly available to most of the population – not just people of driving age/ability or those who own a car -- and have positive health benefits. However, in the absence of data, it is difficult to quantify how many people use these modes, how travel patterns have changed over time, and to calculate crash risk exposure to contextualize trends in vulnerable road user injuries and deaths.

To support these and other use cases, bike/ped counts are typically collected using a combination of permanently-installed continuous counters and short-term counters that can be regularly deployed to different locations. Continuous counters are used to monitor traffic at important locations over time and are typically segregated into "factor groups" based on the types of facilities, modes, and patterns they capture, which can be later used to help estimate annual average daily traffic from short-term counts. In addition to AADT and factor groups, important statics include directional distribution (DD) and peak hour (K Factor).

However, as is the case with motorized traffic, non-motorized volumes are often needed at locations beyond what can be reasonably covered by permanent and portable counters. Emerging estimated/crowdsourced bike/ped products – including several data vendors selling ancillary products in the Coalition's TDM -- seek to solve this problem by estimating bike/ped volumes for large geographic scales using LBS data and other input sources. However, accurate and representative bicycle and pedestrian counts are needed to calibrate these products, which means that that their quality is linked to the availability and accuracy of underlying count programs (especially permanent counters).

Given the importance of collecting bike/ped counts – to support agency use cases *and* help improve estimated/crowdsource bike/ped products -- the project team initially set out to summarize the state of practice for setting up a robust non-motorized count program. However, a preliminary literature review and conversations with CAPABLE committee members indicated that this information is already well documented, and that it was not worth reproducing existing efforts. For completeness, several of the most relevant documents related to bike/ped counting are listed below.

- Bicycle and Pedestrian Count Programs: Summary of Practice and Key Resources (link)
- NCHRP Report 797: Guidebook on Pedestrian and Bicycle Volume Data Collection (link)
- NCHRP Web-Only Document 229: Methods and Technologies for Pedestrian and Bicycle Volume Data Collection: Phase 2 (link)
- 2022 Traffic Monitoring Guide (link)
- Coding Nonmotorized Station Location Information in the 2016 Traffic Monitoring Guide Format (link)
- Guide to Bike and Pedestrian Counting Programs (Portland State University, link)
- Collecting Network-wide Bicycle and Pedestrian Data: A Guidebook for When and Where to Count (Washington State DOT research report, link)
- Improving the Amount and Availability of Pedestrian and Bicyclist Count Data in Texas (TTI research report, link)
- Bicycle and Pedestrian Data Collection Manual (Minnesota DOT research report, link)

The remainder of this section will focus on the state of bike/ped counting efforts *within The Eastern Transportation Coalition* and characterize the state of industry bike/ped data products – including both count equipment/software and estimates of bike/ped activity.

Inventory of TETC Count Programs

The project team conducted an inventory of TETC agency count programs -- both at the state and local level – to characterize the extent to which agencies are collecting bike/ped counts, understand the motivating use cases, and summarize core challenges. The findings are based on results collected from an online survey and follow-up discussion with the steering committee.

Survey Methodology

A survey instrument was prepared using Qualtrics software and administered to state and local agency representatives across the Coalition (link to survey questions). Core questions required respondents to indicate, among other questions, the type and duration of data collected (bicycle/pedestrian/micromobility, continuous vs short term counts), type of equipment technology used, how counts are stored, standardized, and shared, key applications/use cases for the data, and primary challenges. These mandatory questions were a combination of multiple choice (select all that apply), True/False (Yes/No), and ranking options, allowing results to be aggregated and compared. Additionally, optional free-response questions were provided to allow respondents to add detail and context, if desired.

Representatives from all 19 State/District DOTs in the Coalition were contacted to determine whether their state manages a state-level bike/ped count program. For states with such programs, the state representative was asked to complete the survey or distribute it to a colleague with appropriate knowledge of the program. For states without state-level count programs, the representative was asked if they knew of any local agencies (e.g., city, MPO, RPC) within the state that collect bike/ped count data. The project team did a second round of outreach to local agencies and distributed the survey to appropriate contacts in these organizations.

Survey Results and Discussion

The results are summarized below in three tables. Table 1 summarizes the top-level inventory of bike/ped count programs across all Coalition states, while Tables 2 and 3 provide detailed survey responses for agencies with count programs, with Table 2 focusing on state-level count programs and Table 3 dealing with local agencies (cities, MPOs, and RPCs). Criteria for inclusion in the survey table include the agency regularly collecting bike or ped data with both permanent and short-term counters and having a representative willing to complete a detailed survey. This list is not exhaustive, although efforts were made to track down representatives from all agencies that were identified.

Overall, Table 1 highlights that **there are eight states with some type of state-level bike/ped count program– of which six are fairly mature**. Additionally, it shows that there is a wide range of programs represented in the Coalition; some states have no bike/ped count activity (e.g., Alabama, Maine), others have robust, statewide programs (e.g., North Carolina, Florida), while still others have local agencies with highly mature programs (e.g., Pennsylvania and New Jersey with DVRPC in the Philadelphia metro area). In the case of both statewide and local count programs, data collection is usually executed at the local level, but some state DOTS serve as an organizing agent, while others may be unaware of count activity in their state.

Table 1 – Inventory of bike/ped count programs in TETC states/districts

State	Do bike & ped count programs exist?		Are permanent counters deployed?	Comments
	State/DOT Level	City/Local Level		
Alabama	No	No	-	
Connecticut	No	Yes	Yes	Count data is currently limited to several trail counters managed by the University of Connecticut. In the past, contractors used to collect short-term counts.
Delaware	Yes	Yes	Yes	Delaware DOT has a count program consisting of both permanent and short-term counters.
Florida	Yes	Yes	Yes	Florida DOT has a statewide count program consisting of permanent and short-term counts. It relies on coordination and data collection by local agencies across the state but ensures that all data ends up in a central repository.
Georgia	No	No	-	Short-duration counts are collected periodically by GDOT for specific projects, but no coordinated program currently exists
Kentucky	No	Yes	Yes	Lexington Area MPO has a count program
Maine	No	No	-	
Maryland	No	Yes	Yes	Montgomery County has permanent trail counters, Baltimore Metro Council and City of Baltimore also collect counts
Massachusetts	Yes	Yes	Yes	MassDOT has a limited count program consisting of both permanent and short-duration counters.
New Hampshire	No	Yes	Yes	NHDOT announced plans for statewide programs, but currently count programs exist at the planning commission level. Examples include: - Southern New Hampshire Planning Commission (permanent counters on trails, sidewalks, and intersections) - Nashua RPC (regular short-term count deployments)
New Jersey	No	Yes	Yes	DVRPC has a mature count program with permanent and short-term counters (some counters are in NJ). North Jersey Transportation Planning Authority conducted a large pilot bike/ped counting project with the intent to later implement more permanently.
New York	Yes	Yes	Yes	NYSDOT has a small count program and is looking to start aggregating counts from cities/agencies. City/MPO level count programs with permanent counters include NYC and Binghampton MPO.
North Carolina	Yes	Yes	Yes	North Carolina DOT has a mature, statewide count program. Counters were initially managed by a research center at NC State, but over time ownership was transferred to local agencies, with all data available statewide in a central database
Pennsylvania	No	Yes	Yes	DVRPC has count program with a wide array of permanent counters and regularly collects short-term counts.
Rhode Island	No	Yes	Yes	City of Providence has a count program.
Tennessee	No	Yes	No	Several cities and MPOs have limited count programs, but few are active or systematically designed
Vermont	Yes	Yes	Yes	VTrans has a very small count program that is mostly inactive and relies on periodic collection by regional partners. Looking to reinvigorate program in 2024.
Virginia	Yes	Yes	Yes	VDOT manages a count program with permanent and short-term counts in coordination with local agencies.
Washington DC	Yes	Yes	Yes	DDOT manages a count program with both permanent and short-term counts.

Table 2 -- Survey responses from agencies with statewide count programs

Question	North Carolina DOT	Florida DOT	Massachusetts DOT	Virginia DOT	Delaware DOT	District DOT
Type of Data collected *(select any)*						
Pedestrian: Permanent Counts	X	X	X	X	X	X
Pedestrian: Short-term Counts	X	X	X	X	X	
Bike/Micromobility: Permanent Counts	X	X	X	X	X	X
Bike/Micromobility: Short-term Counts	X	X	X	X	X	X
Count Technology: Pedestrian *(select any)*						
Inductance loops						
Magnetometers						
Pressure sensor						
Radar sensor						
Seismic sensor						
Video (automated)		X	X			
Video (manual)		X				
Infrared sensor	X	X	X	X	X	X
Pneumatic tube		X				
Manual count		X			X	
Count Technology: Bike/Micromobility *(select any)*						
Inductance loops	X	X	X		X	X
Magnetometers						
Pressure sensor			X			
Radar sensor						
Seismic sensor						
Video (automated)		X	X			
Video (manual)		X				
Infrared sensor		X		X	X	
Pneumatic tube	X	X	X	X	X	
Manual count		X			X	X
Are bike/ped counts subject to QC/QA?	Yes	Yes	Yes	Yes	Partially	Partially
Primary means of data management *(select one)*						
Spreadsheet		X				
Internal Database/software						
Data management software included with count equipment	X		X	X	X	X
Publicly available count management software						
External Contractor						
Other (specify)						
Are archived data available publicly?	Yes	Yes	Partially	Partially	Partially	Yes
Standardized formats supported *(select any)*						
BikePed Portal						X

National Bike and Ped Safety Data Clearinghouse						
TMAS (TMG 2016 format)		X				
Other (specify)	Vendor-specific (Eco-Visio)		Vendor-specific (MS2)	Vendor-specific (Eco-Visio)	Vendor-specific (Eco-Visio)	
Methods used to estimate bike/ped volumes *(select any)*						
Applying temporal factors to short-term counts	X	X		X	X	
Modeling -- Direct demand		X			X	
Modeling -- Data fusion (including 3rd party sources)		X				
Purchasing 3rd party data sources to be used directly		X				
Other (specify)		X	MS2 Defaults			
Use Cases						
(Scale of 1-5, 1 = Not important, 5 = Extremely important)						
Monitor trends in activity over time	4	5	4	5	4	4
Conduct before/after studies	3	5	4	5	3	4
Prioritize projects	4	5	4	4	4	3
Analyze safety	4	5	4	2	4	3
Research travel behavior	3	4	4	3	4	2
Develop/Calibrate demand models	3	4	4	2	4	2
Policy Decisions *(select any)*						
Facility design		X			X	X
Land use planning		X				
Operations		X	X	X	X	
Maintenance		X				
Address safety concerns	X	X	X			
Investment opportunities	X	X	X	X	X	X
Other (specify)		Signal timing	Future study areas			
Is program meeting agency needs?	No	Partially	Partially	Yes	Partially	Partially
Would access to bike/ped data from other agencies/states/regions be useful?	Not sure	Yes	Not sure	Yes	Not sure	Yes
How challenging are the following factors?						
(Scale of 1-5, 1 = Not challenging, 5 = Extremely Challenging)						
Funding	3	5	3	2	2	2
Staffing	5	5	3	1	5	3
Understanding & Implementing Best Practices	3	3	4	2	4	4
Equipment Maintenance	5	3	4	3	5	5
Ensuring data quality	3	4	4	4	5	4
Data Management	3	4	4	2	5	2
Data sharing/reporting	3	4	4	2	3	2

Table 3 -- Survey responses from local agencies with count programs

Question	DVRPC (PA/NJ)	City of Providence (RI)	Lexington Area MPO (KY, IN)	Baltimore Metro Council (MD)	City of Salisbury (MD)	Montgomery County DOT (MD)
Type of Data collected *(select any)*						
Pedestrian: Permanent Counts	X		X		X	
Pedestrian: Short-term Counts	X	X		X	X	
Bike/Micromobility: Permanent Counts	X	X	X		X	
Bike/Micromobility: Short-term Counts	X	X		X	X	
Count Technology: Pedestrian *(select any)*						
Inductance loops						
Magnetometers						
Pressure sensor			X			
Radar sensor					X	
Seismic sensor						
Video (automated)					X	
Video (manual)	X			X		
Infrared sensor	X					X
Pneumatic tube						
Manual count						
Other		X				
Count Technology: Bike/Micromobility *(select any)*						
Inductance loops	X	X			X	X
Magnetometers						
Pressure sensor			X			
Radar sensor					X	
Seismic sensor						
Video (automated)					X	
Video (manual)	X			X		
Infrared sensor						
Pneumatic tube	X	X				
Manual count						X
Are bike/ped counts subject to QC/QA?	Yes	Partially	Yes	Yes	Yes	No
Primary means of data management *(select one)*						
Spreadsheet		X		X		
Internal Database/software	X					
Data management software included with count equipment			X		X	
Publicly available count management software						
External Contractor						
Other (specify)						X
Are archived data available publicly?	Yes	No	Partially	No	Partially	Partially
Standardized formats supported *(select any)*						
BikePed Portal			X			
National Bike and Ped Safety Data Clearinghouse						

TMAS (TMG 2016 format)	X			X		
Other (specify)		Vendor-specific (Eco-Visio)				Vendor-Specific (Eco-Visio)
Methods used to estimate bike/ped volumes *(select any)*						
Applying temporal factors to short-term counts	X	X				X
Modeling -- Direct demand						
Modeling -- Data fusion (including 3rd party sources)		X				
Purchasing 3rd party data sources to be used directly			X			
Other					X	
Use Cases						
(Scale of 1-5, 1 = Not important, 5 = Extremely important)						
Monitor trends in activity over time	4	3	3	4	5	5
Conduct before/after studies	5	4	1	4	4	5
Prioritize projects	2	3	2	2	4	3
Analyze safety	4	4	4	4	4	2
Research travel behavior	5	3	4	3	3	3
Develop/Calibrate demand models	1	2	3	3	3	2
Policy Decisions *(select any)*						
Facility design		X			X	X
Land use planning			X		X	
Operations					X	
Maintenance			X			
Address safety concerns		X			X	
Investment opportunities		X				
Other (specify)	Planning support			Local jurisdictions	Route prioritization	
Is program meeting agency needs?	Partially	No	Partially	No	Partially	Partially
Would access to bike/ped data from other agencies/states/regions be useful?	Not sure	Yes	Not sure	Yes	Yes	Not sure
How challenging are the following factors?						
(Scale of 1-5, 1 = Not challenging, 5 = Extremely Challenging)						
Funding	4	4	3	4	2	3
Staffing	3	4	5	5	2	3
Understanding & Implementing Best Practices	2	2	4	4	3	4
Equipment Maintenance	4	4	3	1	2	5
Ensuring data quality	3	3	3	4	4	3
Data Management	2	2	4	4	3	4
Data sharing/reporting	2	2	5	5	3	2

Follow-up communication with points of contact from each Coalition state also highlighted that many programs are centered in cities and MPOs and not coordinated at the state level. In some cases, cities or MPOs establish their own local programs for local planning purposes, but data does not filter up to the state DOT, and thus bike/ped data may exist within the state but not be stored in a centralized location or be widely available. The most mature state-level programs (especially North Carolina and Florida) rely on strong collaboration between the state and local level and have mechanisms for storing locally collected data in centralized databases.

The following subsections look at several of the result categories in more detail. Tables 2-3 highlight that there are different priorities, use cases, and core challenges faced by each of the agencies. While there are similarities in sensor technologies used for data collection, data management and standards are not shared across agencies. One commonality observed across many programs is that data management is often handled through software provided by the count equipment manufacturer. While such software may be proprietary, if many agencies use the same company (e.g., Eco Counter with their Eco-Visio platform and API), there may be low-hanging opportunities to obtain data in a common, if vendor-specific, format.

What type of data is collected?

Each of the six state-level programs reported collecting both pedestrian and bike counts, with all agencies using both permanent and short-term counters for counting cyclists and all but one using both types of counters for pedestrians (DDOT did not use short-term counters for pedestrians). Local agencies were more varied; two cities/MPOs collected bike and pedestrian counts using both permanent and short-term counters, while the others collected only a subset. For example, Baltimore Metro council currently only collects short-term counts for bike and pedestrian traffic.

What technology is used for counting?

The top count technology for pedestrian counting was infrared sensors (used by 8 of the 12 agencies), followed by manual and automated video-based solutions. Bicycle and other micromobility counters used a wider variety of detection technologies, with inductance loops the most popular (9 agencies), followed by pneumatic tubes, manual and automated video solutions, and manual counters. Follow-up conversations with the steering committee and other stakeholders suggest that automated video technology is fast-evolving and expected to become more widely adopted.

One such quote included the following (regarding automated video counting):

"This new innovative technology is likely to revolutionize the way old traffic data volume counting programs report and calculate statistics in the coming decades. These new video datasets allow for trajectory lines of travelers by type and are already being used in Florida to discover behaviors and determine new findings that a spreadsheet, pie chart, or other existing older technologies cannot provide."

How is count data managed, standardized, and shared?

The most common way that agencies managed count data was through software provided by the count equipment manufacturer (7 agencies), followed by spreadsheets (3 agencies), and with one agency using an internally developed database. Common standards were not generally supported, although a few agencies prepared their data in FHWA's Traffic Monitoring Guide

(TMG) format. All but a few agencies performed some type of QA/QC on the data, although the survey didn't ask for specifics, and discussions with the CAPABLE committee pointed out that this is an area of inconsistency. Most agency data was either fully or partially publicly available – at least upon request.

The fact that data is typically managed through vendor software is both an opportunity and potential pitfall. While such count equipment software may be proprietary, if many agencies use the same company (e.g., EcoCounter with their Eco-Visio platform and API), there may be low-hanging opportunities to obtain data in a common format. However, this requires locking into a single vendor and does not solve the problem of integrating counts from different count equipment manufacturers. This may be problematic for a state agency looking to aggregate counts from local agencies in the state, or for agencies looking to integrate new technology (e.g., video analytics) alongside legacy sensors and obtain a unified view of their counts.

What are the most important use cases?

Agencies were asked to rank a list of use cases on a scale of 1-5, with 5 being most important and 1 the least important. By summing the total scores across all responding agencies, the ordering of use cases was as follows (avg score in parentheses).

- Monitor trends in activity over time (4.2)
- Conduct before/after studies (3.9)
- Analyze safety (3.6)
- Research travel behavior (3.4)
- Prioritize projects (3.1)
- Calibrate demand models (2.8)

It should be noted that responses to this question probably depend on who was asked; data collection staff, planners, and safety engineers within the same organization may have different perspectives on what is most important.

In a related question, agencies were asked about the policy decisions that are made based on the data. Although similar in nature to the prior question, it is focused on the specific policy outcomes (e.g., what decisions are made from monitoring trends in activity, before/after studies, etc.). The top responses were identifying investment opportunities (seven agencies), designing facilities (six), addressing safety concerns (five), and operations.

What are the most significant challenges?

Agencies were asked to rank a set of statements on a scale of 1-5, with 1 being the least challenging and 5 extremely challenging. The following values represent average scores across all agencies, ranked from most-to-least challenging.

- Equipment maintenance (4.0)
- Ensuring data quality (4.0)
- Staffing (3.9)
- Data management (3.4)
- Funding (3.3)
- Understanding & implementing best practices (3.3)
- Data sharing / reporting (3.3)

These results show that maintaining count equipment, quality assurance, and staffing are the most significant pain points for most agencies – sentiments that were also echoed in the free response questions. This highlights the fact that developing a count program is much more than the initial deployment; they require experienced staff to perform ongoing maintenance / adjustment and assess the quality of data produced – a process that can be resource intensive.

How are count programs funded?

This question was not directly asked on the survey, but one agency brought up program funding in the free response section, and the topic was subsequently discussed by the steering committee.

Many agencies reported using State Planning and Research (SPR) funds as a primary source for supporting bike/ped count programs. One state noted that they initially used SPR funds for this purpose, but after a number of years the work was no longer considered research, and thus they had to find other funding mechanisms, including tech transfer funds at the state level. Other funding sources included additional Federal programs (including Safe Streets and Safe Routes to School), philanthropic grants (used heavily by DVRPC), health systems, and transit funding.

Industry Bike/Ped Data Products

Commercial bike/ped products tend to fall into one of two categories: (1) estimated bike/ped volumes – typically using LBS data and other sources as model inputs, which correspond to ancillary products sold in the Coalition's Traffic Data Marketplace, and (2) bike/ped counters and associated software for collecting and managing bike/ped pedestrian counts. It should be emphasized that these two areas are not independent, as estimated data relies on the availability and quality of counts for calibration.

Estimated / crowdsourced bike/ped data

Estimated / crowdsourced bike/ped products provide estimates of population-level bicycle and pedestrian activity at a variety of spatial and temporal scales. While details differ by vendor, in general, statistical inference techniques are used to create models that leverage LBS-based sample data and other input sources to estimate overall activity using high-quality ground counts as calibration sources.

Depending on the use cases and availability of input data, estimates may be provided at granular spatial scales, such as road links or intersections, or more aggregate levels, such as city, census tract, or other zone level. Likewise, estimates may be produced at different temporal scales, including hourly, daily, monthly, or annual measures of activity. Most vendor products are accompanied by a web interface to allow for ad-hoc querying, downloading of data, and a visual exploration of results.

A critical source of data used to produce estimated/crowdsourced bike/ped volumes comes from Location Based Services (LBS) data, which is derived from smart phones and other location-aware devices. Notably, **LBS data was significantly disrupted (reduced) in 2022 due to major policy changes at Apple and Google (Android)**. Although the amount of LBS data available took a significant hit, the LBS data market has been slowly recovering. Even so, as LBS source data returns, vendors estimating bike/ped activity still need accurate and geographically representative calibration data to maximize its value. Good calibration data is needed even in the pre-2022, rich LBS data environment, and now that the source data has become constricted, its importance is even greater.

Two vendors, StreetLight Data and Replica, sell estimated/crowdsourced bike/ped data in the TDM Marketplace, and have products that are representative of this type of estimation approach (albeit with different approaches). Other sources of crowd-sourced bike/ped data include Strava Metro, a non-commercial product available to cities that uses location data from a fitness app, and Ride Report, which focuses on shared mobility ridership in cities (e.g., bike and scooter share programs). Neither Strava Metro nor Ride Report make inferences about overall biking or walking volumes, but may be useful for identifying patterns in micromobility travel behavior.

StreetLight and Replica were invited to participate in a CAPABLE steering committee meeting to discuss their product offerings, communicate their perspective on the state of bike/ped estimation, and answer questions from committee members. The following subsections summarize key aspects of their products and highlight their perspectives on what the Coalition can do to improve the accuracy of estimates.

StreetLight Data

StreetLight Data's core bike/ped product is called the Active Transportation Monitor, which reports annual bicycle and pedestrian metrics at the census tract level. StreetLight also offers historical bike/ped metrics at a segment level, for 2019 through April 2022. Active Transportation Monitor metrics can be obtained via interactive web platform and include volume estimates and mode share estimates for pedestrian, bicycle, and vehicle modes. Segment-level historical metrics from 2019 through April 2022 can be obtained via API or interactive web platform through a variety of analyses, including origin-destination and zone activity analyses to focus on trips starting in, ending in, or passing through a certain area, and include volume estimates, trip attributes such as speed and travel time for the entire trip, and inferred demographic information.

Figure 1 shows the Active Transportation Monitor web interface, which supports interactive queries, including comparisons between two different time periods. Common use cases for their data include characterizing corridor usage patterns, mapping crash exposure, running before/after studies to evaluate project success, and monitoring macro trends for mode shift.

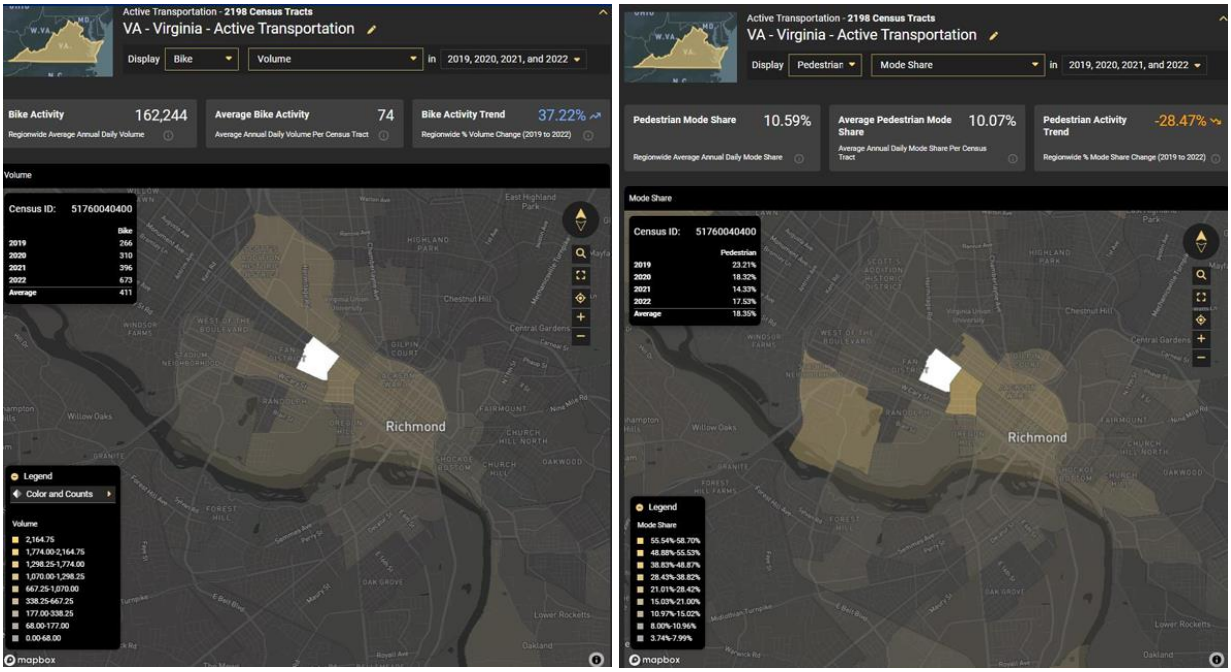

Figure 1: Screenshot of StreetLight's Active Transportation Monitor web interface

StreetLight's methodology for estimating these bike/ped metrics is documented in a publicly available white paper (link), and uses several data sources, including LBS trip data (representing a sample of trips across different modes), the Next Gen National Household Travel Survey (Next Gen NHTS) dataset, OpenStreetMap, and US Census geographies. Their estimation modeling process involves (1) determining modal split for metropolitan statistical areas (MSAs) using LBS and Next Gen NHTS data, (2) estimating total active transportation volumes at the MSA level using the modal splits and vehicle volume estimates, and (3) splitting the total active transportation volumes into bike and ped volumes at the tract level using LBS data. Their whitepaper includes internal validation tests against other data sources, including the American Community Survey.

It should be noted that StreetLight's bike/ped products were impacted by the disruption to LBS data in April 2022. For months up to and including April 2022, their bike/ped offerings include more spatial and temporal granularity, including volume estimates at the road segment level and sub-annual reporting. These segment-level, monthly metrics are still available and used for historical estimates of bike/ped activity at the segment level, but do not include data after April 2022. After the disruption to the LBS data supply chain, they released a new product for post-April 2022 data that focuses on census tract-level spatial granularity, while also introducing features to identify trends over time. LBS data has been steadily recovering, and StreetLight indicated that at some point they plan to update their granular reporting options.

StreetLight's Perspective

StreetLight emphasized that what impacts their data quality the most – even more than the status of the input LBS data – is the availability of high-quality calibration counts. They provided the following comments regarding the type of count data needed to improve estimation accuracy

"We would recommend placement in top metro areas, where non-recreational biking is expected. Biking tends to be in concentrated areas, and the counter data is most likely to be most actionable by cities with a substantial volume of biking activity.

If ETC has funding for a limited number of counters, it is recommended that the investment be focused on one or potentially two cities. For modeling bike activity, a range of 10-50+ counters per city is needed. Adding a collection of new counters to a city (especially one which had few or none before) will be substantially higher impact than adding only one counter to multiple cities.

Counters should be placed in areas where high biking activity is expected, and in areas where bike infrastructure does and does not exist. Placement should take into consideration a variety of typologies/zoning (dense urban, suburban, etc.). Of course, the best insights for placement in a specific location will come from the city planners for where bike activity is expected, and or planned for the near future (which will support before/after studies and trend effects)."

Replica

Replica estimates bike and pedestrian volumes (alongside other modes) using an activity-based model, resulting in highly granular measures of bike/ped activity at the road segment level. Customers can interact with the data through a web interface, export results into common data and GIS formats, and query programmatically by accessing an online database. Figure 2 shows an example of bike and pedestrian volumes visualized through the web interface, which helps communicate how active trips are distributed across a city. Common use cases for their data include safety analysis, determining where to deploy active transportation infrastructure, identifying mode shift opportunities, and communicating trends and opportunities to the public.

Replica's activity-based model leverages many datasets, which can be classified into the following categories: location data (including, but not limited to LBS data), consumer and resident data, the build environment, economic activity (including credit card transactions), and ground truth counts. At a high level, their estimation approach starts with a synthetic population ("replica") of everyone in the US census, models travel behavior in terms of activity sequences and locations (using survey data and other granular datasets), and then determines mode choice for each trip by considering the route along different available modes. Large-scale simulations are used to apply these behavioral models to the synthetic population, ultimately resulting in mode-specific volumes along each road segment (including bike/ped volumes). The outputs are then compared with local data – including 'reference' counts – to calibrate the model.

Although Replica's estimation engine produces estimates for bike/ped volumes under existing conditions, its modeling approach is designed to also consider latent demand (i.e., travel demand that is not currently realized). Replica explicitly estimates demand as part of its modeling framework and then assigns resulting trips to specific travel modes (such as biking or walking) using a multi-modal route choice model that considers the available infrastructure. Using this framework, they can evaluate the impact of policy decisions that would change routing and mode choice outcomes (e.g., adding active transportation infrastructure that results in safe routing options for cyclists and pedestrians), and in so doing, estimate latent demand.

Due to nature of Replica's modeling approach – particularly the fact that travel mode is determined by a behavioral model rather than inferred directly from LBS data -- their product was not highly impacted by the April 2022 LBS data disruption.

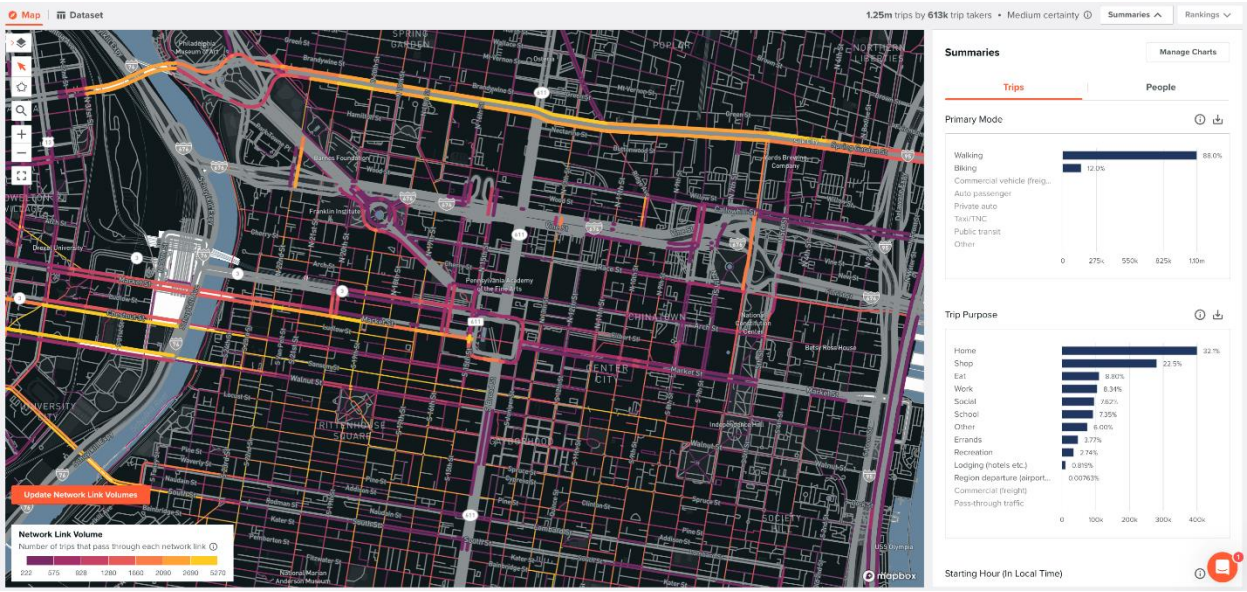

Figure 2: Screenshot of Replica's web interface to visualize active transportation trips

Replica's Perspective

Replica emphasized the importance of distinguishing between actual bike/ped counts and demand for trips (the latter of which is a particular focus for their product). Additionally, they cautioned that applying the same expectations for count accuracy to that of automobile travel is unrealistic (e.g., 30% error for 100 bicyclists is very different than 30% error on 100k vehicular AADT) and that necessary accuracy is driven by the use cases. They suggested that the Coalition or another entity could play a role in defining the acceptable level of error for key use cases. With respect to improving the accuracy of estimates, Replica noted the following:

"We also see value in exploring the development of a more precise model for estimating bike or pedestrian activity. However, given our observations of regional behavioral differences, we believe that focusing on a local scale would be most helpful. In this regard, we recommend partnering with an academic institution to explore specific questions with Replica data. These questions could include testing the feasibility of a local-scale approach, determining the optimal number of counters needed based on local context, and further refining the steps required for a

city or local entity to develop a precise active estimation model using baseline data, and calibrating or augmenting it with counter/computer vision inputs."

With respect to the adequacy of current pedestrian and bicycle count data to support modeling:

"Our observation is that more centralized storage, standardization, and a greater quantity of high quality non-motorized count data is needed to improve precision of modeled estimates."

Bike/Ped Counters

In contrast to estimated/crowdsourced bike/ped products, bike/ped counters directly measure (rather than estimate or model) bicycle and pedestrian counts at specific locations. Counters form the backbone of many agencies' micromobility travel monitoring systems and serve as a critical source of calibration data for emerging estimation products. Bike/ped count equipment encompasses a number of technologies, including inductance loops, pneumatic tubes, radar, video-based, infrared sensors, and others – each of which has pros and cons. Figure 3, which was produced by FHWA and included in the 2022 Traffic Monitoring Guide (TMG), enumerates key count technologies and their suitability for counting different types of micromobility modes (although video-based traffic counting was less mature several years when the guide was put together). The accuracy of data collected depends on several factors, including the underlying count technology, what type of users need to be counted, the deployment location, installation method, and the quality of the equipment.

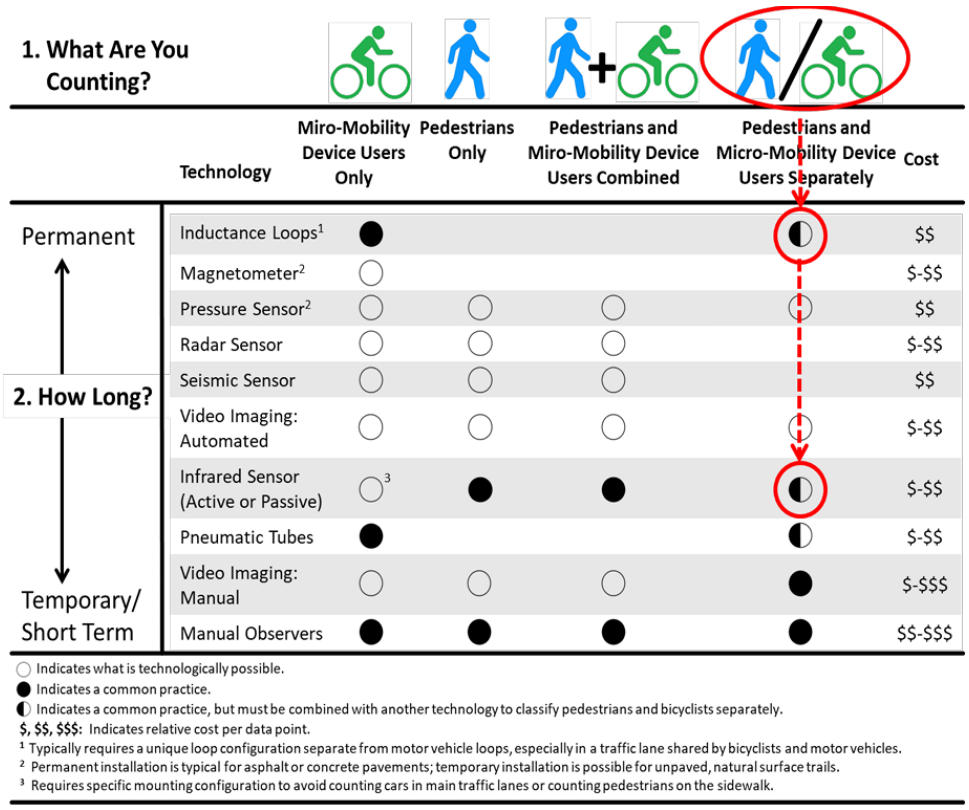

Figure 3: Bike/Ped count technologies (source: 2022 TMG).

The following list highlights several key data vendors that sell count equipment – some of whom also include associated software or APIs for accessing and archiving the data. Based on survey results across The Eastern Transportation Coalition and discussion with steering committee members, Eco-Counter has the majority of market share, with nearly all agencies using their sensors (of which there are several types) for permanent counters, and many also using them for short-term collection. The other vendors listed here were identified based on survey responses.

- **Eco-Counter**: Eco-Counter offers several different services to states/MPOs, including bike/ped count equipment, software for accessing and managing the data, maintenance and technical assistance, and data services. Their equipment offerings include a number of different sensors for various applications, such as the Pyro (pyroelectric pedestrian counter) Zelt (inductive loop sensors for bike and scooter), MULTI (combination of technologies to separately capture pedestrians, bikes, and scooters) and CITIX-AI (video/AI solution capturing all modes). Along with count equipment hardware, they offer the Eco-Visio web platform and public web pages that allow data to be stored, queried, and shared publicly, as well as an API that makes count data available programmatically.

  Figure 4 shows three variations of the Eco-Counter MULTI sensor, each of which is capable of counting multiple modes of micromobility traffic (i.e., pedestrians, bikes, and scooters). The left image shows the Urban MULTI sensor, which is permanently installed in urban environments, and uses a combination of inductive loops (bike/scooter) and pyroelectric sensors (pedestrians). Most of the equipment is contained in a post that is installed along the road or path, while inductive loops are installed under the pavement. The middle image shows the Natural MULTI, which is requires a similar permanent installation as the Urban MULTI, but is geared towards rural settings. Finally, the right image shows the Mobile MULTI sensor, which enables short-term multi-modal data collection without cutting into the pavement.

Figure 4: Eco-Counter MULTI sensors that count bike/ped/scooter traffic.

Figure 5 shows an example of Eco-Counter's public dashboard for Montgomery County, MD. This interface, which can be produced for any sensor deployment (with purchase of appropriate subscription service), provides a simple way to share data with the public.

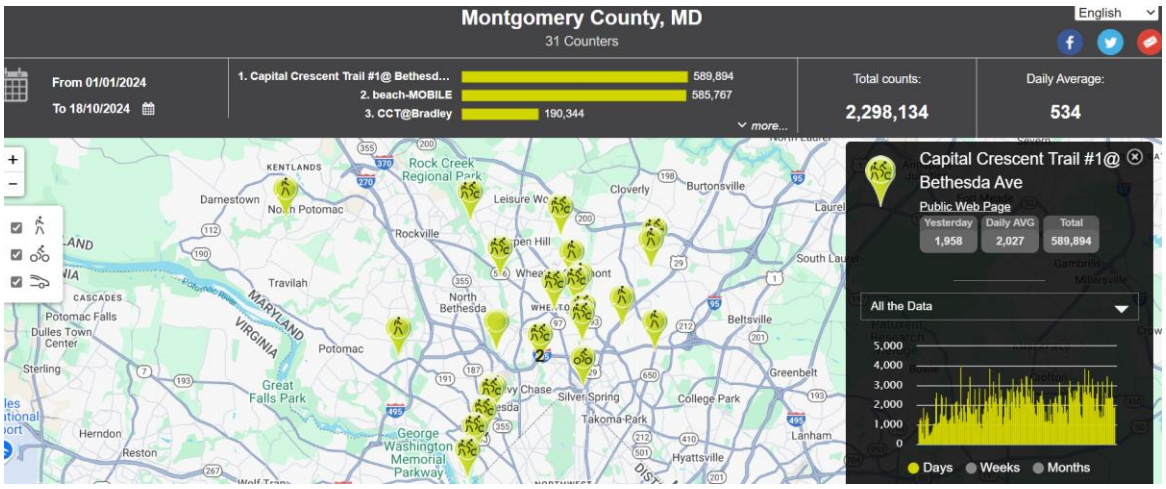

Figure 5: Sample Eco-Counter dashboard

- **TRAFx**: TRAFx's Trail Counter uses infrared sensors to count the number of people on trails, paths, and sidewalks. This system counts the total number of people -- including pedestrians, cyclists, horseback riders, and other micromobility users, but does not segregate results by mode. TRAFx provides a web-based data management system called DataNet to view, analyze, and communicate results. Figure 6 shows an image of the sensor (left) and software interface (right).

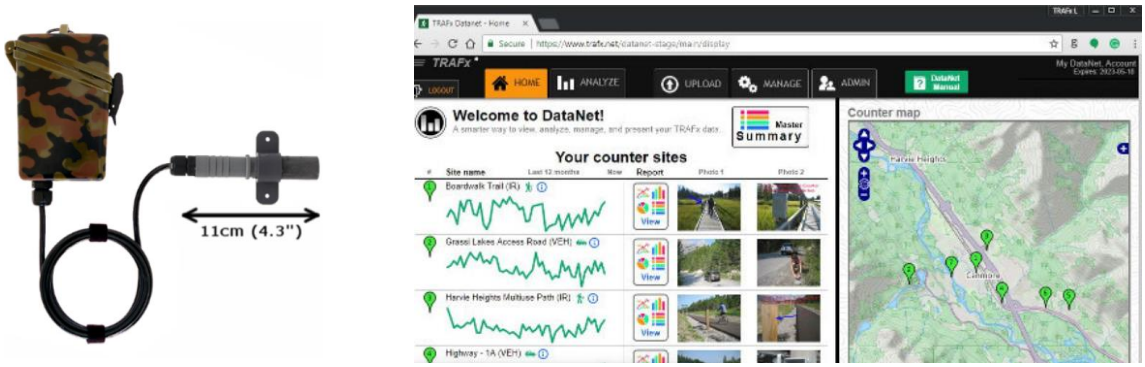

Figure 6: Trafx Trail counter (left) and DataNet web platform (right).

- **JAMAR**: The JAMAR TRAX Pinnacle counter uses pneumatic tubes to count traffic on bike trails, dedicated bike lanes, vehicle lanes, or mixed-use lanes. It is intended to be a general counter that can record either vehicle or bicycle traffic and can appropriately distinguish bicycle counts from vehicular modes. Figure 7 shows an example of the TRAX Pinnacle control device (left) and software interface (right).

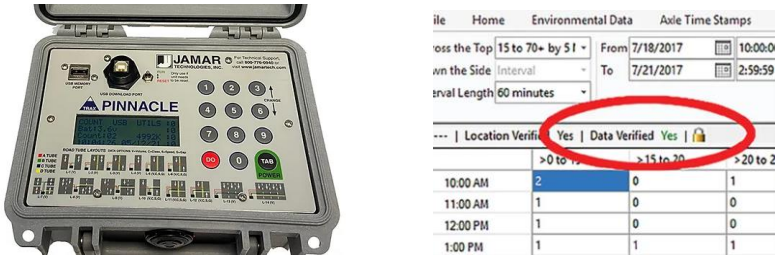

Figure 7: JAMAR TRAX Pinnacle counter and data verification software.

- **Q-Free:** Q-Free's Hi-Trac CMU system uses piezoelectric and pyroelectric sensors to separately count cyclist and pedestrian traffic along either dedicated cycleways or in mixed traffic conditions. Figure 8 shows a typical installation, which requires placing sensors within the pavement.

Figure 8: Q-Free Hi-Trac CMU counter.

- **Miovision**: The Miovision Scout Plus is a portable video data collection system that uses on-board analytics (i.e., computer vision and machine learning techniques) to extract counts and other traffic information from the video feed. This system is intended to capture both vehicle and micromobility traffic and can report counts separately by mode. Figure 9 shows the Scout Plus control box mounted to a pole.

Figure 9: Miovision Scout Plus video-based counter.

Advancement of artificial intelligence (AI) and machine learning (ML) to interpret imagery has the potential to revolutionize bike/ped counting by improving its efficiency and cost. Although deploying video cameras with corresponding count analytics is currently more expensive than traditional bike/ped counters, it is expected that costs will come down as the technology becomes more ubiquitous. Additionally, if agencies can use existing cameras that are deployed for other traffic monitoring purposes, there may be an opportunity to increase the amount of bike/ped data collected without additional equipment installation or maintenance costs. Furthermore, as traffic monitoring at intersection deploys advanced sensors (video, radar and LiDAR), data from these systems can typically report on non-motorized activity along with vehicle detection, counts, and trajectory counting. Finally, an added benefit is that camera-based solutions are non-intrusive and keep staff off the roadway, resulting in a safer way to monitor traffic.

Figure 10 shows an example of vehicle and pedestrian locations extracted from a video feed by a member of the project team. In contrast to camera solutions that either (1) handle object detection, tracking, and counting behind the scenes and output counts (typically as part of an integrated hardware/software system) or (2) simply record video, requiring a separate solution to extract counts (often via software-as-a-service model), this camera system runs an object detection algorithm automatically and provides mode-specific locations programmatically 'out-of-the-box'. Although this is a simple proof-of-concept, the results were encouraging – particularly at a low price point without recurring subscription, – and highlighted how quickly technology is advancing in this area. It appears that video-based systems will be an important part of bike/ped counting in the near future.

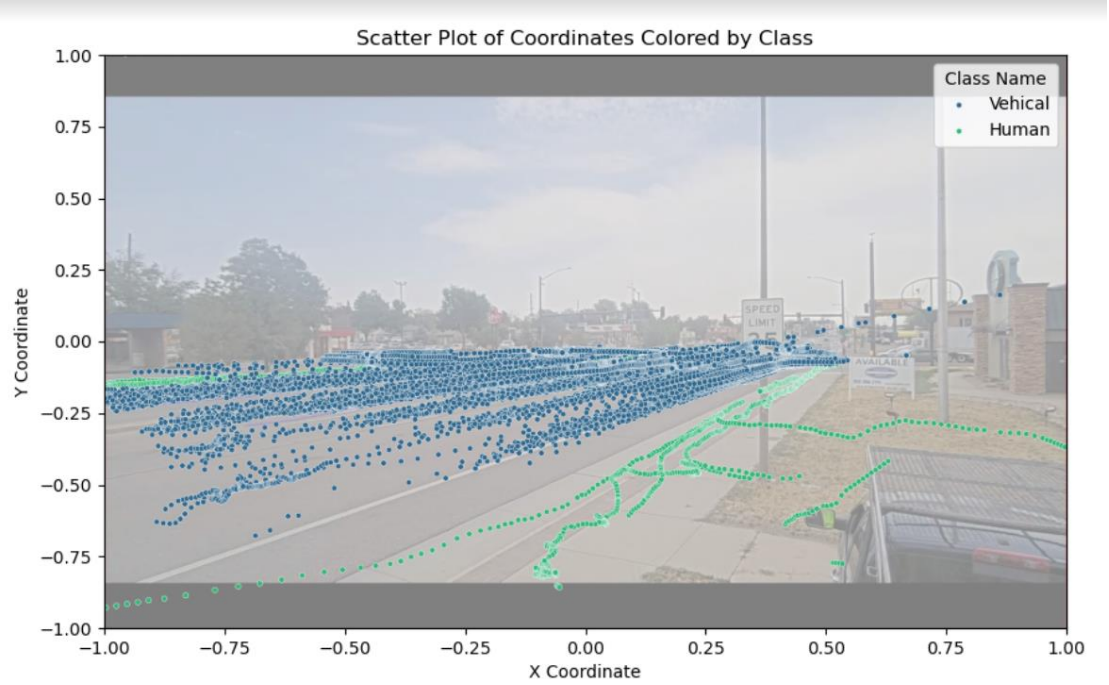

Figure 10: Example of pedestrian and vehicle object detection. (Source: NREL)

Takeaways and Lessons Learned

The following takeaways about the state of bike/ped data emerged from the inventory of Coalition count programs and review of commercial bike/ped data offerings, as well as conversations with the CAPABLE steering committee.

- **The accuracy of estimated / crowdsourced bike/ped products depends on count data collected by governmental agencies, non-profit organizations, and agency support consultants.** Even with high quality LBS and other crowd-sourced input data sources, volume estimation products require accurate, standardized counts to calibrate their models, and vendors have identified the lack of widely available calibration count data as a key impediment to improving accuracy. Increasing the number of counters and working to standardize/aggregate existing counts across agencies would be beneficial (albeit resource intensive).

- **Bike/ped count programs *do* exist in the Coalition and beyond, but data tends to be siloed across organizations and stored in different formats**. Data collection is usually executed by local agencies (cities and MPOs), but the extent to which this information is shared with state DOTs or other entities varies state-by-state. A handful of state DOTs manage count programs at the state level and work closely with local jurisdictions to collect and aggregate data, while others are unaware what data, if any, is being collected by local agencies in their state. Aggregating data from different silos is challenging without common data formats and processes.

- **Agencies face institutional challenges and resource constraints**. In the absence of a federal mandate, many agencies do not prioritize bike/ped data collection, which often includes labor-intensive standardization efforts ("data wrangling"). As such, there is significant variation in the data collected across states and metro areas, depending on priorities, budgets, and presence of local champions within agencies. Even states that manage to prioritize bike/ped data collection often do not have a clear understanding of how converting their natively stored data into another standardized format would benefit their organization.

- **Efforts to aggregate bike/ped data are in their infancy.** FHWA is trying to be a national aggregator of bike/ped data, but there are challenges to widespread adoption – in part because submission is voluntary and can be technically demanding. To date, only a handful of jurisdictions have submitted their data to FHWA's Travel Monitoring and Analysis System (TMAS). BikePed Portal (managed by Portland State) contracts with state and local agencies from across the country to manage bike/ped counts, but it too has limited adoption. Several DOTs have developed sophisticated processes to aggregate data collected by agencies across their states (e.g., TxDOT's Bike Ped Count Exchange), but these systems are limited to single states and are not currently being merged with other datasets.

- **Technology advancements are making computer vision-based data collection more attractive.** Technology, particularly video image processing technology enabled by AI/ML, is significantly impacting the cost and quality of bike/ped data collection. If trends continue, it is anticipated that non-motorized count programs will increasingly use this technology due to significantly lower costs as compared to traditional counting methods.

However, in the absence of data collection standards (with each vendor having their own proprietary format), this may make it more difficult to aggregate data across organizations and regions.

- **Better mapping and location referencing are needed for bike/ped infrastructure.** Location referencing bicycle and pedestrian facilities is varied, with many efforts simply borrowing from roadway referencing standards, augmenting either a linear referencing system or a mapping standard (such as OpenStreetMap or a commercial map provider). Industry is beginning to acknowledge that georeferencing and mapping for bicycle and pedestrian facilities needs to be more sophisticated, providing not only a logical mapping (such as a sidewalk exists along the roadway) but attributes associated with the facility (e.g., width of the sidewalk). Bicycle facilities likewise have evolved rapidly in the last decade with many cities adopting bicycle lanes (either painted or with physical barriers), and it is critical to convey that information about the infrastructure along with use and demand information. In both instances, dedicated bike/ped facilities (not adjoining a roadway) are critical to reflect as part of the network. The National Collaboration on Bicycle, Pedestrian and Accessibility Infrastructure Data (NC-BPAID), led by the Bureau of Transportation Statistics, has recently begun an effort to develop bike/ped infrastructure data standards.

3. Bike/Ped Data Schemas

As detailed in Section 2, bike/ped counts are collected within Coalition agencies, but often exist within silos and are stored using different formats and QA/QC processes, making them difficult to combine. While a few state DOTs work to standardize bike/ped data across local agencies (e.g., NCDOT, FDOT), they are the exception rather than the rule, and few options exist to standardize data at the "super-regional" or national level.

Counting bicycles, pedestrians, and other micromobility users is more challenging than counting motorized traffic. At a given location, traffic may need to be counted for multiple facilities (e.g., sidewalks, bike lanes, roads), multiple travel directions per facility, and multiple modes of travel (pedestrians, bicyclists, other micromobility modes). Additionally, unlike motorized traffic, which travels predictably along the direction of the roadway, pedestrians and other micromobility can move less predictably – at times including perpendicular to the direction of travel. Figure 11, taken from the Minneapolis Street Guide, shows an example configuration that has two sidewalks and two protected bike lanes, with each bike lane limited to a single direction. Recording bike/ped volumes along this roadway would require appropriately describing and assigning mode-specific counts to each facility and direction of travel.

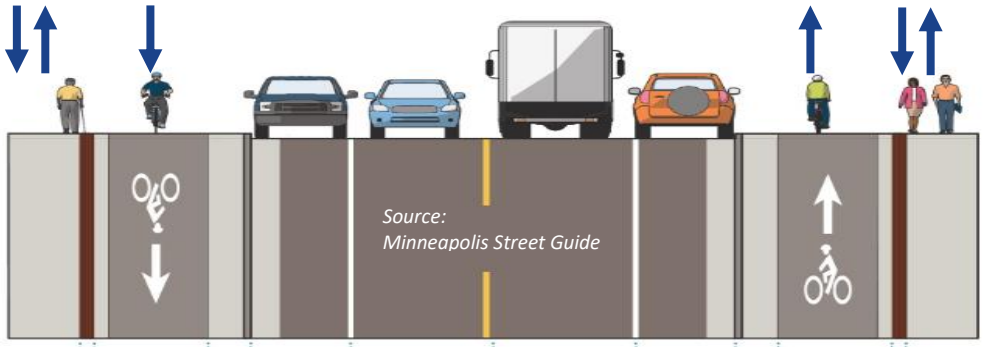

Figure 11: Example roadway with multiple micromobility facilities

The remainder of this section focuses on three existing data schemas that are used to store and communicate bike/ped data in a common format: (1) Federal Highway Administration's TMG format, (2) Portland State's BikePed Portal, and (3) TxDOT's Bike Ped Count Exchange (BP|CX). Note that all three schemas are not directly comparable; TMG is a data format/standard, while BikePed Portal and BP|CX are data management systems. However, each provides a detailed schema that describes how bike/ped counts and associated facility location information are stored. A key objective is to summarize "essential" information that is common to each data schema, as well as key distinctions for each.

At a high level, each of the data schemas reviewed in this section includes two main components, which are highlighted in Figure 12: (1) location metadata, which communicates where each count takes place, and (2) count data. Common attributes for location metadata include a unique location identifier, the latitude/longitude, a description of the facility type, travel direction, travel mode, and other road/path attributes. The count data is generally more straightforward; in most cases, it simply reports the observed traffic counts during a specified time period for location IDs defined in the location metadata, although in some cases supplemental information (e.g., helmet usage, weather) is also included. The remainder of the data schema review focuses on the former

component, location metadata, as describing count locations tends to be more complicated. Once a location is appropriately defined, it can be linked with resulting count data.

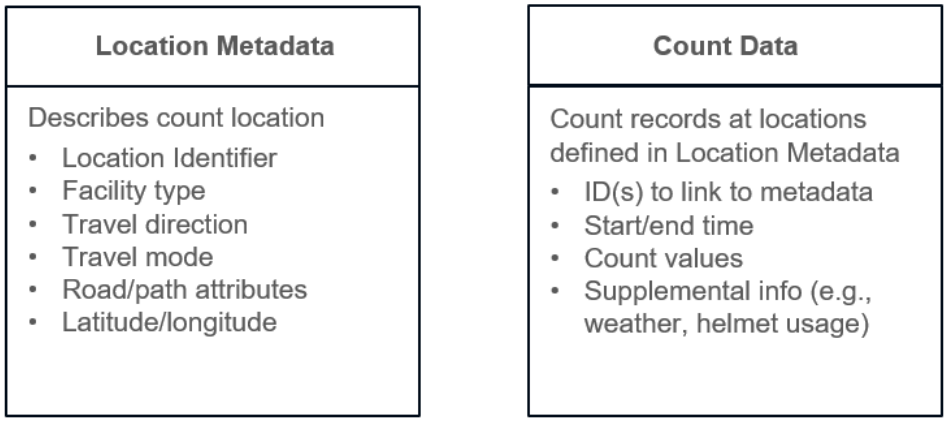

Figure 12: Two components of bike/ped count data schemas: location metadata and counts.

Schema Comparison

Traffic Monitoring Guide (TMG)

The TMG is a Federal Highway Administration (FHWA) format used for submitting count data to the Travel Monitoring and Analysis System (TMAS). This data is used by FHWA primarily for analysis and policy development purposes. Although the program has historically focused on motorized counts, it has been adapted over time to include micromobility traffic, including bike and pedestrian counts.

The TMG schema comes in two versions: TMG 2016 (for current submission to TMAS 2.8) and TMG 2022 (for submission to TMAS 3.0, starting in 2025). During the writing of this document, FHWA released an announcement that they will be updating both the TMG 2016 and 2022 formats to simplify the requirements for data submission. The motivation for these changes was based on feedback from agencies who were struggling to submit non-motorized count data to TMAS.

In all versions, the TMG bike/ped data schema is communicated through two files: a Station file and a Volume/Count file, which correspond to the 'location metadata' and 'count data' components of Figure 12, respectively. The station file describes the count station's attributes, including location, orientation relative to roadway, directionality, functional class, while the Volume file reports the traffic counts for each station defined in the Station file.

Location metadata

The TMG Station file is a text file that describes count locations referenced in the Volume file, with each record representing a unique station, mode of travel, and direction. This text file is defined with very specific formatting requirements; the 2016 TMG version expects 239 text characters per row, which are divided into 35 data fields (attributes) – each of which is allocated a specific set of characters. For example, the Station ID field uses six characters and is defined from character positions 7-12. Assembling a single row of the Station file involves (1) determining the proper

value for each attribute, (2) justifying and padding each attribute's text string based on its specified width, and (3) concatenating the attributes in the proper order. The 2022 TMG makes the formatting process easier by introducing delimiters to specify the breaks between attributes in each row.

Setting aside the formatting nuances, the 2016 TMG includes 35 data attributes, of which 15 are considered mandatory. Figure 13 shows the mandatory attributes, several of which are discussed below. For a full description of each attribute, please see the 2016 TMG or "Coding Nonmotorized Station Location Information in the 2016 Traffic Monitoring Guide.

Field	Width	Description	Type[1]
1	1	Nonmotorized station/location record identifier (L)	C
2	2	State FIPS Code	C
3	3	County FIPS Code	C
4	6	Station ID	C
5	**2**	Functional classification of the roadway	C
6	1	Direction of route	C
7	1	Location of count relative to roadway	C
8	1	Direction of movement	C
9	1	Facility type	C
11	1	Type of count	C
12	1	Method of counting	C
14	4	Year of Data	C
22	4	Year station established	C
25	8	Latitude	C
26	9	Longitude	C

Figure 13: Mandatory attributes in the 2016 Traffic Monitoring Guide format

- **Station ID**: A station represents a count site on a road/path – *not the physical sensor/device.* Different modes (e.g., bicyclists, pedestrians) and directional counts can share the same Station ID, and are differentiated by other attributes.
- **Functional Class**: Uses FHWA functional classification system (Classes 1-7, ranging from Interstates to Local roads), plus two additional classes: trail/shared use & general activity count.
- **Direction of route, direction of count relative to roadway, and Direction of movement**: Three attributes are needed to locate count facilities and define directional movements at each location. The first attribute is 'Direction of route', which defines the orientation of the road route in terms of cardinal/ordinal direction (e.g., N, NE, E, SE). This direction serves as a reference point from which the second and third attributes (Direction of count relative to roadway and Direction of movement) are measured. Figure 14 (taken from *Coding Nonmotorized Station Location Information in the 2016 Traffic Monitoring Guide Format)* shows these attributes visually.

  A key challenge, however, is that *the overall direction of the route may not correspond to the local orientation of the road*. In other words, a North-South route (route direction = North) may locally run East-West, in which case count facility and movement direction attributes would be calculated relative to North. This can make encoding and decoding values (e.g., "traffic on the side of the road for the listed route direction" or "traffic monitored perpendicular to the route, from left to right") unintuitive.

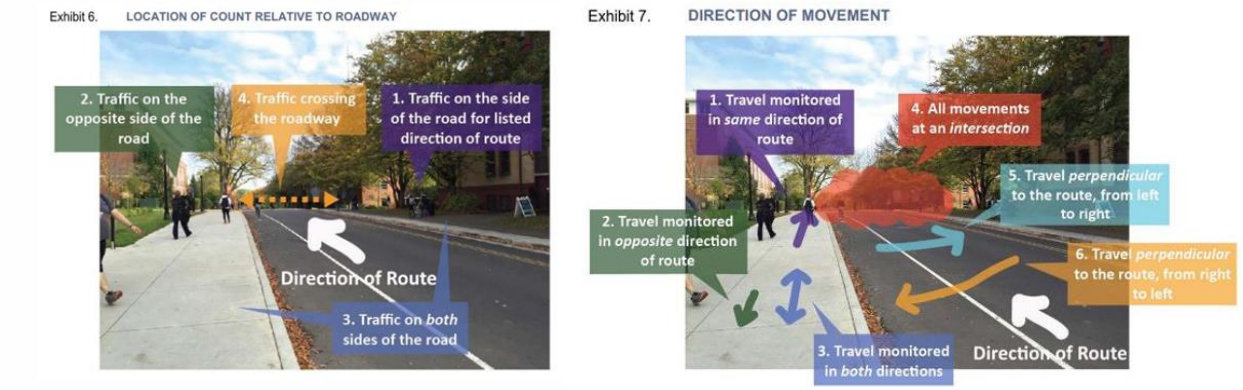

Figure 14: TMG 'Direction of Count relative to roadway' (left) and 'Direction of movement' (right)

- **Facility Type**: Uses nine defined facility types (e.g., trail, in roadway right-of-way, sidewalk, striped bike lane)
- **Type of count:** Uses nine defined user types, including pedestrians, bicyclists, people in wheelchairs.
- **Method of counting:** Manual, portable, or continuous counter.
- **Latitude/Longitude:** Geodetic coordinates for the road/path – not the sensor used to measure traffic.

Overall, the TMG is sufficiently robust to describe just about any count scenario, but can be complex to code correctly. Key challenges include describing where count facilities are located and the direction that users are traveling – both of which are measured relative to route direction, which may not reflect the local orientation of the road.

Upcoming Changes

At the time of writing, FHWA announced that they were in the process of making changes to the TMG format to make submission simpler. The most notable update is changing seven of the (previously 15) mandatory fields to "optional" -- including functional class, location of count relative to roadway, direction of movement, and facility type. Additionally, several fields are being renamed for clarity, and attributes related to directionality are no longer measured relative to the orientation of the route. Finally, FHWA has provided updated guidance on naming Station IDs, requiring separate records be created for each mode of travel and direction of movement.

Overall, these changes appear promising in terms of lowering the barrier to submitting bike/ped data to TMAS, and address at least part of the difficulty with describing locations/directionality. However, at least one committee member voiced concerns that by making important fields optional rather than mandatory, critical information will be missing from the TMAS database and render the count data less useful for certain applications.

BikePed Portal

BikePed Portal is a national data repository and management system for bicycle and pedestrian count data, which was developed and currently overseen by Portland State University's Transportation Research and Education Center (TREC). Agency count data, which can be imported from a variety of count sources, is standardized, quality-checked, and archived in a centralized database. Users interact with the data through a web interface, which supports analysis, visualization, and other features – including data export in a standardized format. BikePed Portal offers these bike/ped data management and analysis services based on a transparent pricing model, which includes a per-sensor setup fee and an annual license fee for data management and analysis services, including the ability to determine correction factors and estimate annual average daily non-motorized traffic (AADNMT).

Figure 15 shows a visual depiction of the BIkePed Portal schema, which includes both metadata and data components (corresponding to 'location metadata' and 'counts' in Figure 12, respectively). Whereas the TMG uses a flat file to the describe count locations, BikePed portal's schema has a nested structure and shows the relationship between different elements – an approach well suited to storing data in a relational database.

Overall, the information content in the BikePed Portal metadata section is very similar to the TMG Station file. For example, *Segment Area* is very similar to Station in TMG terminology, *Facilities* map to TMG facilities, and *Flows* correspond to TMG Travel Movements and Count Type fields. However, BikePed Portal's approach describes the direction travel movements more naturally, with *Flows* capturing the cardinal direction of travel (instead of calculating directionality relative to route orientation, as is the case for the TMG).

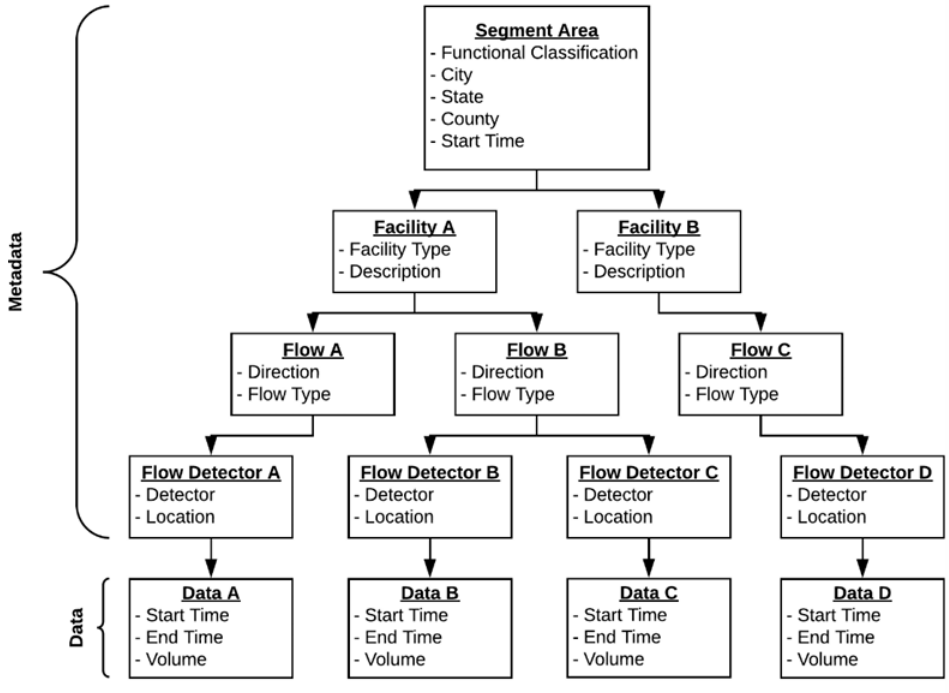

Figure 15: BikePed Portal data schema

Texas Bicycle and Pedestrian Count Exchange (BP|CX)

BP|CX is a data management and analytics platform for storing and analyzing bike/ped counts in Texas, which was developed by Texas A&M Transportation Institute (TTI) in coordination with Texas DOT. It serves as a common repository for TxDOT and local agencies' bike/ped counts across the state and is intended to suit the needs of both local agencies – who need to manage their own data – and the state DOT, which is interested in analyzing data across the whole state. Users interact with the BP|CX through a web platform, which has different levels of access for importing and reviewing data, assessing data quality, and exploring results.

The BP|CX data schema was built on top of the TMG format, and thus organized the same way, using both Station and Volume files. However, BP|CX made several additions to the basic set of TMG Station file attributes to make the files more useful for TxDOT purposes. Figure 16 summarizes the attributes in the Station file, with green records representing mandatory fields for both TxDOT and TMG, and blue records showing the additional mandatory attributes for TxDOT. Key additions to the TMG include *Travel Direction,* which captures travel direction in terms of cardinal direction (not just relative to route direction), and *Flow ID TxDOT*, which combines several fields, including *Travel Direction*, to create a more meaningful id. This essentially creates a single primary key for the station file, making it straightforward to identify a unique location, type of count, and direction of travel.

Table 1. Summary of Attributes in Station Description File.

#	Attribute Name	TxDOT Required?	TMG Required?	Maximum Length in TxDOT Database
1	State	Yes	Yes	Enumerated values
2	TxDOT District	Yes	N.A.	Enumerated values
3	County	Yes	Yes	Enumerated values
4	City/Area	Yes	N.A.	None
5	Station ID TMG	Yes	Yes	6 characters
6	Station Name	Yes	N.A.	None
7	Flow ID TxDOT	Yes	N.A.	14 characters
8	Station ID Agency	No	N.A.	None
9	Travel Direction	Yes	N.A.	Enumerated values
10	Functional Classification	Yes	Yes	Enumerated values
11	Direction of Route	Yes	Yes	Enumerated values
12	Location of Count Relative to Roadway	Yes	Yes	Enumerated values
13	Direction of Movement	Yes	Yes	Enumerated values
14	Facility Type	Yes	Yes	Enumerated values
15	Intersection	No	No	Enumerated values
16	Type of Count	Yes	Yes	Enumerated values
17	Method of Counting	Yes	Yes	Enumerated values
18	Type of Sensor	Yes	No	Enumerated values
19	Year of Data	Yes	Yes	4 characters
20	Factor Group 1	No	No	1 character
21	Factor Group 2	No	No	1 character
22	Factor Group 3	No	No	1 character
23	Factor Group 4	No	No	1 character
24	Factor Group 5	No	No	1 character
25	Primary Count Purpose	No	No	Enumerated values

Figure 16: Texas BP|CX Station file containing location metadata.

TTI Experience with Data Wrangling

Several members of the CAPABLE steering committee emphasized that challenges associated with standardizing bike/ped data are not just technical, and that significant agency resources and local champions are needed to "wrangle" the data into a common format, verify quality, and provide access to different stakeholders. To this end, representatives from Texas Transportation Institute (TTI) joined a CAPABLE committee meeting to discuss their experience over the last 7-

8 years building and managing the BP|CX for TxDOT – a process that required working with local agencies across a large state and creating a unified view of bike/ped data in the state.

TTI defined the concept of "data wrangling" in two parts: (1) *gathering data* from multiple sources and agencies – a process that requires determining and confirming metadata, and (2) *processing data* to a unified standard and reviewing quality. To facilitate data gathering, they developed a centralized open/crowdsourced platform that local agencies could easily use, created direct API links to Eco-Counter (the most widely used data collection equipment deployed throughout the state) while also supporting other formats, and created software tools to support metadata collection. After agency data is in their system, they automatically process the data to a common format, send data through a robust quality review process – including data visualization and automated QC tests, optionally perform factoring, and compute performance measures. Based on experience doing this for 7-8 years, they shared the following lessons about data wrangling: (1) It's not easy, (2) prioritize and commit to maintenance, (3) permanent counts are critical, (4) local partners make it easier, and (5) quality review is also critical.

Summary Findings

The following points summarize top-level findings from the project team's review of the TMG, BikePed Portal, and Texas BP|CX data schemas.

- **There are more similarities than differences between schemas.** Texas BP|CX was built on top of the TMG format, and thus is highly similar to the TMG – albeit with some additions to better suit TxDOT's needs. BikePed Portal does not organize data in the same flat file format and uses different terminology; however, the information captured in their schema is extremely similar to the TMG.
- **Describing directionality is a critical issue and can be a pain point when coding locations.** The 2016 TMG requires several data attributes to describe the location of facilities and the direction of movement and uses the orientation of a route (which may not reflect the local orientation of the road) as a point of reference – making it difficult to code correctly and interpret. However, updates to the TMG (in process at the time of writing) are expected to make this process easier. BikePed Portal and Texas BP|CX have developed other ways to handle directionality, introducing the concept of directional "flows".
- **FHWA is the process of updating the TMG format** for bike/ped data, which includes reducing the number of fields deemed mandatory. Many of the draft changes appear to be useful improvements; however, there is concern that making certain fields optional may make the resulting TMAS dataset less useful for specific applications.
- **A common interchange format is needed for data sharing.** Even though each of the schemas reviewed contain similar information, they currently aren't inter-operable. Although each can export data in the TMG format, as mentioned in the previous point, the TMG format is evolving and may not contain essential fields. As the Coalition thinks about how to represent bike/ped data in a future procurement, it would be beneficial to define an interchange format that captures essential information and is inter-operable with the TMG.
- **Data wrangling is part of standardization and can be resource intensive**. A stable interchange standard would make sharing bike/ped count datasets easier, but it is expected that someone – either agency staff or an external resource – will need to coordinate with local agencies to contextualize and integrate their locally-collected data. There is a technical aspect to this, but adequate staffing is equally important.

4. Opportunities and Next Steps

In light of the current challenges associated with existing bike/ped count programs and vendor products in the Coalition, the project team recognizes that the entire bike/ped data collection ecosystem must be considered and proposes the following framework, shown in Figure 17, for enhancements to the availability, quality, and cost of bike/ped data.

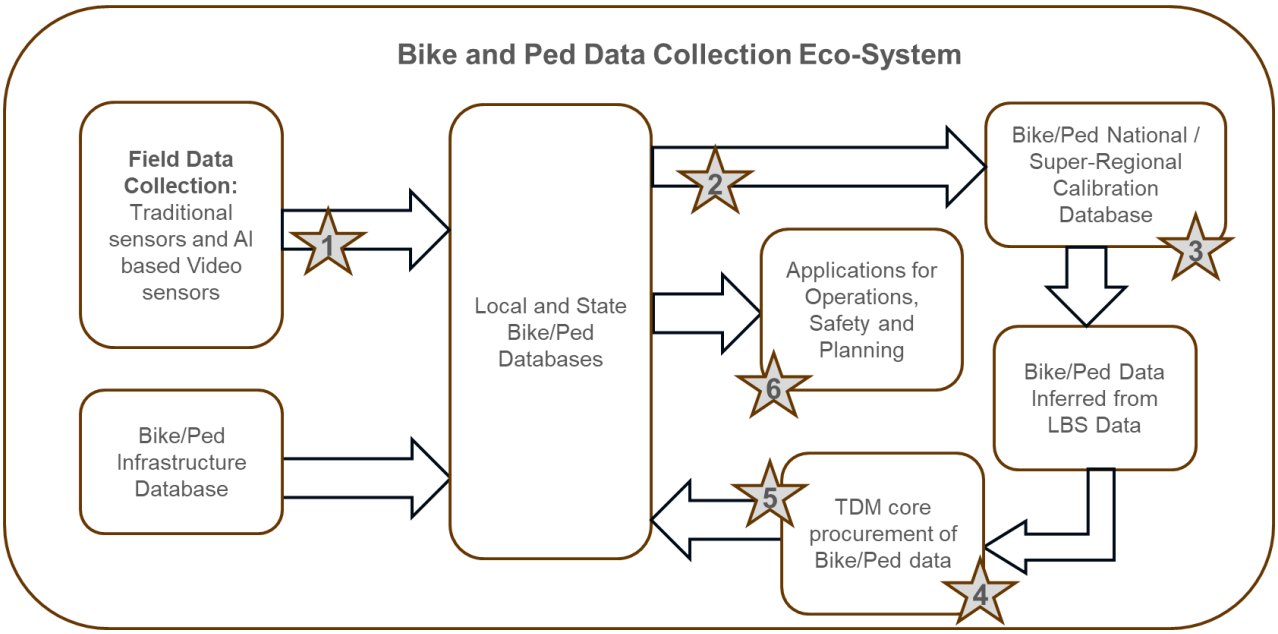

Figure 17: Ecosystem for bike/ped data and opportunities for Coalition efforts

The components of this ecosystem include:

- **Field data collection:** Comprised of the entities that provide the equipment and services to collect bike/ped count data in the field.
- **Bike/ped infrastructure database:** Describes the locations and attributes of bike/ped infrastructure, making sure to capture the varying levels of bicycle facility types (dedicated paths, bike lanes, and shared lanes) as well as the various facility attributes that support pedestrian activity.
- **Local and State bike/ped database:** Contains the count data collected by individual state or local agencies. These databases are typically not designed to be combined with count data from other municipalities.
- **Applications of bike/ped Data**: Bike/ped data collected by state and local agencies are used to support local objectives, which may include operations, safety, and planning application areas.
- **Bike/ped calibration database**: This is a database that aggregates local and state level data to a national or super-regional dataset with consistent format, schema, and quality. FHWA's TMG is a potential candidate (as the TMAS data sets for traffic volume serve this purpose), as is a Coalition-developed database schema. Such a dataset can be easily ingested by third third-party data vendors that utilize LBS and other information sources

to infer bike/ped activity across larger networks and areas. This is recognized as a missing element in the current ecosystem.

- **Bike/Ped data inferred from LBS data**: These third third-party information brokers attempt to provide (or infer) the level of bike/ped activity at a segment or area level. They rely on general behavioral patterns derived from smart phones and other behavioral data available at scale or modeled at scale. These vendors rely on high-quality bike/ped count data from local sources to calibrate estimates of bike/ped activity.

Opportunities for Coalition Involvement

The opportunities for Coalition involvement, influence, and programs are numbered below based on the data flow in Figure 17. They represent possible actions by the Coalition to improve the bike/ped data ecosystem.

1) **Develop data standards for use with emerging technologies such as video-based count and classification solutions** AI/ML learning sensing technologies are already impacting bike/ped data collection, and such technology is expected to continue to proliferate. Based on the relatively inexpensive cost of deploying camera technology combined with the current trajectory of AI/ML, there is reason to believe this approach will dominate the industry in future years. As new products and vendors come online, it is an opportunity to normalize/standardize data format interfaces to influence the industry towards standards-based data collection. This includes not only types and formats of data, but also quality control and quality assurance, and georeferencing methods that are interchangeable. The latter is critical so that jurisdictions are not trapped by the cost of translating bike/ped data from one vendor specific format to others (as is currently prevalent in traffic data streams), and to allow for vendor interoperability as much as possible.
2) **Support/develop/collaborate toward database interchange formats:** Aggregation of bike/ped data toward state, regional, or even national databases for various purposes, inclusive of a common calibration database (see #3) will be enabled by data interchange formats that are open and accessible to all. Similar to #1, such standards will allow more seamless (and thus less resource intensive) data transmission, avoid proprietary vendor formats, and can be implemented through procurement specifications (see #4 and #5 below). These standards are closely related to #1 above, with respect to QA/QC and georeferencing, but are applied to aggregate data from local and state sources to regional and national sources thereby creating a multi-state or national calibration database (#3).
3) **Develop a calibration database for access by third third-party sources:** Access to a large quantity of high-quality calibration data is a key resource currently missing from the ecosystem. The Coalition can either sponsor such a database or collaborate toward a national resource. Aspects of this resource include:
    - The data would incorporate the standards described in #1 and #2 and make broad-based, multi-state data available.
    - The calibration database may physically house the data or may have indexes to distributed data sources at the state or regional level.
    - QA/QC standards would be applied such that only high confidence data is available—data would be scored based on level of confidence-quality.

4) **Transportation Data Marketplace bike/ped core data item:** Currently bike/ped data through the Eastern Transportation Coalition (TETC) Transportation Data Marketplace (TDM) is available only as an ancillary product from vendors providing primarily vehicle-based or vehicle movement data. The current TDM was procured in 2022 when bike/ped data products were emerging, and no requirements related to data standards, schemas or accuracy are currently in effect. This effort would develop and execute a procurement effort that encompasses sensor technology and field data collection services, as well as third third-party data vendors that infer bike/ped activity from LBS data. The procurement would include standards (activity #1 & #2), incorporate open geo-referencing and require common QC/QA protocols. Subsequently, validation activities (see activity #5) would be developed to ensure quality. This would transition bike/ped data from an ancillary product to a core data item and invite vendors to respond to a specification-based RFP to be added to the TDM family of products.

5) **Validation of TDM procured bike/ped data:** Along with developing bike/ped data into a core product, the Coalition would bring bike/ped data into its validation program in a systematic way to monitor quality, accuracy, and specification adherence of data from vendors. This would be inclusive of third third-party aggregators that leverage LBS data for broader coverage of networks and areas, as well as vendors of physical sensing equipment and associated data collection services.

6) **Applications of bike/ped data:** With the increased emphasis on the safety of vulnerable road users, and heightened sense of importance of non-motorized forms of transportation to enhance access, quality of life, and decrease green-house gases, transportation agencies are faced with the responsibility to create and carry out programs that require the type of bike/ped data discussed in this white paper. Possible implementation support from the Coalition includes (1) creating educational materials reflecting the practices of leading practitioners, (2) developing a working group for non-motorized transportation (inclusive of bike/ped data), and (3) working with existing groups such as those that support RITIS application enhancements to develop Coalition specific bike/ped applications.

Recommended Next Steps

The project team presented the bike/ped data ecosystem and opportunities for Coalition involvement to the steering committee. Based on committee feedback and additional stakeholder conversations, the project team recommends the following next steps, which will position the Coalition to procure higher quality bike/ped data in the anticipated 2026 TDM procurement.

- **Establish an ongoing Coalition bike/ped committee** to ensure that data formats, specs, and future procurement of bike/ped data in the TDM align with Coalition member needs. The steering committee formed for the CAPABLE was highly engaged and provided invaluable feedback on the use cases, challenges, and insight into the nuances of bike/ped data collection within the Coalition. Re-forming – and perhaps expanding – the committee is a critical first step to any follow-up work. In addition to providing feedback on data formats and specifications, the committee will also be invited to inform/consult on Traffic Data Marketplace (TDM) data validation activities related to bike/ped data.

- **Create a consistent data schema for bike/ped count data**, to be developed in coordination with practitioners and enforced through the TETC Transportation Data Marketplace procurement. The *bike/ped count schema* would unambiguously describe bike and pedestrian counts through a prescribed data format, making it easier to exchange count information and populate databases in a standardized way.

  Future data collection services through the TDM would be expected to adhere to this format (via RFP requirements), but it would also serve as a common format that legacy data could be translated to, and that agencies could use when hiring consultants to assist with data collection. It could also be used to develop a calibration database that would enable TDM vendors to improve estimation accuracy.

- **Establish Coalition resources to support agencies with bike/ped data collection and analysis.** This recommendation addresses the fact that bike/ped data collection is often (though not always) undertaken at the local level, and many small agencies/municipalities do not have internal resources or prior experience "wrangling" the data to common formats or implementing QA/QC methods. In response, the plan should be developed to:
  - Identify how the Coalition can support agencies with "data wrangling" and management
  - Provide ongoing education on best practices for bike/ped data collection and analysis (e.g., emerging technologies, QA/QC methods). This may include written materials, webinars, or a training program.

Ultimately, improving the availability, quality, and cost of bike/ped data in the Coalition requires considering bike/ped data landscape holistically. Agency count programs serve as a critical component of this landscape, supporting diverse use cases at the local and state level, while also representing the key source of calibration data for modeled products that can generate bike/ped volume estimates at scale. Supporting agencies with these data collection efforts and making it easier to standardize and share the resulting counts – the intent of the recommendations outlined above – represent practical next steps to improve bike/ped data in the Coalition.