APPENDIX A

PEER AGENCY SURVEY RESPONSES AND INTERVIEWS

1 PEER AGENCY SURVEY OVERVIEW

This chapter provides the results from an online survey sent to a total of 35 partner agencies, with 22 of them responding. This chapter intends to:

- Provide the reader with the survey responses about the agency's use and experience with Big Data sources.
- Provide a summary of the discussion from the eight (8) follow-up interviews with peer agencies.

The online survey was designed using the Qualtrics online survey tool with questions developed to elicit responses from peer agencies. The online survey asked about how peer agencies are using Big Data used by each peer agency, what they are being used for, how they are acquired and processed, and what challenges agencies face with their adoption.

Nine of the 22 agencies that responded to the online survey were identified for follow-up phone interviews based on their responses, geographic location in relation to the metropolitan Washington region, and comparable metropolitan regional size. Of the nine interviews requested, eight peer agencies were interviewed, and their responses were used to further inform the state-of-practice review. The peer agencies interviewed as part of the state-of-practice review are indicated in **Table 1**.

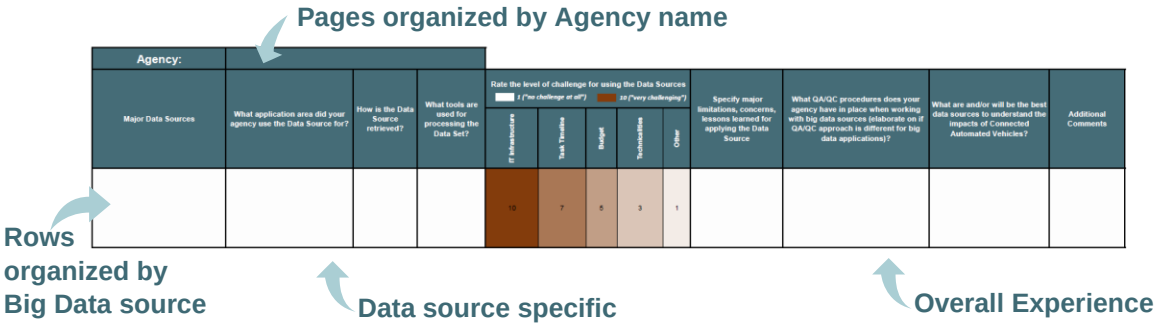
Table 1 | Online Survey Respondents

No.	Agency Name	Metro Area	State(s)
1	Atlanta Regional Commission (ARC)*	Atlanta	GA
2	Chicago Metropolitan Agency for Planning (CMAP)*	Chicago	IL
3	Capital Region Planning Commission (CRPC)	Baton Rouge	LA
4	Delaware Valley Regional Planning Commission (DVRPC)	Philadelphia	PA, DE
5	East-West Gateway Council of Governments (EWGCOG)	Saint Louis	MO
6	Fredericksburg Area Metropolitan Planning Organization (FAMPO)	Fredericksburg	VA
7	Houston-Galveston Area Council (HGAC)	Houston	TX
8	Maricopa Association of Governments (MAG)*	Phoenix	AZ
9	Metro (Portland, Oregon)*	Portland	OR
10	Miami-Dade Transportation Planning Organization [Florida Metropolitan Planning Organization Advisory Council (MPOAC)]*	Miami, Tampa, Orlando	FL
11	Mid-Ohio Regional Planning Commission (MORPC)	Columbus	OH
12	New York Metropolitan Transportation Council (NYMTC)*	New York	NY
13	North Central Texas Council of Governments (NCTCOG)	Dallas/Fort Worth	TX
14	North Jersey Transportation Planning Authority (NJTPA)	Newark	NJ
15	Northeast Ohio Areawide Coordinating Agency (NOACA)	Cleveland	OH
16	Pima Association of Governments (PAG)	Tucson	AZ
17	Regional Transportation Commission of Southern Nevada (RTCSNV)	Las Vegas	NV
18	Southeast Michigan Council of governments (SEMCOG)*	Detroit	MI
19	Southern California Association of Governments (SCAG)*	Los Angeles	CA
20	Virginia Department of Transportation (VDOT)	Virginia statewide	VA
21	Wasatch Front Regional Council (WFRC)	Salt Lake City	UT
22	Anonymous**	Anonymous	-

** Peer agencies who participated in a follow-up phone interview.*
*** Respondents were given option of not identifying themselves although they could provide a response.*

1.1 PEER AGENCY SURVEY RESPONSES

For the 22 agencies that responded to the online survey, each agency's response is populated in a matrix on the following pages. The online survey asked about Big Data used by each peer agency along with more specifics such as what they are used for, how they are acquired and processed, and what challenges agencies face with their adoption. Lastly, respondents could provide comments on overall experience with Big Data including, QA/QC procedures, data sources to understand Connected and Automated Vehicles, and any additional comments. The matrix format, along with specific questions, are outlined below:



BIG DATA SOURCES

The major data sources (rows of the matrix) the respondents collectively use is listed below:

Contextual data

- Disaggregate Census dataset (Specify, e.g. PUMS)
- Government record (Specify, e.g. appraisal, business licensing, QCEW)
- InfoUSA (Data Axle USA)/Dun & Bradstreet business listing
- Google Places
- Co-Star Data
- LIDAR and Other Airborne Data
- Job Postings Data

Location based service (LBS) and Navigational-gps data used

- National Performance Management Research Dataset (NPMRDS)/INRIX
- HERE
- StreetLight Data
- American Transportation Research Institute (ATRI) truck GPS data
- Google Travel Time

Other Big Data

- Transit Automatic Vehicle Location
- Farecard/Toll Data
- Social media (specify. Facebook, LinkedIn, etc.)
- Continuous traffic data
- Other data sources (Open Response)

Data sources that are included in the above list, but not utilized by a specific agency are listed below the matrix.

DATA SOURCE SPECIFIC INFORMATION

The subsequent questions in the survey aimed to gain insight into respondent's experience with each individual Big Data source. The questions for each data source were the same for each agency and are listed below:

- What application area did your agency use the data source for?
- How is the data source retrieved?
- What tools are used for processing the data set?
- Rate the level of challenge for using the data sources (1 = "not challenging", 10 = "very challenging")
- Specify major limitations, concerns, lessons learned for applying the data source

OVERALL EXPERIENCE

- What QA/QC procedures does your agency have in place when working with Big Data sources (elaborate on if QA/QC approach is different for Big Data applications)?
- What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?
- Additional Comments

1.2 PEER AGENCY INTERVIEW SUMMARY

Of the nine interviews requested, eight peer agencies were interviewed, and their responses were used to further inform the state-of-practice review. The purpose of these phone interviews was to gather more information and clarify responses to survey questions. The interviews were scheduled for one hour, and agencies were encouraged to share as many thoughts as they wished during the interview.

Table 2 | Peer Agencies Interviewed

No.	Agency Name	Metro Area	State(s)
1	Atlanta Regional Commission (ARC)	Atlanta	GA
2	Chicago Metropolitan Agency for Planning (CMAP)	Chicago	IL
3	Maricopa Association of Governments (MAG)	Phoenix	AZ
4	Metro (Portland, Oregon)	Portland	OR
5	Miami-Dade Transportation Planning Organization [Florida Metropolitan Planning Organization Advisory Council (MPOAC)]	Miami, Tampa, Orlando	FL
6	New York Metropolitan Transportation Council (NYMTC)	New York	NY
7	Southeast Michigan Council of governments (SEMCOG)	Detroit	MI
8	Southern California Association of Governments (SCAG)	Los Angeles	CA

PEER AGENCY PHONE INTERVIEW QUESTIONS

Questions were structured to gather insight on each peer agency's experience using Big Data for application areas related to the TPB needs. The interviewer guided the conversation in a hybrid fashion to address all major questions identified from the survey feedback and enable fluid conversation based on the interviewed agency's direct experience with Big Data.

1. *Please describe the challenges, limitations, and caveats with applying Big Data in travel demand modeling. Consider the following areas of travel demand modeling.*

    - Obtaining origin-destination ("O-D") patterns by mode.
    - Obtaining geographic O-D for trips that use a specific road segment.
    - Obtaining better information about vehicle travel speeds and volumes.
    - Better understanding the dynamically-priced toll lanes, such as the High Occupancy Toll ("HOT") lanes, including who is using them and the rate at which they are being used per week.
    - Understanding the volume breakdown between the adjacent general-purpose lane and dynamically-priced lanes.
    - Identifying trip purpose / type of destination.
    - Obtaining estimates of through travel and external travel (external-to-internal and internal-to-external), including external transit travel.
    - Using O-D data to validate model estimation derived from household travel surveys.

2. *Please describe the challenges, limitations, and caveats with applying Big Data in understanding and analyzing TNCs. Consider the following areas of TNCs.*

    - Quantifying the magnitude and O-D patterns of TNC trips.
    - Assessing the existing relationship between TNCs and transit travel (i.e., understanding where TNCs support/enhance vs. compete with transit travel).
    - Understanding how and where (across the region, by land use, by residential and commercial density) TNCs are influencing Vehicle Miles of Travel (VMT).
    - Temporal distribution of TNC ridership (peak hours, weekday/weekend, monthly, etc.).

- Demography of TNC riders in the region, including but not limited to income, race/ethnicity, and gender.

3. *Please describe the challenges, limitations, and caveats with applying Big Data in travel survey and traffic count program. Consider the following areas of interest.*
   - Obtaining permanent/temporary/hourly traffic count data.
   - Providing a measure of the day-to-day variability of traffic counts.
   - Validating and supplementing traffic counts collected by State DOTs.
   - Obtaining vehicle classification estimates to supplement traffic counts collected by State DOTs.
   - Data on longitudinal traffic counts (traffic counts by certain times of day over years / overlaid with different land use developments and policies such as HOT lanes).
   - Integrating Big Data with household travel surveys to provide both long-distance and local travel behavior data on a recurring basis.
   - Validating data collected from household travel surveys (i.e., National Household Travel Survey (NHTS), Regional Travel Survey, etc.).

4. *Please describe the challenges, limitations, and caveats with applying Big Data in System Performance/Congestion Management Process. Consider the following areas of interest.*
   - Monitoring and/or evaluating transportation system performance.
   - Obtaining regional and/or project-specific congestion management information.
   - Understanding impacts of congestion management strategies.
   - Recognizing trends in new congestion management strategies and beginning to predict/track their impacts as well.
   - Understanding impacts of specific events, such as major traffic incidents, severe weather events, or major scheduled special events, as well as the impacts of strategies employed to manage transportation impacts of these events.
5. *Please describe the challenges, limitations, and caveats with applying Big Data in other areas of research (e.g., Multimodal, Transit, Freight, Travel Demand Management, Equity, etc.).*
   - Obtaining information about existing and forecast regional freight and commercial vehicle travel.
   - Obtaining information about intercity bus travel.
   - Identifying the destination choice and mode choice for areas with high concentrations of low-income and/or minority residents.
   - Understanding interactions between traffic, traffic congestion, and transit bus operations.
   - Monitoring real-time traffic/transit/rail data.
   - Evaluating how Big Data could be used for transportation-related project selection.
   - Understanding how the use of alternative modes for commuting (bike, walk, ridesharing, transit, car/vanpool, telework) affects the overall network in terms of major highways and arterials.
   - Evaluating the influence of employer-based TDM policies on travel behavior (e.g. obtaining more insight about the influence of teleworking policies on trip generation).
6. *Please describe the challenges and strategies to overcome these challenges in implementing Big Data regarding the aforementioned programs. Consider the ratings you provided in the survey as they relate to IT infrastructure, task timeline, budget/cost, technicalities, others?*

7. *Please describe the QA/QC procedures your agency has in place when working with Big Data sources (elaborate on if QA/QC approach is different for Big Data applications)?*

8. *Please describe any cost-saving strategies in prescribing/purchasing/developing Big Data analytics.*

9. Provide any additional comments about Big Data that are not covered in questions above.

PEER AGENCY PHONE INTERVIEW SUMMARY

Most of the individuals who volunteered to participate in the interview on behalf of their organizations were travel demand modelers. These interviewees are considered advanced users of conventional and Big Data and rigorous about the data quality. The conversation relating to Big Data related challenges and limitations within the traffic demand modeling realm was particularly valuable and insightful. They were able to share experiences based on their familiarity and experience with the data. Due to the interviewee's expertise in the travel demand models, they did not have sufficient knowledge about the Big Data applications in other program areas (e.g., data collection, performance monitoring, and congestion management).

With Big Data sources still new to most metropolitan planning organizations (MPO) agencies, agencies maintain a high level of interest due to the potential advantages while also maintaining a healthy level of skepticism about the quality of the data. In general, the skepticism outweighs the interest, resulting in the level of Big Data usage in the travel demand modeling domain remaining relatively low with a few exceptions:

- Big Data sources have been widely accepted as the 'replacement' of external travel surveys for the development of external (E-E, E-I and I-E) trip tables. In large part, if not entirely, this is because performing a traditional intercept survey at external stations has become overwhelmingly challenging and, in many states, is deemed unconstitutional. Using the Origin-Destination (O-D) pattern offered by Big Data sources in conjunction with the traffic counts collected at the external stations to derive external travel pattern is the best, and likely the only feasible, option for many MPO and state transportation agencies. The Big Data sources need to be reviewed on a fairly frequent basis (every couple of years) as the O-D pattern offered by the Big Data through one-time purchase is subject to changes due to land use and transportation system changes.
- Many MPOs use INRIX/HERE data either through a purchase acquisition or through a MPMRDS arrangement for light model speed/travel time calibration and validation, mainly at corridor level.
- QCEW data for employment database development if accessible.

On the planning side, either at the regional level or mostly at the corridor project level, many agencies indicated that corridor projects are not their responsibility (except for a few who conducted one-off projects). Interviewees were aware that Big Data is increasingly utilized in projects at the local and state DOT levels.

In terms of the challenges and limitations of Big Data, all MPOs, except for New York Metropolitan Transportation Council (NYMTC), agreed that the technical challenge from Big Data application is minimal. Furthermore, no MPOs designated a data analyst position for Big Data applications. One common challenge heard during the interviews is related to the GIS process required to make the Traffic Message Channel (TMC) based travel time/speed reporting from HERE/INRIX and previous version of NPMRDS consistent with reporting denominator (e.g., model highway networks, etc.) for travel demand modeling and other MPO functions.

Some interviewed MPOs identified the data acquisition cost (or upfront purchase cost) as one major challenge, particularly when a continuous subscription for a relatively long time periods is desired. Although, agencies have varying stances towards these types of purchases. For example, a continuous subscription of Streetlight data for a relatively long time period was deemed important by Metroplan in Orlando, Florida while deemed cost-prohibitive by Atlanta Regional Commission. MPOs expressed frustrations over the non-transparent pricing policies most Big Data vendors provide. However, two agencies indicated that they had worked out budgetary plans to use Transportation Management Area (TMA) funds to overcome the cost

challenge. For the freely available NPMRDS product, most agencies noted the network coverage is not extensive enough to support many modeling and planning studies.

The limitations of Big Data sources regarding data validation and transparency reduce their applications at the MPO agencies. Many interviewed agencies clearly expressed their concerns with the validation/quality of datasets from Big Data sources. These concerns primarily arose from a lack of a thorough understanding of how the data samples were collected, processed, and expanded by the data vendors such as StreetLight and AirSage.

Several agencies noted that when comparing the Big Data results with the other collected data, such as StreetLight AADT vs. traffic counts, and StreetLight O-D vs. O-D from HTS and travel model estimates, the discrepancy was noted but could not be evaluated to conclude the factors contributing to the discrepancy or determine which data set is more accurate (e.g., StreetLight vs. NTS). The bottom line is that the agencies which conducted the comparison studies all chose not to use the Big Data to replace the data they collect using traditional methods. Another agency that reviewed Streetlight/AirSage AADT and vehicle classification data noted that it did not reach the level and standards for implementation.

Many agencies currently have aspirations on the prospect of applying Big Data sources to provide TNC analytics; however, hardly any of the agencies interviewed collected and/or used any data to analyze TNC and its impact on the transportation system. A couple of agencies, including NYMTC and Chicago Metropolitan Agency for Planning (CMAP), are considering how to use the TNC data collected through the agreements between Uber, Lyft, and Taxis and the state/local governments in their regions, but no plan or progress was reported thus far. NYMTC noted they had no in-house expertise to analyze the TNC data that they had access to.

One agency noted that the ITS approach to use traffic signal and surveillance devices to collect and process the video data to become traffic counts to support planning/modeling is emerging and promising. However, IT solutions could be challenging in data archiving, processing, and reporting. With FHWA vehicle classifications no longer a good option and no Big Data solution yet, the vehicle classification process continues to be a challenge.

Agency:	**Atlanta Regional Commission**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
Disaggregate Census dataset (Specify, e.g. PUMS)	• Regional and corridor level OD patterns • Managed lane time of day usage • Trip purpose and destination type • Telecommuting trends • Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Freight and commercial vehicle travel • Travel behavior of special population group • Traffic dynamics and interactions with transit • Data validation and integration of HTS	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	6	6	6	6	1	• timeliness of data	• GIS automated procedures	• unknown at this time§but INRIX very likely	• ground-truth benchmarking
Government record (Specify, e.g. appraisal, business licensing, QCEW)	• Regional and corridor level OD patterns • Other model calibration/validation • Travel behavior of special population group	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	7	7	7	7	1	• limitations on data use			
InfoUSA/Dun & Bradstreet business listing	• Regional and corridor level OD patterns • Other model calibration/validation	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	4	4	4	4	1	• limited geographical details			
Co-Star Data	• Regional and corridor level OD patterns • Other model calibration/validation	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	4	3	3	3	1	• limited application to employment data trends			
LIDAR and Other Airborne Data	• Regional and corridor level OD patterns	• Purchase/Subscription	• Database tools (including GIS)	3	3	3	3	1	• limited coverage in GIS			
National Performance Management Research Dataset (NPMRDS)/INRIX	• Regional and corridor level OD patterns • Other model calibration/validation • Day-to-day traffic variation • System performance monitoring/evaluation	• Purchase/Subscription	• Database tools (including GIS) • Custom computer coded scripts	7	7	7	7	1	• INRIX is expensive and costly			
HERE	• Regional and corridor level OD patterns • Other model calibration/validation • Hourly counts and count validation • Day-to-day traffic variation	• Purchase/Subscription	• Database tools (including GIS) • Custom computer coded scripts	3	3	3	3	1	• became outdated			
StreetLight Data	• Regional and corridor level OD patterns • Other model calibration/validation • TDM policy evaluation • Telecommuting trends • Day-to-day traffic variation • System performance monitoring/evaluation	• Purchase/Subscription	• Database tools (including GIS) • Custom computer coded scripts	7	7	7	7	1	• expensive very limited application to specific prioject			

Agency:	**Atlanta Regional Commission**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")* **IT Infrastructure**	**Task Timeline**	**Budget**	**Technicalities**	**Other**	**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
American Transportation Research Institute (ATRI) truck GPS data	• Regional and corridor level OD patterns • Freight and commercial vehicle travel	• Purchase/Subscription	• Database tools (including GIS) • Custom computer coded scripts	3	3	3	3	1	• actually pretty good			
Google Travel Time	• Regional and corridor level OD patterns • Other model calibration/validation • Hourly counts and count validation • Day-to-day traffic variation • System performance monitoring/evaluation	• Purchase/Subscription	• Database tools (including GIS) • Custom computer coded scripts	8	8	8	8	1	• difficult to obtain and download			
Transit Automatic Vehicle Location	• Other model calibration/validation • System performance monitoring/evaluation	• Collected In-house • Purchase/Subscription • Shared cost with other agencies	• Database tools (including GIS) • Custom computer coded scripts	5	5	5	5	1	• AVL pretty useful with GTFS			
Farecard/Toll Data	• Regional and corridor level OD patterns • Managed lane time of day usage • Trip purpose and destination type • Other model calibration/validation • Hourly counts and count validation • Day-to-day traffic variation • System performance monitoring/evaluation	• Collected In-house	• Database tools (including GIS) • Custom computer coded scripts	3	3	3	3	1	• toll data from state tolling agency fare data from MARTA			
Continuous traffic data	• Regional and corridor level OD patterns • Managed lane time of day usage • Day-to-day traffic variation	• Collected In-house	• Database tools (including GIS) • Custom computer coded scripts	4	4	4	4	1	• data from GDOT fairly unique			
Other data sources	• Regional and corridor level OD patterns • External travel attributes • Day-to-day traffic variation	• Purchase/Subscription	• Database tools (including GIS) • Custom computer coded scripts	7	7	7	7	1	• Airsage data to update externals			

Data sources not included by Atlanta Regional Commission

Google Places, Job Postings Data, Social media

Agency:	**Anonymous Agency**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
Disaggregate Census dataset (Specify, e.g. PUMS)	• Regional and corridor level OD patterns • Trip purpose and destination type • External travel attributes • TNC OD characteristics§e.g. time/location • Hourly counts and count validation • Vehicle classification • System performance monitoring/evaluation • Freight and commercial vehicle travel	• Purchase/Subscription	• Database tools (including GIS) • Custom computer coded scripts	7	8	8	6	1	• Sample size of the Disaggregate data might be an issue	• Regular QA/QC procedures	• Micro-simulation data	• n/a
Government record (Specify, e.g. appraisal, business licensing, QCEW)	• Other model calibration/validation • TNC OD characteristics§e.g. time/location • TDM policy evaluation • Vehicle classification • System performance monitoring/evaluation • Freight and commercial vehicle travel	• Purchase/Subscription	• Custom computer coded scripts	7	8	8	6	1	• none			
Google Places	• Other model calibration/validation • TNC OD characteristics§e.g. time/location • TDM policy evaluation • Vehicle classification • System performance monitoring/evaluation • Freight and commercial vehicle travel	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel)	6	6	5	5	1	• none			
National Performance Management Research Dataset (NPMRDS)/INRIX	• Other model calibration/validation • TNC OD characteristics§e.g. time/location • TDM policy evaluation • Hourly counts and count validation • System performance monitoring/evaluation • Other (specify§economic development§etc.)	• Shared cost with other agencies	• Database tools (including GIS) • Custom computer coded scripts	6	6	6	7	1	• Data accuracy and reliability might be an issue			
StreetLight Data	• Regional and corridor level OD patterns • External travel attributes	• Shared cost with other agencies	• Database tools (including GIS) • Custom computer coded scripts	7	7	7	7	1	• none			
American Transportation Research Institute (ATRI) truck GPS data	• Regional and corridor level OD patterns • Freight and commercial vehicle travel	• Purchase/Subscription	• Database tools (including GIS) • Custom computer coded scripts	7	7	8	8	1	• The sample size of the ATRI data needs to be further validated			
Google Travel Time	• Other model calibration/validation • Other (specify§economic development§etc.)	• Collected In-house • Other (specify)	• Spreadsheets (e.g. Microsoft Excel)	5	5	5	5	1	• none			
Continuous traffic data	• Other model calibration/validation • Hourly counts and count validation • Vehicle classification • System performance monitoring/evaluation	• Collected In-house • Purchase/Subscription	• Database tools (including GIS) • Custom computer coded scripts	7	7	7	7	1	• none			

Data sources not included by Anonymous Agency

InfoUSA/Dun & Bradstreet business listing, Co-Star Data, LIDAR and Other Airborne Data, Job Postings Data, HERE, Transit Automatic Vehicle Location, Farecard/Toll Data, Social media, Other data sources

Agency:	**Capital Region Planning Commission (CRPC)**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
Disaggregate Census dataset (Specify, e.g. PUMS)	• Regional and corridor level OD patterns • Trip purpose and destination type • Other model calibration/validation • TNC OD characteristics§e.g. time/location • TDM policy evaluation • Data validation and integration of HTS	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Other (specify)	9	3	10	1	1	• Great and helpful used in the decision making process	• meetings with stakeholders	• no ideas	• None
Government record (Specify, e.g. appraisal, business licensing, QCEW)	• Regional and corridor level OD patterns • Managed lane time of day usage • Other model calibration/validation • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Freight and commercial vehicle travel • Data validation and integration of HTS	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Other (specify)	8	3	10	1	1	• useful and Helpful in decision making process			
InfoUSA/Dun & Bradstreet business listing	• Regional and corridor level OD patterns • Trip purpose and destination type	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	5	2	5	1	1	• accuracy is one of concerns.....			
Google Places	• Regional and corridor level OD patterns • External travel attributes	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	5	2	5	1	1	• None			
LIDAR and Other Airborne Data	• Regional and corridor level OD patterns	• Other (specify)	• Database tools (including GIS) • Other (specify)	5	2	5	1	1	• None			
National Performance Management Research Dataset (NPMRDS)/INRIX	• Regional and corridor level OD patterns • Trip purpose and destination type • External travel attributes • Other model calibration/validation • TNC OD characteristics§e.g. time/location • Assess the impact of TNC on other modes and traffic congestion • Hourly counts and count validation • Vehicle classification • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Freight and commercial vehicle travel	• Shared cost with other agencies	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	5	2	5	1	1	• None			
HERE	• Regional and corridor level OD patterns	• Collected In-house	• Database tools (including GIS)	5	2	5	1	1	• None			
StreetLight Data	• Regional and corridor level OD patterns	• Other (specify)	• Database tools (including GIS)	5	1	1	1	1	• accuracy			

Agency:	Capital Region Planning Commission (CRPC)										
Major Data Sources	What application area did your agency use the Data Source for?	How is the Data Source retrieved?	What tools are used for processing the Data Set?	Rate the level of challenge for using the Data Sources (1 ("no challenge at all") – 10 ("very challenging"))					Specify major limitations, concerns, lessons learned for applying the Data Source	What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?	What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?	Additional Comments
				IT Infrastructure	Task Timeline	Budget	Technicalities	Other				
American Transportation Research Institute (ATRI) truck GPS data	• Regional and corridor level OD patterns • Trip purpose and destination type • External travel attributes • Other model calibration/validation • Freight and commercial vehicle travel	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	5	1	5	1	1	• None			
Google Travel Time	• Regional and corridor level OD patterns • External travel attributes • Assess the impact of TNC on other modes and traffic congestion	• Other (specify)	• Database tools (including GIS)	5	1	5	1	1	• None			
Transit Automatic Vehicle Location	• Regional and corridor level OD patterns • TNC rider demographics	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	5	2	5	1	1	• None			
Continuous traffic data	• Regional and corridor level OD patterns • Assess the impact of TNC on other modes and traffic congestion • System performance monitoring/evaluation	• Shared cost with other agencies	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	5	2	5	1	1	• None			

Data sources not included by Capital Region Planning Commission (CRPC)

Co-Star Data, Job Postings Data, Farecard/Toll Data, Social media, Other data sources

Agency:	**Chicago Metropolitan Agency for Planning**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
Disaggregate Census dataset (Specify, e.g. PUMS)	• Other model calibration/validation	• Other (specify)	• Custom computer coded scripts	1	4	3	8		• Generally a good data set§some concerns with sample size	• Varies by dataset	• they don't exist	
Government record (Specify, e.g. appraisal, business licensing, QCEW)	• Regional and corridor level OD patterns • Trip purpose and destination type • Other model calibration/validation • Telecommuting trends	• Other (specify)	• Database tools (including GIS) • Custom computer coded scripts	1	7	1	4		• QCEW under severe state restrictions LEHD headquarters issue			
InfoUSA/Dun & Bradstreet business listing	• External travel attributes • Other model calibration/validation • Freight and commercial vehicle travel	• Purchase/Subscription	• Database tools (including GIS) • Custom computer coded scripts	2	2	8	2		• Expensive Duplicate and/or obsolete entries			
Co-Star Data	• Other (specify§economic development§etc.)	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	2	4	5	3		• Does not report residential or institutional developments			
LIDAR and Other Airborne Data	• Other (specify§economic development§etc.)	• Other (specify)	• Database tools (including GIS) • Custom computer coded scripts	7	3	1	7		• Data processing			
National Performance Management Research Dataset (NPMRDS)/INRIX	• Regional and corridor level OD patterns • Other model calibration/validation • System performance monitoring/evaluation	• Other (specify)	• Database tools (including GIS) • Custom computer coded scripts	1	3	1	5		• Only covers NHS			
HERE	• Other model calibration/validation • System performance monitoring/evaluation	• Other (specify)	• Database tools (including GIS) • Custom computer coded scripts	1	2	1	4					
American Transportation Research Institute (ATRI) truck GPS data	• Regional and corridor level OD patterns • External travel attributes • Freight and commercial vehicle travel	• Purchase/Subscription	• Database tools (including GIS) • Custom computer coded scripts	4	6	3	6		• Only class 7 & 8 trucks			
Continuous traffic data	• Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification	• Other (specify)	• Database tools (including GIS) • Custom computer coded scripts	10	6	4	7		• mostly limited to expressways			
Other data sources	• Other (specify§economic development§etc.)	• Purchase/Subscription	• Database tools (including GIS)	1	2	4	2					

Data sources not included by Chicago Metropolitan Agency for Planning

Google Places, Job Postings Data, StreetLight Data, Google Travel Time, Transit Automatic Vehicle Location, Farecard/Toll Data, Social media

Agency:	**Delaware Valley Regional Planning Commission**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
National Performance Management Research Dataset (NPMRDS)/INRIX	• Other model calibration/validation • System performance monitoring/evaluation • Freight and commercial vehicle travel	• Collected In-house • Purchase/Subscription • Shared cost with other agencies	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	3	7	5	4		• The licensing of an INRIX data purchase has taken a long time	• Checking if data looks reasonable vs known counts and model results	• Possibly data from cell phones location-based services	

Data sources not included by Delaware Valley Regional Planning Commission

Disaggregate Census dataset

(Specify, e.g. PUMS), Government record, InfoUSA/Dun & Bradstreet business listing, Google Places, Co-Star Data, LIDAR and Other Airborne Data, Job Postings Data, HERE, StreetLight Data, American Transportation Research Institute (ATRI) truck GPS data , Google Travel Time, Transit Automatic Vehicle Location, Farecard/Toll Data, Social media, Continuous traffic data, Other data sources

Agency:	East West Council of Governments (Missouri)											
Major Data Sources	What application area did your agency use the Data Source for?	How is the Data Source retrieved?	What tools are used for processing the Data Set?	Rate the level of challenge for using the Data Sources (1 ("no challenge at all") – 10 ("very challenging"))					Specify major limitations, concerns, lessons learned for applying the Data Source	What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?	What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?	Additional Comments
				IT Infrastructure	Task Timeline	Budget	Technicalities	Other				
Government record (Specify, e.g. appraisal, business licensing, QCEW)	• External travel attributes	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	1	7	3	6	1	• Internal staff skills and staff turn over	• in-house and use RITIS	• FHWA§reports§with details and tools to get MPOs started	• Got HERE data from IDOT§had questions that were sent to IDOT§they could not respond never got response from the consultant
InfoUSA/Dun & Bradstreet business listing	• Trip purpose and destination type • Other (specify§economic development§etc.)	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	1	8	3	9	1	• free availability for D&B			
LIDAR and Other Airborne Data	• Other (specify§economic development§etc.)	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Custom computer coded scripts	1	5	3	6	1	• data size and manipulation§availability			
National Performance Management Research Dataset (NPMRDS)/INRIX	• Regional and corridor level OD patterns • Day-to-day traffic variation • Travel behavior of special population group	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Custom computer coded scripts	1	6	4	6	1	• QA/QC			
HERE	• System performance monitoring/evaluation • Impacts and trend of congestion management strategy	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	1	5	3	5	1	• Staff skills			
Google Travel Time	• Other (specify§economic development§etc.)	• Collected In-house	• Custom computer coded scripts	1	4	3	4	1	• not many free options			

Data sources not included by East West Council of Governments (Missouri)

Disaggregate Census dataset
(Specify, e.g. PUMS), Google Places, Co-Star Data, Job Postings Data, StreetLight Data, American Transportation Research Institute (ATRI) truck GPS data , Transit Automatic Vehicle Location, Farecard/Toll Data, Social media, Continuous traffic data, Other data sources

Agency:	Fredericksburg Area Metropolitan Planning Organization (Virginia)											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
National Performance Management Research Dataset (NPMRDS)/INRIX	• Regional and corridor level OD patterns • Trip purpose and destination type • External travel attributes • Hourly counts and count validation • Day-to-day traffic variation • Impacts and trend of congestion management strategy	• Collected In-house • Shared cost with other agencies	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	4	3	4	3	1	• The data is harder to use within GIS. For our uses§the INRIX dataset has primarily been used for traffic data and congestion analysis for display in GIS or through tables and graphs.	• Using multiple methods of data collection such as the use of StreetLight Data and Inrix for the same project.	• Streetlight Data is a useful tool today for understanding individuals movements and likely will be in the future as individuals will continue to use GPS location data.	
HERE	• Hourly counts and count validation • Day-to-day traffic variation	• Other (specify)		5	5	10	5	1	• Attempted to use HERE data but was cost prohibative			
StreetLight Data	• Regional and corridor level OD patterns • Trip purpose and destination type • External travel attributes • TNC rider demographics • Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification • Impacts and trend of congestion management strategy • Freight and commercial vehicle travel	• Collected In-house • Shared cost with other agencies	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	2	2	1	4	1	• No major concerns with StreetLight Data. Only issue being data manipulation from certain StreetLight project types into GIS.			
Social media (specify. Facebook, LinkedIn, etc.)	• Other (specify§economic development§etc.)	• Other (specify)	• Other (specify)	1	1	1	1	1	• Used for Public Involvement.			

Data sources not included by Fredericksburg Area Metropolitan Planning Organization (Virginia)

Disaggregate Census dataset

(Specify, e.g. PUMS), Government record, InfoUSA/Dun & Bradstreet business listing, Google Places, Co-Star Data, LIDAR and Other Airborne Data, Job Postings Data, American Transportation Research Institute (ATRI) truck GPS data , Google Travel Time, Transit Automatic Vehicle Location, Farecard/Toll Data, Continuous traffic data, Other data sources

Agency:	**Houston-Galveston Area Council**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")* **IT Infrastructure**	**Task Timeline**	**Budget**	**Technicalities**	**Other**	**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
Disaggregate Census dataset (Specify, e.g. PUMS)	• Trip purpose and destination type • Other model calibration/validation • Impacts and trend of congestion management strategy • Data validation and integration of HTS	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	8	6	5	7					
Government record (Specify, e.g. appraisal, business licensing, QCEW)	• Regional and corridor level OD patterns • Trip purpose and destination type • Other model calibration/validation • System performance monitoring/evaluation • Travel behavior of special population group	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	3	4	5	3					
InfoUSA/Dun & Bradstreet business listing	• External travel attributes • TNC OD characteristics§e.g. time/location • Assess the impact of TNC on other modes and traffic congestion • Hourly counts and count validation • Day-to-day traffic variation	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	7	8	8	7	1				
National Performance Management Research Dataset (NPMRDS)/INRIX	• Other model calibration/validation • TDM policy evaluation • Day-to-day traffic variation	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	6	5	5	4	1				
StreetLight Data	• Regional and corridor level OD patterns • Managed lane time of day usage • Hourly counts and count validation	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	5	6	6	6	1				
American Transportation Research Institute (ATRI) truck GPS data	• External travel attributes • Freight and commercial vehicle travel	• Purchase/Subscription	• Database tools (including GIS)	4	4	5	5	1				
Farecard/Toll Data	• Managed lane time of day usage • Other model calibration/validation • Hourly counts and count validation • Data validation and integration of HTS	• Purchase/Subscription	• Database tools (including GIS)	3	2	2	3	1				
Continuous traffic data	• Regional and corridor level OD patterns • Trip purpose and destination type • Hourly counts and count validation • Day-to-day traffic variation • Data validation and integration of HTS	• Purchase/Subscription	• Database tools (including GIS)	3	2	2	3	1				
Other data sources	• Other model calibration/validation • TNC OD characteristics§e.g. time/location • TDM policy evaluation • Hourly counts and count validation • Day-to-day traffic variation • Impacts and trend of congestion management strategy • Project selection	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	3	4	4	5	1				

Data sources not included by Houston-Galveston Area Council

Google Places, Co-Star Data, LIDAR and Other Airborne Data, Job Postings Data, HERE, Google Travel Time, Transit Automatic Vehicle Location, Social media

Agency:	**Maricopa Association of Governments (Florida)**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
Disaggregate Census dataset (Specify, e.g. PUMS)	• Regional and corridor level OD patterns • Trip purpose and destination type • External travel attributes • Other model calibration/validation • Assess the impact of TNC on other modes and traffic congestion • TDM policy evaluation • Day-to-day traffic variation • Inter-city bus travel • Traffic dynamics and interactions with transit • Data validation and integration of HTS	• Collected In-house • Shared cost with other agencies	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts							• QA/QC procedure is defined for a specific data set.		
Government record (Specify, e.g. appraisal, business licensing, QCEW)	• External travel attributes • Other model calibration/validation • TDM policy evaluation • Other (specify§economic development§etc.)	• Collected In-house • Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts									
National Performance Management Research Dataset (NPMRDS)/INRIX	• Other model calibration/validation • Hourly counts and count validation • Day-to-day traffic variation • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Special event impacts • Freight and commercial vehicle travel • Project selection	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts • Other (specify)	6	6	6	6	6				
HERE	• Other model calibration/validation • Hourly counts and count validation • Day-to-day traffic variation • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Special event impacts • Freight and commercial vehicle travel • Project selection	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts • Other (specify)	6	6	6	6	6				
StreetLight Data	• External travel attributes • Other model calibration/validation • Freight and commercial vehicle travel	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded									
American Transportation Research Institute (ATRI) truck GPS data	• External travel attributes • Vehicle classification • Freight and commercial vehicle travel	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts									
Google Travel Time	• Other model calibration/validation • System performance monitoring/evaluation	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	5	5	5	5	5				
Continuous traffic data	• External travel attributes • Other model calibration/validation • TDM policy evaluation • Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Special event impacts • Freight and commercial vehicle travel • Project selection	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	6	6	6	6	6				

Agency:	**Maricopa Association of Governments (Florida)**										
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
Other data sources	• Regional and corridor level OD patterns • Trip purpose and destination type • External travel attributes • Other model calibration/validation • TDM policy evaluation • System performance monitoring/evaluation • Data validation and integration of HTS	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts • Other (specify)	6	6	6	6	6				

Data sources not included by Maricopa Association of Governments (Florida)

InfoUSA/Dun & Bradstreet business listing, Google Places, Co-Star Data, LIDAR and Other Airborne Data, Job Postings Data, Transit Automatic Vehicle Location, Farecard/Toll Data, Social media

Agency:	**Metro (Portland, Oregon)**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
Disaggregate Census dataset (Specify, e.g. PUMS)	• Other model calibration/validation • Other (specify§economic development§etc.)	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Custom computer coded scripts	3	6	3	2	1		• extensive review to identify potential data errors§outliers	• results from pilots and robust SP surveys	
LIDAR and Other Airborne Data	• Other (specify§economic development§etc.)	• Shared cost with other agencies	• Custom computer coded scripts	5	6	5	4	1				
National Performance Management Research Dataset (NPMRDS)/INRIX	• Other model calibration/validation • System performance monitoring/evaluation	• Shared cost with other agencies	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	3	6	3	4	1				
HERE	• Other model calibration/validation • System performance monitoring/evaluation	• Shared cost with other agencies	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	3	6	3	4	1				
StreetLight Data	• Other model calibration/validation	• Other (specify)	• Other (specify)	2	4	2	2	1				
Google Travel Time	• Other model calibration/validation	• Collected In-house	• Custom computer coded scripts	2	6	2	5	1				
Continuous traffic data	• Other model calibration/validation	• Purchase/Subscription • Shared cost with other agencies	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	3	6	4	3	1				

Data sources not included by Metro (Portland, Oregon)

Government record, InfoUSA/Dun & Bradstreet business listing, Google Places, Co-Star Data, Job Postings Data, American Transportation Research Institute (ATRI) truck GPS data , Transit Automatic Vehicle Location, Farecard/Toll Data, Social media, Other data sources

Agency:	**Miami-Dade Transportation Planning Organization**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
Disaggregate Census dataset (Specify, e.g. PUMS)	• Regional and corridor level OD patterns • Trip purpose and destination type • External travel attributes • Other model calibration/validation • Telecommuting trends • Hourly counts and count validation • Day-to-day traffic variation • System performance monitoring/evaluation • Freight and commercial vehicle travel • Inter-city bus travel • Travel behavior of special population group • Data validation and integration of HTS	• Shared cost with other agencies	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	5	5	5	5	5			• Connected Autonomous Vehicles adoption rate by geographic area	
InfoUSA/Dun & Bradstreet business listing	• Regional and corridor level OD patterns • Managed lane time of day usage • Trip purpose and destination type • External travel attributes • Other model calibration/validation	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	5	5	5	5	5				
Google Places	• Regional and corridor level OD patterns • Trip purpose and destination type • External travel attributes • Other model calibration/validation	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	5	5	5	5	5				
National Performance Management Research Dataset (NPMRDS)/INRIX	• Regional and corridor level OD patterns • Managed lane time of day usage • Trip purpose and destination type • External travel attributes • Other model calibration/validation • TNC OD characteristics§e.g. time/location • TNC rider demographics • Assess the impact of TNC on other modes and traffic congestion • Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Special event impacts • Freight and commercial vehicle travel • Inter-city bus travel • Travel behavior of special population group • Traffic dynamics and interactions with transit • Project selection • Data validation and integration of HTS	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	5	5	5	5	5				
StreetLight Data	• Regional and corridor level OD patterns • Managed lane time of day usage • Trip purpose and destination type • External travel attributes • Other model calibration/validation • TNC OD characteristics§e.g. time/location • TNC rider demographics • Assess the impact of TNC on other modes and traffic congestion • Telecommuting trends • Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Special event impacts • Freight and commercial vehicle travel • Inter-city bus travel • Travel behavior of special population group • Traffic dynamics and interactions with transit • Data validation and integration of HTS	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	5	5	5	5	5				

Agency:	Miami-Dade Transportation Planning Organization											
Major Data Sources	What application area did your agency use the Data Source for?	How is the Data Source retrieved?	What tools are used for processing the Data Set?	Rate the level of challenge for using the Data Sources *1 ("no challenge at all")* *10 ("very challenging")*					Specify major limitations, concerns, lessons learned for applying the Data Source	What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?	What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?	Additional Comments
				IT Infrastructure	Task Timeline	Budget	Technicalities	Other				
American Transportation Research Institute (ATRI) truck GPS data	• Regional and corridor level OD patterns • Managed lane time of day usage • Trip purpose and destination type • External travel attributes • Other model calibration/validation • Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Special event impacts • Freight and commercial vehicle travel • Data validation and integration of HTS	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	5	5	5	5	5				
Google Travel Time	• Regional and corridor level OD patterns • Managed lane time of day usage • Trip purpose and destination type • External travel attributes • Other model calibration/validation • Day-to-day traffic variation • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Special event impacts • Freight and commercial vehicle travel • Inter-city bus travel • Travel behavior of special population group • Traffic dynamics and interactions with transit • Data validation and integration of HTS	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	5	5	5	5	5				

Data sources not included by Miami-Dade Transportation Planning Organization

Government record, Co-Star Data, LIDAR and Other Airborne Data, Job Postings Data, HERE, Transit Automatic Vehicle Location, Farecard/Toll Data, Social media, Continuous traffic data, Other data sources

Agency:	**Mid-Ohio Regional Planning Commission**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
Disaggregate Census dataset (Specify, e.g. PUMS)	• External travel attributes	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	1	5	1	5	1				
Government record (Specify, e.g. appraisal, business licensing, QCEW)	• External travel attributes	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	1	8	1	8	1				
InfoUSA/Dun & Bradstreet business listing	• External travel attributes • Other (specify§economic development§etc.)	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	1	8	1	8	1				
Google Places	• External travel attributes • Other (specify§economic development§etc.)	• Purchase/Subscription	• Custom computer coded scripts	2	8	2	8	1				
National Performance Management Research Dataset (NPMRDS)/INRIX	• Other model calibration/validation • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Special event impacts • Project selection	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	3	6	1	6	1				
HERE	• Other model calibration/validation • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Special event impacts • Project selection	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	2	6	1	6	1				
StreetLight Data	• Regional and corridor level OD patterns • Other model calibration/validation • TDM policy evaluation • Day-to-day traffic variation • Special event impacts • Freight and commercial vehicle travel	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	2	6	4	6	1				
Google Travel Time	• Other (specify§economic development§etc.)	• Purchase/Subscription	• Custom computer coded scripts	1	6	1	6	1				
Transit Automatic Vehicle Location	• Other model calibration/validation	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	1	7	1	7	1				
Continuous traffic data	• Other model calibration/validation • Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification • System performance monitoring/evaluation • Project selection	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	1	6	1	4	1				

Data sources not included by Mid-Ohio Regional Planning Commission

Co-Star Data, LIDAR and Other Airborne Data, Job Postings Data, American Transportation Research Institute (ATRI) truck GPS data , Farecard/Toll Data, Social media, Other data sources

Agency:	**New York Metropolitan Transportation Council (NYMTC)**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")* **IT Infrastructure**	**Task Timeline**	**Budget**	**Technicalities**	**Other**	**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
Disaggregate Census dataset (Specify, e.g. PUMS)	• Regional and corridor level OD patterns • Other model calibration/validation • Travel behavior of special population group • Data validation and integration of HTS	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	2	2	3	3	7	• Visualization and validity	• We identify comparable data sets and validate the results.	• NPMRDS data can provide change in traffic speed information; ACS data can provide sprawl in population and change in commute distance information; NTD data can provide change in transit ridership data	• This an exciting and hopeful time for transportation and the way in which we move people and goods. Big data can provide an opportunity to empirically measure travel behavior that was previously not measurable. Now§more than ever§transportation planners should have access to high quality and economical data. However§transportation planners and engineers are not data scientists and programmers and most planning organizations are not setup to develop big data talent and to offer competitive compensation§incentive§and recognition to recruit and keep this talent within the organization.
National Performance Management Research Dataset (NPMRDS)/INRIX	• Other model calibration/validation • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Other (specify§economic development§etc.)	• Other (specify)	• Database tools (including GIS) • Custom computer coded scripts	5	6	1	9		• Data Size			
StreetLight Data	• Other (specify§economic development§etc.)	• Purchase/Subscription	• Other (specify)	1	3	10	4		• It has mode agnostic data; was not able to effectively validate with AADT			

Data sources not included by New York Metropolitan Transportation Council (NYMTC)

Government record, InfoUSA/Dun & Bradstreet business listing, Google Places, Co-Star Data, LIDAR and Other Airborne Data, Job Postings Data, HERE, American Transportation Research Institute (ATRI) truck GPS data , Google Travel Time, Transit Automatic Vehicle Location, Farecard/Toll Data, Social media, Continuous traffic data, Other data sources

Agency:	**North Central Texas Council of Governments**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
Disaggregate Census dataset (Specify, e.g. PUMS)	• Regional and corridor level OD patterns • Managed lane time of day usage • Trip purpose and destination type • External travel attributes • Other model calibration/validation • TNC OD characteristics§e.g. time/location • TNC rider demographics • Assess the impact of TNC on other modes and traffic congestion • TDM policy evaluation • Telecommuting trends • Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Special event impacts • Freight and commercial vehicle travel • Inter-city bus travel • Travel behavior of special population group • Traffic dynamics and interactions with transit • Project selection • Data validation and integration of HTS	• Collected In-house • Purchase/Subscription • Shared cost with other agencies	• Database tools (including GIS)	8	7	7	8	1			• WEBINAR§PAPER§CONFERENCE§DEMONSTRATION	
Government record (Specify, e.g. appraisal, business licensing, QCEW)	• Regional and corridor level OD patterns • Managed lane time of day usage • Trip purpose and destination type • External travel attributes • TNC rider demographics • TDM policy evaluation • Telecommuting trends • Hourly counts and count validation • Impacts and trend of congestion management strategy • Travel behavior of special population group • Data validation and integration of HTS	• Collected In-house • Purchase/Subscription • Shared cost with other agencies	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	8	7	7	7	1				
Google Places	• Regional and corridor level OD patterns • Managed lane time of day usage • Trip purpose and destination type • External travel attributes • Other model calibration/validation • TNC OD characteristics§e.g. time/location • TNC rider demographics • Assess the impact of TNC on other modes and traffic congestion • TDM policy evaluation • Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Special event impacts • Freight and commercial vehicle travel • Inter-city bus travel • Travel behavior of special population group • Traffic dynamics and interactions with transit • Project selection • Data validation and integration of HTS	• Collected In-house	• Database tools (including GIS)	5	5	5	5	1				

Agency:	**North Central Texas Council of Governments**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
National Performance Management Research Dataset (NPMRDS)/INRIX	• Regional and corridor level OD patterns • Managed lane time of day usage • Trip purpose and destination type • External travel attributes • Other model calibration/validation • TNC OD characteristics§e.g. time/location • TNC rider demographics • Assess the impact of TNC on other modes and traffic congestion • TDM policy evaluation • Telecommuting trends • Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Special event impacts • Freight and commercial vehicle travel • Inter-city bus travel • Travel behavior of special population group • Traffic dynamics and interactions with transit • Project selection • Data validation and integration of HTS	• Collected In-house	• Database tools (including GIS)	4	4	4	4	1				
HERE	• Regional and corridor level OD patterns • Managed lane time of day usage • Trip purpose and destination type • External travel attributes • Other model calibration/validation • TNC OD characteristics§e.g. time/location • TNC rider demographics • Assess the impact of TNC on other modes and traffic congestion • TDM policy evaluation • Telecommuting trends • Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Special event impacts • Freight and commercial vehicle travel • Inter-city bus travel • Travel behavior of special population group • Traffic dynamics and interactions with transit • Project selection • Data validation and integration of HTS	• Purchase/Subscription	• Database tools (including GIS)	5	5	5	5	1				
StreetLight Data	• Other (specify§economic development§etc.)	• Other (specify)		7	7	7	7	1				
American Transportation Research Institute (ATRI) truck GPS data	• External travel attributes • Other model calibration/validation • Hourly counts and count validation • Freight and commercial vehicle travel	• Purchase/Subscription	• Database tools (including GIS)	4	4	4	4	1				

Agency:	**North Central Texas Council of Governments**										
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
Google Travel Time	• Regional and corridor level OD patterns • Managed lane time of day usage • Trip purpose and destination type • External travel attributes • Other model calibration/validation • TNC OD characteristics§e.g. time/location • TNC rider demographics • Assess the impact of TNC on other modes and traffic congestion • TDM policy evaluation • Telecommuting trends • Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Special event impacts • Freight and commercial vehicle travel • Inter-city bus travel • Travel behavior of special population group • Traffic dynamics and interactions with transit • Project selection • Data validation and integration of HTS	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel)	3	3	3	3	1			
Social media (specify. Facebook, LinkedIn, etc.)	• Other (specify§economic development§etc.)	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel)	2	2	2	2	1			

Data sources not included by North Central Texas Council of Governments

InfoUSA/Dun & Bradstreet business listing, Co-Star Data, LIDAR and Other Airborne Data, Job Postings Data, Transit Automatic Vehicle Location, Farecard/Toll Data, Continuous traffic data, Other data sources

Agency:	North Jersey Transportation Planning Authority											
Major Data Sources	What application area did your agency use the Data Source for?	How is the Data Source retrieved?	What tools are used for processing the Data Set?	Rate the level of challenge for using the Data Sources (1 ("no challenge at all") – 10 ("very challenging"))					Specify major limitations, concerns, lessons learned for applying the Data Source	What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?	What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?	Additional Comments
				IT Infrastructure	Task Timeline	Budget	Technicalities	Other				
Disaggregate Census dataset (Specify, e.g. PUMS)	• Other model calibration/validation	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	1	3	1	3	1		• We don't have a standard QA/QC procedure for big datasets although each dataset might bring with it the need for QA. For example§LEHD (employment) data is often checked against other employment datasets (BEA§QCEW)to gauge changes in employment. We have found errors in LEHD on major employment locations. Errors were also found in the Census on a major group quarters location (incorrect Block Group).	• We don't know yet although we have surmised that AVs might operate similar to ride hailing companies. I would be helpful to receive data from the ride hailing companies (Uber and Lyft) on the activities of their users. We have also worked with companies that collect cell phone O-D data in trying to impute ride hailing data (and theoretically estimate AV use from that) via App usage.	• We have explored the use of cell phone based big data for uses such as to supplement travel surveys. We are still unsure how helpful they will be given some of their limitations such as the manner of imputing demographics§the difficulties in differentiating transit or non-motorized modes§and the difficulties in guessing trip purposes. Still§this seems like a rich dataset that might prove useful to us in the future
Government record (Specify, e.g. appraisal, business licensing, QCEW)	• Other model calibration/validation	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel)	1	3	1	2	1				
InfoUSA/Dun & Bradstreet business listing	• Other (specify§economic development§etc.)	• Purchase/Subscription	• Database tools (including GIS)	1	8	3	3	5	• Purchased a subscription that limited the number of records that could be downloaded at one time. Some of the data fields such as employment were estimates.			
Co-Star Data	• Other (specify§economic development§etc.)	• Purchase/Subscription	• Database tools (including GIS)	1	5	3	3	1	• Required a process to link buildings to tenants.			
National Performance Management Research Dataset (NPMRDS)/INRIX	• System performance monitoring/evaluation	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Other (specify)	3	5	1	8	1	• Still Learning.			

Data sources not included by North Jersey Transportation Planning Authority

Google Places, LIDAR and Other Airborne Data, Job Postings Data, HERE, StreetLight Data, American Transportation Research Institute (ATRI) truck GPS data , Google Travel Time, Transit Automatic Vehicle Location, Farecard/Toll Data, Social media, Continuous traffic data, Other data sources

Agency:	Northeast Ohio Areawide Coordinating Agency										
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
Disaggregate Census dataset (Specify, e.g. PUMS)	• Other model calibration/validation • TDM policy evaluation • Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Freight and commercial vehicle travel • Inter-city bus travel	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	3	3	3	3	3			• Household income and car ownership	
Google Places	• Other model calibration/validation	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel)	4	3	3	4	4				
National Performance Management Research Dataset (NPMRDS)/INRIX	• Trip purpose and destination type • Other model calibration/validation • Freight and commercial vehicle travel	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel)	5	5	3	4	4				
StreetLight Data	• Regional and corridor level OD patterns • Trip purpose and destination type • External travel attributes • Other model calibration/validation • Hourly counts and count validation • Day-to-day traffic variation	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel)	5	5	3	5	5				

Data sources not included by Northeast Ohio Areawide Coordinating Agency

Government record, InfoUSA/Dun & Bradstreet business listing, Co-Star Data, LIDAR and Other Airborne Data, Job Postings Data, HERE, American Transportation Research Institute (ATRI) truck GPS data , Google Travel Time, Transit Automatic Vehicle Location, Farecard/Toll Data, Social media, Continuous traffic data, Other data sources

Agency:	**Pima Association of Governments (Arizona)**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
Disaggregate Census dataset (Specify, e.g. PUMS)	• Trip purpose and destination type • Other model calibration/validation • Travel behavior of special population group	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	1	1	1	3		• Limited sample size.	• Different data source typically require different data QA/QC procedure. Census and Government Data: Not much we do consider QA/QC other than aggregation level or individual record review with some samples. Employment Data: It does need to integrate other data sources considering the priority of data sources and need to develop a methodology how to review and stitch different sources of data especially with sampling and the priority. GPS data (ATRI or High resolution Metropia data): Since the data is typically geo-coded point data§it is only useful through a procedure to get the intended objective§for example§segment- or route-level performance measure and activity location identification. It typically requires a procedural QA/QC in model development like map matching algorithm for segment- or route-level performance measures§more sample data for development and validation as well as filtering out erroneous or unnecessary portion. StreetLight Data: This data is useful and easy to handle for multiple validation projects§especially for O-D trip data or AADT. For AADT§traffic count data including segment and turning movement count data is useful to evaluate the data. For O-D trip data§local housing unit and employment data could be used for evaluating the trips§especially for HBW trips.	• Not sure. I think It depends on purpose. As a MPO§understanding CAV behavior§network performance and market share in future§is critical. GPS trip data could be useful to evaluate the transportation system especially for indicating network performance. But forecasting in future should be limited to scenario-based model tests with various assumptions. SP survey data could be useful.	
Government record (Specify, e.g. appraisal, business licensing, QCEW)	• Other model calibration/validation • Other (specify§economic development§etc.)	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	1	5	2	3		• Take some time to understand the data§especially definition and quality of data			
InfoUSA/Dun & Bradstreet business listing	• Other model calibration/validation • Other (specify§economic development§etc.)	• Purchase/Subscription • Shared cost with other agencies	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	1	7	2	7		• Need to develop a method of data QA/QC with additional data§such as Google Places and government data including university and school district data.			
Google Places	• Other model calibration/validation • Other (specify§economic development§etc.)	• Purchase/Subscription	• Custom computer coded scripts	2	2	4	7		• This is a good tool to evaluate the existence of business. More specific lessons learned described in TRR publication (https://journals.sagepub.com/doi/full/10.1177/0361198119852068)			
LIDAR and Other Airborne Data	• Other (specify§economic development§etc.)	• Purchase/Subscription	• Database tools (including GIS)	5	8	9	5					
National Performance Management Research Dataset (NPMRDS)/INRIX	• Other model calibration/validation • System performance monitoring/evaluation	• Other (specify)	• Database tools (including GIS) • Custom computer coded scripts	4	3	1	6		• Good tool for performance measurement but limited to get routing speed and travel time§instead of segment (TMC) level.			
StreetLight Data	• Regional and corridor level OD patterns • Trip purpose and destination type • External travel attributes • Other model calibration/validation • Hourly counts and count validation • Day-to-day traffic variation • Travel behavior of special population group	• Shared cost with other agencies	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	5	7	8	7		• This tool is pretty useful for numerous studies but need to pay attention for QA/QC§for example§trip purpose could be skewed more on HBO and traffic volume could be over- and under-estimated.			
American Transportation Research Institute (ATRI) truck GPS data	• Freight and commercial vehicle travel	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	3	6	7	7		• >1 minute resolution is good for highway assessment and trip pattern analysis but limited to analyze more in-depth arterial analysis (routing is not accurate with >1 minute resolution) and have small truck population for home-delivery.			
Other data sources	• Other model calibration/validation • System performance monitoring/evaluation	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	5	9	7	9		• High resolution GPS data (from voluntary survey using Metropia App.) useful to calibration model VDF but require to develop a tool§map matching and analysis in case of using raw GPS data. Model described through https://trid.trb.org/view/1439478.			

Data sources not included by Pima Association of Governments (Arizona)

Co-Star Data, Job Postings Data, HERE, Google Travel Time, Transit Automatic Vehicle Location, Farecard/Toll Data, Social media, Continuous traffic data

Agency:	Regional Transportation Commission of Southern Nevada											
Major Data Sources	What application area did your agency use the Data Source for?	How is the Data Source retrieved?	What tools are used for processing the Data Set?	Rate the level of challenge for using the Data Sources (1 ("no challenge at all") – 10 ("very challenging"))					Specify major limitations, concerns, lessons learned for applying the Data Source	What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?	What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?	Additional Comments
				IT Infrastructure	Task Timeline	Budget	Technicalities	Other				
Disaggregate Census dataset (Specify, e.g. PUMS)	• Other model calibration/validation • Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification • System performance monitoring/evaluation • Impacts and trend of congestion management strategy	• Other (specify)	• Database tools (including GIS)	1	1	1	1	1				
InfoUSA/Dun & Bradstreet business listing	• Other model calibration/validation • Travel behavior of special population group • Data validation and integration of HTS	• Purchase/Subscription	• Database tools (including GIS)	1	1	5	3	1				
National Performance Management Research Dataset (NPMRDS)/INRIX	• Other model calibration/validation • System performance monitoring/evaluation • Data validation and integration of HTS	• Purchase/Subscription	• Database tools (including GIS)	7	5	1	6	1				
HERE	• Regional and corridor level OD patterns • Other model calibration/validation • Data validation and integration of HTS • Other (specify§economic development§etc.)	• Purchase/Subscription • Shared cost with other agencies	• Database tools (including GIS) • Custom computer coded scripts	6	5	2	4	1				
Transit Automatic Vehicle Location	• Other model calibration/validation • Data validation and integration of HTS	• Collected In-house	• Database tools (including GIS)	4	4	1	3	1				

Data sources not included by Regional Transportation Commission of Southern Nevada

Government record, Google Places, Co-Star Data, LIDAR and Other Airborne Data, Job Postings Data, StreetLight Data, American Transportation Research Institute (ATRI) truck GPS data , Google Travel Time, Farecard/Toll Data, Social media, Continuous traffic data, Other data sources

Agency:	**SouthEast Michigan Council of Governments (SEMCOG)**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")* **IT Infrastructure**	**Task Timeline**	**Budget**	**Technicalities**	**Other**	**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
Disaggregate Census dataset (Specify, e.g. PUMS)	• Regional and corridor level OD patterns • Other model calibration/validation	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	2	5	1	2	1	• not enough samples for some market segments	• understanding the underlying algorithm; reasonableness check for certain locations; comparison between big data and other data sources (survey§etc.)	• N/A	• N/A
Government record (Specify, e.g. appraisal, business licensing, QCEW)	• Other (specify§economic development§etc.)	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	2	5	1	2	1	• limitations on sharing the data with other agencies			
National Performance Management Research Dataset (NPMRDS)/INRIX	• Regional and corridor level OD patterns • Other model calibration/validation • System performance monitoring/evaluation • Impacts and trend of congestion management strategy	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	3	3	1	3	1	• concerns on consistence/accuracy of the data on arterial roads			
StreetLight Data	• Regional and corridor level OD patterns • External travel attributes • Other model calibration/validation • TDM policy evaluation	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	2	5	9	3	1	• concerns on overall data quality and bias on the samples			
Continuous traffic data	• Other model calibration/validation • Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification • System performance monitoring/evaluation • Impacts and trend of congestion management strategy	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	2	5	1	3	1	• Limited locations with ATR counts			
Other data sources	• Regional and corridor level OD patterns • Other model calibration/validation • TDM policy evaluation • Other (specify§economic development§etc.)	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	2	5	3	2	1	• overall data quality			

Data sources not included by SouthEast Michigan Council of Governments (SEMCOG)

InfoUSA/Dun & Bradstreet business listing, Google Places, Co-Star Data, LIDAR and Other Airborne Data, Job Postings Data, HERE, American Transportation Research Institute (ATRI) truck GPS data , Google Travel Time, Transit Automatic Vehicle Location, Farecard/Toll Data, Social media

Agency:	**Southern California Association of Governments**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
Disaggregate Census dataset (Specify, e.g. PUMS)	• Regional and corridor level OD patterns • Trip purpose and destination type • Other model calibration/validation • TDM policy evaluation • Telecommuting trends	• Collected In-house	• Other (specify)	3	2	1	1		• We've been using PUMS data for many years. Don't have concern.	• Will summary the data§and compared with historical patterns.	• Does such data exist?	
Government record (Specify, e.g. appraisal, business licensing, QCEW)	• Other (specify§economic development§etc.)	• Purchase/Subscription	• Other (specify)	2	2	3	2					
InfoUSA/Dun & Bradstreet business listing	• Other (specify§economic development§etc.)	• Purchase/Subscription	• Other (specify)	2	2	3	2		• InfoUSA: Need to clean headquarter data			
National Performance Management Research Dataset (NPMRDS)/INRIX	• Other model calibration/validation • System performance monitoring/evaluation	• Collected In-house	• Other (specify)	2	4	2	3					
StreetLight Data	• External travel attributes	• Purchase/Subscription	• Other (specify)	3	3	6	3					
American Transportation Research Institute (ATRI) truck GPS data	• Freight and commercial vehicle travel	• Other (specify)										
Farecard/Toll Data	• Other (specify§economic development§etc.)	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel)	2	3	1	2					
Continuous traffic data	• Other model calibration/validation	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel)	2	2	1	2					
Other data sources	• Trip purpose and destination type • Other model calibration/validation • TDM policy evaluation • Telecommuting trends	• Collected In-house	• Other (specify)	2	2	2	2					

Data sources not included by Southern California Association of Governments

Google Places, Co-Star Data, LIDAR and Other Airborne Data, Job Postings Data, HERE, Google Travel Time, Transit Automatic Vehicle Location, Social media

Agency:	**Virginia Department of Transportation**										
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
Disaggregate Census dataset (Specify, e.g. PUMS)	• Regional and corridor level OD patterns • Other model calibration/validation • Day-to-day traffic variation	• Shared cost with other agencies	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	8	7	9	7			• multiple data source comparing local expert review historical data comparing reasonable check	• GPS tracking	
InfoUSA/Dun & Bradstreet business listing	• External travel attributes • Other model calibration/validation	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	9	9	7	9					
Job Postings Data	• Trip purpose and destination type • Other model calibration/validation	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	8	5	4	7					
National Performance Management Research Dataset (NPMRDS)/INRIX	• External travel attributes • Other model calibration/validation • System performance monitoring/evaluation • Impacts and trend of congestion management strategy	• Shared cost with other agencies	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	9	8	6	7					
HERE	• External travel attributes • Other model calibration/validation	• Purchase/Subscription	• Database tools (including GIS)	8	6	5	8					
StreetLight Data	• Regional and corridor level OD patterns • Trip purpose and destination type • External travel attributes • Other model calibration/validation	• Purchase/Subscription	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	9	9	5	9					
Continuous traffic data	• Managed lane time of day usage • External travel attributes • Other model calibration/validation • Hourly counts and count validation • Vehicle classification • System performance monitoring/evaluation • Impacts and trend of congestion management strategy • Special event impacts • Freight and commercial vehicle travel • Project selection	• Collected In-house	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Custom computer coded scripts	6	7	9	8					

Data sources not included by Virginia Department of Transportation

Government record, Google Places, Co-Star Data, LIDAR and Other Airborne Data, American Transportation Research Institute (ATRI) truck GPS data , Google Travel Time, Transit Automatic Vehicle Location, Farecard/Toll Data, Social media, Other data sources

Agency:	**Wasatch Front Regional Council (Utah)**											
Major Data Sources	**What application area did your agency use the Data Source for?**	**How is the Data Source retrieved?**	**What tools are used for processing the Data Set?**	**Rate the level of challenge for using the Data Sources** *1 ("no challenge at all")* *10 ("very challenging")*					**Specify major limitations, concerns, lessons learned for applying the Data Source**	**What QA/QC procedures does your agency have in place when working with big data sources (elaborate on if QA/QC approach is different for big data applications)?**	**What are and/or will be the best data sources to understand the impacts of Connected Automated Vehicles?**	**Additional Comments**
				IT Infrastructure	**Task Timeline**	**Budget**	**Technicalities**	**Other**				
Disaggregate Census dataset (Specify, e.g. PUMS)	• Other (specify§economic development§etc.)	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Other (specify)	3	3	3	4	3				
Government record (Specify, e.g. appraisal, business licensing, QCEW)	• Other (specify§economic development§etc.)	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	2	2	1	2	7				
HERE	• Other (specify§economic development§etc.)	• Shared cost with other agencies	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS) • Other (specify)	2	2	1	8	1				
Google Travel Time	• Other (specify§economic development§etc.)	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Other (specify)	3	1	2	2	1				
Continuous traffic data	• Managed lane time of day usage • External travel attributes • Other model calibration/validation • Hourly counts and count validation • Day-to-day traffic variation • Vehicle classification	• Other (specify)	• Spreadsheets (e.g. Microsoft Excel) • Database tools (including GIS)	3	2	1	3	1				

Data sources not included by Wasatch Front Regional Council (Utah)

InfoUSA/Dun & Bradstreet business listing, Google Places, Co-Star Data, LIDAR and Other Airborne Data, Job Postings Data, National Performance Management Research Dataset (NPMRDS)/INRIX , StreetLight Data, American Transportation Research Institute (ATRI) truck GPS data , Transit Automatic Vehicle Location, Farecard/Toll Data, Social media, Other data sources

APPENDIX B

EVALUATION DOCUMENTATION

Table B-1 | Big Data Evaluation Table for the Travel Demand Forecasting Research Area

Big Data Evaluation Table | Travel Demand Forecasting									
Product	**Applicability to Research Area**	**Data Reliability and Validity**	**Data Coverage**	**Resource Requirements (Technical Staff and Training Requirements)**	**Resource Requirements (Technology, IT)**	**Data Sharing Restrictions**	**Cost**	**Overall Recommendation**	**Comments on Scoring** *** Details of subresearch area applicability are provided in Chapter 3**
INRIX	+	+	+	+	+	+	-	Yes	• **Applicability**: INRIX is a provider of vehicle probe data for segment-level congestion analytics as well as origin-destinations (O-Ds) for customized zones. INRIX has various product offerings: probe data (segment-level speed/travel time/annual average daily traffic [AADT] estimates), O-D summaries (trip starts/ends), and trip paths (includes waypoints along routes and is a very large dataset). • **Reliability/Validity**: INRIX is the underlying data source for National Performance Management Research Data Set (NPMRDS), which provides region-wide travel speeds and volume estimates for National Highway System (NHS) facilities. Therefore, the data product has undergone numerous validation efforts and is widely accepted in the industry. The Eastern Transportation Coalition (ETC) also has conducted extensive validation of various INRIX products, most notably their validations of travel speeds, they are currently sponsoring ongoing efforts to validate INRIX's ubiquitous traffic volume data. • **Coverage**: The underlying data sources are a combination of location-based services (LBS), global positioning system (GPS) from local delivery fleets and long-haul trucks, and connected vehicles (e.g. Audi/BMW). INRIX has been increasing their data providers to increase their penetration rate for passenger vehicles as well as temporal resolution (e.g. pings every 3-5 seconds). • **Resource Requirements**: Raw data, especially the O-D (raw trip paths) dataset, is enormous and requires significant data processing, analysis, and storage expertise. These individual device pings are simply a latitude/longitude/timestamp and are not mapped to a specific facility. However, INRIX has a partnership with regional integrated transportation information System/the University of Maryland (RITIS/UMD) and most agencies take advantage of the RITIS data analytics platform to store and process INRIX data. This can be done through an online graphical user interface (GUI) and aggregated result files can be downloaded in summary images or Excel files. • **Data Sharing**: INRIX data purchased by the Maryland Department of Transportation (MDOT), District Department of Transportation (DDOT), and Virginia Department of Transportation (VDOT) could be shared with the National Capital Region Transportation Planning Board (TPB) and vise-versa. If matching data sources from partner agencies in the metropolitan Washington region were purchased, these data sources could be merged for the TPB's research needs within RITIS. • **Cost**: The TPB is already investing in INRIX vehicle probe data for speed and congestion data; however, the high cost for O-D and sub-AADT volume data may be a barrier to product adoption. INRIX vehicle probe data and congestion information is available to the TPB via partner agency agreements. • **Overall Recommendation**: Yes, INRIX should be considered as an option for addressing the TPB's travel demand forecasting needs given its widespread applicability and suite of analysis tools, including O-D analytics.
StreetLight	+	+	+	+	+	+	-	Yes	• **Applicability**: StreetLight Data is an online platform for O-D or segment-based analytics based on mobile device data. Similar to INRIX, StreetLight offers a suite of analysis tools: AADT estimation, O-D, O-D with middle filter (through a midpoint location), O-D to preset geography (e.g. transportation analysis zones [TAZs], zip codes, census block groups), top routes between O-D's, and a congestion diagnostics tool for auto-generating insights. • **Reliability/Validity**: Various whitepapers are available on StreetLight's website for a variety of applications (e.g., validation of AADTs, turning movements). VDOT recently completed an evaluation of various products offered by StreetLight, including AADT, O-D estimates, traffic link volumes, turning movement volumes, and truck traffic. Evaluation results were mixed depending on the product used and volume levels being measured, with larger errors often associated with lower volumes and shorter time periods. This evaluation also provides a literature review of other recent validation efforts of StreetLight. A third-party validation recently completed by Fehr and Peers for hourly turning movement counts showed that 90 percent of locations were effectively replicated by StreetLight. • **Coverage**: StreetLight's underlying mobile data sources are mainly LBS-based (cell phone apps); INRIX is one of their underlying data sources. This mobile device data is integrated with underlying contextual data such as census demographics to provide additional insights. • **Resource Requirements**: StreetLight's online platform (StreetLight InSight) has a visualization feature for exploring and summarizing data. No raw data (individual trip data) is available; comma separated values (CSVs) and shapefiles can be downloaded from the online platform. Excel and geographic information system (GIS),if desired, are typically sufficient for further analysis, although larger or more complex queries from StreetLight may require large CSVs to be processed via a scripting tool such as R or Python in order to be usable in Excel. • **Data Sharing**: A DOT with a subscription (such as VDOT, MDOT, DDOT) can grant access to underlying metropolitan planning organizations (MPOs) covered by that DOT. • **Cost**: StreetLight offers three different packages: Essentials, Advanced, and Multi-Mode. These packages can be purchased as a subscription or on a project-by-project basis. Subscription pricing is based on the population of the coverage area (e.g., VDOT's subscription to StreetLight data is estimated to cost more than $500,000).

Big Data Evaluation Table | Travel Demand Forecasting

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									• **Overall Recommendation**: Yes, StreetLight should be considered as an option for addressing the TPB's travel demand forecasting needs given its widespread applicability and suite of analysis tools, including O-D analytics.
Teralytics	+	+	+	+	+	-	+	Yes	• **Applicability**: Teralytics is an online platform for O-D analytics based on cell phone tower triangulation data. • **Reliability/Validity**: Teralytics claim to have a less biased sample than LBS-based analytic providers as they "sit behind the firewall of all major mobile phone carriers"; they claim sufficient market share among all demographics, ethnic groups, income levels, and age groups. No validation whitepapers are provided on their website. Teralytics claims to have a high level of accuracy given its deep penetration rate, but all benchmarking appears to be done internally. One study was identified noting a limited validation effort of Teralytics "because Teralytics relies on a single data source with a considerably large penetration rate". This study showed a distribution of light rail transit (LRT) trips by time-of-day estimated by Teralytics to be consistent with a "general understanding of transit trip-making patterns" for the LRT system. • **Coverage**: Cell tower triangulation has a high sample rate (estimated at 15-35percent of population) but a lower spatial resolution. Thus, this data is not usable for route-level analysis (individual trips cannot be mapped to specific roads), but it is usable at the census tract or even TAZ level. Segment-level traffic count estimates (AADTs or turning movement counts) are not estimated via this platform. • **Resource Requirements**: Similar to StreetLight, data is accessed through an online visualization platform and viewed through the user interface (UI); data can also be downloaded via CSV and analyzed in Excel or GIS. • **Data Sharing**: No sharing of data is allowed outside of the purchasing agency except with consulting firms doing a project with that agency using the data. A purchasing agency can share derivatives (e.g., analysis results). A purchasing agency could negotiate a unique data sharing agreement as needed. • **Cost**: Generally, Teralytics is less expensive than products providing route-level analytics (e.g., INRIX and StreetLight) while providing a robust sample for understanding larger-scale travel patterns. Pricing is provided on an individual project basis or via subscription (e.g., 1 year of unlimited use and 1 year's worth of data). Pricing is based on population of the coverage area. The estimated subscription cost for standard out of the box use of the platform for an area with a population approximately the size of the metropolitan Washington Metropolitan statistical area (MSA) is approximately $50,000. Custom data sets for a population of the same size are likely in the $70,000-90,000 range, cost depends on the level of customization. • **Overall Recommendation**: Yes, Teralytics should be considered as an option for addressing the TPB's travel demand forecasting needs given its widespread applicability and suite of analysis tools, including O-D analytics.
Locus (Cambridge Systematics)	+	N/A	+	+	+	-	-	Yes	• **Applicability**: Locus is a product of LBS-based travel analytic products provided by Cambridge Systematics (consulting firm that has been the TPB's travel demand model developer). Four separate products are offered in addition to the custom analyses: O-D tables (expanded/validated by travel purpose/time-of-day), a transit competitiveness dashboard, a geofence analysis of activity around activity centers, and survey assist to supplement traditional Household Travel Survey (HTS) data. • **Reliability/Validity**: Validation is a trade secret. • **Coverage**: Locus is built through a partnership with PlaceIQ for the underlying LBS dataset (mainly from cell phone apps). • **Resource Requirements**: This product is essentially a combination of consultant services and software, it is a customizable/tailored solution. Data can be sliced as needed, with analyses conducted on an as-needed basis in addition to the products described in the next bullet. Analyses are not constrained by the available inputs and options on an online platform. Limited data storage and processing are required on the agency end, as analyses are being conducted as part of consultant services, and data is accessed via dashboards. • **Data Sharing**: Data licensing and sharing is flexible; MPO sharing upward to a DOT level would be an additional fee. • **Cost**: Base product (trip tables), including consultant services, is approximately $150,000 to $200,000; data would be available at the census tract level. These trip tables would provide O-D flows by trip purpose and time of day. Consulting labor fees are the main driver of cost for additional custom analyses, including building custom dashboards. • **Overall Recommendation**: Yes, Locus should be considered as an option for addressing the TPB's travel demand forecasting needs given its widespread applicability and suite of analysis tools, including O-D analytics.
Replica	+	N/A	+	+	+	+	+	Yes	• **Applicability**: Replica is an online platform for aggregate-level mobility, economic activity, and COVID-19 trend data as well as detailed travel pattern data for select regions, including the Baltimore-Washington region which covers the TPB planning area. O-D data, represented by a synthetic population, is available for all major purposes (work, home, eat, shop, school, social, recreation, errands, lodging, pass-through, commercial and other) and modes (driving, auto passenger, taxi/TNC, transit, walk, bike and commercial vehicles). O-D data can be further filtered and partitioned based on several other variables, such as trip start time, distance and duration. The data for individual trips from the synthetic population data can be downloaded for post-processing outside of the Replica platform. Replica provides highway traffic volume estimates based on the OSM street network. • **Reliability/Validity**: Validation results can be found mostly from the quality reports prepared by Replica when calibrating the activity-based model using ground truth data collected from each region. • **Coverage**: The aggregate-level Trends module covers the entire US, and the more detailed Places module covers select regions, including the Baltimore-Washington region which includes the TPB planning area.

Big Data Evaluation Table | Travel Demand Forecasting

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									• **Resource Requirements**: The data platform is straightforward to use for viewing, charting, tabulating and mapping data, developing standard data reports, and performing data analytics (e.g. select-link analysis). The data platform also supports the download of detailed data for custom applications. Although the size of some downloaded data files may be large, the files can be analyzed further in Excel, ArcGIS, and other commonly used software tools. • **Data Sharing**: Data licensing and sharing is flexible; annual subscription supports unlimited user licenses within an MPO if the MPO single access option is selected or unlimited user licenses within both an MPO and its member organizations with the MPO regional access option. Consultants working for the MPO are regarded as the extended MPO staff and also can be granted temporary license for data access. • **Cost**: Data cost is relatively reasonable when compared with similar data such as StreetLight Data. The total cost is based on the population of the largest city in the MPO region and the choice of either the single access option or regional access option. For TPB, the annual subscription cost is estimated to range between $104,000 and $173,000 for 2 years of full data access. • **Overall Recommendation**: Yes, Replica should be considered as an option for addressing the TPB's travel demand forecasting needs given its widespread applicability and suite of analysis tools, including O-D, transit, and active transportation analytics.
Uber Movement	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: This dataset provides aggregated zone-to-zone travel time data derived from Uber trips freely available for the Washington, D.C., area, but only at the TAZ (District of Columbia only) or census tract (roughly the area contained by the I-495 Capital Beltway) level. • **Overall Recommendation**: No, Uber Movement should not be considered as an option for addressing the TPB's Travel Demand Forecasting needs given that it does not provide O-D data, and the zone-to-zone travel time data it does provide is only within the District of Columbia and based on sampled Uber trips.
Taxi/TNC Trip Data	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: The District of Columbia has aggregated data for private vehicles-for-hire, such as taxis and transportation network companies (TNCs), including Uber and Lyft, as information is required to be reported to Department of For-Hire Vehicles (DFHV) and DDOT under D.C. law (§ 50–301.29a. General requirements for private vehicles-for-hire). • **Overall Recommendation**: No, Taxi/TNC trip data should not be considered as an option for addressing the TPB's travel demand forecasting needs given that it is only limited to the District of Columbia .
Strava Metro	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Strava Metro is an aggregator of bicycle and pedestrian movements using data obtained from the Strava user mobile app (for tracking bicycle/running/walking trips). Strava Metro is their product offering for planners and provides the relative level of activity (separated out into bicycle/pedestrian) along various facilities. • **Overall Recommendation**: No, Strava Metro should not be considered as an option for addressing the TPB's travel demand forecasting needs as it does not provide O-D data, and the level of bicycle and pedestrian activity shown is relative (not quantified).
Disaggregate Census Data	+	+	+	+	+	+	+	Yes	• **Applicability**: The U.S. Census Bureau publishes large demographic datasets on travel behavior and population demographics. Decennial Census data (e.g., 2020 Census) is usually made available by the Census Bureau less than a year after the survey. Other major Census product uses are the Public Use Microdata Sample (PUMS) from the American Community Survey (ACS) and Census Transportation Planning Product (CTPP). Quarterly Census of Employment and Wages (QCEW) is a government data product used for a variety of purposes, such as economic security monitoring and labor statistic reporting. • **Reliability/Validity**: National Cooperative Highway Research Program *(NCHRP) 08-36, Task 127, Employment Data for Planning - A Resource Guide*, includes detailed discussion of QCEW data, how it is compared with the other employment data products as well as how the quality and granularity of QCEW data is compared from state to state. Peer agencies such as Southern California Association of Governments (SCAG), Chicago Metropolitan Agency for Planning (CMAP), Atlanta Regional Commission (ARC), Maricopa Association of Governments (MAG) and many others use the QCEW data to help develop the employment databases for their regional travel demand models. • **Coverage**: QCEW provides both detailed disaggregate records and summaries of employment count and wage information for more than 95 percent of U.S. jobs, available at the county, MSA, state, and national levels by industry, and by establishment and/or headquarter locations. • **Resource Requirements**: The product contains a relatively small number of data records that are easy to analyze using conventional data tools. • **Data Sharing/Cost**: The disaggregate records of QCEW can only be acquired through an agreement with the state Department of Economic Security (DES) that may require a fee, which is usually nominal. • **Overall Recommendation**: Yes, Disaggregate Census Data should be considered as an option for addressing the TPB's travel demand forecasting needs given its applicability toward creating zonal demographic information.
Household Travel Survey	+	+	+	+	+	+	+	Yes	• **Applicability**: The Household Travel Survey (HTS) reveals insightful information with behavioral characteristics of travelers and the relationship between travel decisions and travelers' demographic backgrounds. Traditionally, HTS collects data that pertains to household, person, vehicle, and trip characteristics on selected travel day(s) and usually is conducted every 5 to 10 years by MPOs, state DOTs, and the Federal Highway Administration (FHWA) through the National Household Travel Survey (NHTS) program. All surveyed peer agencies stated the usage of HTS data to support travel demand modeling and transportation planning studies. • **Reliability/Validity**: NHTS provides a wealth of information regarding national HTS through published survey data, reports, and other documentations. MAG, Tampa, FL; Baton Rouge, LA; along with several others, recently conducted GPS-assisted HT. MAG's household travel survey report describes in much detail how the survey was administered, data was collected, and processed for the Phoenix area in Arizona. All surveyed peer agencies stated the usage of HTS data to support travel demand modeling and transportation planning studies.

Big Data Evaluation Table | Travel Demand Forecasting

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									• **Coverage**: Dependent upon the survey and agency conducting the survey, VDOT will participate in the add-on program of the first NextGen HTS to collect additional 10,000 household samples throughout the state of Virginia, including areas of Northern Virginia in the metropolitan Washington, region. • **Resource Requirements**: HTS data usually comes with the sample expansion factors for households, persons, and trips that should be applied to represent the regional travel pattern. Other data processing requirements are all analysis specific. These analyses can be performed with traditional data processing tools. • **Data Sharing**: Agreements can be made to share survey data from the agency initiating the survey. • **Cost**: HTS data can be collected either through participating in the add-on program of NHTS at a unit price of $200-$250 per completed household sample, or by other data collection contractors with the cost ranging in general between $150-$350 per sample. • **Overall Recommendation**: Yes, HTS should be considered as an option for addressing the TPB's travel d forecasting needs, as the HTS is a widely-accepted source of data for travel demand model development.
InfoUSA/Dun & BradStreet Business Listings	+	+	-	+	+	+	-	Yes	• **Applicability**: InfoUSA/Dun & Bradstreet business listing data is among the most commonly used data products to support the development of employment databases for both regional land use, travel demand modeling, and economic development activities. Both InfoUSA and Dun & Bradstreet data products provide detailed established location-based business information including employment size and industry sector. • **Reliability/Validity**: No validation white papers are provided on their website. Both InfoUSA and Dun & Bradstreet data claim to have a high level of data accuracy and coverage. *NCHRP 08-36, Task 127, Employment Data for Planning - A Resource Guide*, provides some coverage of both data products. Many peer agencies, including ARC, CMAP, Southeast Michigan Council of Governments (SEMCOG) and Pima Association of Governments (PAG), indicated the usage of either or both of the data products for employment data development. • **Coverage**: InfoUSA and Dun & Bradstreet business listing data are very similar to QCEW data in terms of applicable research areas at MPOs; however, they do not provide as much detailed wages information as QCEW. The archived data from InfoUSA/Dun & Bradstreet is a good data source to help understand the historical trend of local and regional employment and industries as well as the formation, closure, transformation, and relocation of individual businesses. • **Resource Requirements**: The data size is relatively small and is easy to analyze in Excel or other similar tools. • **Data Sharing/Cost**: The cost of either dataset is generally based on the number of business records and number of data attributes included in the regional dataset. Discrete purchase versus continuous subscription, data customization, the overall size of the data, among many other factors contribute to the overall cost. Data sharing would be negotiated upon data purchase. • **Overall Recommendation**: Yes, InfoUSA/Dun & Bradstreet Business Listings should be considered as an option for addressing the TPB's travel demand forecasting needs given its applicability toward creating zonal demographic and employment information.
CoStar Data	+	+	-	+	+	+	-	Yes	• **Applicability**: CoStar is a database of commercial real estate transactions. CoStar data can be used to help understand the real estate development patterns in the region, provide input and parameters to land use modeling, and derive employment type and activities. • **Reliability/Validity**: Data is verified with phone calls and market research. Their market-brochure suggested 47,000 calls are made daily to continually update their database. ARCused CoStar data to assist with regional modeling work, especially as a data source of average monthly rents by type of multifamily units for land use modeling. MAG also used CoStar data for socioeconomic modeling activities. • **Coverage**: CoStar provides a database with 114 billion square feet of tracked real estate and five million properties tracked across every sector. Underwriting reports are completed at the property level, covering approximately 470,000 multifamily homes and 700,000 office spaces. • **Resource Requirements**: A user interface is provided to query the information for specific commercial real-estate records and the query results can be downloaded into a format that can be easily handled by Excel and other tools. • **Data Sharing/Cost**: Cost of CoStar data is subscription based and the usage of the data is limited to the designated personnel included in the subscription agreement which covers a year at a minimum. Detailed pricing information can only be acquired through providing many usage specifics to the sales team of CoStar, including the term of subscription, number of assigned data users, geography to be covered as well as the nature of the purchase agency (government, nonprofit, etc.). Data sharing would be negotiated upon data purchase. • **Overall Recommendation**: Yes, CoStar Data should be considered as an option for addressing the TPB's travel demand forecasting needs given its applicability toward creating zonal demographic and employment information.
Google Places	+	+	-	+	+	+	-	Yes	• **Applicability**: Google Places provides information on location and type of places such as offices, parks, restaurants, and transit stops. This product offers an alternative approach to verify and supplement the business listing data acquired from the other sources (e.g., InfoUSA, Dun & Bradstreet, and QCEW), particularly about the presence, location, and type of businesses to better inform regional land use and travel demand models. • **Reliability/Validity**: According to their website, updates are performed regularly with 25 million updates each day to provide accurate real-time location information. PAG uses Google Places data as a routine part of the standard processes to develop the socioeconomic dataset and statewide employment database. A few other agencies also use this product but in a less formal fashion. • **Coverage**: According to their website, the database is comprised of 150 million places. • **Resource Requirements**: Google Places is housed in the Google Cloud platform with a web-based UI that is similar to Google Maps.

Big Data Evaluation Table | Travel Demand Forecasting

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									• **Data Sharing/Cost**: As advertised on their website, pricing for this service is pay-as-you-go; therefore, it is a function of use for each product. Cost is calculated differently for each type of query made within Google Places. For example, query of "Place Details" (i.e., name and address) is approximately $17 per 1000 requests. • **Overall Recommendation**: Yes, Google Places should be considered as an option for addressing the TPB's travel demand forecasting needs given that it provides location and type information for offices, parks, restaurants, and transit stops.
Continuous Traffic Count Station/Sensor Data	+	+	+	+	+	+	+	Yes	• **Applicability**: Local and state DOTs typically have a series of permanent or temporary devices set up for collecting traffic counts, and likely, vehicle classifications and spot speeds. From this data, agencies can estimate the average daily number of vehicles traversing roadway segments and vehicle miles traveled. • **Reliability/Validity**: This data source is commonly used as the benchmark to validate Big Data products. • **Coverage**: Coverage is dependent on the number of stations implemented by partner agencies within their jurisdictions. • **Resource Requirements**: Permanent count devices/sensors may have communications set up to allow for real-time transfer of count data to a central server. While this data, as well as data from short-term counts, is stored and organized to an extent within each DOT's count data management systems, real-time feeds for traffic counts are not typically publicly available, and publicly-available historic archive count data is typically in a post-processed format (e.g. PDF, Excel, Shapefile) that provides summary information such as AADT and daily truck percentages. • **Data Sharing/Cost**: This data would be provided by partner agencies at their discretion. Most count data is open to the public in various formats. Access to raw count data would need to be discussed on an agency-by-agency basis. Aggregated traffic count data for the region is already being compiled by TPB and is provided on the Regional Transportation Data Clearinghouse (RTDC). • **Overall Recommendation**: Yes, Continuous Traffic Count Station/Sensor Data should be considered as an option for addressing the TPB's travel demand forecasting needs, as this data is currently used in validation of volumes at screenline and individual link levels in model development and calibration efforts.
Automated Traffic Signal Performance Measures (ATSPM)	+	+	-	-	-	+	-	Yes	• **Applicability**: An emerging field of Big Data analytics in the transportation industry involves performance measures for traffic signals based on high resolution (up to 1/10th second) event data. This data, such as detector calls, phase changes, and transit signal priority/preemption can be used to provide analytics on operations at individual signals or along entire arterial corridors. Automated traffic signal performance measures (ATSPMs) are emerging as an alternative to the traditional traffic engineering practice of ad-hoc turning movement count data collection, traffic signal timing, and arterial operations analysis. The research team is not aware of any agencies using ATSPM data for travel demand modeling applications. • **Reliability/Validity**: This data source is pulling directly from infrastructure (e.g., signal cabinet and detectors); therefore, its validity and reliability is dependent on the functionality of those pieces of infrastructure (e.g., are all detectors working). Readings from this system could be used to detect infrastructure malfunctions. • **Coverage**: ATSPMs are typically an undertaking at the individual agency level for traffic signals that the agency owns and maintains. An ATSPM system requires traffic signal controllers that can generate the high-resolution event logs, vehicle detection (for most measures), communications from the signal (e.g., fiber or cellular communications), a server to gather and store the data in a database, and software to convert the data into usable information and dashboards. In the region, VDOT has deployed ATSPM systems in its Northern Region. It currently includes 12 signals on Route 50 in Fairfax County. • **Resource Requirements**: Various software packages are available for storing and analyzing the data, including open-source software (initially developed by Utah DOT) or private third-party software products, such as Miovision, Live Traffic Data, and Verizon. • **Data Sharing/Cost**: This data would be provided by partner agencies at their discretion. • **Overall Recommendation**: ATSPM data should be considered as an option for addressing the TPB's travel demand forecasting needs. However, at this time the research team notes the limited coverage within the region and likely significant processing demands required for any sort of integration in a travel demand model setting.
Transit Data from On-Board ITS Devices	+	+	+	-	+	+	+	Yes	• **Applicability**: Transit agencies employ a variety of ITS devices for monitoring the status of their system and improving system performance. The most applicable sources of transit intelligent transportation system (ITS) data for travel demand modeling is O-D data from regional SmarTrip cards Washington Metropolitan Area Transit Authority [WMATA] maintains the SmarTrip farecard data for the region). This data can be used to understand O-D travel patterns across the Metrorail system as well as some (but not all) bus rides, including transfers to/from Metrorail. Boarding and alighting data for bus systems in the region also can be applicable to model validation but do not provide O-D information. WMATA farecard data provides O-D's for all Metrorail trips as well as any bus-to-rail or rail-to-bus transfers using a SmarTrip card. Note that this data does not provide where riders are ultimately starting or ending their trips (the first-mile/last-mile problem). • **Reliability/Validity/Coverage**: O-D data for bus trips is much more challenging to obtain, as many agencies do not have Automated passenger counters (APCs) (or if they do, they do not have APCs calibrated/validated). Regardless, farebox and APC data provide information of where riders board (or also depart, in the case of APCs), but do not connect origins and destinations together; this data source does not provide first-mile/last-mile information.

Big Data Evaluation Table | Travel Demand Forecasting

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									• **Resource Requirements**: Typically, the data from these systems is available via the vendor software for agencies to process further and analyze. However, depending on the age of the software and/or investment level by an agency, the data from these systems may be in a very raw format and require a significant amount of processing and analysis to gain insights. • **Data Sharing/Cost**: This data would be provided by partner agencies at their discretion. • **Overall Recommendation**: Yes, Transit Data from On-Board ITS devices should be considered as an option for addressing the TPB's travel demand forecasting needs. The most obvious application would be farecard O-D data from the Metrorail system, which could be used for mode choice and transit assignment validation efforts.
RITIS	+	+	N/A	+	+	+	+	Yes	• **Applicability**: RITIS from the Center for Advanced Transportation Technology (CATT) Lab at the University of Maryland (UMD) is an "automated data fusion and dissemination system that provides an enhanced overall view of the transportation network. Participating agencies are able to view transportation and emergency management information through innovative visualizations and use it to improve their operations and emergency preparedness". RITIS ingests standardized data from outside sources, including mobile device data from providers such as INRIX, HERE, and TomTom as well as data on weather, traffic incidents, and other sources of data such as agency closed-circuit television (CCTV) and detector feeds. • **Reliability/Validity**: The RITIS platform is used by numerous agencies to process a variety of Big Data products. The validity of the processed outputs are largely dependent on the underlying validity of the ingested Big Data products themselves. • **Coverage**: Coverage would be dependent on the data source acquired and provided to RITIS for analytics. • **Resource Requirements**: RITIS provides a user-friendly web-based interface for performing analytics. Result files can be downloaded from this interface as summary images or Excel workbooks. • **Data Sharing/Cost**: Access to various tools, as well as the underlying data being processed by those tools, is dependent upon each state's or agency's contract with the CATT Lab. Use of RITIS is free-of-charge when purchasing more than $100,000 of INRIX data. • **Overall Recommendation**: Yes, RITIS should be considered as an option for addressing the TPB's travel demand forecasting needs given its interoperability with INRIX data as well as its ability to ingest data from other sources, such as transit automatic vehicle location (AVL) data.
Moonshadow/DB4IoT	+	+	N/A	-	+	+	-	Yes	• **Applicability**: Moonshadow is a provider of connected vehicle data analytics. Through a partnership with underlying providers such as INRIX and wejo, they have developed a database and visualization suite called DB4IoT. This product serves as both a backend (data storage/processing engine for raw data from vehicles and infrastructure) and a frontend (dashboard for visualizing data). Moonshadow/DB4IoT currently ingests data from a variety of sources including: connected vehicle data (with underlying providers such as INRIX and wejo), mobile application LBS data (with underlying providers such as Unacast and X-Mode Social), and customer data (i.e., data feed unique to specific agencies, such as public transportation feeds, micromobility companies, WiFi/Bluetooth field data collection devices, and traffic counters). • **Reliability/Validity**: Moonshadow/DB4IoT is a newer platform that performs many of the same functions as RITIS. Its competitive advantage is its processing speed, which allows for queries and analyses to be run in a few seconds, rather than requiring the user to check back in to see when an analysis is done running (as currently needed with the StreetLight and RITIS tool suites). Moonshadow is headquartered in Oregon and noted having a "large presence" with clients in western Europe as well as having worked with some agencies in the New York City area. • **Coverage**: Coverage would be dependent on the data source acquired and provided to Moonshadow for analytics. • **Resource Requirements**: The research team feels that this product currently offers the most potential to users who are very familiar with the underlying datasets and comfortable setting up custom analyses and viewing outputs as a heat map; its end-user analytics features are not as refined and developed as a platform such as RITIS. It appears to be most applicable for project-specific use cases, rather than system-wide performance monitoring at this time. No modal breakdown is provided at this time and data is provided as raw trip counts (i.e., not scaled up to estimate total traffic on a facility). • **Data Sharing/Cost**: Agencies purchase separate underlying datasets and then pay Moonshadow a set-up fee plus a percentage of the data purchase cost (typically 30 percent, although the pricing would be structured differently if agencies are coming in with data that they have already purchased separately but need Moonshadow to process). Essentially, for INRIX and wejo, Moonshadow is a value-added reseller of their data. Pricing also is dependent upon the amount of time an agency wants access to the data and the number of users accessing the data. • **Overall Recommendation**: Yes, Moonshadow/DB4IoT should be considered as an option for addressing the TPB's travel demand forecasting needs at this time. However, the research team notes its additional cost as compared to RITIS and unproven track record for system-wide use in the U.S.
Swiftly	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Swiftly is a vendor of transit analytics using feeds from agency systems; they claim to be "hardware-agnostic and software-centered" for consuming and analyzing feeds from various AVL providers. They assist public transit agencies in enhancing their transit service by analyzing on-time performance and identifying operational issues. • **Overall Recommendation**: No, Swiftly should not be considered as an option for addressing the TPB's travel demand forecasting needs. It is a tool geared more toward applications assessing transit operations and performance as opposed to understanding regional travel within a transit system.

Big Data Evaluation Table | Travel Demand Forecasting									
Product	**Applicability to Research Area**	**Data Reliability and Validity**	**Data Coverage**	**Resource Requirements (Technical Staff and Training Requirements)**	**Resource Requirements (Technology, IT)**	**Data Sharing Restrictions**	**Cost**	**Overall Recommendation**	**Comments on Scoring** *** Details of subresearch area applicability are provided in Chapter 3**
Moovit	-	N/A	N/A	N/A	N/A	N/A	N/A	**No**	• **Applicability**: Moovit is a mobility software company that offers both user-facing products (e.g., rider trip planning tools) and analytics for transit agencies. They offer two main products: Moovit Urban Mobility Analytics (MUMA) and Moovit Mobility as a Service (MaaS). Due to their multimodal focus, they have as much experience with MPOs as they do with transit agencies. • **Overall Recommendation**: No, Moovit should not be considered as an option for addressing the TPB's travel demand forecasting needs. It is a tool geared more toward applications assessing transit operations and performance as opposed to understanding regional travel within a transit system.
Emerging Data Sources	N/A	N/A	N/A	N/A	N/A	N/A	N/A	**N/A**	• The emerging data sources evaluated as part of this study are non-exhaustive and geared towards the following research areas with limited data availability: transit and non-motorized travel, TNCs, and connected and automated vehicles.

Table B-2 | Big Data Evaluation Table for the Travel Demand Management Research Area

Big Data Evaluation Table | Travel Demand Management									
Product	**Applicability to Research Area**	**Data Reliability and Validity**	**Data Coverage**	**Resource Requirements (Technical Staff and Training Requirements)**	**Resource Requirements (Technology, IT)**	**Data Sharing Restrictions**	**Cost**	**Overall Recommendation**	**Comments on Scoring** *** Details of subresearch area applicability are provided in Chapter 3**
INRIX	+	+	+	+	+	+	-	Yes	• **Applicability**: INRIX is a provider of vehicle probe data for segment-level congestion analytics as well as O-D's for customized zones. INRIX has various product offerings: probe data (segment-level speed/travel time/AADT estimates), O-D summaries (trip starts/ends), and trip paths (includes waypoints along routes and is a very large dataset). • **Reliability/Validity**: INRIX is the underlying data source for NPMRDS, which provides region-wide travel speeds and volume estimates for National Highway System (NHS) facilities. Therefore, the data product has undergone numerous validation efforts and is widely accepted in the industry. ETC also has conducted extensive validation of various INRIX products, most notably their validations of travel speeds. They are currently sponsoring ongoing efforts to validate INRIX's ubiquitous traffic volume data. • **Coverage**: The underlying data sources are a combination of LBS, GPS from local delivery fleets and long-haul trucks, and connected vehicles (e.g. Audi/BMW). INRIX has been increasing their data providers to increase their penetration rate for passenger vehicles as well as temporal resolution (e.g. pings every 3-5 seconds). • **Resource Requirements**: Raw data, especially the O-D (raw trip paths) dataset, is enormous and requires significant data processing, analysis, and storage expertise. These individual device pings are simply a latitude/longitude/timestamp and are not mapped to a specific facility. However, INRIX has a partnership with RITIS/UMD and most agencies take advantage of the RITIS data analytics platform to store and process INRIX data. This can be done through an online GUI and aggregated result files can be downloaded in summary images or Excel files. • **Data Sharing**: INRIX data purchased by MDOT, DDOT, and VDOT could be shared with the TPB and vise-versa. If matching data sources from partner agencies in the metropolitan Washington region were purchased, these data sources could be merged for the TPB's research needs within RITIS. • **Cost**: The TPB is already investing in INRIX vehicle probe data for speed and congestion data; however, the high cost for O-D and sub-AADT volume data may be a barrier to product adoption. INRIX vehicle probe data and congestion information is available to the TPB via partner agency agreements. • **Overall Recommendation**: Yes, INRIX should be considered as an option for addressing the TPB's travel demand management needs, as INRIX O-D data could be explored to highlight opportunities for multimodal shifts.
StreetLight	+	+	+	+	+	+	-	Yes	• **Applicability**: StreetLight Data is an online platform for O-D or segment-based analytics based on mobile device data. Similar to INRIX, StreetLight offers a suite of analysis tools: AADT estimation, O-D, O-D with middle filter (through a midpoint location), O-D to preset geography (e.g. TAZs, zip codes, census block groups), top routes between O-D's, and a congestion diagnostics tool for auto-generating insights. • **Reliability/Validity**: Various whitepapers are available on StreetLight's website for a variety of applications (e.g., validation of AADTs, turning movements). VDOT recently completed an evaluation of various products offered by StreetLight, including AADT, O-D estimates, traffic link volumes, turning movement volumes, and truck traffic. Evaluation results were mixed depending on the product used and volume levels being measured, with larger errors often associated with lower volumes and shorter time periods. This evaluation also provides a literature review of other recent validation efforts of StreetLight. A third-party validation recently completed by Fehr and Peers for hourly turning movement counts showed that 90 percent of locations were effectively replicated by StreetLight. • **Coverage**: StreetLight's underlying mobile data sources are mainly LBS-based (cell phone apps); INRIX is one of their underlying data sources. This mobile device data is integrated with underlying contextual data such as census demographics to provide additional insights. • **Resource Requirements**: StreetLight's online platform (StreetLight InSight) has a visualization feature for exploring and summarizing data. No raw data (individual trip data) is available; CSVs and shapefiles can be downloaded from the online platform. Excel and GIS (if desired) are typically sufficient for further analysis, although larger or more complex queries from StreetLight may require large CSVs to be processed via a scripting tool such as R or Python to be usable in Excel. • **Data Sharing**: A DOT with a subscription (such as VDOT, MDOT, DDOT) can grant access to underlying MPOs covered by that DOT. • **Cost**: StreetLight offers three different packages: Essentials, Advanced, and Multi-Mode. These packages can be purchased as a subscription or on a project-by-project basis. Subscription pricing is based on the population of the coverage area (e.g., VDOT's subscription to StreetLight data is estimated to cost more than $500,000). • **Overall Recommendation**: Yes, StreetLight should be considered as an option for addressing the TPB's travel demand management needs, as its trip and traveler attributes datasets can potentially be used to identify trips amenable to a potential mode shift. Additionally, analytics provided for individual days of the week can be used to understand travel pattern differences across the work week. StreetLight also offers the option to analyze existing bicycle and pedestrian travel.
Teralytics	+	+	+	+	+	-	+	Yes	• **Applicability**: Teralytics is an online platform for O-D analytics based on cell phone tower triangulation data. • **Reliability/Validity**: Teralytics claim to have a less biased sample than LBS-based analytics providers as they "sit behind the firewall of all major mobile phone carriers"; they claim sufficient market share among all demographics, ethnic groups, income levels, and age groups. No validation whitepapers are provided on their website. Teralytics claims to have a high level of accuracy given its deep penetration rate, but all benchmarking appears to be done internally. One study was identified noting a limited validation effort of Teralytics "because Teralytics relies on a single data

Big Data Evaluation Table | Travel Demand Management

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									source with a considerably large penetration rate". This study showed a distribution of LRT trips by time-of-day estimated by Teralytics to be consistent with a "general understanding of transit trip-making patterns" for the LRT system. • **Coverage**: Cell tower triangulation has high sample rate (estimated at 15-35 percent of population) but a lower spatial resolution. Thus, this data is not usable for route-level analysis (individual trips cannot be mapped to specific roads), but it is usable at the census tract or even TAZ level. Segment-level traffic count estimates (AADTs or turning movement counts) are not estimated via this platform. • **Resource Requirements**: Similar to StreetLight, data is accessed through an online visualization platform and viewed through the UI; data also can be downloaded via CSV and analyzed in Excel or GIS. • **Data Sharing**: No sharing of data is allowed outside of the purchasing agency except with consulting firms doing a project with that agency using the data. A purchasing agency can share derivatives (e.g., analysis results). A purchasing agency could negotiate a unique data sharing agreement as needed. • **Cost**: Generally, Teralytics is less expensive than products providing route-level analytics (e.g., INRIX and StreetLight) while providing a robust sample for understanding larger-scale travel patterns. Pricing is provided on an individual project basis or via subscription (e.g., one year of unlimited use and one year's worth of data). Pricing is based on population of coverage area. The estimated subscription cost for standard out of the box use of the platform for an area with a population approximately the size of metropolitan Washington MSA is approximately $50,000. Custom data sets for a population of the same size are likely in the $70,000-$90,000 range; cost depends on the level of customization. • **Overall Recommendation**: Yes, Teralytics should be considered as an option for addressing the TPB's travel demand management needs as it is able to ascertain trip purpose and trip frequency for O-D data. This data would be valuable for identifying work-related trips as well as identifying the number of trips being made between O-D pairs on different days of the week.
Locus (Cambridge Systematics)	+	N/A	+	+	+	-	-	Yes	• **Applicability**: Locus is a product of LBS-based travel analytic products provided by Cambridge Systematics (consulting firm that has been the TPB's travel demand model developer). Four separate products are offered in addition to the custom analyses: O-D tables (expanded/validated by travel purpose/time-of-day), a transit competitiveness dashboard, a geofence analysis of activity around activity centers, and survey assist to supplement traditional HTS data. • **Reliability/Validity**: Validation is a trade secret. • **Coverage**: Locus is built through a partnership with PlaceIQ for the underlying LBS dataset (mainly from cell phone apps). • **Resource Requirements**: This product is essentially a combination of consultant services and software; it is a customizable/tailored solution. Data can be sliced as needed, with analyses conducted on an as-needed basis in addition to the products described in the next bullet. Analyses are not constrained by the available inputs and options on an online platform. Limited data storage and processing are required on the agency end, as analyses are being conducted as part of consultant services; data accessed via dashboards. • **Data Sharing**: Data licensing and sharing is flexible; MPO sharing upward to a DOT-level would be an additional fee. • **Cost**: Base product (trip tables), including consultant services, is approximately $150,000 to $200,000; data would be available at the census tract level. These trip tables would provide O-D flows by trip purpose and time of day. Consulting labor fees are the main driver of cost for additional custom analyses, including building custom dashboards. • **Overall Recommendation**: Yes, Locus should be considered as an option for addressing the TPB's travel demand management needs, as its O-D tables could be analyzed to better understand travel patterns that are amenable to mode shifts and telework programs.
Replica	+	N/A	+	+	+	+	+	Yes	• **Applicability**: Replica is an online platform for aggregate-level mobility, economic activity, and COVID-19 trend data as well as detailed travel pattern data for select regions, including the Baltimore-Washington region which covers the TPB planning area. O-D data, represented by a synthetic population, is available for all major purposes (work, home, eat, shop, school, social, recreation, errands, lodging, pass-through, commercial and other) and modes (driving, auto passenger, taxi/TNC, transit, walk, bike and commercial vehicles). O-D data can be further filtered and partitioned based on several other variables, such as trip start time, distance and duration. The data for individual trips from the synthetic population data can be downloaded for post-processing outside of the Replica platform. The granularity of the synthetic trip data (purpose, mode, length and duration) and demographic information of travelers (income, age, gender and race/ethnicity) are particularly useful and valuable to identify trips amenable to understanding the impacts of travel demand management policies and strategies from both transportation system management and environmental justice perspectives. • **Reliability/Validity**: Validation results can be found mostly from the quality reports prepared by Replica when calibrating the activity-based model using ground truth data collected from each region. • **Coverage**: The aggregate-level Trends module covers the entire US, and the more detailed Places module covers select regions, including the Baltimore-Washington region which includes the TPB planning area. • **Resource Requirements**: The data platform is straightforward to use for viewing, charting, tabulating and mapping data, developing standard data reports, and performing data analytics (e.g. select-link analysis). The data platform also supports the download of detailed data for custom applications. Although the size of some downloaded data files may be large, the files can be analyzed further in Excel, ArcGIS, and other commonly used software tools.

Big Data Evaluation Table | Travel Demand Management

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									• **Data Sharing**: Data licensing and sharing is flexible; annual subscription supports unlimited user licenses within an MPO if the MPO single access option is selected or unlimited user licenses within both an MPO and its member organizations with the MPO regional access option. Consultants working for the MPO are regarded as the extended MPO staff and also can be granted temporary license for data access. • **Cost**: Data cost is relatively reasonable when compared with similar data such as StreetLight Data. The total cost is based on the population of the largest city in the MPO region and the choice of either the single access option or regional access option. For TPB, the annual subscription cost is estimated to range between $104,000 and $173,000 for 2 years of full data access. • **Overall Recommendation**: Yes, Replica should be considered as an option for addressing the TPB's travel demand management needs given its widespread applicability and suite of analysis tools, including O-D, transit, and active transportation analytics, as well as the synthetic population trip table.
Uber Movement	-	N/A	N/A	N/A	N/A	N/A	N/A	**No**	• **Applicability**: This dataset provides aggregated zone-to-zone travel time data derived from Uber trips freely available for the metropolitan Washington., area, but only at the TAZ (District of Columbia only) or census tract (roughly the area contained by the I-495 Capital Beltway) level. • **Overall Recommendation**: No, Uber Movement should not be considered as an option for addressing the TPB's travel demand management needs. This dataset is simply a set of zone-to-zone travel times based on Uber trip data for zones limited to the District of Columbia.
Taxi/TNC Trip Data	-	N/A	N/A	N/A	N/A	N/A	N/A	**No**	• **Applicability**: The District of Columbia has aggregated data for private vehicles-for-hire, such as taxis and TNCs, including Uber and Lyft, as information is required to be reported to DFHV and DDOT under D.C. law (§ 50–301.29a. General requirements for private vehicles-for-hire). • **Overall Recommendation**: No, Taxi/TNC trip data should not be considered as an option for addressing the TPB's travel demand management needs given that it is only limited to the District of Columbia.
Strava Metro	-	N/A	N/A	N/A	N/A	N/A	N/A	**No**	• **Applicability**: Strava Metro is an aggregator of bicycle and pedestrian movements using data obtained from the Strava user mobile app (for tracking bicycle/running/walking trips). Strava Metro is their product offering for planners and provides the relative level of activity (separated out into bicycle/pedestrian) along various facilities. • **Overall Recommendation**: No, Strava Metro should not be considered as an option for addressing the TPB's travel demand management needs as it does not provide O-D data, and the level of bicycle and pedestrian activity shown is relative (not quantified).
Disaggregate Census Data	+	+	+	+	+	+	+	Yes	• **Applicability**: The US Census Bureau publishes large demographic datasets on travel behavior and population demographics. Decennial Census data (e.g., 2020 Census) is usually made available by the Census Bureau less than a year after the survey. Other major Census product uses are PUMS from (ACS and CTPP. QCEW is a government data product used for a variety of purposes, such as economic security monitoring and labor statistic reporting. • **Reliability/Validity**: *NCHRP 08-36, Task 127, Employment Data for Planning - A Resource Guide*, includes detailed discussion of QCEW data, how it is compared with the other employment data products as well as how the quality and granularity of QCEW data is compared from state to state. Peer agencies such as SCAG, CMAP, ARC, MAG and many others use the QCEW data to help develop the employment databases for their regional travel demand models. • **Coverage**: QCEW provides both detailed disaggregate records and summaries of employment count and wage information for more than 95 percent of U.S. jobs, available at the county, MSA, state and national levels by industry, and by establishment and/or headquarter locations. • **Resource Requirements**: The product contains a relatively small number of data records that are easy to analyze using conventional data tools. • **Data Sharing/Cost**: The disaggregate records of QCEW can only be acquired through an agreement with the state DES that may require a fee, which is usually nominal. • **Overall Recommendation**: Yes, Disaggregate Census Data should be considered as an option for addressing the TPB's travel demand management needs. Census Journey-to-Work data can potentially be used to identify high-frequency home-work pairs that may be amenable to a modal shift. QCEW data can potentially be used to understand detailed employment information and possible locations to target for employer-based travel demand management policies on travel behavior.
Household Travel Survey	+	+	+	+	+	+	+	Yes	• **Applicability**: HTS reveals insightful information with behavioral characteristics of travelers and the relationship between travel decisions and travelers' demographic backgrounds. Traditionally, HTS collects the data that pertains to household, person, vehicle, and trip characteristics on selected travel day(s) and usually is conducted every 5 to 10 years by MPOs, state DOTs, and FHWA through the NHTS program. All surveyed peer agencies stated the usage of household travel survey data to support travel demand modeling and transportation planning studies. • **Reliability/Validity**: NHTS provides a wealth of information regarding national household travel survey through the published survey data, reports, and other documentations. MAG, Tampa, FL, Baton Rouge, LA; along with several others, recently conducted GPS-assisted household travel surveys. MAG's household travel survey report describes in much detail how the survey was administered, data was collected and processed for the Phoenix area in Arizona. All surveyed peer agencies stated the usage of household travel survey data to support travel demand modeling and transportation planning studies. • **Coverage**: Coverage is dependent upon the survey and agency conducting the survey. VDOT will participate in the add-on program of the first NextGen HTS to collect additional 10,000 household samples throughout the state of Virginia, including areas of Northern Virginia in the metropolitan Washington region.

Big Data Evaluation Table | Travel Demand Management

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									• **Resource Requirements**: HTS data usually comes with the sample expansion factors for households, persons, and trips that should be applied to represent the regional travel pattern. Other data processing requirements are all analysis specific. These analyses can be performed with traditional data processing tools. • **Data Sharing**: Agreements can be made to share survey data from the agency initiating the survey. • **Cost**: Household travel survey data can be collected either through participating in the add-on program of NHTS at a unit price of $200-$250 per completed household sample or by other data collection contractors with the cost ranging in general between $150-$350 per sample. • **Overall Recommendation**: Yes, HTS should be considered as an option for addressing the TPB's travel demand management needs. This traditional and long-standing survey is valuable in understanding the makeup of a community to better define current and forecast future travel demand. Self-reporting travel diaries can be used to better understand telecommuting habits.
InfoUSA/Dun & BradStreet Business Listings	+	+	-	+	+	+	-	Yes	• **Applicability**: InfoUSA/Dun & Bradstreet business listing data is among the most commonly used data products to support the development of employment databases for both regional land use, travel demand modeling, and economic development activities. Both InfoUSA and Dun & Bradstreet data products provide detailed establishment location-based business information including employment size and industry sector. • **Reliability/Validity**: No validation white papers are provided on their website. Both InfoUSA and Dun & Bradstreet data claim to have a high level of data accuracy and coverage. *NCHRP 08-36, Task 127, Employment Data for Planning - A Resource Guide*, provides some coverage of both data products. Many peer agencies, including ARC, CMAP, SEMCOG, and PAG, indicated the usage of either or both of the data products for employment data development. • **Coverage**: InfoUSA and Dun & BradStreet business listing data are very similar to QCEW data in terms of applicable research areas at MPOs. However, they do not provide as detailed wages information as QCEW. The archived data from InfoUSA/Dun & BradStreet is a good data source to help understand the historical trend of local and regional employment and industries as well as the formation, closure, transformation, and relocation of individual businesses. • **Resource Requirements**: The data size is relatively small and is easy to analyze in Excel or other similar tools. • **Data Sharing/Cost**: The cost of either dataset is generally based on the number of business records and number of data attributes included in the regional dataset. Discrete purchase versus continuous subscription, data customization, and the overall size of the data among many other factors contribute to the overall cost. Data sharing would be negotiated upon data purchase. • **Overall Recommendation**: Yes, InfoUSA/Dun & Bradstreet Business Listings should be considered as an option for addressing the TPB's traveld management needs, as it offers detailed employment information for potentially targeting specific TDM applications or programs.
CoStar Data	+	+	-	+	+	+	-	Yes	• **Applicability**: CoStar is a database of commercial real estate transactions. CoStar data can be used to help understand the real estate development patterns in the region, provide input and parameters to land use modeling, and derive employment type and activities. • **Reliability/Validity**: Data is verified with phone calls and market research. Their market-brochure suggested 47,000 calls are made daily to continually update their database. ARC used CoStar data to assist with regional modeling work, especially as a data source of average monthly rents by type of multifamily units for land use modeling. MAG also used CoStar data for socioeconomic modeling activities. • **Coverage**: CoStar provides a database with 114 billion square feet of tracked real estate and 5 million properties tracked across every sector. Underwriting reports are completed at the property level, covering approximately 470,000 multifamily homes and 700,000 office spaces. • **Resource Requirements**: A user interface is provided to query the information for specific commercial real estate records and the query results can be downloaded into a format easily handled by Excel and other tools. • **Data Sharing/Cost**: Cost of CoStar data is subscription based and the usage of the data is limited to the designated personnel included in the subscription agreement which covers a year at a minimum. Detailed pricing information can only be acquired through providing many usage specifics to the sales team of CoStar, including the term of subscription, number of assigned data users, geography to be covered as well as the nature of the purchase agency (government, nonprofit, etc.). Data sharing would be negotiated upon data purchase. • **Overall Recommendation**: Yes, CoStar Data should be considered as an option for addressing the TPB's travel demand management needs, as it offers detailed demographic information for potentially targeting specific TDM applications or programs.
Google Places	+	+	-	+	+	+	-	Yes	• **Applicability**: Google Places provides information on location and type of places such as offices, parks, restaurants, and transit stops. This product offers an alternative approach to verify and supplement the business listing data acquired from the other sources (e.g., InfoUSA, Dun & Bradstreet, and QCEW), particularly about the presence, location, and type of the businesses to better inform regional land use and travel demand models. • **Reliability/Validity**: According to their website, updates are performed regularly with 25 million updates each day to provide accurate real-time location information. PAG uses Google Places data as a routine part of the standard processes to develop the socioeconomic dataset and statewide employment database. A few other agencies also use this product but in a less formal fashion. • **Coverage**: According to their website, the database is comprised of 150 million places. • **Resource Requirements**: Google Places is housed in the Google Cloud platform with a web-based user-interface that is similar to Google Maps. • **Data Sharing/Cost**: As advertised on their website, pricing for this service is pay-as-you-go; therefore, it is a function of use for each product. Cost is calculated differently for each type of query made within Google Places. For example, query of “Place Details” (i.e., name and address) is approximately $17 per 1000 requests.

Big Data Evaluation Table | Travel Demand Management

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									• **Overall Recommendation**: Yes, Google Places should be considered as an option for addressing the TPB's travel demand management needs, as it offers detailed place-based information for potentially targeting specific TDM applications or programs.
Continuous Traffic Count Station/Sensor Data	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Local and state DOTs typically have a series of permanent or temporary devices set up for collecting traffic counts, and likely vehicle classifications and spot speeds. From this data, agencies can estimate the average daily number of vehicles traversing roadway segments and vehicle miles traveled. • **Overall Recommendation**: No, Continuous Traffic Count Station/Sensor Data should not be considered as an option for addressing the TPB's travel demand management needs. This is not an applicable data source for this research area.
Automated Traffic Signal Performance Measures (ATSPM)	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: An emerging field of Big Data analytics in the transportation industry involves performance measures for traffic signals based on high resolution (up to 1/10th second) event data. This data, such as detector calls, phase changes, and transit signal priority/preemption can be used to provide analytics on operations at individual signals or along entire arterial corridors. ATSPMs are emerging as an alternative to the traditional traffic engineering practice of ad-hoc turning movement count data collection, traffic signal timing, and arterial operations analysis. The research team is not aware of any agencies using ATSPM data for travel demand modeling applications. • **Overall Recommendation**: No, ATSPM data should not be considered as an option for addressing the TPB's travel demand management needs. This is not an applicable data source for this research area.
Transit Data from On-Board ITS Devices	+	+	+	-	+	+	+	Yes	• **Applicability**: Transit agencies employ a variety of ITS devices for monitoring the status of their system and improving system performance. The most applicable sources of transit ITS data for travel demand modeling is O-D data from regional SmarTrip cards (WMATA maintains the SmarTrip farecard data for the region). Therefore, this data can be used to understand O-D travel patterns across the Metrorail system as well as some (but not all) bus rides, including transfers to/from Metrorail. Boarding and alighting data for bus systems in the region also can be applicable to model validation but do not provide O-D information. WMATA farecard data provides O-D's for all Metrorail trips as well as any bus-to-rail or rail-to-bus transfers using a SmarTrip card. Thus, O-D data is available for trips within the Metrorail system. Note that this data does not provide where riders are ultimately starting or ending their trips (the first-mile/last-mile problem). • **Reliability/Validity/Coverage**: O-D data for bus trips is much more challenging to obtain, as many agencies do not have APCs (or if they do, they do not have APCs calibrated/validated). Regardless, farebox and APC data provide information of where riders board (or depart, in the case of APCs), but do not connect origins and destinations together; this data source does not provide first-mile/last-mile information. • **Resource Requirements**: Typically, the data from these systems is available via the vendor software for agencies to process further and analyze. However, depending on the age of the software and/or investment level by an agency, the data from these systems may be in a very raw format and require a significant amount of processing and analysis to gain insights. • **Data Sharing/Cost**: This data would be provided by partner agencies at their discretion. • **Overall Recommendation**: Yes, Transit Data from On-Board ITS devices should be considered as an option for addressing the TPB's travel demand management needs. The most obvious application would be farecard O-D data from the Metrorail system. This database could be used to support travel demand management program areas by providing reliable transit O-D data that could be paired with other Big Data products (e.g., mobile device-based travel pattern data) to identify propensity for modal shifts.
RITIS	+	+	N/A	+	+	+	+	Yes	• **Applicability**: RITIS from the CATT Lab at UMD is an "automated data fusion and dissemination system that provides an enhanced overall view of the transportation network. Participating agencies are able to view transportation and emergency management information through innovative visualizations and use it to improve their operations and emergency preparedness". RITIS ingests standardized data from outside sources, including mobile device data from providers such as INRIX, HERE, and TomTom as well as data on weather, traffic incidents, and other sources of data such as agency CCTV and detector feeds. • **Reliability/Validity**: The RITIS platform is used by numerous agencies to process a variety of Big Data products. The validity of the processed outputs are largely dependent on the underlying validity of the ingested Big Data products themselves. • **Coverage**: Coverage would be dependent on the data source acquired and provided to RITIS for analytics. • **Resource Requirements**: RITIS provides a user-friendly web-based interface for performing analytics. Result files can be downloaded from this interface as summary images or Excel workbooks. • **Data Sharing/Cost**: Access to various tools, as well as the underlying data being processed by those tools, is dependent upon each state's or agency's contract with the CATT Lab. Use of RITIS is free-of-charge when purchasing more than $100,000 of INRIX data. • **Overall Recommendation**: Yes, RITIS should be considered as an option for addressing the TPB's travel demand management needs. If INRIX O-D data is procured for the region, this platform would presumably be used to set up the analyses. This O-D data could be used to identify major O-D pairs contributing to the most constrained locations within the highway system as locations to target for travel demand management.
Moonshadow/DB4IoT	+	+	N/A	-	+	+	-	Yes	• **Applicability**: Moonshadow is a provider of connected vehicle data analytics. Through a partnership with underlying providers such as INRIX and wejo, they have developed a database and visualization suite called DB4IoT. This product serves as both a backend (data storage/processing engine for raw data from vehicles and infrastructure) and a frontend (dashboard for visualizing data). Currently DB4IoT ingests data from a variety of sources including: connected vehicle data (with underlying providers such as INRIX and wejo), mobile application LBS data (with underlying

Big Data Evaluation Table | Travel Demand Management

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									providers such as Unacast and X-Mode Social), and customer data (i.e., data feed unique to specific agencies, such as public transportation feeds, micromobility companies, WiFi/Bluetooth field data collection devices, and traffic counters). • **Reliability/Validity**: Moonshadow/DB4IoT is a newer platform that performs many of the same functions as RITIS. Its competitive advantage is its processing speed, which allows for queries and analyses to be run in a few seconds, rather than requiring the user to check back in to see when an analysis is done running (as currently needed with the StreetLight and RITIS tool suites). Moonshadow is headquartered in Oregon and noted having a large presence with clients in western Europe as well as having worked with some agencies in the New York City area. • **Coverage**: Coverage would be dependent on the data source acquired and provided to Moonshadow for analytics. • **Resource Requirements**: The research team feels that this product currently offers the most potential to users who are very familiar with the underlying datasets and comfortable setting up custom analyses and viewing outputs as a heat map; its end-user analytics features are not as refined and developed as a platform such as RITIS. It appears to be most applicable for project-specific use cases, rather than system-wide performance monitoring at this time. No modal breakdown is provided at this time and data is provided as raw trip counts (i.e., not scaled up to estimate total traffic on a facility). • **Data Sharing/Cost**: Agencies purchase separate underlying datasets and then pay Moonshadow a set-up fee plus a percentage of the data purchase cost (typically 30 percent, although the pricing would be structured differently if agencies are coming in with data that they have already purchased separately but need Moonshadow to process). Essentially, for INRIX and wejo, Moonshadow is a value-added reseller of their data. Pricing also is dependent upon the amount of time an agency wants access to the data and the number of users accessing the data. • **Overall Recommendation**: Yes, Moonshadow/DB4IoT should be considered as an option for addressing the TPB's travel demand management needs at this time, as it can provide detailed visualizations of trips destined for a specific location or using a specific roadway facility. However, the research team notes its additional cost as compared to RITIS and unproven track record for system-wide use in the U.S.
Swiftly	-	N/A	N/A	N/A	N/A	N/A	N/A	**No**	• **Applicability**: Swiftly is a vendor of transit analytics using feeds from agency systems. They claim to be "hardware-agnostic and software-centered" for consuming and analyzing feeds from various AVL providers. They assist public transit agencies in enhancing their transit service by analyzing on-time performance and identifying operational issues. • **Overall Recommendation**: No, Swiftly should not be considered as an option for addressing the TPB's travel demand management needs. It is a tool geared more toward applications assessing transit operations and performance as opposed to understanding regional travel within a transit system.
Moovit	-	N/A	N/A	N/A	N/A	N/A	N/A	**No**	• **Applicability**: Moovit is a mobility software company that offers both user-facing products (e.g., rider trip planning tools) and analytics for transit agencies. They offer two main products: MUMA and MaaS. Due to their multimodal focus, they have as much experience with MPOs as they do with transit agencies. • **Overall Recommendation**: No, Moovit should not be considered as an option for addressing the TPB's travel demand management needs. It is a tool geared more toward applications assessing transit operations and performance as opposed to understanding regional travel within a transit system.
Emerging Data Sources	N/A	N/A	N/A	N/A	N/A	N/A	N/A	**N/A**	• The emerging data sources evaluated as part of this study are non-exhaustive and geared towards the following research areas with limited data availability: transit and non-motorized travel, transportation network companies, and connected and automated vehicles.

Table B-3 | Big Data Evaluation Table for the System Performance and Congestion Management Research Area

Big Data Evaluation Table | System Performance and Congestion Management									
Product	**Applicability to Research Area**	**Data Reliability and Validity**	**Data Coverage**	**Resource Requirements (Technical Staff and Training Requirements)**	**Resource Requirements (Technology, IT)**	**Data Sharing Restrictions**	**Cost**	**Overall Recommendation**	**Comments on Scoring** *** Details of subresearch area applicability are provided in Chapter 3**
INRIX	+	+	+	+	+	+	-	Yes	• **Applicability**: INRIX is a provider of vehicle probe data for segment-level congestion analytics as well as O-D's for customized zones. INRIX has various product offerings: probe data (segment-level speed/travel time/AADT estimates), O-D summaries (trip starts/ends), and trip paths (includes waypoints along routes and is a very large dataset). • **Reliability/Validity**: INRIX is the underlying data source for NPMRDS, which provides region-wide travel speeds and volume estimates for NHS facilities. The data product has undergone numerous validation efforts and is widely accepted in the industry. The ETC also has conducted extensive validation of various INRIX products, most notably their validations of travel speeds; they are currently sponsoring ongoing efforts to validate INRIX's ubiquitous traffic volume data. • **Coverage**: The underlying data sources are a combination of LBS, GPS from local delivery fleets and long-haul trucks, and connected vehicles (e.g. Audi/BMW). INRIX has been increasing their data providers to increase their penetration rate for passenger vehicles as well as temporal resolution (e.g. pings every 3-5 seconds). • **Resource Requirements**: Raw data, especially the O-D (raw trip paths) dataset, is enormous and requires significant data processing, analysis, and storage expertise. These individual device pings are simply a latitude/longitude/timestamp and are not mapped to a specific facility. However, INRIX has a partnership with RITIS/UMD and most agencies take advantage of the RITIS data analytics platform to store and process INRIX data. This can be done through an online GUI and aggregated result files can be downloaded in summary images or Excel files. • **Data Sharing**: INRIX data purchased by MDOT, DDOT, and VDOT could be shared with the TPB and vise-versa. If matching data sources from partner agencies in the metropolitan Washington region were purchased, these data sources could be merged for the TPB's research needs within RITIS. • **Cost**: The TPB is already investing in INRIX vehicle probe data for speed and congestion data; however, the high cost for O-D and sub-AADT volume data may be a barrier to product adoption. INRIX vehicle probe data and congestion information is available to the TPB via partner agency agreements. • **Overall Recommendation**: Yes, INRIX should be considered as an option for addressing the TPB's system performance and congestion management needs given its widespread applicability and suite of analysis tools. Data can be queried at a regional or corridor-specific level in support of performance measurement or project-specific effort, including data for multiple time periods to track the impact of congestion management strategies. Trends in regional traffic and congestion can be obtained looking at annual performance compared to previous years. Certain tools within RITIS, such as the Bottleneck Ranking or User Delay Cost tools, can be used to target areas for congestion management. Data can be queried for individual days to understand the impacts of specific non-recurring events and compared to average days.
StreetLight	+	+	+	+	+	+	-	Yes	• **Applicability**: StreetLight Data is an online platform for O-D or segment-based analytics based on mobile device data. Similar to INRIX, StreetLight offers a suite of analysis tools: AADT estimation, O-D, O-D with middle filter (through a midpoint location), O-D to preset geography (e.g. TAZs, zip codes, census block groups), top routes between O-D's, and a congestion diagnostics tool for auto-generating insights. • **Reliability/Validity**: Various whitepapers are available on StreetLight's website for a variety of applications (e.g., validation of AADTs, turning movements). VDOT recently completed an evaluation of various products offered by StreetLight, including AADT, O-D estimates, traffic link volumes, turning movement volumes, and truck traffic. Evaluation results were mixed depending on the product used and volume levels being measured, with larger errors often associated with lower volumes and shorter time periods. This evaluation also provides a literature review of other recent validation efforts of StreetLight. A third-party validation recently completed by Fehr and Peers for hourly turning movement counts showed that 90 percent of locations were effectively replicated by StreetLight. • **Coverage**: StreetLight's underlying mobile data sources are mainly LBS-based (cell phone apps); INRIX is one of their underlying data sources. This mobile device data is integrated with underlying contextual data such as census demographics to provide additional insights. • **Resource Requirements**: StreetLight's online platform (StreetLight InSight) has a visualization feature for exploring and summarizing data. No raw data (individual trip data) is available; CSVs and shapefiles can be downloaded from the online platform. Excel and GIS (if desired) are typically sufficient for further analysis, although larger or more complex queries from StreetLight may require large CSVs to be processed via a scripting tool such as R or Python to be usable in Excel. • **Data Sharing**: A DOT with a subscription (such as VDOT, MDOT, DDOT) can grant access to underlying MPOs covered by that DOT. • **Cost**: StreetLight offers three different packages: Essentials, Advanced" and Multi-Mode. These packages can be purchased as a subscription or on a project-by-project basis. Subscription pricing is based on the population of the coverage area (e.g., VDOT's subscription to StreetLight data is estimated to cost more than $500,000). • **Overall Recommendation**: Yes, StreetLight should be considered as an option for addressing the TPB's system performance and congestion management needs given its widespread applicability and suite of analysis tools. Data can be queried at a regional or corridor-specific level in support of performance measurement or project-specific effort, including data for multiple time periods to track the impact of congestion management strategies. Trends in regional traffic and congestion can be obtained looking at annual performance compared to previous years. Certain tools within StreetLight, such as the congestion diagnostics tools, can be used to target areas for congestion management. Data can be queried for individual days (at an additional cost) to understand the impacts of specific non-recurring events and compared to average days.

Big Data Evaluation Table | System Performance and Congestion Management

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
Teralytics	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Teralytics is an online platform for O-D analytics based on cell phone tower triangulation data. • **Overall Recommendation**: No, Teralytics should not be considered as an option for addressing the TPB's system performance and congestion management needs, as it is not able to provide analytics for individual facilities.
Locus (Cambridge Systematics)	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Locus is a product of LBS-based travel analytics product provided by Cambridge Systematics (consulting firm that has been the TPB's travel demand model developer). Four separate products are offered in addition to the custom analyses: O-D tables (expanded/validated by travel purpose/time-of-day), a transit competitiveness dashboard, a geofence analysis of activity around activity centers, and survey assist to supplement traditional HTS data. • **Overall Recommendation**: No, Locus should not be considered as an option for addressing the TPB's system performance and congestion management needs, as it is not able to provide analytics for individual facilities.
Replica	-	N/A	+	+	+	+	+	No	• **Applicability**: Replica is an online platform for aggregate-level mobility, economic activity, and COVID-19 trend data as well as detailed travel pattern data for select regions, including the Baltimore-Washington region which covers the TPB planning area. O-D data, represented by a synthetic population, is available for all major purposes (work, home, eat, shop, school, social, recreation, errands, lodging, pass-through, commercial and other) and modes (driving, auto passenger, taxi/TNC, transit, walk, bike and commercial vehicles). Replica provides estimated traffic volume by vehicle classes that can be used to help with system performance and congestion management process. At this time, the platform does not provide information related to congestion on individual links and facilities. • **Reliability/Validity**: Validation results can be found mostly from the quality reports prepared by Replica when calibrating the activity-based model using ground truth data collected from each region. • **Coverage**: The aggregate-level Trends module covers the entire US, and the more detailed Places module covers select regions, including the Baltimore-Washington region which includes the TPB planning area. Replica provides estimated traffic volume by vehicle classes that can be used to help with system performance and congestion management process. At this time, the platform does not provide information related to congestion on individual links and facilities. • **Resource Requirements**: The data platform is straightforward to use for viewing, charting, tabulating and mapping data, developing standard data reports, and performing data analytics (e.g. select-link analysis). The data platform also supports the download of detailed data for custom applications. Although the size of some downloaded data files may be large, the files can be analyzed further in Excel, ArcGIS, and other commonly used software tools. • **Data Sharing**: Data licensing and sharing is flexible; annual subscription supports unlimited user licenses within an MPO if the MPO single access option is selected or unlimited user licenses within both an MPO and its member organizations with the MPO regional access option. Consultants working for the MPO are regarded as the extended MPO staff and also can be granted temporary license for data access. • **Cost**: Data cost is relatively reasonable when compared with similar data such as StreetLight Data. The total cost is based on the population of the largest city in the MPO region and the choice of either the single access option or regional access option. For TPB, the annual subscription cost is estimated to range between $104,000 and $173,000 for 2 years of full data access. • **Overall Recommendation**: No, at this time Replica should not be considered as an option for addressing the TPB's system performance and congestion management needs given that its data represents modeled trips for a synthetic population, and speed/congestion data is not available through the platform at this time.
Uber Movement	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: This dataset provides aggregated zone-to-zone travel time data derived from Uber trips freely available for the metropolitan Washingtonarea, but only at the TAZ (District of Columbia only) or census tract (roughly the area contained by the I-495 Capital Beltway) level. • **Overall Recommendation**: No, Uber Movement should not be considered as an option for addressing the TPB's system performance and congestion management needs. This dataset is not applicable to the research needs within this research area.
Taxi/TNC Trip Data	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: The District of Columbia has aggregated data for private vehicles-for-hire, such as taxis and TNCs, including Uber and Lyft, as information is required to be reported to DFHV and DDOT under D.C. law (§ 50–301.29a. General requirements for private vehicles-for-hire). • **Overall Recommendation**: No, Taxi/TNC trip data should not be considered as an option for addressing the TPB's system performance and congestion management needs. This dataset is not applicable to the research needs within this research area.
Strava Metro	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Strava Metro is an aggregator of bicycle and pedestrian movements using data obtained from the Strava user mobile app (for tracking bicycle/running/walking trips). Strava Metro is their product offering for planners and provides the relative level of activity (separated out into bicycle/pedestrian) along various facilities. • **Overall Recommendation**: No, Strava Metro should not be considered as an option for addressing the TPB's system performance and congestion management needs. This dataset is not applicable to the research needs within this research area.
Disaggregate Census Data	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: The US Census Bureau publishes large demographic datasets on travel behavior and population demographics. Decennial Census data (e.g., 2020 Census) is usually made available by the Census Bureau less than a year after the survey. Other major Census product uses are PUMS from ACS and CTPP. QCEW is a government data product used for a variety of purposes, such as economic security monitoring and labor statistic reporting. • **Overall Recommendation**: No, Disaggregate Census Data should not be considered as an option for addressing the TPB's system performance and congestion management needs. This dataset is not applicable to the research needs within this research area.

Big Data Evaluation Table | System Performance and Congestion Management

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
Household Travel Survey	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: HTS reveals insightful information with behavioral characteristics of travelers and the relationship between travel decisions and travelers' demographic backgrounds. Traditionally, HTS collects the data that pertains to household, person, vehicle, and trip characteristics on selected travel day(s) and usually is conducted every 5 to 10 years by MPOs, state DOTs, and FHWA through the NHTS program. All surveyed peer agencies stated the usage of household travel survey data to support travel demand modeling and transportation planning studies. • **Overall Recommendation**: No, HTS should not be considered as an option for addressing the TPB's system performance and congestion management needs. This dataset is not applicable to the research needs within this research area.
InfoUSA/Dun & BradStreet Business Listings	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: InfoUSA/Dun & Bradstreet business listing data is among the most commonly used data products to support the development of employment databases for both regional land use, travel demand modeling, and economic development activities. Both InfoUSA and Dun & Bradstreet data products provide detailed establishment location-based business information including employment size and industry sector. • **Overall Recommendation**: No, InfoUSA/Dun & Bradstreet business listings should not be considered as an option for addressing the TPB's system performance and congestion management needs. This dataset is not applicable to the research needs within this research area.
CoStar Data	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: CoStar is a database of commercial real estate transactions. CoStar data can be used to help understand the real estate development patterns in the region, provide input and parameters to land use modeling, and derive employment type and activities. • **Overall Recommendation**: No, CoStar Data should not be considered as an option for addressing the TPB's system performance and congestion management needs. This dataset is not applicable to the research needs within this research area.
Google Places	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Google Places provides information on location and type of places such as offices, parks, restaurants, and transit stops. This product offers an alternative approach to verify and supplement the business listing data acquired from the other sources (e.g., InfoUSA, Dun & Bradstreet, and QCEW), particularly about the presence, location, and type of the businesses, to better inform regional land use and travel demand models. • **Overall Recommendation**: No, Google Places should not be considered as an option for addressing the TPB's system performance and congestion management needs. This dataset is not applicable to the research needs within this research area.
Continuous Traffic Count Station/Sensor Data	+	+	+	+	+	+	+	Yes	• **Applicability**: Local and state DOTs typically have a series of permanent or temporary devices set up for collecting traffic counts, and likely vehicle classifications and spot speeds. From this data, agencies can estimate the average daily number of vehicles traversing roadway segments and vehicle miles traveled. • **Overall Recommendation**: Yes, Continuous Traffic Count Station/Sensor Data should be considered as an option for addressing the TPB's system performance and congestion management needs. The regional CMP includes congestion monitoring and data consolidation activities such as aggregation and consolidation of regional traffic count data. Aggregated traffic count data for the region is already being compiled by TPB and is provided on the RTDC.
Automated Traffic Signal Performance Measures (ATSPM)	+	+	-	-	-	+	-	Yes	• **Applicability**: An emerging field of Big Data analytics in the transportation industry involves performance measures for traffic signals based on high resolution (up to 1/10th second) event data. This data, such as detector calls, phase changes, and transit signal priority/preemption can be used to provide analytics on operations at individual signals or along entire arterial corridors. ATSPMs are emerging as an alternative to the traditional traffic engineering practice of ad-hoc turning movement count data collection, traffic signal timing, and arterial operations analysis. The research team is not aware of any agencies using ATSPM data for travel demand modeling applications. • **Overall Recommendation**: Yes, ATSPM data should be considered as an option for addressing the TPB's system performance and congestion management needs. However, at this time the research team notes the limited coverage within the region and likely significant processing demands required for any sort of integration in a system performance setting. The regional CMP includes an assessment of delay at signalized intersections. This assessment currently accounts for regional surveys of the status of signal optimization activities. As agencies migrate to upgraded signal controllers and improved communications, ATSPMs offer the opportunity for a more in-depth assessment of arterial delay and performance. Georgia DOT is currently using ATSPMs to assess corridor, zonal, and regional arterial performance in the Atlanta region.
Transit Data from On-Board ITS Devices	+	+	+	-	+	+	+	Yes	• **Applicability**: Transit agencies employ a variety of ITS devices for monitoring the status of their system and improving system performance. The most applicable sources of transit ITS data for travel demand modeling is O-D data from regional SmarTrip cards (WMATA maintains the SmarTrip farecard data for the region). This data can be used to understand O-D travel patterns across the Metrorail system as well as some (but not all) bus rides, including transfers to/from Metrorail. Boarding and alighting data for bus systems in the region also can be applicable to model validation but do not provide O-D information. WMATA farecard data provides O-Ds for all Metrorail trips as well as any bus-to-rail or rail-to-bus transfers using a SmarTrip card. Thus, O-D data is available for trips within the Metrorail system. Note that this data does not provide where riders are ultimately starting or ending their trips (the first-mile/last-mile problem). • **Reliability/Validity/Coverage**: O-D data for bus trips is much more challenging to obtain, as many agencies do not have APCs (or if they do, they do not have APCs calibrated/validated). Regardless, farebox and APC data provide information of where riders board (or also depart, in the case of APCs), but do not connect origins and destinations together; this data source does not provide first-mile/last-mile information. • **Resource Requirements**: Typically, the data from these systems is available via the vendor software for agencies to process further and analyze. However, depending on the age of the software and/or investment level by an agency, the data from these systems may be in a very raw format and require a significant amount of processing and analysis to gain insights. • **Data Sharing/Cost**: This data would be provided by partner agencies at their discretion.

Big Data Evaluation Table | System Performance and Congestion Management

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									• **Overall Recommendation**: Yes, Transit Data from On-Board ITS devices should be considered as an option for addressing the TPB's system performance and congestion management needs. The regional CMP includes an assessment of congestion on transit systems. Currently, this is assessed looking at travel speeds on transit-significant roads (using INRIX data for those roads) as well as average bus travel speeds from agency AVL data. However, likely given the processing challenges with combining AVL data from several sources, the most recent region-wide bus speeds data cited in the CMP is from 2011-2012. The CMP also assesses crowding on transit vehicles, including buses and rail. WMATA is able to provide this data for the Metrobus and Metrorail system using passenger boarding data and their own in-house analyses (for example, WMATA has an internal tool called the Line Load App to monitor passenger loads and crowdedness at Metrorail stations). Data from transit ITS systems is likely to continue to feed this region-wide analysis.
RITIS	+	+	N/A	+	+	+	+	Yes	• **Applicability**: RITIS from the CATT Lab at UMDis an "automated data fusion and dissemination system that provides an enhanced overall view of the transportation network. Participating agencies are able to view transportation and emergency management information through innovative visualizations and use it to improve their operations and emergency preparedness". RITIS ingests standardized data from outside sources, including mobile device data from providers such as INRIX, HERE, and TomTom as well as data on weather, traffic incidents, and other sources of data such as agency CCTV and detector feeds. • **Reliability/Validity**: The RITIS platform is used by numerous agencies to process a variety of Big Data products. The validity of the processed outputs are largely dependent on the underlying validity of the ingested Big Data products themselves. • **Coverage**: Coverage would be dependent on the data source acquired and provided to RITIS for analytics. • **Resource Requirements**: RITIS provides a user-friendly web-based interface for performing analytics. Result files can be downloaded from this interface as summary images or Excel workbooks. • **Data Sharing/Cost**: Access to various tools as well as the underlying data being processed by those tools, is dependent upon each state's or agency's contract with the CATT Lab. Use of RITIS is free-of-charge when purchasing more than $100,000 of INRIX data. • **Overall Recommendation**: Yes, RITIS should be considered as an option for addressing the TPB's system performance and congestion Management needs. This platform is already being used regionally to obtain link-level speed data for the CMP process. Should TPB continue to use INRIX data for this process, this data will likely be accessed through RITIS.
Moonshadow/DB4IoT	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Moonshadow is a provider of connected vehicle data analytics. Through a partnership with underlying providers such as INRIX and wejo, they have developed a database and visualization suite called DB4IoT. This product serves as both a backend (data storage/processing engine for raw data from vehicles and infrastructure) and a frontend (dashboard for visualizing data). Currently, DB4IoT ingests data from a variety of sources including: connected vehicle data (with underlying providers such as INRIX and wejo), mobile application LBS data (with underlying providers such as Unacast and X-Mode Social), and customer data (i.e., data feed unique to specific agencies, such as public transportation feeds, micromobility companies, WiFi/Bluetooth field data collection devices, and traffic counters). • **Overall Recommendation**: No, Moonshadow/DB4IoT should not be considered as an option for addressing the TPB's system performance and congestion management needs, as its end-user platform appears currently geared toward project-specific analyses, rather than system-wide performance measurement.
Swiftly	+	+	N/A	+	+	+	+	Yes	• **Applicability**: Swiftly is a vendor of transit analytics using feeds from agency systems; they claim to be "hardware-agnostic and software-centered" for consuming and analyzing feeds from various AVL providers. They assist public transit agencies in enhancing their transit service by analyzing on-time performance and identifying operational issues. • **Reliability/Validity**: In Baltimore, MD, Swiftly helped MTA improve on-time performance from around 65 percent averages in 2018 to 80 percent averages in early 2020. Their modules track transit headways, speeds, travel times, and runtimes by route, stop, day, and trip metrics that are monitored there routinely or in real-time. • **Coverage**: Coverage would be dependent on the transit data provided to Swiftly for analytics. • **Resource Requirements**: Swiftly provides a user-friendly dashboard that can be used to visualize analyzed data. • **Data Sharing/Cost**: An annual subscription to Swiftly could cost between $100,000 and $600,000, depending on the size of the agency's transit-fleet. Further discussion with sales representatives would be needed to determine an agency-specific quote and identify potential for data sharing. • **Overall Recommendation**: Yes, Swiftly should be considered as an option for addressing the TPB's system performance and congestion management needs. Tools that aggregate transit data across multiple agencies could be used to streamline analyses such as the CMP process for assessing region-wide bus speeds.
Moovit	+	+	N/A	+	+	+	+	Yes	• **Applicability**: Moovit is a mobility software company that offers both user-facing products (e.g., rider trip planning tools) and analytics for transit agencies. They offer two main products: MUMA and MaaS. Due to their multimodal focus, they have as much experience with MPOs as they do with transit agencies. • **Reliability/Validity**: In early 2020, Moovit announced a partnership with Cubic Transportation Systems, a company providing ticketing, fare collection, and management solutions. Through this partnership, Moovit will incorporate its MaaS application programming interfaces (APIs) with Cubic's Mobile Suite to offer an integrated traveler experience enabling Moovit's multimodal trip planning and Cubic's mobile payment and ticketing capabilities. They have been working with transit agencies to roll out this platform in Boston, Chicago, Los Angeles, Miami, New York, San

Big Data Evaluation Table | System Performance and Congestion Management									
Product	**Applicability to Research Area**	**Data Reliability and Validity**	**Data Coverage**	**Resource Requirements (Technical Staff and Training Requirements)**	**Resource Requirements (Technology, IT)**	**Data Sharing Restrictions**	**Cost**	**Overall Recommendation**	**Comments on Scoring** *** Details of subresearch area applicability are provided in Chapter 3**
									Francisco, and Washington, D.C. These upcoming implementation efforts offer case studies that could be used to inform platform reliability and validity. • **Coverage**: Coverage would be dependent on the transit data provided to Swiftly for analytics. • **Resource Requirements**: Moovit provides a user-friendly dashboard that can be used to visualize analyzed data. • **Data Sharing\Cost**: Due to their multimodal focus, they have as much experience with MPOs as they do with transit agencies. Though they favor annual contracts as revenue streams, they have mentioned in the past to be willing to do small pilot projects at no cost. Further discussion with sales representatives would be needed to determine an agency-specific quote and identify potential for data sharing. • **Overall Recommendation**: Yes, Moovit should be considered as an option for addressing the TPB's system performance and congestion management needs. Tools that aggregate transit data across multiple agencies could be used to streamline analyses such as the CMP process for assessing region-wide bus speeds.
Emerging Data Sources	N/A	N/A	N/A	N/A	N/A	N/A	N/A	**N/A**	• The emerging data sources evaluated as part of this study are non-exhaustive and geared towards the following research areas with limited data availability: transit and non-motorized travel, TNCs, and connected and automated vehicles.

Table B-4 | Big Data Evaluation Table for the Transit and Non-Motorized Travel Research Area

Big Data Evaluation Table | Transit and Non-Motorized Travel									
Product	**Applicability to Research Area**	**Data Reliability and Validity**	**Data Coverage**	**Resource Requirements (Technical Staff and Training Requirements)**	**Resource Requirements (Technology, IT)**	**Data Sharing Restrictions**	**Cost**	**Overall Recommendation**	**Comments on Scoring** *** Details of subresearch area applicability are provided in Chapter 3**
INRIX	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: INRIX is a provider of vehicle probe data for segment-level congestion analytics as well as O-D's for customized zones. INRIX has various product offerings: probe data (segment-level speed/travel time/AADT estimates), O-D summaries (trip starts/ends), and trip paths (includes waypoints along routes and is a very large dataset). • **Overall Recommendation**: No, INRIX should not be considered as an option for addressing the TPB's Transit and Non-Motorized Travel needs, as it does not currently provide any data for non-auto (car or truck) trips.
StreetLight	+	+	+	+	+	+	-	Yes	• **Applicability**: StreetLight Data is an online platform for O-D or segment-based analytics based on mobile device data. Similar to INRIX, StreetLight offers a suite of analysis tools: AADT estimation, O-D, O-D with middle filter (through a midpoint location), O-D to preset geography (e.g. TAZs, zip codes, census block groups), top routes between O-D's, and a congestion diagnostics tool for auto-generating insights. StreetLight recently began offering O-D analytics for bicycles and pedestrians. • **Reliability/Validity**: Various whitepapers are available on StreetLight's website for a variety of applications (e.g., validation of AADTs, turning movements). VDOT recently completed an evaluation of various products offered by StreetLight, including AADT, O-D estimates, traffic link volumes, turning movement volumes, and truck traffic. Evaluation results were mixed depending on the product used and volume levels being measured, with larger errors often associated with lower volumes and shorter time periods. This evaluation also provides a literature review of other recent validation efforts of StreetLight. A third-party validation recently completed by Fehr and Peers for hourly turning movement counts showed that 90 percent of locations were effectively replicated by StreetLight. • **Coverage**: StreetLight's underlying mobile data sources are mainly LBS-based (cell phone apps); INRIX is one of their underlying data sources. This mobile device data is integrated with underlying contextual data such as census demographics to provide additional insights. • **Resource Requirements**: StreetLight's online platform (StreetLight InSight) has a visualization feature for exploring and summarizing data. No raw data (individual trip data) is available; CSVs and shapefiles can be downloaded from the online platform. Excel and GIS (if desired) are typically sufficient for further analysis, although larger or more complex queries from StreetLight may require large CSVs to be processed via a scripting tool such as R or Python to be usable in Excel. • **Data Sharing**: A DOT with a subscription (such as VDOT, MDOT, DDOT) can grant access to underlying MPOs covered by that DOT. • **Cost**: StreetLight offers three different packages: Essentials, Advanced, and Multi-Mode. These packages can be purchased as a subscription or on a project-by-project basis. Subscription pricing is based on the population of the coverage area (e.g., VDOT's subscription to StreetLight data is estimated to cost more than $500,000). Additional cost would be expected to get the Multi-Mode package that includes bicycle and pedestrian metrics. • **Overall Recommendation**: Yes, StreetLight should be considered as an option for addressing the TPB's transit and non-motorized travel needs. This product can be used to estimate the amount of bicycle or pedestrian activity at a given location as well as the origins and destinations of that activity.
Teralytics	+	+	+	+	+	-	+	Yes	• **Applicability**: Teralytics is an online platform for O-D analytics based on cell phone tower triangulation data. • **Reliability/Validity**: Teralytics claim to have a less biased sample than LBS-based analytics providers as they "sit behind the firewall of all major mobile phone carriers"; they claim sufficient market share among all demographics, ethnic groups, income levels, and age groups. No validation whitepapers are provided on their website. Teralytics claims to have a high level of accuracy given its deep penetration rate, but all benchmarking appears to be done internally. One study was identified noting a limited validation effort of Teralytics "because Teralytics relies on a single data source with a considerably large penetration rate". This study showed a distribution of LRT trips by time-of-day estimated by Teralytics to be consistent with a "general understanding of transit trip-making patterns" for the LRT system. • **Coverage**: Cell tower triangulation has a high sample rate (estimated at 15-35 percent of population) but a lower spatial resolution. Thus, this data is not usable for route-level analysis (individual trips cannot be mapped to specific roads), but it is usable at the census tract or even TAZ level. Segment-level traffic count estimates (AADTs or turning movement counts) are not estimated via this platform. • **Resource Requirements**: Similar to StreetLight, data is accessed through an online visualization platform and viewed through the UI; data can also be downloaded via CSV and analyzed in Excel or GIS. • **Data Sharing**: No sharing of data is allowed outside of the purchasing agency except with consulting firms doing a project with that agency using the data. A purchasing agency can share derivatives (e.g., analysis results). A purchasing agency could negotiate a unique data sharing agreement as needed. • **Cost**: Generally, Teralytics is less expensive than products providing route-level analytics (e.g., INRIX and StreetLight) while providing a robust sample for understanding larger-scale travel patterns. Pricing is provided on an individual project basis or via subscription (e.g., one year of unlimited use and one year's worth of data). Pricing is based on the population of the coverage area. The estimated subscription cost for standard out of the box use of platform for an area with a population approximately the size of the metropolitan Washington,. MSA is approximately $50,000. Custom data sets for a population of the same size are likely in the $70,000-90,000 range; cost depends on the level of customization.

Big Data Evaluation Table | Transit and Non-Motorized Travel

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									• **Overall Recommendation**: Yes, Teralytics should be considered as an option for addressing the TPB's transit and non-motorized travel needs, as it is applicable to TPB's research need of understanding transit and rail demand in that it can separate out trips for modes such as long-distance rail, subway, and planes.
Locus (Cambridge Systematics)	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Locus is a product of LBS-based travel analytics product provided by Cambridge Systematics (consulting firm that has been the TPB's travel demand model developer). Four separate products are offered in addition to the custom analyses: O-D tables (expanded/validated by travel purpose/time-of-day), a transit competitiveness dashboard, a geofence analysis of activity around activity centers, and survey assist to supplement traditional HTS data. • **Overall Recommendation**: No, Locus should not be considered as an option for addressing the TPB's transit and non-motorized travel needs, as data is not broken out into various modes and simply represents all multimodal trips.
Replica	+	N/A	+	+	+	+	+	Yes	• **Applicability**: Replica is an online platform for aggregate-level mobility, economic activity, and COVID-19 trend data as well as detailed travel pattern data for select regions, including the Baltimore-Washington region which covers the TPB planning area. O-D data, represented by a synthetic population, is available for all major purposes (work, home, eat, shop, school, social, recreation, errands, lodging, pass-through, commercial and other) and modes (driving, auto passenger, taxi/TNC, transit, walk, bike and commercial vehicles). O-D data can be further filtered and partitioned based on several other variables, such as trip start time, distance and duration. The data for individual trips from the synthetic population data can be downloaded for post-processing outside of the Replica platform. Replica's synthetic population data address multimodal aspects of the transportation system and provide detailed travel pattern data and insights on transit and active travel modes. Transit, bicycle, and pedestrian O-Ds are also available for Census geographies, cities, counties and TAZs. However, the available data is not yet broken down by some "new" modes such as scooters. In addition to transit and active travel mode O-D data that can be partitioned based on demographic information of travelers and other variables as mentioned above, Replica provides public transit passenger boarding/alighting counts at stations/stops and onboard passenger counts. Similarly, transit boarding data can be classified or filtered based on demographic variables. These variables offer additional insights to understand the transit travel patterns and transit system usage of any targeted demographic groups on the entire transit network and for selected transit routes/stations/stops. • **Reliability/Validity**: Validation results can be found mostly from the quality reports prepared by Replica when calibrating the activity-based model using ground truth data collected from each region. • **Coverage**: The aggregate-level Trends module covers the entire US, and the more detailed Places module covers select regions, including the Baltimore-Washington region which includes the TPB planning area. • **Resource Requirements**: The data platform is straightforward to use for viewing, charting, tabulating and mapping data, developing standard data reports, and performing data analytics (e.g. select-link analysis). The data platform also supports the download of detailed data for custom applications. Although the size of some downloaded data files may be large, the files can be analyzed further in Excel, ArcGIS, and other commonly used software tools. • **Data Sharing**: Data licensing and sharing is flexible; annual subscription supports unlimited user licenses within an MPO if the MPO single access option is selected or unlimited user licenses within both an MPO and its member organizations with the MPO regional access option. Consultants working for the MPO are regarded as the extended MPO staff and also can be granted temporary license for data access. • **Cost**: Data cost is relatively reasonable when compared with similar data such as StreetLight Data. The total cost is based on the population of the largest city in the MPO region and the choice of either the single access option or regional access option. For TPB, the annual subscription cost is estimated to range between $104,000 and $173,000 for 2 years of full data access. • **Overall Recommendation**: Yes, Replica should be considered as an option for addressing the TPB's transit and active travel needs given its detailed modeling across a large number of transit and active travel modes and rich dataset supporting trips for these modes.
Uber Movement	+	+	+	+	+	+	+	Yes	• **Applicability**: This dataset provides aggregated zone-to-zone travel time data derived from Uber trips freely available for the metropolitan Washington area, but only at the TAZ (District of Columbia only) or census tract (roughly the area contained by the I-495 Capital Beltway) level. Uber recently began offering a beta mobility heatmap product showing color-coded links throughout the District of Columbia and Arlington County based on the density of Uber-owned mobility devices (e.g. Uber Jump bikes and scooters). • **Reliability/Validity**: Data solely is based on Uber or Uber-device (e.g. Jump bikes and scooters) trips. • **Coverage**: Coverage is limited to the District of Columbia and Arlington County for the mobility heatmap. • **Resource Requirements**: Data can be viewed through the online platform. • **Data Sharing/Cost**: Freely available online. • **Overall Recommendation**: Yes, Uber Movement should be considered as an option for addressing the TPB's transit and non-motorized travel needs. The mobility heatmap can help provide an improved understanding of the concentration and relative magnitude of mobility devices within the metropolitan Washington area. Mobility heatmap data could be explored to highlight demand of micromobility activities and opportunities for improved multimodal facilities if overlaying with land use and other transportation network data.
Taxi/TNC Trip Data	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: The District of Columbia has aggregated data for private vehicles-for-hire, such as taxis and TNCs, including Uber and Lyft, as information is required to be reported to DFHV and DDOT under D.C. law (§ 50–301.29a. General requirements for private vehicles-for-hire).

Big Data Evaluation Table | Transit and Non-Motorized Travel

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									• **Overall Recommendation**: No, Taxi/TNC Trip Data should not be considered as an option for addressing the TPB's transit and non-motorized travel needs. This dataset is not applicable to the research needs within this research area.
Strava Metro	+	+	+	+	+	+	+	Yes	• **Applicability**: Strava Metro is an aggregator of bicycle and pedestrian movements using data obtained from the Strava user mobile app (for tracking bicycle/running/walking trips). Strava Metro is their product offering for planners and provides the relative level of activity (separated out into bicycle/pedestrian) along various facilities. • **Reliability/Validity**: While their dataset is derived from users of their mobile app, they offer a whitepaper on their website claiming to have a representative sample of the overall population for counts/route choice[1]. However, a few of the agencies interviewed as part of this evaluation raised concern with the representativeness of Strava application users; suggesting there may be an over-representation of recreational cyclists over commuting cyclists. • **Coverage**: Coverage is provided for any facility on which people bike or walk. • **Resource Requirements**: Strava Metro requires no GIS expertise to access and use; however, data downloads for professionals familiar with GIS software are available, depending on the purchased package. A demo dataset with representative data from Denver, CO is available on their website. • **Data Sharing/Cost**: Data sharing and cost is unclear to the research team at this time. Agencies can apply for access to the online platform via a form on Strava's website. • **Overall Recommendation**: Yes, Strava Metro should be considered as an option for addressing the TPB's transit and non-motorized travel needs. Separate datasets are provided for bicycle and pedestrian activity. Strava claims to provide a representative sample of the overall national population for counts/route choice. This dataset could be useful for investment prioritization and understanding bicycle and pedestrian demand.
Disaggregate Census Data	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: The US Census Bureau publishes large demographic datasets on travel behavior and population demographics. Decennial Census data (e.g., 2020 Census) is usually made available by the Census Bureau less than a year after the survey. Other major Census product uses are PUMS from ACS and CTPP. QCEW is a government data product used for a variety of purposes, such as economic security monitoring and labor statistic reporting. • **Overall Recommendation**: No, Disaggregate Census Data should not be considered as an option for addressing the TPB's transit and non-motorized travel needs. This dataset is not applicable to the research needs within this research area.
Household Travel Survey	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: HTS reveals insightful information with behavioral characteristics of travelers and the relationship between travel decisions and travelers' demographic backgrounds. Traditionally, HTS collects the data that pertains to household, person, vehicle, and trip characteristics on selected travel day(s) and usually is conducted every 5 to 10 years by MPOs, state DOTs, and FHWA through the NHTS program. All surveyed peer agencies stated the usage of household travel survey data to support travel demand modeling and transportation planning studies. • **Overall Recommendation**: No, HTS should not be considered as an option for addressing the TPB's transit and non-motorized travel needs.
InfoUSA Dun & BradStreet Business Listings	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: InfoUSA/Dun & Bradstreet business listing data is among the most commonly used data products to support the development of employment databases for both regional land use, travel demand modeling, and economic development activities. Both InfoUSA and Dun & Bradstreet data products provide detailed establishment location-based business information including employment size and industry sector. • **Overall Recommendation**: No, InfoUSA/Dun & Bradstreet Business Listings should not be considered as an option for addressing the TPB's transit and non-motorized travel needs. This dataset is not applicable to the research needs within this research area.
CoStar Data	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: CoStar is a database of commercial real estate transactions. CoStar data can be used to help understand the real estate development patterns in the region, provide input and parameters to land use modeling, and derive employment type and activities. • **Overall Recommendation**: No, CoStar Data should not be considered as an option for addressing the TPB's transit and non-motorized travel needs. This dataset is not applicable to the research needs within this research area.
Google Places	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Google Places provides information on location and type of places such as offices, parks, restaurants, and transit stops. This product offers an alternative approach to verify and supplement the business listing data acquired from other sources (e.g., InfoUSA, Dun & Bradstreet, and QCEW), particularly about the presence, location, and type of the businesses to better inform regional land use and travel demand models. • **Overall Recommendation**: No, Google Places should not be considered as an option for addressing the TPB's transit and non-motorized travel needs. This dataset is not applicable to the research needs within this research area.
Continuous Traffic Count Station/Sensor Data	+	+	+	+	+	+	+	Yes	• **Applicability**: Local and state DOTs typically have a series of permanent or temporary devices set up for collecting traffic counts, and likely, vehicle classifications and spot speeds. From this data, agencies can estimate the average daily number of vehicles traversing roadway segments and vehicle miles traveled. • **Overall Recommendation**: Yes, Continuous Traffic Count Station/Sensor Data should be considered as an option for addressing the TPB's transit and non-motorized travel needs.

[1] https://cdn2.hubspot.net/hubfs/1979139/Bike%20Counter%20Correlation.pdf

Big Data Evaluation Table | Transit and Non-Motorized Travel

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
Automated Traffic Signal Performance Measures (ATSPM)	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: An emerging field of Big Data analytics in the transportation industry involves performance measures for traffic signals based on high resolution (up to 1/10th second) event data. This data, such as detector calls, phase changes, and transit signal priority/preemption can be used to provide analytics on operations at individual signals or along entire arterial corridors. ATSPMs are emerging as an alternative to the traditional traffic engineering practice of ad-hoc turning movement count data collection, traffic signal timing, and arterial operations analysis. The research team is not aware of any agencies using ATSPM data for travel demand modeling applications . • **Overall Recommendation**: No, ATSPM Data should not be considered as an option for addressing the TPB's transit and non-motorized travel needs. This dataset is not applicable to the research needs within this research area.
Transit Data from On-Board ITS Devices	+	+	+	-	+	+	+	Yes	• **Applicability**: Transit agencies employ a variety of ITS devices for monitoring the status of their system and improving system performance. The most applicable sources of transit ITS data for travel demand modeling is O-D data from regional SmarTrip cards (WMATA maintains the SmarTrip farecard data for the region). This data can be used to understand O-D travel patterns across the Metrorail system as well as some (but not all) bus rides, including transfers to/from Metrorail. Boarding and alighting data for bus systems. In the metropolitan Washington region it also can be applicable to model validation but does not provide O-D information. WMATA farecard data provides O-Ds for all Metrorail trips as well as any bus-to-rail or rail-to-bus transfers using a SmarTrip card. Thus, O-D data is available for trips within the Metrorail system. Note that this data does not provide where riders are ultimately starting or ending their trips (the first-mile/last-mile problem). • **Reliability/Validity/Coverage**: O-D data for bus trips is much more challenging to obtain, as many agencies do not have APCs (or if they do, they do not have APCs calibrated/validated). Regardless, farebox and APC data provide information of where riders board (or also depart, in the case of APCs), but do not connect origins and destinations together; this data source does not provide first-mile/last-mile information. • **Resource Requirements**: Typically, the data from these systems is available via the vendor software for agencies to process further and analyze. However, depending on the age of the software and/or investment level by an agency, the data from these systems may be in a very raw format and require a significant amount of processing and analysis to gain insights. • **Data Sharing/Cost**: This data would be provided by partner agencies at their discretion. • **Overall Recommendation**: Yes, Transit Data from On-Board ITS devices should be considered as an option for addressing the TPB's transit and non-motorized travel needs. There are several different datasets that would be applicable for this research area. TPB has used aggregated AVL data in the past to understand regional bus travel speeds and hotspots as part of its congestion management process (CMP). Across the region, WMATA SmarTrip card data can be used to understand O-D patterns across the Metrorail system as well as some (but not all) bus rides, including transfers to/from Metrorail. APC data can be used to monitor bus demand, including intercity bus ridership. In a planning context, APC data can be strengthened when joined to AVL data, allowing for calculations of vehicle load.
RITIS	+	+	N/A	+	+	+	+	Yes	• **Applicability**: RITIS from the CATT Lab at UMD is an “automated data fusion and dissemination system that provides an enhanced overall view of the transportation network. Participating agencies are able to view transportation and emergency management information through innovative visualizations and use it to improve their operations and emergency preparedness”. RITIS ingests standardized data from outside sources, including mobile device data from providers such as INRIX, HERE, and TomTom as well as data on weather, traffic incidents, and other sources of data such as agency CCTV and detector feeds. • **Reliability/Validity**: The RITIS platform is used by numerous agencies to process a variety of Big Data products. The validity of the processed outputs are largely dependent on the underlying validity of the ingested Big Data products themselves. • **Coverage**: Coverage would be dependent on the data source acquired and provided to RITIS for analytics. • **Resource Requirements**: RITIS provides a user-friendly web-based interface for performing analytics. Result files can be downloaded from this interface as summary images or Excel workbooks. • **Data Sharing/Cost**: Access to various tools, as well as the underlying data being processed by those tools, is dependent upon each state’s or agency’s contract with the CATT Lab. Use of RITIS is free-of-charge when purchasing more than $100,000 of INRIX data. • **Overall Recommendation**: Yes, RITIS should be considered as an option for addressing the TPB's transit and non-motorized travel needs. RITIS could potentially be used to ingest transit data from the region; while the research team is unaware of any transit-specific tools within the platform, many users throughout the region have familiarity with the platform.
Moonshadow/DB4IoT	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Moonshadow is a provider of connected vehicle data analytics. Through a partnership with underlying providers such as INRIX and wejo, they have developed a database and visualization suite called DB4IoT. This product serves as both a backend (data storage/processing engine for raw data from vehicles and infrastructure) and a frontend (dashboard for visualizing data). Currently, DB4IoT ingests data from a variety of sources including: connected vehicle data (with underlying providers such as INRIX and wejo), mobile application LBS data (with underlying providers such as Unacast and X-Mode Social), and customer data (i.e., data feed unique to specific agencies, such as public transportation feeds, micromobility companies, WiFi/Bluetooth field data collection devices, and traffic counters). • **Overall Recommendation**: No, Moonshadow/DB4IoT should not be considered as an option for addressing the TPB’s transit and non-motorized travel needs. No modal breakdown is provided at this time.
Swiftly	+	+	N/A	+	+	+	+	Yes	• **Applicability**: Swiftly is a vendor of transit analytics using feeds from agency systems; they claim to be “hardware-agnostic and software-centered” for consuming and analyzing feeds from various AVL providers. They assist public transit agencies in enhancing their transit service by analyzing on-time performance and identifying operational issues.

Big Data Evaluation Table | Transit and Non-Motorized Travel

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									• **Reliability/Validity**: In Baltimore, MD, Swiftly helped MTA improve on-time performance from around 65 percent averages in 2018 to 80 percent averages in early 2020. Their modules track transit headways, speeds, travel times, and runtimes by route, stop, day, and trip metrics that are monitored there routinely or in real-time. • **Coverage**: Coverage would be dependent on the transit data provided to Swiftly for analytics. • **Resource Requirements**: Swiftly provides a user-friendly dashboard that can be used to visualize analyzed data. • **Data Sharing/Cost**: An annual subscription to Swiftly could cost between $100,000 and $600,000, depending on the size of the agency's transit-fleet. Further discussion with sales representatives would be needed to determine an agency-specific quote and identify potential for data sharing. • **Overall Recommendation**: Yes, Swiftly should be considered as an option for addressing the TPB's transit and non-motorized travel needs. Their modules track transit headways, speeds, travel times, and runtimes by route, stop, day, and trip metrics that are monitored there routinely or in real-time.
Moovit	+	+	N/A	+	+	+	+	Yes	• **Applicability**: Moovit is a mobility software company that offers both user-facing products (e.g., rider trip planning tools) and analytics for transit agencies. They offer two main products: MUMA and MaaS. Due to their multimodal focus, they have as much experience with MPOs as they do with transit agencies. • **Reliability/Validity**: In early 2020, Moovit announced a partnership with Cubic Transportation Systems, a company providing ticketing, fare collection, and management solutions. Through this partnership, Moovit will incorporate its MaaS APIs with Cubic's Mobile Suite to offer an integrated traveler experience enabling Moovit's multimodal trip planning and Cubic's mobile payment and ticketing capabilities. They have been working with transit agencies to roll out this platform in Boston, Chicago, Los Angeles, Miami, New York, San Francisco, and Washington, D.C. These upcoming implementation efforts offer case studies that could be used to inform platform reliability and validity. • **Coverage**: Coverage would be dependent on the transit data provided to Swiftly for analytics. • **Resource Requirements**: Moovit provides a user-friendly dashboard that can be used to visualize analyzed data. • **Data Sharing\Cost**: Due to their multimodal focus, they have as much experience with MPOs as they do with transit agencies. Though they favor annual contracts as revenue streams, they have mentioned in the past to be willing to do small pilot projects at no cost. Further discussion with sales representatives would be needed to determine an agency-specific quote and identify potential for data sharing. • **Overall Recommendation**: Yes, Moovit should be considered as an option for addressing the TPB's transit and non-motorized travel needs. This data could enhance the understanding of how alternative commuting modes or change in commuting behavior (e.g., bike, walk, transit, rideshare, car/vanpool, and teleworking) affect the overall network with their focus on integrating real-time traffic, transit, and rail demand to one platform to provide insights into multimodal mobility.
Emerging Data Sources	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Yes	• The emerging data sources evaluated as part of this study are non-exhaustive and geared towards the following research areas with limited data availability: transit and non-motorized travel, TNCs, and connected and automated vehicles.

Table B-5 | Big Data Evaluation Table for the Transportation Network Companies Research Area

Big Data Evaluation Table | Transportation Network Companies									
Product	**Applicability to Research Area**	**Data Reliability and Validity**	**Data Coverage**	**Resource Requirements (Technical Staff and Training Requirements)**	**Resource Requirements (Technology, IT)**	**Data Sharing Restrictions**	**Cost**	**Overall Recommendation**	**Comments on Scoring** *** Details of subresearch area applicability are provided in Chapter 3**
INRIX	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: INRIX is a provider of vehicle probe data for segment-level congestion analytics as well as O-D's for customized zones. INRIX has various product offerings: probe data (segment-level speed/travel time/AADT estimates), O-D summaries (trip starts/ends), and trip paths (includes waypoints along routes and is a very large dataset). • **Overall Recommendation**: No, INRIX should not be considered as an option for addressing the TPB's transportation network companies research area needs. It is not able to provide a breakdown for modes beyond cars and trucks.
StreetLight	+	+	+	+	+	+	-	Yes	• **Applicability**: StreetLight Data is an online platform for O-D or segment-based analytics based on mobile device data. Similar to INRIX, StreetLight offers a suite of analysis tools: AADT estimation, O-D, O-D with middle filter (through a midpoint location), O-D to preset geography (e.g. TAZs, zip codes, census block groups), top routes between O-D's, and a congestion diagnostics tool for auto-generating insights. • **Reliability/Validity**: Various whitepapers are available on StreetLight's website for a variety of applications (e.g., validation of AADTs, turning movements). VDOT recently completed an evaluation of various products offered by StreetLight, including AADT, O-D estimates, traffic link volumes, turning movement volumes, and truck traffic. Evaluation results were mixed depending on the product used and volume levels being measured, with larger errors often associated with lower volumes and shorter time periods. This evaluation also provides a literature review of other recent validation efforts of StreetLight. A third-party validation recently completed by Fehr and Peers for hourly turning movement counts showed that 90 percent of locations were effectively replicated by StreetLight. • **Coverage**: StreetLight's underlying mobile data sources are mainly LBS-based (cell phone apps); INRIX is one of their underlying data sources. This mobile device data is integrated with underlying contextual data such as census demographics to provide additional insights. • **Resource Requirements**: StreetLight's online platform (StreetLight InSight) has a visualization feature for exploring and summarizing data. No raw data (individual trip data) is available; CSVs and shapefiles can be downloaded from the online platform. Excel and GIS (if desired) are typically sufficient for further analysis, although larger or more complex queries from StreetLight may require large CSVs to be processed via a scripting tool such as R or Python to be usable in Excel. • **Data Sharing**: A DOT with a subscription (such as VDOT, MDOT, DDOT) can grant access to underlying MPOs covered by that DOT. • **Cost**: StreetLight offers three different packages: Essentials, Advanced, and Multi-Mode. These packages can be purchased as a subscription or on a project-by-project basis. Subscription pricing is based on the population of the coverage area (e.g., VDOT's subscription to StreetLight data is estimated to cost more than $500,000). • **Overall Recommendation**: Yes, StreetLight should be considered as an option for addressing the TPB's transportation network companies research area needs. Customized analysis could provide data to infer the magnitude and the O-D patterns of TNC trips. Such information may be used to understand how TNCs are influencing vehicle miles traveled and the relationship with transit use.
Teralytics	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Teralytics is an online platform for O-D analytics based on cell phone tower triangulation data. • **Overall Recommendation**: No, Teralytics should not be considered as an option for addressing the TPB's transportation network companies research area needs. It is not able to provide a breakdown of this specific mode.
Locus (Cambridge Systematics)	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Locus is a product of LBS-based travel analytics product provided by Cambridge Systematics (consulting firm that has been the TPB's travel demand model developer). Four separate products are offered in addition to the custom analyses: O-D tables (expanded/validated by travel purpose/time-of-day), a transit competitiveness dashboard, a geofence analysis of activity around activity centers, and survey assist to supplement traditional HTS data. • **Overall Recommendation**: No, Locus should not be considered as an option for addressing the TPB's transportation network companies research area needs. It is not able to provide a breakdown of this specific mode.
Replica	+	N/A	+	+	+	+	+	Yes	• **Applicability**: Replica is an online platform for aggregate-level mobility, economic activity, and COVID-19 trend data as well as detailed travel pattern data for select regions, including the Baltimore-Washington region which covers the TPB planning area. O-D data, represented by a synthetic population, is available for all major purposes (work, home, eat, shop, school, social, recreation, errands, lodging, pass-through, commercial and other) and modes (driving, auto passenger, taxi/TNC, transit, walk, bike and commercial vehicles). O-D data can be further filtered and partitioned based on several other variables, such as trip start time, distance and duration. The data for individual trips from the synthetic population data can be downloaded for post-processing outside of the Replica platform. • **Reliability/Validity**: Validation results can be found mostly from the quality reports prepared by Replica when calibrating the activity-based model using ground truth data collected from each region. Although Replica uses taxi/TNC as a separate mode for travel pattern data, the data accuracy and applicability may need to go through a more rigorous data validation from independent data users. • **Coverage**: The aggregate-level Trends module covers the entire US, and the more detailed Places module covers select regions, including the Baltimore-Washington region which includes the TPB planning area. • **Resource Requirements**: The data platform is straightforward to use for viewing, charting, tabulating and mapping data, developing standard data reports, and performing data analytics (e.g. select-link analysis). The data platform also supports the download of detailed data for custom applications. Although the size of some downloaded data files may be large, the files can be analyzed further in Excel, ArcGIS, and other commonly used software tools.

Big Data Evaluation Table | Transportation Network Companies

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									• **Data Sharing**: Data licensing and sharing is flexible; annual subscription supports unlimited user licenses within an MPO if the MPO single access option is selected or unlimited user licenses within both an MPO and its member organizations with the MPO regional access option. Consultants working for the MPO are regarded as the extended MPO staff and also can be granted temporary license for data access. • **Cost**: Data cost is relatively reasonable when compared with similar data such as StreetLight Data. The total cost is based on the population of the largest city in the MPO region and the choice of either the single access option or regional access option. For TPB, the annual subscription cost is estimated to range between $104,000 and $173,000 for 2 years of full data access. • **Overall Recommendation**: Yes, Replica should be considered as an option for addressing the TPB's transportation network companies research area needs given its ability to model this mode and its ability to view the synthetic population trip data and understand trips on this mode. However, given the novelty of the product, the data accuracy and applicability may need to go through a more rigorous data validation from independent data users.
Uber Movement	+	+	-	+	+	+	+	Yes	• **Applicability**: This dataset provides aggregated zone-to-zone travel time data derived from Uber trips freely available for the Washington, D.C., area, but only at the TAZ (District of Columbia only) or census tract (roughly the area contained by the I-495 Capital Beltway) level. • **Reliability/Validity**: Data solely is based on Uber trips. • **Coverage**: Coverage is limited to the District of Columbia for data at the TAZ level or the geographic area roughly within the I-495 Capital Beltway for data at the census tract level. • **Resource Requirements**: Data can be viewed through the online platform. • **Data Sharing / Cost**: Freely available online. • **Overall Recommendation**: Yes, Uber Movement should be considered as an option for addressing the TPB's transportation network companies research area needs. Note that this dataset currently only provides travel time information and does not provide quantification of the number of trips, in addition to the geographic coverage limitations described above.
Taxi/TNC Trip Data	+	+	+	+	+	+	+	Yes	• **Applicability**: The District of Columbia has aggregated data for private vehicles-for-hire, such as taxis and TNCs, including Uber and Lyft, as information is required to be reported to DFHV and DDOT under D.C. law (§ 50–301.29a. General requirements for private vehicles-for-hire). • **Reliability/Validity**: Data solely is reported for any taxi or TNC trip starting or ending in the District of Columbia. • **Coverage**: Coverage is limited to the District of Columbia. Origin and destination points of trips are provided. • **Resource Requirements**: Additional data processing is required to visualize and analyze raw trip data. The D.C. government has been developing a data lake for integrating data from various feeds. • **Data Sharing/Cost**: This data is being provided to the D.C. government per legal requirements. WMATA and COG can request access to the data for specific purposes that are agreed upon between these agencies and the District government • **Overall Recommendation**: Yes, Taxi/TNC Trip Data should be considered as an option for addressing the TPB's transportation network companies research area needs.
Strava Metro	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Strava Metro is an aggregator of bicycle and pedestrian movements using data obtained from the Strava user mobile app (for tracking bicycle/running/walking trips). Strava Metro is their product offering for planners and provides the relative level of activity (separated out into bicycle/pedestrian) along various facilities. • **Overall Recommendation**: No, Strava Metro should not be considered as an option for addressing the TPB's transportation network companies research area needs. This dataset is not applicable to the research needs within this research area.
Disaggregate Census Data	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: The US Census Bureau publishes large demographic datasets on travel behavior and population demographics. Decennial Census data (e.g., 2020 Census) is usually made available by the Census Bureau less than a year after the survey. Other major Census product uses are PUMS from ACS and CTPP. QCEW is a government data product used for a variety of purposes, such as economic security monitoring and labor statistic reporting. • **Overall Recommendation**: No, Disaggregate Census Data should not be considered as an option for addressing the TPB's transportation network companies research area needs. This dataset is not applicable to the research needs within this research area.
Household Travel Survey	+	+	+	+	+	+	+	Yes	• **Applicability**: HTS reveals insightful information with behavioral characteristics of travelers and the relationship between travel decisions and travelers' demographic backgrounds. Traditionally, HTS collects the data that pertains to household, person, vehicle, and trip characteristics on selected travel day(s) and usually is conducted every 5 to 10 years by MPOs, state DOTs, and FHWA through the NHTS program. All surveyed peer agencies stated the usage of household travel survey data to support travel demand modeling and transportation planning studies. • **Reliability/Validity**: NHTS provides a wealth of information regarding national household travel survey through the published survey data, reports, and other documentations. MAG, Tampa, FL; Baton Rouge, LA; along with several others recently conducted GPS-assisted household travel surveys. MAG's household travel survey report describes in much detail how the survey was administered, and how data was collected and processed for the Phoenix area in Arizona. All surveyed peer agencies stated the usage of HTS data to support travel demand modeling and transportation planning studies. • **Coverage**: Coverage is dependent upon the survey and agency conducting the survey. VDOT will participate in the add-on program of the first NextGen HTS to collect additional 10,000 household samples throughout the state of Virginia, including Northern Virginia in the metropolitan Washington. region.

Big Data Evaluation Table | Transportation Network Companies

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									• **Resource Requirements**: HTS data usually comes with the sample expansion factors for households, persons, and trips that should be applied to represent the regional travel pattern. Other data processing requirements are all analysis specific. These analyses can be performed with traditional data processing tools. • **Data Sharing**: Agreements can be made to share survey data from the agency initiating the survey. • **Cost**: HTS data can be collected either through participating in the add-on program of NHTS at a unit price of $200-$250 per completed household sample or by other data collection contractors with the cost ranging in general between $150-$350 per sample. • **Overall Recommendation**: Yes, HTS should be considered as an option for addressing the TPB's transportation network companies research area needs, as it provides TNC rider demographic information and trip purpose information that is not accessible from the previously discussed private vehicle-for-hire data.
InfoUSA Dun & BradStreet Business Listings	-	N/A	N/A	N/A	N/A	N/A	N/A	**No**	• **Applicability**: InfoUSA/Dun & Bradstreet business listing data is among the most commonly used data products to support the development of employment databases for both regional land use, travel demand modeling, and economic development activities. Both InfoUSA and Dun & Bradstreet data products provide detailed establishment location-based business information including employment size and industry sector. • **Overall Recommendation**: No, InfoUSA Dun & Bradstreet Business Listings should not be considered as an option for addressing the TPB's Transportation Network Companies research area needs. This dataset is not applicable to the research needs within this research area.
CoStar Data	-	N/A	N/A	N/A	N/A	N/A	N/A	**No**	• **Applicability**: CoStar is a database of commercial real estate transactions. CoStar data can be used to help understand the real estate development patterns in the region, provide input and parameters to land use modeling, and derive employment type and activities. • **Overall Recommendation**: No, CoStar Data should not be considered as an option for addressing the TPB's transportation network companies research area needs. This dataset is not applicable to the research needs within this research area.
Google Places	-	N/A	N/A	N/A	N/A	N/A	N/A	**No**	• **Applicability**: Google Places provides information on location and type of places such as offices, parks, restaurants, and transit stops. This product offers an alternative approach to verify and supplement the business listing data acquired from the other sources (e.g., InfoUSA, Dun & Bradstreet, and QCEW), particularly about the presence, location, and type of businesses, to better inform regional land use and travel demand models. • **Overall Recommendation**: No, Google Places should not be considered as an option for addressing the TPB's transportation network companies research area needs. This dataset is not applicable to the research needs within this research area.
Continuous Traffic Count Station/Sensor Data	-	N/A	N/A	N/A	N/A	N/A	N/A	**No**	• **Applicability**: Local and state DOTs typically have a series of permanent or temporary devices set up for collecting traffic counts, and likely, vehicle classifications and spot speeds. From this data, agencies can estimate the average daily number of vehicles traversing roadway segments and vehicle miles traveled. • **Overall Recommendation**: No, Continuous Traffic Count Station/Sensor Data should not be considered as an option for addressing the TPB's transportation network companies research area needs. This dataset is not applicable to the research needs within this research area.
Automated Traffic Signal Performance Measures (ATSPM)	-	N/A	N/A	N/A	N/A	N/A	N/A	**No**	• **Applicability**: An emerging field of Big Data analytics in the transportation industry involves performance measures for traffic signals based on high resolution (up to 1/10th second) event data. This data, such as detector calls, phase changes, and transit signal priority/preemption can be used to provide analytics on operations at individual signals or along entire arterial corridors. ATSPMs are emerging as an alternative to the traditional traffic engineering practice of ad-hoc turning movement count data collection, traffic signal timing, and arterial operations analysis. The research team is not aware of any agencies using ATSPM data for travel demand modeling applications. • **Overall Recommendation**: ATSPM data should not be considered as an option for addressing the TPB's transportation network companies research area needs. This dataset is not applicable to the research needs within this research area.
Transit Data from On-Board ITS Devices	-	N/A	N/A	N/A	N/A	N/A	N/A	**No**	• **Applicability**: Transit agencies employ a variety of ITS devices for monitoring the status of their system and improving system performance. The most applicable sources of transit ITS data for travel demand modeling is O-D data from regional SmarTrip cards (WMATA maintains the SmarTrip farecard data for the region). Therefore, this data can be used to understand O-D travel patterns across the Metrorail system as well as some (but not all) bus rides, including transfers to/from Metrorail. Boarding and alighting data for bus systems in the region can also be applicable to model validation but do not provide O-D information. WMATA farecard data provides O-D's for all Metrorail trips as well as any bus-to-rail or rail-to-bus transfers using a SmarTrip card. Thus, O-D data is available for trips within the Metrorail system. Note that this data does not provide where riders are ultimately starting or ending their trips (the first-mile/last-mile problem). • **Overall Recommendation**: No, Transit Data from On-Board ITS devices should not be considered as an option for addressing the TPB's transportation network companies research area needs. This dataset is not applicable to the research needs within this research area.
RITIS	-	N/A	N/A	N/A	N/A	N/A	N/A	**No**	• **Applicability**: RITIS from the CATT Lab at UMD is an "automated data fusion and dissemination system that provides an enhanced overall view of the transportation network. Participating agencies are able to view transportation and emergency management information through innovative visualizations and use it to improve their operations and emergency preparedness". RITIS ingests standardized data from outside sources, including mobile device data from providers such as INRIX, HERE, and TomTom as well as data on weather, traffic incidents, and other sources of data such as agency CCTV and detector feeds. • **Overall Recommendation**: No, RITIS should not be considered as an option for addressing the TPB's transportation network companies research area needs. This dataset is not applicable to the research needs within this research area.

Big Data Evaluation Table | Transportation Network Companies

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
Moonshadow/DB4IoT	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Moonshadow is a provider of connected vehicle data analytics. Through a partnership with underlying providers such as INRIX and wejo, they have developed a database and visualization suite called DB4IoT. This product serves as both a backend (data storage/processing engine for raw data from vehicles and infrastructure) and a frontend (dashboard for visualizing data). Currently DB4IoT ingests data from a variety of sources including: connected vehicle data (with underlying providers such as INRIX and wejo), mobile application LBS data (with underlying providers such as Unacast and X-Mode Social), and customer data (i.e., data feed unique to specific agencies, such as public transportation feeds, micromobility companies, WiFi/Bluetooth field data collection devices, and traffic counters). • **Overall Recommendation**: No, Moonshadow/DB4IoT should not be considered as an option for addressing the TPB's transportation network companies research area needs. This dataset is not applicable to the research needs within this research area.
Swiftly	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Swiftly is a vendor of transit analytics using feeds from agency systems; they claim to be "hardware-agnostic and software-centered" for consuming and analyzing feeds from various AVL providers. They assist public transit agencies in enhancing their transit service by analyzing on-time performance and identifying operational issues. • **Overall Recommendation**: No, Swiftly should not be considered as an option for addressing the TPB's transportation network companies research area needs. This dataset is not applicable to the research needs within this research area.
Moovit	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Moovit is a mobility software company that offers both user-facing products (e.g., rider trip planning tools) and analytics for transit agencies. They offer two main products: MUMA and MaaS. Due to their multimodal focus, they have as much experience with MPOs as they do with transit agencies. • **Overall Recommendation**: No, Moovit should not be considered as an option for addressing the TPB's transportation network companies research area needs. This dataset is not applicable to the research needs within this research area.
Emerging Data Sources	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Yes	• The emerging data sources evaluated as part of this study are non-exhaustive and geared towards the following research areas with limited data availability: transit and non-motorized travel, TNCs, and connected and automated vehicles.

Table B-6 | Big Data Evaluation Table for the Traffic Counts Research Area

Big Data Evaluation Table | Traffic Counts									
Product	**Applicability to Research Area**	**Data Reliability and Validity**	**Data Coverage**	**Resource Requirements (Technical Staff and Training Requirements)**	**Resource Requirements (Technology, IT)**	**Data Sharing Restrictions**	**Cost**	**Overall Recommendation**	**Comments on Scoring** *** Details of sub-research area applicability are provided in Chapter 3**
INRIX	+	+	+	+	+	+	-	Yes	• **Applicability**: INRIX is a provider of vehicle probe data for segment-level congestion analytics as well as O-D's for customized zones. INRIX has various product offerings: probe data (segment-level speed/travel time/AADT estimates), O-D summaries (trip starts/ends), and trip paths (includes waypoints along routes and is a very large dataset). • **Reliability/Validity**: INRIX is the underlying data source for NPMRDS, which provides region-wide travel speeds and volume estimates for NHS facilities. The data product has undergone numerous validation efforts and is widely accepted in the industry. The ETC also has conducted extensive validation of various INRIX products, most notably their validations of travel speeds. They are currently sponsoring ongoing efforts to validate INRIX's ubiquitous traffic volume data. • **Coverage**: The underlying data sources are a combination of LBS, GPS from local delivery fleets and long-haul trucks, and connected vehicles (e.g. Audi/BMW). INRIX has been increasing their data providers to increase their penetration rate for passenger vehicles as well as temporal resolution (e.g. pings every 3-5 seconds). • **Resource Requirements**: Raw data, especially the O-D (raw trip paths) dataset, is enormous and requires significant data processing, analysis, and storage expertise. These individual device pings are simply a latitude/longitude/timestamp and are not mapped to a specific facility. However, INRIX has a partnership with RITIS/UMD and most agencies take advantage of the RITIS data analytics platform to store and process INRIX data. This can be done through an online GUI and aggregated result files can be downloaded in summary images or Excel files. • **Data Sharing**: INRIX data purchased by MDOT, DDOT, and VDOT could be shared with the TPB and vise-versa. If matching data sources from partner agencies in the metropolitan Washington region were purchased, these data sources could be merged for the TPB's research needs within RITIS. • **Cost**: The TPB is already investing in INRIX vehicle probe data for speed and congestion data; however, the high cost for O-D and sub-AADT volume data may be a barrier to product adoption. INRIX vehicle probe data and congestion information is available to the TPB via partner agency agreements. • **Overall Recommendation**: Yes, INRIX should be considered as an option for addressing the TPB's traffic counts research area needs. INRIX's Volume Profile data can hypothetically provide counts for all facilities in a region. This dataset allows for segment-level volumes to be queried for individual days and intervals throughout the day. INRIX also provides AADT estimates for roadway segments as part of the NPMRDS dataset.
StreetLight	+	+	+	+	+	+	-	Yes	• **Applicability**: StreetLight Data is an online platform for O-D or segment-based analytics based on mobile device data. Similar to INRIX, StreetLight offers a suite of analysis tools: AADT estimation, O-D, O-D with middle filter (through a midpoint location), O-D to preset geography (e.g. TAZs, zip codes, census block groups), top routes between O-D's, and a congestion diagnostics tool for auto-generating insights. • **Reliability/Validity**: Various whitepapers are available on StreetLight's website for a variety of applications (e.g., validation of AADTs, turning movements). VDOT recently completed an evaluation of various products offered by StreetLight, including AADT, O-D estimates, traffic link volumes, turning movement volumes, and truck traffic. Evaluation results were mixed depending on the product used and volume levels being measured, with larger errors often associated with lower volumes and shorter time periods. This evaluation also provides a literature review of other recent validation efforts of StreetLight. A third-party validation recently completed by Fehr and Peers for hourly turning movement counts showed that 90 percent of locations were effectively replicated by StreetLight. • **Coverage**: StreetLight's underlying mobile data sources are mainly LBS-based (cell phone apps); INRIX is one of their underlying data sources. This mobile device data is integrated with underlying contextual data such as census demographics to provide additional insights. • **Resource Requirements**: StreetLight's online platform (StreetLight InSight) has a visualization feature for exploring and summarizing data. No raw data (individual trip data) is available; CSVs and shapefiles can be downloaded from the online platform. Excel and GIS (if desired) are typically sufficient for further analysis, although larger or more complex queries from StreetLight may require large CSVs to be processed via a scripting tool such as R or Python to be usable in Excel. • **Data Sharing**: A DOT with a subscription (such as VDOT, MDOT, DDOT) can grant access to underlying MPOs covered by that DOT. • **Cost**: StreetLight offers three different packages: Essentials, Advanced, and Multi-Mode. These packages can be purchased as a subscription or on a project-by-project basis. Subscription pricing is based on the population of the coverage area (e.g., VDOT's subscription to StreetLight data is estimated to cost more than $500,000). • **Overall Recommendation**: Yes, StreetLight should be considered as an option for addressing the TPB's traffic counts research area needs. StreetLight's AADT tool can hypothetically provide AADTs for all facilities in a region. Counts for segments or turning movements are available at intervals as granular as a single hour. It is unclear at this time if any agencies are using StreetLight for widespread traffic count data collection, especially at a sub-daily level.
Teralytics	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Teralytics is an online platform for O-D analytics based on cell phone tower triangulation data. • **Overall Recommendation**: No, Teralytics should not be considered as an option for addressing the TPB's traffic counts research area needs as it is not able to provide analytics for individual facilities.

Big Data Evaluation Table | Traffic Counts

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of sub-research area applicability are provided in Chapter 3
Locus (Cambridge Systematics)	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Locus is a product of LBS-based travel analytics product provided by Cambridge Systematics (consulting firm that has been the TPB's travel demand model developer). Four separate products are offered in addition to the custom analyses: O-D tables (expanded/validated by travel purpose/time-of-day), a transit competitiveness dashboard, a geofence analysis of activity around activity centers, and survey assist to supplement traditional HTS data. • **Overall Recommendation**: No, Locus should not be considered as an option for addressing the TPB's traffic counts research area needs as it is not able to provide analytics for individual facilities.
Replica	+	N/A	+	+	+	+	+	Yes	• **Applicability**: Replica is an online platform for aggregate-level mobility, economic activity, and COVID-19 trend data as well as detailed travel pattern data for select regions, including the Baltimore-Washington region which covers the TPB planning area. O-D data, represented by a synthetic population, is available for all major purposes (work, home, eat, shop, school, social, recreation, errands, lodging, pass-through, commercial and other) and modes (driving, auto passenger, taxi/TNC, transit, walk, bike and commercial vehicles). O-D data can be further filtered and partitioned based on several other variables, such as trip start time, distance and duration. The data for individual trips from the synthetic population data can be downloaded for post-processing outside of the Replica platform. Replica provides highway traffic volume estimates based on the OSM street network. Note that AADT or intersection turning movement counts (TMC) are not available from Replica in the same manner as from StreetLight Data, although Replica provides model estimated traffic volumes at the link level. • **Reliability/Validity**: Validation results can be found mostly from the quality reports prepared by Replica when calibrating the activity-based model using ground truth data collected from each region. • **Coverage**: The aggregate-level Trends module covers the entire US, and the more detailed Places module covers select regions, including the Baltimore-Washington region which includes the TPB planning area. • **Resource Requirements**: The data platform is straightforward to use for viewing, charting, tabulating and mapping data, developing standard data reports, and performing data analytics (e.g. select-link analysis). The data platform also supports the download of detailed data for custom applications. Although the size of some downloaded data files may be large, the files can be analyzed further in Excel, ArcGIS, and other commonly used software tools. • **Data Sharing**: Data licensing and sharing is flexible; annual subscription supports unlimited user licenses within an MPO if the MPO single access option is selected or unlimited user licenses within both an MPO and its member organizations with the MPO regional access option. Consultants working for the MPO are regarded as the extended MPO staff and also can be granted temporary license for data access. • **Cost**: Data cost is relatively reasonable when compared with similar data such as StreetLight Data. The total cost is based on the population of the largest city in the MPO region and the choice of either the single access option or regional access option. For TPB, the annual subscription cost is estimated to range between $104,000 and $173,000 for 2 years of full data access. • **Overall Recommendation**: Yes, Replica should be considered as an option for addressing the TPB's traffic counts needs given its modeled traffic volumes on all roadway links in the detailed Places module. However, while traffic counts are validated based on publicly-available DOT count data, the traffic volumes represented in a Replica model are based on the outputs of an activity-based travel demand model.
Uber Movement	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: This dataset provides aggregated zone-to-zone travel time data derived from Uber trips freely available for the Washington, D.C., area, but only at the TAZ (District of Columbia only) or census tract (roughly the area contained by the I-495 Capital Beltway) level. • **Overall Recommendation**: No, Uber Movement should not be considered as an option for addressing the TPB's traffic counts research area needs. This dataset is not applicable to the research needs within this research area.
Taxi/TNC Trip Data	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: The District of Columbia has aggregated data for private vehicles-for-hire, such as taxis and TNCs, including Uber and Lyft, as information is required to be reported to DFHV and DDOT under D.C. law (§ 50–301.29a. General requirements for private vehicles-for-hire). • **Overall Recommendation**: No, Taxi/TNC Trip Data should not be considered as an option for addressing the TPB's traffic counts research area needs. This dataset is not applicable to the research needs within this research area.
Strava Metro	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Strava Metro is an aggregator of bicycle and pedestrian movements using data obtained from the Strava user mobile app (for tracking bicycle/running/walking trips). Strava Metro is their product offering for planners and provides the relative level of activity (separated out into bicycle/pedestrian) along various facilities. • **Overall Recommendation**: No, Strava Metro should not be considered as an option for addressing the TPB's traffic counts research area needs. This dataset is not applicable to the research needs within this research area.
Disaggregate Census Data	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: The US Census Bureau publishes large demographic datasets on travel behavior and population demographics. Decennial Census data (e.g., 2020 Census) is usually made available by the Census Bureau less than a year after the survey. Other major Census product uses are PUMS from (ACS and CTPP. QCEW is a government data product used for a variety of purposes, such as economic security monitoring and labor statistic reporting. • **Overall Recommendation**: No, Disaggregate Census Data should not be considered as an option for addressing the TPB's traffic counts research area needs. This dataset is not applicable to the research needs within this research area.
Household Travel Survey	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: HTS reveals insightful information with behavioral characteristics of travelers and the relationship between travel decisions and travelers' demographic backgrounds. Traditionally, HTS collects the data that pertains to household, person, vehicle, and trip characteristics on

Big Data Evaluation Table | Traffic Counts

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of sub-research area applicability are provided in Chapter 3
									selected travel day(s) and usually is conducted every 5 to 10 years by MPOs, state DOTs, and FHWA through the NHTS program. All surveyed peer agencies stated the usage of household travel survey data to support travel demand modeling and transportation planning studies. • **Overall Recommendation**: No, HTS should not be considered as an option for addressing the TPB's traffic counts research area needs. This dataset is not applicable to the research needs within this research area.
InfoUSA Dun & BradStreet Business Listings	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: InfoUSA/Dun & Bradstreet business listing data is among the most commonly used data products to support the development of employment databases for both regional land use, travel demand modeling, and economic development activities. Both InfoUSA and Dun & Bradstreet data products provide detailed establishment location-based business information including employment size and industry sector. • **Overall Recommendation**: No, InfoUSA/Dun & Bradstreet Business Listings should not be considered as an option for addressing the TPB's traffic counts research area needs. This dataset is not applicable to the research needs within this research area.
CoStar Data	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: CoStar is a database of commercial real estate transactions. CoStar data can be used to help understand the real estate development patterns in the region, provide input and parameters to land use modeling, and derive employment type and activities. • **Overall Recommendation**: No, CoStar Data should not be considered as an option for addressing the TPB's traffic counts research area needs. This dataset is not applicable to the research needs within this research area.
Google Places	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Google Places provides information on location and type of places such as offices, parks, restaurants, and transit stops. This product offers an alternative approach to verify and supplement the business listing data acquired from the other sources (e.g., InfoUSA, Dun & Bradstreet, and QCEW), particularly about the presence, location, and type of the businesses to better inform regional land use and travel demand models. • **Overall Recommendation**: No, Google Places should not be considered as an option for addressing the TPB's traffic counts research area needs. This dataset is not applicable to the research needs within this research area.
Continuous Traffic Count Station/Sensor Data	+	+	+	+	+	+	+	Yes	• **Applicability**: Local and state DOTs typically have a series of permanent or temporary devices set up for collecting traffic counts, and likely vehicle classifications and spot speeds. From this data, agencies can estimate the average daily number of vehicles traversing roadway segments and vehicle miles traveled. • **Overall Recommendation**: Yes, Continuous Traffic Count Station/Sensor Data should be considered as an option for addressing the TPB's traffic counts research area needs. In the absence of a Big Data solution for ubiquitous count data across a region, permanent count stations provide potential control points for calibration and validating Big Data solutions. Aggregated traffic count data for the region is already being compiled by TPB and is provided on the RTDC.
Automated Traffic Signal Performance Measures (ATSPM)	+	+	-	-	-	+	-	Yes	• **Applicability**: An emerging field of Big Data analytics in the transportation industry involves performance measures for traffic signals based on high resolution (up to 1/10th second) event data. This data, such as detector calls, phase changes, and transit signal priority/preemption can be used to provide analytics on operations at individual signals or along entire arterial corridors. ATSPMs are emerging as an alternative to the traditional traffic engineering practice of ad-hoc turning movement count data collection, traffic signal timing, and arterial operations analysis. The research team is not aware of any agencies using ATSPM data for travel demand modeling applications. • **Overall Recommendation**: ATSPM data should be considered as an option for addressing the TPB's traffic counts research area needs. ATSPMs offer the opportunity for collection of turning movement counts at signalized intersections based on high-resolution detector data. However, at this time the research team notes the limited coverage within the region and likely significant processing demands required for any sort of integration in a travel demand model setting.
Transit Data from On-Board ITS Devices	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Transit agencies employ a variety of ITS devices for monitoring the status of their system and improving system performance. The most applicable sources of transit ITS data for travel demand modeling is O-D data from regional SmarTrip cards (WMATA maintains the SmarTrip farecard data for the region). This data can be used to understand O-D travel patterns across the Metrorail system as well as some (but not all) bus rides, including transfers to/from Metrorail. Boarding and alighting data for bus systems in the region also can be applicable to model validation but do not provide O-D information. WMATA farecard data provides O-D's for all Metrorail trips as well as any bus-to-rail or rail-to-bus transfers using a SmarTrip card. Thus, O-D data is available for trips within the Metrorail system. Note that this data does not provide where riders are ultimately starting or ending their trips (the first-mile/last-mile problem). • **Overall Recommendation**: No, Transit Data from On-Board ITS devices should not be considered as an option for addressing the TPB's traffic counts research area needs. This dataset is not applicable to the research needs within this research area.
RITIS	+	+	N/A	+	+	+	+	Yes	• **Applicability**: RITIS from the CATT Lab at UMD is an "automated data fusion and dissemination system that provides an enhanced overall view of the transportation network. Participating agencies are able to view transportation and emergency management information through innovative visualizations and use it to improve their operations and emergency preparedness". RITIS ingests standardized data from outside sources, including mobile device data from providers such as INRIX, HERE, and TomTom as well as data on weather, traffic incidents, and other sources of data such as agency CCTV and detector feeds. • **Reliability/Validity**: The RITIS platform is used by numerous agencies to process a variety of Big Data products. The validity of the processed outputs are largely dependent upon the underlying validity of the ingested Big Data products themselves. • **Coverage**: Coverage would be dependent upon the data source acquired and provided to RITIS for analytics. • **Resource Requirements**: RITIS provides a user-friendly web-based interface for performing analytics. Result files can be downloaded from this interface as summary images or Excel workbooks.

Big Data Evaluation Table | Traffic Counts

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of sub-research area applicability are provided in Chapter 3
									• **Data Sharing/Cost**: Access to various tools as well as the underlying data being processed by those tools, is dependent upon each state's or agency's contract with the CATT Lab. Use of RITIS is free-of-charge when purchasing more than $100,000 of INRIX data. • **Overall Recommendation**: Yes, RITIS should be considered as an option for addressing the TPB's traffic counts research area needs. Should TPB seek to access INRIX Volume Profile data, this data will likely be accessed through RITIS.
Moonshadow/DB4IoT	-	N/A	N/A	N/A	N/A	N/A	N/A	**No**	• **Applicability**: Moonshadow is a provider of connected vehicle data analytics. Through a partnership with underlying providers such as INRIX and wejo, they have developed a database and visualization suite called DB4IoT. This product serves as both a backend (data storage/processing engine for raw data from vehicles and infrastructure) and a frontend (dashboard for visualizing data). Currently, DB4IoTingests data from a variety of sources including: connected vehicle data (with underlying providers such as INRIX and wejo), mobile application LBS data (with underlying providers such as Unacast and X-Mode Social), and customer data (i.e., data feed unique to specific agencies, such as public transportation feeds, micromobility companies, WiFi/Bluetooth field data collection devices, and traffic counters). • **Overall Recommendation**: No, Moonshadow/DB4IoT should not be considered as an option for addressing the TPB's traffic counts research area needs. Currently, this product is only providing raw trip counts (counts of sampled data), not estimates that have been extrapolated to a full count estimate.
Swiftly	-	N/A	N/A	N/A	N/A	N/A	N/A	**No**	• **Applicability**: Swiftly is a vendor of transit analytics using feeds from agency systems; they claim to be "hardware-agnostic and software-centered" for consuming and analyzing feeds from various AVL providers. They assist public transit agencies in enhancing their transit service by analyzing on-time performance and identifying operational issues. • **Overall Recommendation**: No, Swiftly should not be considered as an option for addressing the TPB's traffic counts research area needs. This dataset is not applicable to the research needs within this research area.
Moovit	-	N/A	N/A	N/A	N/A	N/A	N/A	**No**	• **Applicability**: Moovit is a mobility software company that offers both user-facing products (e.g., rider trip planning tools) and analytics for transit agencies. They offer two main products: MUMA and MaaS. Due to their multimodal focus, they have as much experience with MPOs as they do with transit agencies. • **Overall Recommendation**: No, Moovit should not be considered as an option for addressing the TPB's traffic counts research area needs. This dataset is not applicable to the research needs within this research area.
Emerging Data Sources	N/A	N/A	N/A	N/A	N/A	N/A	N/A	**N/A**	• The emerging data sources evaluated as part of this study are non-exhaustive and geared towards the following research areas with limited data availability: transit and non-motorized travel, TNCs, and connected and automated vehicles.

Table B-7 | Big Data Evaluation Table for Other Research Areas

Big Data Evaluation Table | Other Research Areas									
Product	**Applicability to Research Area**	**Data Reliability and Validity**	**Data Coverage**	**Resource Requirements (Technical Staff and Training Requirements)**	**Resource Requirements (Technology, IT)**	**Data Sharing Restrictions**	**Cost**	**Overall Recommendation**	**Comments on Scoring** *** Details of subresearch area applicability are provided in Chapter 3**
INRIX	+	+	+	+	+	+	-	Yes	• **Applicability**: INRIX is a provider of vehicle probe data for segment-level congestion analytics as well as O-D's for customized zones. INRIX has various product offerings: probe data (segment-level speed/travel time/AADT estimates), O-D summaries (trip starts/ends), and trip paths (includes waypoints along routes and is a very large dataset). • **Reliability/Validity**: INRIX is the underlying data source for NPMRDS, which provides region-wide travel speeds and volume estimates for NHS facilities. The data product has undergone numerous validation efforts and is widely accepted in the industry. The ETC also has conducted extensive validation of various INRIX products, most notably their validations of travel speeds; they are currently sponsoring ongoing efforts to validate INRIX's ubiquitous traffic volume data. • **Coverage**: The underlying data sources are a combination of LBS, GPS from local delivery fleets and long-haul trucks, and connected vehicles (e.g. Audi/BMW). INRIX has been increasing their data providers to increase their penetration rate for passenger vehicles as well as temporal resolution (e.g. pings every 3-5 seconds). • **Resource Requirements**: Raw data, especially the O-D (raw trip paths) dataset, is enormous and requires significant data processing, analysis, and storage expertise. These individual device pings are simply a latitude/longitude/timestamp and are not mapped to a specific facility. However, INRIX has a partnership with RITIS/UMD and most agencies take advantage of the RITIS data analytics platform to store and process INRIX data. This can be done through an online GUI and aggregated result files can be downloaded in summary images or Excel files. • **Data Sharing**: INRIX data purchased by MDOT, DDOT, and VDOT could be shared with the TPB and vise-versa. If matching data sources from partner agencies in the metropolitan Washington. region were purchased, these data sources could be merged for the TPB's research needs within RITIS. • **Cost**: The TPB is already investing in INRIX vehicle probe data for speed and congestion data; however, the high cost for O-D and sub-AADT volume data may be a barrier to product adoption. INRIX vehicle probe data and congestion information is available to the TPB via partner agency agreements. • **Overall Recommendation**: Yes, INRIX should be considered as an option for addressing the TPB's other research area needs. INRIX O-D data could be used to validate findings from the HTS and potentially could be integrated within the HTS to expand the temporal and geographic coverage of the traditional responses. It also could be used to analyze truck travel patterns, as data is available for medium and heavy trucks as separate datasets.
StreetLight	+	+	+	+	+	+	-	Yes	• **Applicability**: StreetLight Data is an online platform for O-D or segment-based analytics based on mobile device data. Similar to INRIX, StreetLight offers a suite of analysis tools: AADT estimation, O-D, O-D with middle filter (through a midpoint location), O-D to preset geography (e.g. TAZs, zip codes, census block groups), top routes between O-D's, and a congestion diagnostics tool for auto-generating insights. • **Reliability/Validity**: Various whitepapers are available on StreetLight's website for a variety of applications (e.g., validation of AADTs, turning movements). VDOT recently completed an evaluation of various products offered by StreetLight, including AADT, O-D estimates, traffic link volumes, turning movement volumes, and truck traffic. Evaluation results were mixed depending on the product used and volume levels being measured, with larger errors often associated with lower volumes and shorter time periods. This evaluation also provides a literature review of other recent validation efforts of StreetLight. A third-party validation recently completed by Fehr and Peers for hourly turning movement counts showed that 90 percent of locations were effectively replicated by StreetLight. • **Coverage**: StreetLight's underlying mobile data sources are mainly LBS-based (cell phone apps); INRIX is one of their underlying data sources. This mobile device data is integrated with underlying contextual data such as census demographics to provide additional insights. • **Resource Requirements**: StreetLight's online platform (StreetLight InSight) has a visualization feature for exploring and summarizing data. No raw data (individual trip data) is available; CSVs and shapefiles can be downloaded from the online platform. Excel and GIS (if desired) are typically sufficient for further analysis, although larger or more complex queries from StreetLight may require large CSVs to be processed via a scripting tool such as R or Python to be usable in Excel. • **Data Sharing**: A DOT with a subscription (such as VDOT, MDOT, DDOT) can grant access to underlying MPOs covered by that DOT. • **Cost**: StreetLight offers three different packages: Essentials, Advanced, and Multi-Mode. These packages can be purchased as a subscription or on a project-by-project basis. Subscription pricing is based on the population of the coverage area (e.g., VDOT's subscription to StreetLight data is estimated to cost more than $500,000). • **Overall Recommendation**: Yes, StreetLight should be considered as an option for addressing the TPB's other research area needs. StreetLight O-D data could be used to validate findings from the HTS and could potentially be integrated within the HTS to expand the temporal and geographic coverage of the traditional responses. The additional information StreetLight provides over INRIX (e.g., income levels, trip purpose, and bicycle/pedestrian O-Ds) could be further advantageous for understanding travel patterns and mode choice based on community demographics. Additionally, StreetLight O-Ds can be extracted for heavy and medium trucks, which may be used for estimating current and future freight and commercial vehicle travel within the region.

Big Data Evaluation Table | Other Research Areas

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
Teralytics	+	+	+	+	+	-	+	Yes	• **Applicability**: Teralytics is an online platform for O-D analytics based on cell phone tower triangulation data. • **Reliability/Validity**: Teralytics claim to have a less biased sample than LBS-based analytics providers as they "sit behind the firewall of all major mobile phone carriers"; they claim sufficient market share among all demographics, ethnic groups, income levels, and age groups. No validation whitepapers are provided on their website. Teralytics claims to have a high level of accuracy given its deep penetration rate, but all benchmarking appears to be done internally. One study was identified noting a limited validation effort of Teralytics "because Teralytics relies on a single data source with a considerably large penetration rate". This study showed a distribution of LRT trips by time-of-day estimated by Teralytics to be consistent with a "general understanding of transit trip-making patterns" for the LRT system. • **Coverage**: Cell tower triangulation has high sample rate (estimated at 15-35 percent of population) but a lower spatial resolution. This data is not usable for route-level analysis (individual trips cannot be mapped to specific roads), but it is usable at the census tract or even TAZ level. Segment-level traffic count estimates (AADTs or turning movement counts) are not estimated via this platform. • **Resource Requirements**: Similar to StreetLight, data is accessed through an online visualization platform and viewed through the UI; data can also be downloaded via CSV and analyzed in Excel or GIS. • **Data Sharing**: No sharing of data is allowed outside of the purchasing agency except with consulting firms doing a project with that agency using the data. A purchasing agency can share derivatives (e.g., analysis results). A purchasing agency could negotiate a unique data sharing agreement as needed. • **Cost**: Generally, Teralytics is less expensive than products providing route-level analytics (e.g., INRIX and StreetLight) while providing a robust sample for understanding larger-scale travel patterns. Pricing is provided on an individual project basis or via subscription (e.g., one year of unlimited use and one year's worth of data). Pricing is based on population of coverage area. The estimated subscription cost for standard out of the box use of the platform for an area with a population approximately the size of the metropolitan Washington MSA is approximately $50,000. Custom data sets for a population of the same size are likely in the $70,000-$90,000 range; cost depends on the level of customization. • **Overall Recommendation**: Yes, Teralytics should be considered as an option for addressing the TPB's other research area needs. Since Teralytics uses cell phone tower triangulation data, they are able to sustain a longer-term trace of individuals to ascertain trip purpose and trip frequency. This data would be valuable for identifying trip purpose as well as identifying the number of trips being made between O-D pairs on different days of the week. These data could be used to validate findings from HTS and could potentially be integrated within the HTS to expand the temporal and geographic coverage of the traditional responses.
Locus (Cambridge Systematics)	+	N/A	+	+	+	-	-	Yes	• **Applicability**: Locus is a product of LBS-based travel analytics product provided by Cambridge Systematics (consulting firm that has been the TPB's travel demand model developer). Four separate products are offered in addition to the custom analyses: O-D tables (expanded/validated by travel purpose/time-of-day), a transit competitiveness dashboard, a geofence analysis of activity around activity centers, and survey assist to supplement traditional HTS data. • **Reliability/Validity**: Validation is a trade secret. • **Coverage**: Locus is built through a partnership with PlaceIQ for the underlying LBS dataset (mainly from cell phone apps). • **Resource Requirements**: This product is essentially a combination of consultant services and software, it is a customizable/tailored solution. Data can be sliced as needed, with analyses conducted on an as-needed basis in addition to the products described in the next bullet. Analyses are not constrained by the available inputs and options on an online platform. Limited data storage and processing are required on the agency end. Analyses are being conducted as part of consultant services and data accessed via dashboards. • **Data Sharing**: Data licensing and sharing is flexible; MPO sharing upward to a DOT level would be an additional fee. • **Cost**: Base product (trip tables), including consultant services, is approximately $150,000 to $200,000; data would be available at the census tract level. These trip tables would provide O-D flows by trip purpose and time of day. Consulting labor fees are the main driver of cost for additional custom analyses, including building custom dashboards. • **Overall Recommendation**: Yes, Locus should be considered as an option for addressing the TPB's other research area needs. Locus O-D trip tables show multimodal trips that show trip purpose and time of day, which could be used to validate findings from HTS and could potentially be integrated within HTS to expand the temporal and geographic coverage of traditional responses. Due to their customizable business model, where Cambridge Systematics is offering both a Big Data product as well as consulting services, it is possible that the resulting data could be designed specifically for comparison and integration with HTS.
Replica	+	N/A	+	+	+	+	+	Yes	• **Applicability**: Replica is an online platform for aggregate-level mobility, economic activity, and COVID-19 trend data as well as detailed travel pattern data for select regions, including the Baltimore-Washington region which covers the TPB planning area. O-D data, represented by a synthetic population, is available for all major purposes (work, home, eat, shop, school, social, recreation, errands, lodging, pass-through, commercial and other) and modes (driving, auto passenger, taxi/TNC, transit, walk, bike and commercial vehicles). O-D data can be further filtered and partitioned based on several other variables, such as trip start time, distance and duration. The data for individual trips from the synthetic population data can be downloaded for post-processing outside of the Replica platform. Given the various interests of the TPB for this research area, it is likely that Replica could be applied to support analyses related to community-specific travel patterns and mode choice. • **Reliability/Validity**: Validation results can be found mostly from the quality reports prepared by Replica when calibrating the activity-based model using ground truth data collected from each region.

Big Data Evaluation Table | Other Research Areas

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
									• **Coverage**: The aggregate-level Trends module covers the entire US, and the more detailed Places module covers select regions, including the Baltimore-Washington region which includes the TPB planning area. • **Resource Requirements**: The data platform is straightforward to use for viewing, charting, tabulating and mapping data, developing standard data reports, and performing data analytics (e.g. select-link analysis). The data platform also supports the download of detailed data for custom applications. Although the size of some downloaded data files may be large, the files can be analyzed further in Excel, ArcGIS, and other commonly used software tools. • **Data Sharing**: Data licensing and sharing is flexible; annual subscription supports unlimited user licenses within an MPO if the MPO single access option is selected or unlimited user licenses within both an MPO and its member organizations with the MPO regional access option. Consultants working for the MPO are regarded as the extended MPO staff and also can be granted temporary license for data access. • **Cost**: Data cost is relatively reasonable when compared with similar data such as StreetLight Data. The total cost is based on the population of the largest city in the MPO region and the choice of either the single access option or regional access option. For TPB, the annual subscription cost is estimated to range between $104,000 and $173,000 for 2 years of full data access. • **Overall Recommendation**: Yes, Replica should be considered as an option for addressing the TPB's other research area needs given its widespread applicability and suite of analysis tools, including O-D, transit, and active transportation analytics.
Uber Movement	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: This dataset provides aggregated zone-to-zone travel time data derived from Uber trips freely available for the Washington, D.C., area, but only at the TAZ (District of Columbia only) or census tract (roughly the area contained by the I-495 Capital Beltway) level. • **Overall Recommendation**: No, Uber Movement should not be considered as an option for addressing the TPB's other research area needs. This dataset is not applicable to the research needs within this research area.
Taxi/TNC Trip Data	+	N/A	+	-	+	-	N/A	Yes	• **Applicability**: The District of Columbia has aggregated data for private vehicles-for-hire, such as taxis and TNCs, including Uber and Lyft, as information is required to be reported to DFHV and DDOT under D.C. law (§ 50–301.29a. General requirements for private vehicles-for-hire). • **Overall Recommendation**: No, Taxi/TNC Trip Data should not be considered as an option for addressing the TPB's other research area needs. This dataset is not applicable to the research needs within this research area.
Strava Metro	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Strava Metro is an aggregator of bicycle and pedestrian movements using data obtained from the Strava user mobile app (for tracking bicycle/running/walking trips). Strava Metro is their product offering for planners and provides the relative level of activity (separated out into bicycle/pedestrian) along various facilities. • **Overall Recommendation**: No, Strava Metro should not be considered as an option for addressing the TPB's other research area needs. This dataset is not applicable to the research needs within this research area.
Disaggregate Census Data	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: The US Census Bureau publishes large demographic datasets on travel behavior and population demographics. Decennial Census data (e.g., 2020 Census) is usually made available by the Census Bureau less than a year after the survey. Other major Census product uses are PUMS from ACS and CTPP. QCEW is a government data product used for a variety of purposes, such as economic security monitoring and labor statistic reporting. • **Overall Recommendation**: No, Disaggregate Census Data should not be considered as an option for addressing the TPB's other research area needs. This dataset is not applicable to the research needs within this research area.
Household Travel Survey	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: HTS reveals insightful information with behavioral characteristics of travelers and the relationship between travel decisions and travelers' demographic backgrounds. Traditionally, HTS collects the data that pertains to household, person, vehicle, and trip characteristics on selected travel day(s) and usually is conducted every 5 to 10 years by MPOs, state DOTs, and FHWA through the NHTS program. All surveyed peer agencies stated the usage of household travel survey data to support travel demand modeling and transportation planning studies. • **Overall Recommendation**: No, HTS should be considered as an option for addressing the TPB's other research area needs. This research area includes datasets that can be used to validate findings from the HTS.
InfoUSA/Dun & BradStreet Business Listings	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: InfoUSA/Dun & Bradstreet business listing data is among the most commonly used data products to support the development of employment databases for both regional land use, travel demand modeling, and economic development activities. Both InfoUSA and Dun & Bradstreet data products provide detailed establishment location-based business information including employment size and industry sector. • **Overall Recommendation**: No, InfoUSA/Dun & Bradstreet Business Listings should not be considered as an option for addressing the TPB's other research area needs. This dataset is not applicable to the research needs within this research area.
CoStar Data	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: CoStar is a database of commercial real estate transactions. CoStar data can be used to help understand the real estate development patterns in the region, provide input and parameters to land use modeling, and derive employment type and activities. • **Overall Recommendation**: No, CoStar Data should not be considered as an option for addressing the TPB's other research area needs. This dataset is not applicable to the research needs within this research area.
Google Places	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Google Places provides information on location and type of places such as offices, parks, restaurants, and transit stops. This product offers an alternative approach to verify and supplement the business listing data acquired from the other sources (e.g., InfoUSA, Dun & Bradstreet, and QCEW), particularly about the presence, location, and type of businesses to better inform regional land use and travel demand models. • **Overall Recommendation**: No, Google Places should not be considered as an option for addressing the TPB's other research area needs. This dataset is not applicable to the research needs within this research area.

Big Data Evaluation Table | Other Research Areas

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation	Comments on Scoring * Details of subresearch area applicability are provided in Chapter 3
Continuous Traffic Count Station/Sensor Data	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Local and state DOTs typically have a series of permanent or temporary devices set up for collecting traffic counts, and likely vehicle classifications and spot speeds. From this data, agencies can estimate the average daily number of vehicles traversing roadway segments and vehicle miles traveled. • **Overall Recommendation**: No Continuous Traffic Count Station/Sensor Data should not be considered as an option for addressing the TPB's other research area needs. This dataset is not applicable to the research needs within this research area.
Automated Traffic Signal Performance Measures (ATSPM)	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: An emerging field of Big Data analytics in the transportation industry involves performance measures for traffic signals based on high resolution (up to 1/10th second) event data. This data, such as detector calls, phase changes, and transit signal priority/preemption can be used to provide analytics on operations at individual signals or along entire arterial corridors. ATSPMs are emerging as an alternative to the traditional traffic engineering practice of ad-hoc turning movement count data collection, traffic signal timing, and arterial operations analysis. The research team is not aware of any agencies using ATSPM data for travel demand modeling applications. • **Overall Recommendation**: No, ATSPM Data should not be considered as an option for addressing the TPB's other research area needs. This dataset is not applicable to the research needs within this research area.
Transit Data from On-Board ITS Devices	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Transit agencies employ a variety of ITS devices for monitoring the status of their system and improving system performance. The most applicable sources of transit ITS data for travel demand modeling is O-D data from regional SmarTrip cards (WMATA maintains the SmarTrip farecard data for the region). Therefore, this data can be used to understand O-D travel patterns across the Metrorail system as well as some (but not all) bus rides, including transfers to/from Metrorail. Boarding and alighting data for bus systems in the region also can be applicable to model validation but do not provide O-D information. WMATA farecard data provides O-D's for all Metrorail trips as well as any bus-to-rail or rail-to-bus transfers using a SmarTrip card. Thus, O-D data is available for trips within the Metrorail system. Note that this data does not provide where riders are ultimately starting or ending their trips (the first-mile/last-mile problem). • **Overall Recommendation**: No, Transit Data from On-Board ITS devices should not be considered as an option for addressing the TPB's other research area needs. This dataset is not applicable to the research needs within this research area.
RITIS	+	N/A	+	+	+	+	+	Yes	• **Applicability**: RITIS from the CATT Lab at UMD is an "automated data fusion and dissemination system that provides an enhanced overall view of the transportation network. Participating agencies are able to view transportation and emergency management information through innovative visualizations and use it to improve their operations and emergency preparedness". RITIS ingests standardized data from outside sources, including mobile device data from providers such as INRIX, HERE, and TomTom as well as data on weather, traffic incidents, and other sources of data such as agency CCTV and detector feeds. • **Overall Recommendation**: Yes, RITIS should be considered as an option for addressing the TPB's other research area needs. Should TPB continue to use INRIX data products, this data will likely be accessed through RITIS
Moonshadow/ DB4IoT	+	N/A	+	-	+	+	-	Yes	• **Applicability**: Moonshadow is a provider of connected vehicle data analytics. Through a partnership with underlying providers such as INRIX and wejo, they have developed a database and visualization suite called DB4IoT. This product serves as both a backend (data storage/processing engine for raw data from vehicles and infrastructure) and a frontend (dashboard for visualizing data). Currently, DB4IoT ingests data from a variety of sources including: connected vehicle data (with underlying providers such as INRIX and wejo), mobile application LBS data (with underlying providers such as Unacast and X-Mode Social), and customer data (i.e., data feed unique to specific agencies, such as public transportation feeds, micromobility companies, WiFi/Bluetooth field data collection devices, and traffic counters). • **Overall Recommendation**: Yes, Moonshadow/DB4IoT should be considered as an option for addressing the TPB's other research area needs as it can be used to view INRIX data and filter down to different types of trips (e.g. truck-specific trips). However, the research team notes its additional cost as compared to RITIS and unproven track record for system-wide use in the U.S.
Swiftly	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Swiftly is a vendor of transit analytics using feeds from agency systems; they claim to be "hardware-agnostic and software-centered" for consuming and analyzing feeds from various AVL providers. They assist public transit agencies in enhancing their transit service by analyzing on-time performance and identifying operational issues. • **Overall Recommendation**: No, Swiftly should not be considered as an option for addressing the TPB's other research area needs.
Moovit	-	N/A	N/A	N/A	N/A	N/A	N/A	No	• **Applicability**: Moovit is a mobility software company that offers both user-facing products (e.g., rider trip planning tools) and analytics for transit agencies. They offer two main products: MUMA and MaaS. Due to their multimodal focus, they have as much experience with MPOs as they do with transit agencies. • **Overall Recommendation**: No, Moovit should not be considered as an option for addressing the TPB's other research area needs.
Emerging Data Sources	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Yes	• The emerging data sources evaluated as part of this study are non-exhaustive and geared towards the following research areas with limited data availability: transit and non-motorized travel, TNCs, and connected and automated vehicles.

APPENDIX C

LOCAL MEETING SUMMARIES

APPENDIX C: LOCAL MEETING SUMMARIES

This appendix provides a summary of the meetings between the research team and technical representatives from local partner agencies in the Washington DC metropolitan region. The intent of this appendix is to provide the reader with an outline of the discussion that occurred. The key agenda items from each meeting are listed below.

- Agency's biggest need(s) relating to Big Data
- Current Big Data investments, experience, and lessons learned
- Interest in pooled funding/sharing of data resources across the region

1.1 DDOT MEETING SUMMARY

The TPB project team met with the District Department of Transportation (DDOT) in May and December 2020. The following sections describe DDOT's experience with various Big Data vendors and DDOT's potential interest in pooled funding of Big Data.

1.1.1 REPLICA

DDOT had access to a free trial version of Replica that ended at the end of December 2020. The platform was geared towards understanding COVID impacts on travel patterns. DDOT indicated that some documentation on data validity is available, but the team has not performed large validation efforts. DDOT also indicated that the level of data provided in this trial version is not suitable for DDOT planning activities. Specifically, travel pattern data is needed at a smaller granularity than Metropolitan Washington Council of Governments (COG) travel demand management (TDMs). Additionally, DDOT anticipates having an internal discussion with OCTO (Office of the Chief Technology Officer) on whether there is appetite to purchase Replica.

1.1.2 STREETLIGHT

DDOT recently reopened conversation with StreetLight due to the Ford Motor Company–Safety Insights Program, of which DC is participating as part of their Vision Zero mission. DDOT had mentioned existing concerns with StreetLight demographics data, specifically how well sampling accounts for underrepresented populations. DDOT indicated that in recent discussions with StreetLight, this topic was discussed and while it is not perfect, DDOT suggested it was better than they thought it was and likely better than what could be achieved otherwise. DDOT also indicated that an agency subscription (similar to VDOT's) was not likely because of the cost; however, more data may be purchased for individual planning studies at smaller quantities.

1.1.3 TERALYTICS

In a brief discussion on Teralytics, DDOT indicated that they are not very interested in cell tower-based data (e.g., AirSaGE) because the data granularity is not sufficient for most of their planning needs. There is less interest at DDOT for data products providing travel patterns at the transportation analysis zone (TAZ) level as compared to the block-group level, based on the type of work they do and projects they consider.

1.1.4 INTRIX OD DATA

DDOT was one of the first agencies to use INRIX OD data. At the time, there were issues conflating the INRIX data to DDOT's Shared Streets Network. However, data now comes conflated to the Shared Streets network, and DDOT has seen proposals from INRIX offering the old data DDOT purchased in the new format. DDOT indicated that a shortcoming of INRIX OD data is that the price was high for not receiving

multimodal data. As part of this effort, DDOT developed a memorandum summarizing all imagined use cases for detailed OD travel patterns data.

1.1.5 DISCUSSION ON COLLABORATIVE PROCUREMENT

DDOT listed a few challenges to collective procurement. The first of these challenges was funding availability as the FY2022 budget has already been formulated and the council began discussion in February 2020—the call for projects generally comes out two years in advance. The second challenge to collective procurement was agreement duration.

1.2 VDOT MEETING SUMMARY

The TPB project team met with VDOT in November 2020. The following sections describe VDOT's experience with various Big Data vendors, VDOT's infrastructure investments, and VDOT's potential interest in pooled funding of Big Data.

1.2.1 STREETLIGHT

VDOT recently updated their agency-wide StreetLight subscription to include Multi-Mode. This decision came from multiple agencies, including the Office of Intermodal Planning and Investment (OIPI). VDOT indicated there is a huge demand for multimodal data; however, the department is unsure of how much validation has been performed on the data and the underlying sample size used. VDOT suggested StreetLight data should not be used in its raw form (rather used to demonstrate relative travel patterns). VDOT also informed the group that the Research Council released a report with guidelines of how StreetLight data should be used for transportation planning projects (referenced in evaluation report). With this new subscription, VDOT indicated that metropolitan planning organizations (MPOs), local jurisdictions, and transit agencies can access the data; however, the data coverage is for the state of Virginia and a 25-mile buffer around the state. Overall, VDOT indicated that StreetLight data has been successful for identifying O-D patterns, cut-through traffic, and understanding traffic shifts for new construction. It also was mentioned that the agency has used StreetLight to investigate issues related to transportation equity using StreetLight's demographics information.

1.2.2 BIG DATA NEEDS

VDOT expressed interest in gathering better data related to bicyclist and pedestrian travel patterns and are concerned about small sample sizes for non-motorized users, even in Northern Virginia. The VDOT central office is interested in understanding safety metrics for micromobility to answer questions such as, "*should scooters be allowed to share bike lanes*?". Further, there is interest in better understanding vehicle occupancy (i.e., number of riders) and vehicle classification (e.g., passenger cars versus heavy vehicles).

1.2.3 FIELD INFRASTRUCTURE INVESTMENTS

VDOT explained that as part of their agreement with StreetLight, data from permanent count stations are provided to StreetLight and used by StreetLight to calibrate their models. They suggested this partnership is very important and critical for data validation. Currently, they do not know of any planned increases to field infrastructure aside from automated traffic signal performance measures (ATSPM), ramp metering, variable speed limits, and the regional multimodal mobility program (RM3P) data project.

1.2.4 DISCUSSION ON COLLABORATIVE PROCUREMENT

VDOT said that further discussion of collaboration would likely need to include the StreetLight contract manager and other leaders at VDOT and OIPI. VDOT also indicated that a larger StreetLight region would be beneficial to many agencies; however, they did not seem to think that VDOT projects regularly would benefit from an expanded region beyond their current 25-mile buffer around the state.

1.3 MDOT-SHA MEETING SUMMARY

The TPB project team met with the Maryland Department of Transportation – State Highway Administration (MDOT-SHA) in May and November 2020. The following sections describe MDOT-SHA's experience with various Big Data vendors, interest in new infrastructure, potential interest in pooled funding of Big Data, and recommendations for future data collection.

1.3.1 INRIX O-D DATA

MDOT-SHA has had a few meetings and webinars with INRIX from which they determined that INRIX's underlying O-D data has been upgraded; however, they have not used the data enough recently to see an improvement. MDOT-SHA has not purchased new trips data recently. A few concerns MDOT-SHA indicated with INRIX O-D Trips Analytics were a limited pool of information, the data are freight-based, there is a lack of rural data, and they are paying INRIX to continue building their tool without knowing what this tool will provide in the future. MDOT-SHA did mention that working with the Center for Advanced Transportation Technology (CATT) Lab is an advantage to using INRIX O-D data because more is understood about the inner-workings of the underlying algorithms (less of a black box).

1.3.2 REPLICA

MDOT-SHA said that BMC/BRTB (Baltimore Metropolitan Council/Baltimore Regional Transportation Board) recently had a conversation with Replica. The Replica tool is at a zonal level and transparency is somewhere in between StreetLight and the CATT Lab. The tool they were pitched included forecasted travel demand in the future. They raised some concern about the cost in relation to the level of detail the data is providing and shared that the cost estimate they received from Replica was population-based, similar to StreetLight.

1.3.3 STREETLIGHT

MDOT-SHA purchased StreetLight data for the Baltimore area and this subscription expired at the end of September 2020. The StreetLight data was used to validate a travel demand model in the Baltimore region. MDOT-SHA discussed that StreetLight's underlying algorithms are not transparent and more of a black box.

1.3.4 LOCUS

MDOT-SHA briefly mentioned Locus data, suggesting that the transparency of this data product is higher than other products due to the structure (consultant firm developing a custom dashboard for the client).

1.3.5 INFRASTRUCTURE

MDOT-SHA mentioned there is interest for installing more count sensors around the state, especially within the Eastern Shore Bay Bridge to better understand congestion trends in the area; however, they did not seem to anticipate funding would be available for this in the near future.

1.3.6 DISCUSSION ON COLLABORATIVE PROCUREMENT

MDOT-SHA expressed interest in collective procurement with cost-sharing. A few barriers that were raised include requirements for competitive procurement (their StreetLight purchase was sole-sourced, not competitive; however, Maryland has placed more scrutiny on procurement), privacy concerns (MDOT-SHA experienced considerable push-back when first procuring Big Data over concerns of "tracking constituents"; these concerns were addressed at the time, but could be raised again), and funding (however, partnering with an MPO and other partner agencies could reduce the burden on a single agency). They mentioned

that partnering with an MPO and transferring money over to the MPO to make the procurement is sometimes easier than the DOT making the procurement independently.

1.3.7 FURTHER DISCUSSION

MDOT-SHA mentioned the National Household Travel Survey (NHTS) NextGen Survey, which includes both the traditional survey and the collection of passive O-D data (collected by the USDOT and discussed in the Evaluation Report). As part of this survey, the University of Maryland (UMD) was awarded a contract to collect the passive O-D data. This data may be available for purchase in the near-future and could be another option for COG to consider, depending on the level of data granularity needed. Further, there is currently a pooled fund study using StreetLight and location-based services (LBS) data to examine whether this data can be used for the HMPS (highway performance monitoring system). Depending on the findings, this could result in a shift of methodologies in the future.

1.3.8 RECOMMENDATIONS

MDOT-SHA suggested that the TPB project team consider the questions *"What are you going to use the data for in your final deliverable? How exactly will it be used?"* to communicate the value of Big Data products with leadership. One big reason this was brought up during discussion was that subscription tools that are readily available are often helpful for answering questions from senior leadership related to real-time current events.

1.4 WMATA MEETING SUMMARY

The TPB project team met with WMATA in November 2020. The following sections describe WMATA's experience with various Big Data vendors and their upcoming procurement efforts.

1.4.1 MOBILITY LANDSCAPE

Most of WMATA's current Big Data investments are based on farecard data and AVL (automatic vehicle location) data; however, they are actively moving towards other parts of the mobility landscape and desire to understand how people are getting to and from stations. They indicated that the Trace Model (matching farecard data with AVL data) is their "bread and butter"; however, in the long-term they are looking for collaboration opportunities with other agencies to further refine the Trace model and share data resources.

1.4.2 CURRENT DATA COLLECTION

Operational data collected by WMATA are used for many things including in-network O-D analyses, crowding, and service planning. However, this data is limited to how people are taking transit and not capturing travel patterns on other modes of travel. Previous projects included using WiFi readers and Bluetooth beacons to monitor travel flow; however, the team indicated these data collection efforts did not result in large benefits.

1.4.3 DC TNC DATA

WMATA has access to the transportation network company (TNC) data and are feeding this data into their Trip Planner to look at the competitiveness of Uber/Lyft versus Transit.

1.4.4 DATA PROCESSING

WMATA uses Oracle for nearly all data storage and processing. They perform manual structured query language (SQL) queries and use tools such as Microsoft Excel and Tableau for data analytics. They are actively expanding their Tableau internal footprint and sharing dashboards across the agency. Eventually, they envision internal dashboards may be shared externally.

1.4.5 STREETLIGHT

StreetLight data was procured by WMATA as part of the late-night mobility study.

1.4.6 UPCOMING PROCUREMENT EFFORTS

WMATA has written a request for proposal (RFP) to solicit competitive procurement of a Big Data product. They indicated that they were willing to pay for the first attempt at procuring data. They are looking for a consulting firm to partner with a vendor to procure data. WMATA reviewed this project's initial list of Big Data products and conducted market research to write the RFP and list of requirements. There was discussion that writing this list of requirements is very challenging; however, to complete the competitive procurement process, it is necessary. They provided a summary of the high-level requirements WMATA is seeking in their Big Data procurement: O-D data with trip length, travel time, full O-D, demographics, trip purpose, mode choice, and granularity to divide by time period. They are very interested in how mode choice changes throughout a trip duration (e.g., collecting data on continuous trips that use different modes: walk to train station, take train, bike to work). WMATA indicated they were looking for census tract/census block group granularity of travel patterns data. Overall, they are trying to better understand customers they are not reaching (i.e., those who are not currently taking public transit). During the discussion, WMATA acknowledged the Black Box nature of Big Data products and specifically mentioned some skepticism in the reliability of demographics data.

1.5 DRPT MEETING SUMMARY

The TPB project team met the Department of Rail and Public Transportation (DRPT) in November 2020. The following sections describe DRPT's experience with various Big Data vendors and their potential interest in pooled funding of Big Data.

1.5.1 TRANSIT BUSINESS INTELLIGENCE

DRPT indicated a need for increased transit business intelligence or integration of data across multiple statewide agencies. Currently, they are not looking at a lot of live data to evaluate operations. They indicated the Virginia RM3P project will be looking to integrate real-time transit congestion and crowding data. Additionally, the Northern Virginia Transportation Authority (NVTA) does more data analytics and recognizes the need for transit performance data to steer planning and funding. For example, operational data is needed for SMART SCALE to provide performance metrics. This team was not familiar with Moovit or Swiftly (transit intelligence products researched as part of the Evaluation Report). Overall, they are looking to move beyond the traditional O-D survey and begin integrating a lot of data sources that can hold up under scrutiny. Right now, the largest ongoing project in Northern Virginia is an upgrade to their fare collection system.

1.5.2 STREETLIGHT

DRPT mentioned that StreetLight is being used in feasibility studies across the state and indicated a few gaps in the StreetLight data, specifically non-motorized travel modes (pedestrians and bicyclists) data availability and reliability. One specific example is a feasibility study conducted in the Staton/Charlottesville region of Virginia for an intercity connector in which they used StreetLight to build O-D patterns. This study took advantage of the Virginia Commonwealth StreetLight subscription. DRPT indicated that the TPB project team would have access to this subscription if the focus of the study area is in Virginia. They also mentioned that while StreetLight is good for understanding general trip travel patterns, it does not provide data for granular transit analyses because it is difficult to identify mode choice. Similarly, this type of data is not accepted for the Federal Transit Administration (FTA) transit planning. DRPT mentioned that these regulations may not be purely inhibiting; rather, they are good rules that are there for a reason and prevent decisions from being made with Black Box data sources.

1.5.3 DISCUSSION ON COLLABORATIVE PROCUREMENT

In a discussion of pooled funding and sharing of resources, DRPT indicated there is always interest in sharing resources, especially in northern Virginia where travel does not get cut off at the state lines. Barriers to a collective investment may include funding avenues (e.g., grants, state funding) and communication with leadership. One potential opportunity for collaboration may be the Capital Beltway Accord (they recommended discussing this further with VDOT). They also mentioned that increased collaboration in sharing agency data would be very valuable; specifically mentioning difficulty coordinating with WMATA to collect data necessary for project planning and funding requests (SMART SCALE).

APPENDIX D

COMPREHENSIVE PROCUREMENT OPTIONS

APPENDIX D: COMPREHENSIVE PROCUREMENT OPTIONS

As described in **Chapter 3: Big Data Product Evaluation**, findings from the independent evaluation of Big Data suggested that numerous Big Data products could be applied within the National Capital Region Transportation Planning Board's (TPB) research areas to add value to ongoing and future efforts. Numerous promising Big Data products were described in **Section 3.3** (by product) and **Section 3.4** (by research area). As shown, multiple Big Data products could be used to serve similar purposes across multiple research areas, each with its own advantages and disadvantages. A summary of the TPB's research areas alongside all potentially applicable Big Data products is provided in **Table D-1**. Using these findings, the following Big Data product options were initially developed for consideration by the TPB Study Working Group (SWG). These initial procurement options are divided by Big Data product type as introduced in **Section 3.3**. As such, the options are labeled as follows:

- Option "MD" for Mobile Device-Based Patterns Data
- Options "SE" for Socioeconomic or Location-Based Data
- Options "EP" for End User Platforms

A detailed discussion of the validity and limitations of each data product can be found in **Section 3.2** and **Appendix B**. ***As noted in Chapter 4, these options were eventually narrowed down to specific recommendations of moving forward with procuring StreetLight Data and/or Replica.***

1.1 MOBILE DEVICE-BASED TRAVEL PATTERNS DATA

The independent evaluation of Big Data products in **Chapter 3** highlighted five promising products within the mobile device-based travel patterns data category. These products were: INRIX, StreetLight, Teralytics, Locus, and Replica. As shown in **Table D-1**, numerous research areas are addressed by these product types. In reviewing the cost and the ability of these products to serve the key research considerations, it was recommended that the TPB select one or two of these Big Data products for further consideration and potential procurement. The following options were initially developed in consideration of each product's applicability and limitations for the TPB's research areas. These options are also shown in **Table D-2**.

1.1.1 OPTION MD-1: INRIX (PROBE DATA AND ORIGIN-DESTINATION SUMMARIES)

In this option, the TPB would use INRIX as its main source of mobile device-based travel patterns data (including Volume Profile and origin-destination [O-D] data), building off the TPB's existing use of INRIX probe (speed/congestion) data. This could all be accessed via the RITIS platform.

KEY CONSIDERATIONS

- Provides vehicle probe data for segment-level congestion analytics, both in real-time and maintains a historic archive.
- Provides O-D trip tables for customized zones and allows the user to perform select link and route analyses.
    - Includes all trips starting, ending, or passing through the selected area
    - Trip purpose data is not provided
    - Trip tables separated by passenger vehicles and trucks only
- Provides directional segment AADTs and sub-hourly volume estimates through their Volume Profile application.

Table D-1 | Big Data Product Applicability by Research Area

Product Category	Product	Travel Demand Forecasting	Travel Demand Management	System Performance and Congestion Management	Transit and Active Travel	Transportation Network Companies	Traffic Counts	Connected and Automated Vehicles	Other Research Areas
Mobile Device-Based Travel Patterns Data	INRIX	**YES**	**YES**	**YES**	NO	NO	**YES**	NO	**YES**
	StreetLight	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	NO	**YES**
	Teralytics	**YES**	**YES**	NO	**YES**	NO	NO	NO	**YES**
	Locus (Cambridge Systematics)	**YES**	**YES**	NO	NO	NO	NO	NO	**YES**
	Uber Movement	NO	NO	NO	**YES**	**YES**	NO	NO	NO
	Replica	**YES**	**YES**	NO	**YES**	**YES**	**YES**	NO	**YES**
	Taxi/TNC Trip Data	NO	NO	NO	NO	**YES**	NO	NO	NO
	Strava Metro	NO	NO	NO	**YES**	NO	NO	NO	NO
Socioeconomic or Location-based Data	Disaggregate Census Data	**YES**	**YES**	NO	**YES**	NO	NO	NO	NO
	Household Travel Survey	**YES**	**YES**	NO	**YES**	**YES**	NO	NO	NO
	InfoUSA	**YES**	**YES**	NO	NO	NO	NO	NO	NO
	Dun & Bradstreet Business Listings	**YES**	**YES**	NO	NO	NO	NO	NO	NO
	CoStar Data	**YES**	**YES**	NO	NO	NO	NO	NO	NO
	Google Places	**YES**	**YES**	NO	NO	NO	NO	NO	NO
Data from Public Infrastructure	Continuous Traffic Count Station/Sensor Data	**YES**	NO	**YES**	**YES**	NO	**YES**	NO	NO
	Automated Traffic Signal Performance Measures (ATSPM)	**YES**	NO	**YES**	NO	NO	**YES**	NO	NO
	Transit Data from On-Board ITS Devices	**YES**	**YES**	**YES**	**YES**	NO	NO	NO	NO
End-User Platforms for Data Analytics	RITIS	**YES**	**YES**	**YES**	**YES**	NO	**YES**	NO	NO
	Moonshadow/DB4IoT	**YES**	**YES**	NO	NO	NO	NO	NO	NO
	Swiftly	NO	NO	**YES**	**YES**	NO	NO	NO	NO
	Moovit	NO	NO	**YES**	**YES**	NO	NO	NO	NO
Emerging Big Data Sources		--	--	--	**YES**	**YES**	--	**YES**	--

** "Yes" indicates the Big Data product is applicable based on the evaluation criteria for the specific research area.*
** "No" indicates the Big Data product is not applicable based on the evaluation criteria for the specific research area.*

Table D-2 | Big Data Product Procurement Option Applicability by Research Area

Product Category	Options		Travel Demand Forecasting	Travel Demand Management	System Performance and Congestion Management	Transit and Active Travel	Transportation Network Companies	Traffic Counts	Other Research Areas
Mobile Device-Based Travel Patterns Data & End-User Platforms for Data Analytics	A	MD-1: INRIX EP-2: RITIS	**YES**	**YES**	**YES**	NO	NO	**YES**	**YES**
	B	MD-2: StreetLight EP-1: Product Specific Platforms	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**
	C	MD-3: Teralytics + INRIX EP-1: Product Specific Platforms EP-2: RITIS	**YES**	**YES**	NO	**YES**	NO	NO	**YES**
	D	MD-4: Locus + INRIX EP-1: Product Specific Platforms EP-2: RITIS	**YES**	**YES**	NO	NO	NO	NO	**YES**
	E	MD-5: Replica	**YES**	**YES**	NO	**YES**	**YES**	**YES**	**YES**
	F	MD-6: TNC	NO	NO	NO	**YES**	**YES**	NO	NO
Socioeconomic or Location-based Data	G	SE-1: InfoUSA (Data Axle USA) + Google Places	**YES**	**YES**	NO	**YES**	NO	NO	NO
	H	SE-2: COG/TPB Similar Products	*To be determined*						
Data from Public Infrastructure	I	Continuous Traffic Count Station/Sensor Data	**YES**	NO	**YES**	**YES**	NO	**YES**	NO
	J	Automated Traffic Signal Performance Measures (ATSPM)	**YES**	NO	**YES**	NO	NO	**YES**	NO
	K	Transit Data from On-Board ITS Devices	**YES**	**YES**	**YES**	**YES**	NO	NO	NO

** "Yes" indicates the Option grouping is applicable based on the evaluation criteria for the specific research area.*
** "No" indicates the Option grouping is not applicable based on the evaluation criteria for the specific research area.*

- The data could be stored and processed within RITIS.
- The TPB can currently freely access INRIX real-time and historic archive data for segment speeds and congestion for the entire region through the RITIS platform; this data is paid for by state Departments of Transportation (DOTs) and was purchased through the Eastern Transportation Coalition (ETC). This includes the NPMRDS probe data archive, which is freely available but technically a separate speed dataset.
- Pricing for the O-D trip tables and the Volume Profiles is based on the provided analytics and geographic area for which the data was purchased.
    - Purchases of $100,000 or more include access to RITIS at no additional cost.
    - Data purchased through the ETC is subject to a 10 percent or more discount.
    - Volume Profile is estimated to cost approximately $60,000 for the Metropolitan Washington Council of Governments (MWCOG) region based on the region's population.
- DOTs also could purchase O-D trip tables and Volume Profiles for their individual states, which could be shared with the TPB. Coordination with the Center for Advanced Transportation Technology (CATT) Lab would need to take place to ensure access to a combined Maryland/Virginia/District of Columbia dataset for TPB staff.
    - These data sets would need to cover the same temporal range and spatial granularity of O-D data (e.g., transportation analysis zone- [TAZ]-level); otherwise, TPB would likely have access to each separately, but they could not be combined into one O-D dataset.

PROCUREMENT STRATEGY

The TPB could go about procuring INRIX data in two ways:

- **Suboption MD-1 (A)**: The TPB purchases O-D data and Volume Profile data that covers the metropolitan Washington region.
- **Suboption MD-1 (B)**: The TPB enters a sharing agreement with local agencies (i.e., District Department of Transportation [DDOT], Maryland Department of Transportation – State Highway Administration [MDOT-SHA], and Virginia Department of Transportation [VDOT]) and coordinates with them to purchase O-D data and Volume Profile data for their entire state or just the metropolitan Washington region. The TPB would need to coordinate that the purchased data from each jurisdiction followed the same spatial and temporal range to ensure their inter-compatibility for the TPB's use. It is noted that VDOT currently has a multiyear subscription of StreetLight Data and is unlikely to opt to purchase INRIX O-D data during the term.

If identified as a viable option, a request for information (RFI) should be submitted to INRIX to fine tune the TPB's unique quote.

Ultimately, the project team did not recommend moving forward with procuring additional INRIX data at this time, noting partner agency feedback on limitations of the O-D data and a lack of nationwide use of the Volume Profile data. The TPB should continue to use INRIX and RITIS for speed and congestion data.

1.1.2 OPTION MD-2: STREETLIGHT

In this option, the TPB would use StreetLight as its main source of mobile device-based travel patterns data. Data would be accessed through the StreetLight online platform. The TPB could potentially solicit for further custom analyses such as an analysis of gig travel that may not be available through the platform. Additionally, region-wide analyses such as region-wide link speeds for the congestion management

process (CMP) may require querying data through StreetLight's application programming interfaces (API) and post-processing using scripting tools.

KEY CONSIDERATIONS

- Provides AADT estimation, O-D, O-D with middle filter (through a midpoint location), O-D to preset geography (e.g., TAZs, zip codes, census block groups), top routes between O-Ds, and a congestion diagnostics tool for auto-generating insights.
- Provides O-D analytics for bicycles and pedestrians, as well as limited bus and rail analytics (new as of 2021), at an additional cost.
- Provides an online interface for user-generated analyses using completely custom zones that can be either geographic areas or individual roadway segments.
- Provides the option of custom analyses of gig travel (e.g., TNCs) at an additional cost.
- In general, StreetLight data is expensive to purchase.
- StreetLight can be purchased with an annual subscription or for individual projects.
    - Subscription costs are based on the population of the coverage area. For comparison, VDOT's current annual subscription (population of approximately 8.5 million) cost is estimated to be more than $500,000.
    - Individual project purchases are priced by the number of zones and types of analysis tools desired, with the least expensive option being approximately $5,000 (10 unique zones and the most basic suite of tools).
- StreetLight offers the following package types:
    - Essentials package tools: average annual daily traffic (AADT), O-D, zone activity, trip attributes, traveler attributes, O-D with preset geography.
    - Advanced package tools: Essentials plus segment analysis, top routes between O-D's, custom specific dates, commercial vehicle metrics, and traffic diagnostics tool for identifying sources of congestion, mode shift potential, etc.
    - "Multi-mode" package tools: Advanced plus bicycle/pedestrian metrics.
- The Multi-mode package is the most expensive and can be two to three times as costly as the Essentials package.

PROCUREMENT STRATEGY

The TPB could go about procuring StreetLight data in numerous ways; below are a few options:

- **Suboption MD-1 (A)** | The TPB purchases a subscription that covers the metropolitan Washington region. Currently, the research team recommends soliciting pricing for all three tiers of tools.
    - The TPB may consider holding off on purchasing the Multi-mode package until more data validation case studies become available.
- **Suboption MD-1 (B)** | The TPB purchases StreetLight data for a set of zones to address specific research needs on specific projects.
- **Suboption MD-1 (C)**: The TPB enters a sharing agreement with local agencies (i.e., DDOT, MDOT-SHA, and VDOT) and coordinates with them to purchase the Essentials or Advanced Analytics package for their entire state or just the metropolitan Washington region. The TPB would need to coordinate with StreetLight to ensure the inter-compatibility of data streams provided to different jurisdictions for the TPB's use.

- VDOT has already purchased a StreetLight data subscription for Advanced Analytics. A strategy would be to negotiate with StreetLight Data to expand the spatial boundary to include metropolitan Washington area for a discrete time period, which would allow TPB to work with DDOT and MDOT-SHA to evaluate the cost/benefit for such regional subscription.

If identified as a viable option, an RFI should be submitted to StreetLight to fine tune the TPB's unique quote.

Ultimately, the project team has recommended moving forward with submitting an RFI to StreetLight and exploring the sub-options noted above for procuring StreetLight Data. A Pilot Program approach for the immediate future has also been recommended.

1.1.3 OPTION MD-3: TERALYTICS AND INRIX (PROBE DATA ONLY)

In this option, the TPB would continue its existing use of INRIX probe (speed/congestion) data, accessed via the Regional Integrated Transportation Information System (RITIS) platform. This data would be supplemented by O-D data at the TAZ level from Teralytics for the region. Although Teralytics would not be granular enough to analyze individual roadway links, its cost at a regional level makes it an appealing low risk investment for obtaining O-D information that could be used in travel demand modeling and other purposes.

KEY CONSIDERATIONS

- Teralytics provides O-D analytics.
  - Data is intended for use at the census tract or TAZ level.
  - Data is not usable at a route-level.
  - Segment-level traffic counts are not estimated from this vendor.
- Teralytics is a less expensive option compared to INRIX and StreetLight and is being used by agencies to investigate larger-scale travel patterns.
- Pricing is provided on an individual project basis or via subscription (e.g., 1 year of unlimited use and one year's worth of data).
  - Pricing is based on population of coverage area.
- The estimated subscription cost for the standard out of the box use of the platform for an area with a population approximately the size of the metropolitan Washington region is $50,000.
  - Custom data sets of the same size are likely in the $70,000 to $90,000 range, with the cost dependent on the level of customization.

PROCUREMENT STRATEGY

Teralytics does not allow sharing of data between agencies. Therefore, the procurement option for the TPB would be to purchase the dataset directly from Teralytics to perform internal analytics.

If identified as a viable option, an RFI should be submitted to StreetLight to fine tune the TPB's unique quote.

Ultimately, the project team did not recommend moving forward with procuring Teralytics data at this time, noting TPB staff feedback on the need for granular, multimodal O-D and volume data. The TPB should continue to use INRIX and RITIS for speed and congestion data.

1.1.4 OPTION MD-4: LOCUS AND INRIX (PROBE DATA)

In this option, the TPB would continue its existing use of INRIX probe (speed/congestion) data, accessed via the RITIS platform. This data would be supplemented by O-D data at customized levels from Locus for the region. Although Locus would not be granular enough to analyze individual roadway links, its cost at a regional level makes it an appealing low risk investment for obtaining O-D information that could be used in travel demand modeling and other purposes. While more expensive than Teralytics, Locus is still much less expensive than StreetLight and includes consultant support for tailored analyses.

KEY CONSIDERATIONS

- Locus is a Big Data product developed by Cambridge Systematics.
- Locus provides a combination of consultant services and data outputs/analytics.
- At a base level, this product provides O-D tables (expanded/validated by travel purpose/time-of-day), a transit competitiveness dashboard, a geofence analysis of activity around activity centers, and survey assist to supplement traditional household travel survey (HTS) data.
    - O-D data is available at the census tract level.
    - O-D data can be aggregated by trip purpose and time of day.
- Additionally, Locus provides a customizable and tailored solution based on data needs. Analyses can be conducted on an as-need bases, and analyses are not constrained by the available inputs and options from the online platform.
- Locus does not provide speeds/travel times, count estimates, or breakdowns by modes/vehicle classes (trucks). This tool is focused on understanding regional travel patterns for all trips across all modes.
- The base product (trip tables), including consultant services, is approximately $150,000 to $200,000.
- Consulting labor fees are the main driver of cost for additional custom analyses, including building custom dashboards.
- Data licensing and sharing is flexible; metropolitan planning organizations (MPO) sharing upward to a DOT level would be an additional fee.

PROCUREMENT STRATEGY

- **Suboption MD-4 (A)** | The TPB enters a contract with Cambridge Systematics to purchase data and develop dashboards specific to their highest priority research areas.
- **Suboption MD-4 (B)** | The TPB coordinates with one or more partner agencies (e.g., DDOT, MDOT, VDOT) and jointly enters a contract with Cambridge Systematics to purchase data and develop dashboards specific to their highest priority research areas. Added fees are associated with data sharing.

If identified as a viable option, an RFI should be submitted to Cambridge Systematics to fine tune the TPB's unique quote.

Ultimately, the project team did not recommend moving forward with procuring Locus data at this time, noting TPB staff feedback on the need for granular, multimodal O-D and volume data. The TPB should continue to use INRIX and RITIS for speed and congestion data.

1.1.5 OPTION MD-5: REPLICA

In this option, the TPB would use Replica as its main source of mobile device-based travel patterns data. Data would be accessed through the Replica online platform through its two modules: (1) Trends for regional or aggregate travel patterns/trends, and (2) Places for local travel details and insights into

demographics of travelers for various geographic levels (e.g., Census Block groups/tracts, zip codes, cities, counties, and TAZs). Replica provides O-D metrics across a variety of modes and for several different trip purposes, and these metrics can be filtered down for individual geographic areas, roadway facilities, and transit routes or stations. The synthetic population trip data available for download through the Places module provides a potentially very rich dataset for downstream analyses related to travel demand management and understanding mode choice.

KEY CONSIDERATIONS

- Replica ingests a wide range of data inputs, including mobile location data similar to what INRIX and StreetLight Data use; publicly available datasets such as the US Census, land use regulations, and real estate transaction data; observed ground-truth, such as traffic counts or transit ridership; and commonly used geographies, such as Census geographies.
- Replica has two modules: Trends and Places. Trends depicts weekly trends of aggregate-level travel patterns, provided at the national and statewide levels as well as for cities, Metropolitan Statistical Areas (MSAs), and census tracts. Places provides much more in-depth travel data for select U.S. metropolitan areas (including the metropolitan Washington region with a model that covers all of Virginia and Maryland). The Places data that Replica provides for a metropolitan area is simulated data, based on an activity-based travel model that Replica has developed for the area using a synthetic population.
- The Trends module provides various metrics for an average weekday within a selected week (e.g. week of April 5, 2021), and these metrics can be compared for two separate selected weeks. This data is available at the census tract level or less granular levels such as cities/MSAs or states.
- The Places module includes a much wider spectrum of travel metrics, and data can be be filtered/partitioned using any of the following variables:
  - Trip origin and/or destination, down to individual census block groups or TAZs
  - Trip start time (individual hours of the day)
  - Trip purpose (work, home, eat, shop, school, social, recreation, errands, lodging, pass-through, commercial and other)
  - Primary mode (Driving, auto passenger, taxi/TNC, transit, walk, bike and commercial vehicles)
  - Distance
  - Duration (travel time)
  - Network link (e.g., Francis Scott Key Bridge, I-66, I-495, I-495 Express Lanes)
  - Transit route
  - Transit boarding and/or alighting station or stop
  - Household income group (in thousands $: 0-15, 15-25, 25-50, 50–75, 75–100, 100–150, 150–200, >200)
  - Age group (< 5, 5–11, 12–17, 18–34, 35–49, 50–64, > 65)
  - Auto availability (0, 1, 2, 3+ household vehicles)
  - Race and ethnicity (White, Black, Hispanic or Latino, Asian, multiple races, Other, American Indian or Alaskan Native, Native Hawaiian or Pacific Islander)
  - Employment status (employed, unemployed, not in the labor force)
  - Work/School location (people with a workplace or school in a geography such as a census tract)
  - Home location (people living in a geography such as a census tract or zip code)
- Places data may be downloaded for various geographic levels, including the following:
  - Block group
  - Census tract

- Zip code
- City
- County
- TAZ

The downloaded Places data includes a table of each trip taken by each member of the synthetic population over the course of an average weekday or weekend day.

- All Trends data and Places insights are accessible in Replica's web-based interface. Replica also generates pre-packaged reports, such as access to jobs, transit ridership by demographic group, or seasonal comparisons. Also, complete trip table downloads are available as an option for those who need to run custom analyses outside of the web app.
- Replica is currently in the process of developing a scenario analysis component to help understand future impacts from alternatives defined across four dimensions — changes in demographics, land use, infrastructure, and behavioral preferences. This scenario tool is not yet available.

PROCUREMENT STRATEGY

- **Suboption MD-5 (A)** | The TPB enters a contract with Replica to purchase data via annual subscription. This could be either via an MPO Single Access or Regional Access subscription (described in **Section 3.2**).
- **Suboption MD-5 (B)** | The TPB does not purchase the data but uses the data analytics and reports via hiring the consultants who are in partnership with Replica to perform travel studies.

Ultimately, the project team has recommended moving forward with submitting an RFI to Replica and exploring the sub-options noted above for procuring Replica data. A Pilot Program approach for the immediate future has also been recommended.

1.1.6 OPTION MD-6: TNC TRIP DATA

In this option, the TPB would seek to obtain access to the transportation network companies (TNC) trip data for the private vehicles-for-hire from the District of Columbia government. The TPB can request access to the data for specific purposes that are agreed upon between TPB and the District government. This will include primarily serving TPB's needs internally due to data sharing limitations. The TPB would need to navigate the legal boundaries around sharing information from the analysis for external use or soliciting help in processing the data. Other option(s) to access TNC trip data are introduced in the end-user platform for data analytics.

Ultimately, the project team has recommended moving forward with developing an agreement with the DC government regarding accessing TNC trip data. This is noted in the project Recommendations in Section 4.3.

1.2 SOCIOECONOMIC OR LOCATION-BASED DATA

The considered Big Data products under the socioeconomic and location-based data type can be divided into two separated discrete categories: (1) public data generated from surveys and (2) Big Data products from independent vendors.

Public data generated from surveys elicited from various levels of government are assumed to be available to the TPB free of charge; these surveys include the Census and the National Household Travel Surveys (NHTS). As described in **Section 1.1**, while these surveys do not meet the typical definition of Big Data, they are considered as part of this evaluation because of their large size and applicability for integrated use with emerging Big Data sources. The data available from the Census and NHTS as well as the regional HTS that TPB conducts on a less frequent basis (e.g., 2018-2019 MWCOG region HTS), are applicable to

numerous research areas—many of which are already leveraging the benefits of these traditional data sources in their ongoing efforts. ***Thus, it is recommended these products be used in their current capacity and new ways to integrate their data into additional research and subresearch areas should be explored. In addition, it is recommended that these traditional products be used as the baseline for data validation as new Big Data products are introduced to the TPB's research areas.***

The second category, Big Data products from independent vendors, includes products that can be used to better understand the real estate makeup of the region. These products provide information about commercial and residential developments. As shown in **Table D-1**, this type of information is perceived to be applicable to a number of the TPB's research areas. Due to each product's similarity with one another, it is recommended that the TPB limit the number of product investments to one. The following options were identified from the evaluation results. ***At this time, TPB staff and partner agency input has led the project team to focus on recommendations for procurement of mobile device-based travel patterns datasets; these additional socioeconomic datasets could be explored in the future.***

1.2.1 OPTION SE-1: INFOUSA AND GOOGLE PLACES

The accuracy of verified baseline employment lays the groundwork and is critical for employment forecasting and the travel demand analysis. Due to the extremely dynamic nature of the employment, it is challenging for the data vendors to maintain the list of regional businesses up to date. The gap between the actual accuracy of some commonly cited business listing data products and that claimed by the data vendors was noted by some external agencies.

In this option, TPB would continue its existing use of InfoUSA data. This data would be verified and supplemented by Google Places data to improve accuracy particularly in terms of the existence, location, and type of the regional businesses.

KEY CONSIDERATIONS

- InfoUSA product is a widely-used employment data source by many state and MPO agencies in support of land use and travel demand modeling and analysis.
- InfoUSA data is repurposed to provide disaggregate and location specific employment data that can be summarized to meet the land use input data needs of travel demand models.
- TPB has used InfoUSA that roughly includes 280,000 records for the TPB Model Region for several years. The TPB provides the individual records to the participating jurisdictions in the TPB Cooperative Forecasting Program, (e.g., DC Office of Planning, Montgomery County Planning Department, Fairfax County Planning and Development, etc.)
- Google Places is a proven data source that may improve the employment data accuracy. It can be leveraged to verify all the businesses in the region or used only for the focus areas that present more stringent requirements on data accuracy.
- To use Google Places at a large scale, a billing account must be created with Google (the data must be purchased), and then data must be pulled via an API using third-party software or user-generated scripts. The data must then be stored and processed (unless a third-party software product is able to do this).
- Retrieved Google Places data has no limitation on data sharing as long as it is not used for commercial purposes.

PROCUREMENT STRATEGY

The procurement strategy for InfoUSA data may remain the same although group purchase with the state agencies, e.g., VDOT and MDOT, usually result in reduced overall cost if the data needs are shared and the institutional barriers do not exist.

Google Places offers standard pricing to end data users.

1.2.2 OPTION SE-2: IDENTIFY SIMILAR DATA PRODUCTS AT COG/TPB

In conducting background research for this evaluation, the project team identified the potential for intra-agency collaboration within COG with respect to Big Data product investments, especially those related to socioeconomic or location-based data.

For example, COG is actively analyzing the labor market to better understand in-demand jobs and skills within the region. As part of this effort, COG reviews real-time job posting data with the intention to identify emerging market-driven changes in industry and occupational demands. To complete this analysis, COG acquired a 12-month license in 2015 for Burning Glass' Labor Insight—a data tool that informs workforce and economic planning with accurate, real-time labor market analysis.[1]

- **Sub-option SE-2 (A)** | Further investigation of internal Big Data investments within COG could be conducted to identify potential data sources that have already been procured and could be repurposed for the TPB's research areas.
- **Sub-option SE-2 (B)** | Intra-agency coordination could be facilitated to collaboratively select a socioeconomic and location-based Big Data product that meets the data needs of multiple programs within COG. Subscriptions to widely accepted socioeconomic datasets, such as InfoUSA, Dun & Bradstreet, or quarterly Census of Employment and Wages (QCEW), by agencies in the metropolitan Washington region may provide a snapshot of the socioeconomic status of the region. It also may offer insights into how population growth, transportation investments, and other regional transportation policy initiatives could incubate and shape the environment for socioeconomic growth

1.3 DATA FROM PUBLIC INFRASTRUCTURE

Data originating with public infrastructure included the following categories of data:

- **Continuous Traffic Count Station and Sensor Data**
  - The TPB continues to maintain and develop an online resource, the Regional Transportation Data Clearinghouse (RTDC). In addition to aviation, land use, and transit data, it includes traffic counts from partner agencies (e.g., VDOT, MDOT-SHA, and DDOT). Finer-grained traffic count data (e.g., at the county level) also can be obtained from partner agencies.
- **Automated Traffic Signal Performance Measures (ATSPM) Data**
  - The research team believes that the use of ATSPM data across the region is limited; VDOT just recently began a deployment across 12 signals.
  - The TPB could work with partner agencies (e.g., VDOT, MDOT-SHA, and DDOT) to help facilitate ATSPM programs to begin collecting data across the region.
- **Transit Data from On-Board Intelligent Transportation System (ITS) Devices**
  - The TPB can coordinate with regional transit agencies (e.g., Washington Metropolitan Area Transit Authority [WMATA], Department of Rail and Public Transportation [DRPT], Maryland Transit Administration [MTA]) to aggregate regional transit-related data for further processing.
  - These data could be collected and then applied in a transit-specific end-user platform (please reference **Section D.1.4**).

[1] www.mwcog.org

In each of these cases, the TPB serves as the facilitator between agencies to share information and identify new data that could be collected simultaneously by multiple jurisdictions to better understand regional travel conditions.

At this time, TPB staff and partner agency input has led the project team to focus on recommendations for procurement of mobile device-based travel patterns datasets; however, as noted in Section 4.4, the project team recommends that the TPB be proactive and function as a regional advocate for Big Data and the importance of maintaining infrastructure.

1.4 END-USER PLATFORMS FOR DATA ANALYTICS

In addition to evaluating Big Data products, the available end-user platforms for data analytics were explored and evaluated.

1.4.1 OPTION EP-1: PRODUCT-SPECIFIC PLATFORMS

Data analytics platforms for mobile device-based travel patterns data are primarily product-specific.

- Product-specific dashboards exist for StreetLight, Teralytics, and Replica that are made available with the product purchase.
- Locus typically develops a custom data analytics platform for the end-user unique to the project needs.
- INRIX uses the RITIS platform, which is made available to agencies at no cost when the INRIX purchase exceeds $100,000[2].

Ultimately, the project team has recommended moving forward with submitting an RFI to StreetLight and Replica and assumes that the TPB would utilize the platforms within one or both of these products. A Pilot Program approach for the immediate future has been recommended.

1.4.2 OPTION EP-2: RITIS

While RITIS maintains an agreement with INRIX, the platform can be used to ingest and explore data from a variety of vendors. RITIS can ingest standardized data from outside sources, including mobile device-based travel patterns data, weather data, traffic incident data, CCTV and detector feeds, transit data, and agency-specific data.

- RITIS standard suite of tools includes: speed and congestion analytics, O-D analytics, and traffic signal analytics.
- RITIS is popular among peer agencies for its suite of tools to visualize travel trends using the historic archive of data.
- INRIX data purchases of $100,000 or more (such as O-D data) are given free access to the RITIS platform.
- Agencies can contract with CATT Lab to ingest additional data sets and potentially develop analytics; for example, MTA is currently paying to house an archive of transit automatic vehicle location (AVL) data, although nothing is being done with this data currently.

Therefore, regardless of the selected Big Data products for procurement, the TPB could consider investing in the RITIS platform to fuse existing and new sources of data for further analysis.

Ultimately, the project team has recommended moving forward with submitting an RFI to StreetLight and Replica, which does not require further investment in RITIS. However, as noted, the TPB would continue to utilize RITIS to access INRIX speed and congestion data.

[2] Interview with INRIX Sales Representative.

1.4.3 OPTION EP-3: MOONSHADOW/DB4IOT

Moonshadow and its database and visualization suite, DB4IoT, are fairly novel in the United States and have generally been used on the west coast and in Europe. While less tested and mature than some of its competitors, DB4IoT does offer some unique capabilities:

- Similar to RITIS, DB4IoT currently ingests data from several providers including INRIX, Unacast, public agency transit data, micromobility companies, WiFi/Bluetooth field data collection devices, and traffic counters.
- DB4IoT offers the ability to ingest wejo connected vehicle data, although traffic signal analytics are not built into the platform.
- The competitive advantage for this platform is its processing speed, which allows queries and analyses to run substantially faster than competitor tools (e.g., StreetLight product platform and RITIS).
- The research team feels that this product currently offers the most potential to users who are very familiar with the underlying datasets and comfortable setting up custom analyses and viewing outputs as a heat map; its end-user analytics features are not as refined and developed as a platform such as RITIS. It appears to be most applicable for project-specific use cases, rather than system-wide performance monitoring at this time.
- Agencies purchase separate underlying datasets and then pay Moonshadow a set-up fee plus a percentage of the data purchase cost (typically 30 percent, although the pricing would be structured differently if agencies are coming in with data that they've already purchased separately but need Moonshadow to process).

Thus, Moonshadow DB4IoT also could be considered as an analytics platform if the TPB elected to move forward with INRIX data, albeit almost certainly less cost-effectively than using RITIS. Further discussions with Moonshadow could be conducted to determine if other Big Data products could be ingested within their platform.

Ultimately, the project team did not recommend moving forward with Moonshadow/DB4IOT at this time given the limitations of the platform compared to other novel platforms such as Replica.

1.4.4 OPTION EP-4: TRANSIT-SPECIFIC PLATFORMS

Another investment option targeting transit-specific research areas is focused on transit-specific end-user data analysis platforms. As mentioned in **Section 4.3**, the TPB serves as the facilitator between regional transit agencies, and within this role, if transit data could be collected from all agencies, an investment could be made into a transit-specific data analytics platform to consolidate the data on a regional level.

- **Suboption EP-4 (A)** | Swiftly is a vendor offering transit analytics using agency AVL data feeds. Their modules track transit headways, speeds, travel times, and runtimes by route, stop, day, and trip metrics that are monitored routinely or in real-time. They assist public transit agencies in enhancing their transit service by analyzing on-time performance and identifying operational issues.
    - Swiftly is currently being used by MTA in Baltimore, Maryland.
- **Suboption EP-4 (B)** | Moovit is a mobility software company that offers both user-facing products (e.g., rider trip planning tools) and analytics for transit agencies. The platform offers insights into multimodal mobility within a region as well as transit O-D matrices, travel times, and modal splits. Their user-facing phone application is among the world's most downloaded in the urban mobility category. Due to their multimodal focus, they have experience working with MPOs and transit agencies.

- In early 2020, Moovit announced a partnership with Cubic Transportation Systems and will be working with transit agencies (i.e., including transit agencies in the metropolitan Washington region.) to unveil an integrated suite of tools.
- Though they favor annual contracts as revenue streams, they have mentioned in the past to be willing to do small pilot projects at no cost.

At this time, TPB staff and partner agency input has led the project team to focus on recommendations for procurement of mobile device-based travel patterns datasets; these additional transit-specific platforms could be explored in the future.

1.5 EMERGING BIG DATA SOURCES

As discussed in **Section 3.2.5**, there are a number of emerging Big Data products that hold potential for future research but have not matured enough for significant use at this time. Therefore, the research team recommends that the TPB take a proactive approach to monitoring the development and progress of these data sources and ensuring regional coordination on their applications.

- **Connected and Automated Vehicle Data (CAV)** | CAV technologies are an area of active research and development. Most experts in the field expect adoption of CAVs to result in significant changes to travel patterns and travel behaviors. However, currently few sources of reliable and representative data exist to represent these anticipated changes.
- **Micromobility Data/Mobility Data Specification**. During the past few years, a variety of private companies have deployed fleets of scooters and bicycles that can be rented via mobile apps. Transportation professionals across the country have desired to have data on the use of these micromobility services to better understand their use and interactivity with other modes. This area may rapidly evolve given data sharing requirements from jurisdictions, proposed data standards, and a wide variety of mobility platforms (e.g., scooters, bicycles) and uses (e.g., transportation, food delivery, package delivery). The more jurisdictions coordinate to develop common data standards and reporting, the more useful this data will be for agencies in the region. The micromobility data obtained in compliance with mobility data specification (MDS) can be integrated to end-user platform for data analytics for real-time reporting along with other Big Data streams.
- **Transportation Network Companies Data**. Third-party companies, such as SharedStreets, are building software to support new ways of managing and sharing TNC and mobility data powered by collaboration between ride-hailing companies and governments. This is an alternative to the TNC datasets the ride hailing companies are required by law to share with the governments or the analytics they offer as a mobility tool (e.g., Uber Movement) to help infer travel behavior information using the underlying TNC data stream. Currently DDOT uses SharedStreets platform to view aggregated TNC data (e.g., hotspots for pickups/drop-offs) and help deploy pickup/drop-off zones (e.g., converting parking spaces).

At this time, TPB staff and partner agency input has led the project team to focus on recommendations for procurement of mobile device-based travel patterns datasets; these additional data sources could be explored in the future.