

1. Report No. FHWA/TX-12/0-6659-1		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle SYNTHESIS OF TXDOT USES OF REAL-TIME COMMERCIAL TRAFFIC DATA				5. Report Date September 2011 Published: January 2012	
				6. Performing Organization Code	
7. Author(s) Dan Middleton, Rajat Rajbhandari, Robert Brydia, Praprut Songchitruksa, Edgar Kraus, Salvador Hernandez, Kelvin Cheu, Vichika Iragavarapu, and Shawn Turner				8. Performing Organization Report No. Report 0-6659-1	
9. Performing Organization Name and Address Texas Transportation Institute The Texas A&M University System College Station, Texas 77843-3135				10. Work Unit No. (TRAIS)	
				11. Contract or Grant No. Project 0-6659	
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Implementation Office P. O. Box 5080 Austin, Texas 78763-5080				13. Type of Report and Period Covered Technical Report: September 2010–August 2011	
				14. Sponsoring Agency Code	
15. Supplementary Notes Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration. Project Title: Synthesis of TxDOT Uses of Real-Time Commercial Traffic Routing Data URL: http://tti.tamu.edu/documents/0-6659-1.pdf					
16. Abstract Traditionally, the Texas Department of Transportation (TxDOT) and its districts have collected traffic operations data through a system of fixed-location traffic sensors, supplemented with probe vehicles using transponders where such tags are already being used primarily for tolling purposes and where their numbers are sufficient. In recent years, private providers of traffic data have entered the scene, offering traveler information such as speeds, travel time, delay, and incident information. The question that this research project answered was the following: Whether TxDOT could and should utilize the data offerings by private sector providers to supplement its own data collection efforts and, if so, how. The research also determined the following: <ul style="list-style-type: none"> • What data are available from private providers (either free or for purchase). • What other states are doing with data from private providers. • Opinions of TxDOT decision makers on the utility of these data sources. • How the data should be normalized, combined, and delivered for TxDOT or other public sector partner agency use. • A recommended path for implementing the TxDOT response. 					
17. Key Words Private Sector Traffic Data, SAFETEA0-LU Section 1201 Requirements, Point Sensor Data			18. Distribution Statement No restrictions. This document is available to the public through NTIS: National Technical Information Service Alexandria, Virginia 22312 http://www.ntis.gov		
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of Pages 122	22. Price

SYNTHESIS OF TXDOT USES OF REAL-TIME COMMERCIAL TRAFFIC DATA

by

Dan Middleton, Ph.D., P.E.
Program Manager
Texas Transportation Institute

Rajat Rajbhandari, Ph.D., P.E.
Associate Research Engineer
Texas Transportation Institute

Robert Brydia
Research Scientist
Texas Transportation Institute

Praprut Songchitruksa, Ph.D.
Associate Research Engineer
Texas Transportation Institute

Edgar Kraus, P.E.
Associate Research Engineer
Texas Transportation Institute

Salvador Hernandez, Ph.D.
Assistant Professor
University of Texas at El Paso

Kelvin Cheu, Ph.D., P.E.
Associate Professor
University of Texas at El Paso

Vichika Iragavarapu
Assistant Transportation Researcher
Texas Transportation Institute

Shawn Turner, P.E.
Senior Research Engineer
Texas Transportation Institute

Report 0-6659-1

Project 0-6659

Project Title: Synthesis of TxDOT Uses of Real-Time Commercial Traffic Routing Data

Performed in cooperation with the
Texas Department of Transportation
and the
Federal Highway Administration

September 2011

Published: January 2012

TEXAS TRANSPORTATION INSTITUTE
The Texas A&M University System
College Station, Texas 77843-3135

DISCLAIMER

The contents of this report reflect the views of the authors, who are solely responsible for the facts and accuracy of the data, the opinions, and the conclusions presented here. The contents do not necessarily reflect the official view or policies of the Texas Department of Transportation (TxDOT) or the Federal Highway Administration (FHWA). This report does not constitute a standard or regulation, and its contents are not intended for construction, bidding, or permit purposes. The use of names or specific products or manufacturers listed herein does not imply endorsement of those products or manufacturers. The engineer in charge of the project was Dan Middleton, P.E. #60764.

The United States Government and the State of Texas do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of this report.

ACKNOWLEDGMENTS

This project was conducted in cooperation with the Texas Department of Transportation and the Federal Highway Administration. The authors gratefully acknowledge the contributions of several persons who made the successful completion of this research possible. This especially includes the project director, Ms. Cynthia Flores. Special thanks are also extended to the following members of the project monitoring committee: Mr. David Fink, Mr. Alex Power, Mr. Mike Wulczyn, Mr. Frank Espinosa, and Mr. Wade Odell of the Texas Department of Transportation.

TABLE OF CONTENTS

	Page
1. Introduction.....	1
1.1 Overview	1
1.2 Background	1
1.3 Objectives.....	1
1.4 Organization of the Report.....	2
2. Refine Work Plan.....	3
2.1 Introduction	3
2.2 Project Monitoring Committee.....	3
2.3 Research Team	3
2.4 Kick-Off Meeting Summary	4
2.5 Action Items Based on the Kick-Off Meeting	5
3. Gather Relevant Information	7
3.1 Introduction	7
3.2 Literature Review and Internet Search.....	7
3.3 Assessing Traffic Information Quality.....	38
3.4 Survey of Private Data Providers and Consumers	39
4. Provide Feedback to TxDOT	55
4.1 Introduction	55
4.2 Methodology	55
4.3 Documentation of TxDOT Input.....	57
5. Develop Opportunity Matrix.....	65
5.1 Introduction	65
5.2 SAFETEA-LU Section 1201.....	66
5.3 Strengths and Weaknesses of Each Method.....	67
5.4 Comprehensive Opportunity Matrix	74
6. Investigate Data Fusion.....	79
6.1 Introduction	79
6.2 Past Studies on Data Fusion	80
6.3 Data Fusion Process and Case Study	85
6.4 Summary	88
7. Summary of Findings and Recommendations	89
7.1 Key Findings	89
7.2 Critical Factors for TXDOT.....	91
7.3 Recommendations/Implementation Strategy	96
8. References.....	97
Appendix A. Data Provider Survey Form.....	99
Appendix B. Data Consumer Survey Form	103
Appendix C. TxDOT Survey Form	107

LIST OF FIGURES

	Page
Figure 1. Snapshot of New Mexico’s 511 System Website Showing Integration of Private Sector Data.....	14
Figure 2. LBS as an Intersection of Technologies.....	15
Figure 3. Basic Components of an LBS.....	16
Figure 4. Reference Markers on the State Highway Network at the Intersection of I-10 and I-610 in Houston.....	19
Figure 5. AirSage WiSE.....	21
Figure 6. Cellint’s Deployment in Atlanta, Georgia.....	23
Figure 7. Cellint’s TrafficSense System Overview.....	24
Figure 8. Example of Cellint’s User Interface.....	25
Figure 9. Houston Traffic Conditions.....	28
Figure 10. SpeedInfo Doppler Radar Sensor.....	30
Figure 11. Quality Indices QKZ_1 and QKZ_2 (18).....	38
Figure 12. Quality Diagram (18).....	39
Figure 13. Electronic Mail Template for Initial Contact.....	42
Figure 14. Electronic Mail Template for Follow-Up Contact.....	43
Figure 15. Travel Time Errors Based on ± 5 Percent Speed Error.....	71
Figure 16. Comparative Costs of Detection.....	73
Figure 17. Framework for Data Fusion.....	79
Figure 18. Functional Steps in the Data Fusion Process.....	80
Figure 19. Architecture of the TSMC Traffic Reporting System.....	81
Figure 20. Data Fusion Technique in Traffic State Estimation.....	82
Figure 21. Data Fusion in Overlapping Disciplines.....	83
Figure 22. Data Fusion Technique Classification.....	84
Figure 23. Conceptual Framework for Integrating INRIX TMC Data with TxDOT Reference Mile Marker Data (El Paso as Case Study).....	86
Figure 24. Data Fusion of INRIX and TxDOT Mile Marker—A Microscopic View.....	87
Figure 25. Cost Comparison for PS versus Radar Fixed Sensor.....	93

LIST OF TABLES

	Page
Table 1. Roles of Each Participating University.....	4
Table 2. Summary of Independent Evaluations of the Completed Deployments.....	10
Table 3. Description of Ongoing Deployments.....	11
Table 4. AirSage Services.....	22
Table 5. Cities Served by SpeedInfo.....	30
Table 6. TomTom Services.....	31
Table 7. List of TTN Coverage Cities.....	33
Table 8. TrafficCast Services.....	34
Table 9. Traffic Data Provider Comparison—Business Model.....	35
Table 10. Provider Primary Data Sources.....	36
Table 11. Traffic Data Provider Comparison—Coverage.....	36
Table 12. Traffic Data Provider Comparison – Services.....	37
Table 13. Summary of Historical Data Consumer Survey Results.....	44
Table 14. Summary of Historical Data Available by Provider.....	47
Table 15. Summary of Types of Data Offered.....	49
Table 16. Information Delivery Requirements for Real-Time System Management Information Program.....	52
Table 17. Q1: Percent of Data Collected In-House.....	58
Table 18. Q2: Value Placed by Districts.....	58
Table 19. Q3: Which of the Following Would Most Enhance Traveler Information?.....	59
Table 20. Q4: Rank of Statements for Districts.....	59
Table 21. Q5: Traffic Counts Not Available from Private Sector.....	59
Table 22. Q6: Is Lack of Per-Lane Data a Limitation.....	60
Table 23. Q7: Conditions for District to Consider Real-Time PS Data.....	61
Table 24. Q8: Conditions to Consider Purchase of Historical PS Data.....	61
Table 25. Q9: Use of Historical PS Data if Purchased.....	62
Table 26. Q10: Would Use of TMC be an Impediment to Using PS Data.....	62
Table 27. Q11: Other Examples of Long-Term Opportunities for PS Data.....	62
Table 28. San Francisco 511 Real-Time Pricing Stipulations.....	66
Table 29. Information Delivery Requirements of Section 1201.....	66
Table 30. Provider Primary Data Sources.....	70
Table 31. Summary Comparison of Data Sources.....	73
Table 32. List of Opportunities Considered in the Study.....	74
Table 33. Strength, Weaknesses, and Opportunities of the Private Sector Data in Relation with ITS Application Areas.....	75
Table 34. Relative Importance of Various Governing Factors.....	77
Table 35. Providers and Consumers Providing Input.....	89
Table 36. Summary of Historical Data Consumer Survey Results.....	89
Table 37. Summary of Historical Data Available by Provider.....	90
Table 38. Information Delivery Requirements of Section 1201.....	91

Table 39. Provider Primary Data Sources.....	92
Table 40. Comparison of Performance of Various Data Sources.....	93
Table 41. Relative Importance of Various Governing Factors.....	95

1. INTRODUCTION

1.1 OVERVIEW

Traditionally, the Traffic Operations Division (TRF) and the districts have collected traffic operations data through a system of fixed-location traffic sensors, supplemented with probe vehicles using transponders where such tags are already being used primarily for tolling purposes and where their numbers are sufficient. TxDOT owns and maintains the traffic sensors and toll tag readers and manages the data that come from these systems. In recent years, private providers of traffic data have entered the scene, offering traveler information such as speeds, travel time, delay, and incident information. The question that Research Project 0-6659 should answer is whether TxDOT could and should utilize the data offerings by private sector providers to supplement its own data collection efforts and, if so, how.

1.2 BACKGROUND

Early deployments of the private sector to collect data from probes in the traffic stream and other available sources through about 2005 resulted in undesirable error rates. Past surveys of departments of transportation indicate that what they desire is speeds within 5 mph of actual speeds, incident detection at least 95 percent of the time, and traveler information at least 85 percent of the time.

As of about 10 years ago, private sources were able to generate speed and travel time errors in the ± 20 percent range, but they have improved significantly. Latencies in the reporting of data have also improved over those years and are now about 4 minutes or less. Today the providers are claiming speed errors (i.e., the difference between the estimated average speed and the actual average speed) of less than 5 mph and up time of about 99 percent. Recent large-scale validations indicate that errors are declining as more and more participants in voluntary programs are increasing the number of probes in the traffic stream.

This research project builds on this background and seeks to update the available information pertaining to the data offerings of private sector providers and to learn what other DOTs are doing with the data. This report includes results of a survey of both providers and users to learn what the providers claim and to determine whether the users' experience is consistent with these claims.

1.3 OBJECTIVES

The research objectives seek to determine:

- What data are available from private providers (either free or for purchase).
- What other states are doing with data from private providers.
- Opinions of TxDOT decision makers on the utility of these data sources.
- How the data should be normalized, combined, and delivered for TxDOT or other public sector partner agencies use.

- A recommended path for implementing the TxDOT response.

1.4 ORGANIZATION OF THE REPORT

This research report consists of seven chapters organized by topic. [Chapter 2](#) addresses refining the work plan based on the kick-off meeting with the Project Monitoring Committee (PMC). There were no substantial changes as a result of these discussions. [Chapter 3](#) includes results of a detailed review of literature and the Internet regarding commercial traffic data. All of the major providers of such traffic data have significant information on their individual websites about the sources of their data. This chapter also has a survey of the providers and consumers of private data. [Chapter 4](#) provides a synthesis of the information gathered in Task 2 (reported in [Chapter 3](#)) to TxDOT stakeholders and solicits their input regarding potential use of private sector data. [Chapter 5](#) presents the findings of an investigation into data fusion issues that TxDOT needs to consider when fusing multiple sources of data. It also presents a case study of data fusion for merging private sector data with TxDOT data. [Chapter 6](#) presents a comprehensive opportunity matrix using the survey information from TxDOT gathered in Task 3 along with information from other states that are using either historical or real-time private sector data. [Chapter 7](#) provides a summary and recommendations.

2. REFINE WORK PLAN

2.1 INTRODUCTION

The objective of Task 1 was to make sure that the research team clearly understood what the sponsor, the Texas Department of Transportation, expected from this research project, primarily in terms of project objectives and scope. Task 1 consisted primarily of the kick-off meeting on September 21, 2010, whereby the research team and the Project Monitoring Committee discussed each task of the project.

Traditionally, the Traffic Operations Division and the districts have collected traffic operations data through a system of fixed-location traffic sensors, supplemented with probe vehicles using transponders where such tags are already being used primarily for tolling purposes and where their numbers are sufficient. TxDOT owns and maintains the traffic sensors and toll tag readers and manages the data that come from these systems. In recent years, private providers of traffic data have entered the scene, offering traveler information such as speeds, travel time, delay, and incident information. The question that Research Project 0-6659 should answer is whether TxDOT could and should utilize the data offerings by private sector providers to supplement its own data collection efforts and, if so, how. Specifically, the research should determine:

- What data are available from private providers (either free or for purchase).
- What other states are doing with data from private providers.
- Opinions of TxDOT decision-makers on the utility of these data sources.
- How the data should be normalized, combined, and delivered for TxDOT or other public sector partner agencies use.
- A recommended path for implementing the TxDOT response.

2.2 PROJECT MONITORING COMMITTEE

The PMC for Research Project 0-6659 consists of the following offices:

- Traffic Operations Division—two members (including the Project Director).
- Research and Technology Implementation (RTI)—two members.
- Houston TranStar—one member.

2.3 RESEARCH TEAM

The research team for this project consists of the Texas Transportation Institute (TTI) as the lead, supported by the University of Texas at El Paso (UTEP). Both universities were represented at the kick-off meeting. [Table 1](#) shows the tasks that the research team had proposed and the lead or support roles for TTI and UTEP.

Table 1. Roles of Each Participating University.

Task	TTI	UTEP
Task 1. Refine Work Plan	Lead	Support
Task 2. Gather Relevant Information	Lead	Support
Task 3. Provide Feedback to TxDOT	Lead	Support
Task 4. Develop Opportunity Matrix	Lead	Support
Task 5. Investigate Data Fusion	Support	Lead
Task 6. Prepare Deliverables	Lead	Support

2.4 KICK-OFF MEETING SUMMARY

One of the goals of the kick-off meeting was to address any questions regarding project goals or any new developments that might have occurred between proposal submission and the start of work. For example, the Project Statement mentioned the Metropolitan Planning Organizations' (MPO) use of data by private providers, although real-time traffic is not a typical MPO function. The research team will determine the need to include one or more MPOs in the list of agencies to be contacted in Task 2.

An RTI representative gave some background for this research project, saying that the idea for the research originated in a Research Oversight Committee (ROC) meeting a year prior to the project's beginning. Even though TxDOT collected traffic data at fixed sites across the state, the TxDOT data did not meet certain data needs. With the increasing availability and use of private-sector data across the country and limited knowledge of how TxDOT might best take advantage of its availability, key TxDOT decision makers put forth this research idea as a Research Management Committee (RMC) project. The idea received a relatively high ranking and ultimately resulted in this research.

The TTI/UTEP proposal suggested including the following states to be contacted as part of the research effort to determine their experience with private-sector data:

- Alabama.
- California.
- North Carolina.
- South Carolina.
- New Jersey.
- Wisconsin.

The PMC recommended adding the State of Florida to this list. A member of the research team suggested adding Washington State. The project agreement states that the research team will contact five of the states.

A PMC member recommended that the research team review Section 1201 of the Safe, Accountable, Flexible Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU) regarding the Real-Time System Management Information Program in conducting this project.

This Section 1201 is now designated as 23CFR511. There was also a suggestion that the research team investigate the progress of 511 activities included in the Integrated Corridor Management (ICM) deployment in Dallas. The TTI Dallas office was involved in the design and implementation of ICM on the US 75 corridor. The research team offered to discuss this deployment with the TTI Dallas group to find out if the latter is planning to use private data as part of the ICM deployment strategy. The research team acknowledged these suggestions by adding them to the early list of project activities.

The research team learned recently that some of the private providers will allow potential users to go onto their website and download sample data. Historical data might be aggregated, so some of the details are missing. Among the sources of private-sector data, there was mention of freight data collected by the American Transportation Research Institute (ATRI) and future origin-destination data that INRIX is planning.

In preparation for beginning Task 5, UTEP researchers hired a graduate student to work on the project. They conducted their part of the research in parallel with other tasks (especially Task 2).

There was a question as to whether researchers would only consider real-time data or if historical data should also be part of the research. The project title includes the words “real-time” so it should take priority. The PMC members agreed, however, indicating that researchers should not exclude historical data.

There was a desire expressed to develop a website for the project so PMC members could have easy access to files and findings. The persons that had access were PMC members and researchers. Subsequent meetings (e.g., between the Project Director and researchers) were documented and made available to the PMC.

There was discussion during the kick-off meeting pertaining to Task 3, Provide Feedback to TxDOT, and the most effective means of getting input from TxDOT decision makers. One option was to physically travel to selected locations to achieve the desired geographic coverage since TxDOT personnel from different areas will likely view the use of private data in different ways. However, a consensus indicated that a better use of project resources involved webinars using the TTI Polycom system. The plan would likely involve two meetings, with the second being a backup meeting for those who could not attend the first one.

2.5 ACTION ITEMS BASED ON THE KICK-OFF MEETING

Here are the changes made on the research plan during the kick-off meeting:

- The research team added MPOs to the list of possible categories to contact.
- The State of Florida and Washington State were added to the list of potential contacts (with the idea that the total number would still be five).
- Researchers investigated Section 1201 of SAFETEA-LU and 511 activities included in the Integrated Corridor Management deployment in Dallas.

- The research team is including both real-time and historical data in this investigation, with emphasis on real-time data.
- Researchers will determine the best means to share information with the PMC during the course of the project.
- Getting input from TxDOT in Task 3 will utilize TTI webinar capabilities.

3. GATHER RELEVANT INFORMATION

3.1 INTRODUCTION

This chapter presents the results of efforts supporting Task 2, Gather Relevant Information, with its three subtasks:

- Subtask 2a. Identify and collect relevant literature.
- Subtask 2b. Interview states with experience with private data providers.
- Subtask 2c. Interview private providers of traffic data.

This document includes background information on the project and task, followed by the results of the literature search and surveys conducted as the main component of this work effort. The report concludes with an appendix containing the survey documents used in support of the work efforts.

3.2 LITERATURE REVIEW AND INTERNET SEARCH

The following information on users and providers of private sector data comes from a literature search and an Internet search. The latter search focused on data provider websites.

3.2.1. Use of Private Sector Data by Other Agencies

Washington State Department of Transportation (WSDOT) examined the accuracy and reliability of private sector data in the Puget Sound Metropolitan Region during the spring of 2007 (Hallenbeck, et al., 2007). Due to a non-disclosure agreement, no information in the report identified the data provider. The research team did not receive any information on the number of probes traveling on roadways of interest, the composition of the probe vehicle fleet, or the specific techniques used to obtain the probe fleet data. Therefore, they could not test the accuracy of the data against ground truth speed and travel time statistic. They also could not judge the combined effects of probe vehicle location/speed data collection methodology data availability and data fusion on the accuracy of reported statistics. The project team conducted the following two tests:

- Compare data from GPS-equipped floating car runs with the privately reported data.
- Compare WSDOT fixed sensor (inductive loop) data with the fused private data.

The overall finding of the study was that private sector data provided overly conservative estimates of roadway speed and performance. The data appeared to underestimate vehicle speed in some conditions and/or locations. However, the private data used for the test should be sufficient to provide a basic indication of congestion (green/yellow/red indications). At the time of the study, the project team did not recommend the use of private data for arterial performance unless the following conditions were met:

- Sufficient probe data were available to provide reliable speed estimates on the roads.
- The vendor used data fusion algorithms that could handle the increased variability in vehicle speeds present on arterials as a result of traffic signals.
- Agencies are open to additional testing of these data sources as improvements are made.

The Texas Transportation Institute conducted a study in 1993 to determine the feasibility of a real-time travel information system involving a public private partnership, with focus on four roadways; US 59, I-45, I-45 HOV lane, and the Hardy Toll Road, in the north corridor of Houston during the peak period (Smalley, et al., 1993). The study recruited 200 probe vehicles to commute to or through the central business district (CBD) during the peak periods and travel through at least four of the seven predetermined station locations on each freeway. Probe drivers used free cell phones to make a brief report to the study office as they passed each station.

The vendors participating in the study were: Metro Traffic Central, Shadow Traffic, and Infobanq Inc. Overall, the study found that the travel time data were sufficient to provide reliable real-time traffic information during the peak period (about two hours). However, the study was unable to obtain a sufficient sample and uniform distribution for the four hours of a.m. and p.m. peak periods. The participation rate of the 200 probes over the one-year period was over 80 percent. The quality and quantity of in-vehicle information necessary to induce diversion was not covered in-depth in this study. Limited results indicated that drivers were reluctant to divert without receiving detailed information on the reason for diversion.

A 2007 study evaluated the procurement procedure for vehicle probe technology through use of the I-95 corridor coalition. The partnership area for the I-95 corridor coalition extends from Maine to Florida (including Washington, DC), along with affiliate members from Canada (Young, 2007). The mission of the coalition was to:

- Reduce congestion.
- Increase safety/security.
- Ensure that the entire transportation network supports economic vitality throughout the region.

Construction and maintenance of the system was outsourced with no particular technology specified for the system. Instead, the Request for Proposals (RFP) was open to any system that would support the broad range of Advanced Traveler Information System (ATIS), Advanced Traffic Management System (ATMS), engineering, and planning applications of the coalition and its members without deploying additional infrastructure in the right-of-way. The evaluation methodology focused on risk management (for both the vendor and the coalition) and demonstration of ability to meet technical specification. Vendors could take advantage of data from existing systems that relied on field assets such as inductive loops, radar, and toll-tag systems. Some highlights of the procurement process were:

- Specifications regarding data quality were determined based on the intended uses of the data. The specifications limited the error in reported speed (and associated travel time) under varying roadway conditions.
- An independent agent had to validate the data service on behalf of the coalition.
- The vendor had to supply a risk assessment. Service that was dependent on third party contracts required evidence of the sustainability of such contracts.

The vendor retained full ownership of data for resale in the commercial market. Minimum data rights are defined to support the intended applications within the coalition. Vendors could propose additional restrictions (or fuller rights to the data) in the proposals. Any additional data rights (or restrictions) were assessed as part of the RPF evaluation process.

The vendor could provide data using any one of a number of common formats, technologies, and data standards. However, the vendor must be able to transform or translate that format into whatever format is needed for integration into coalition members' data systems as part of ancillary consulting services. The ability to transform the data format into Intelligent Transportation Systems (ITS) standard protocols was required.

The base contract (and associated funding) covered the first three years, with options to renew for an additional seven years. Coalition members will provide the supplemental funding to extend the contract beyond the initial three years. Supplemental funding was not guaranteed but based wholly on the success of the project and its critical role in corridor operations.

Coverage included I-95, beltways, parallel freeways, parallel signalized arterials, cross-linking freeways, and cross-linking arterials. The coalition preferred full coverage on all road classes for a limited geographical area rather than coverage of only higher class facilities along the whole corridor. Evaluation and award of the contract was based on the best value for the coalition.

A 2010 study validated data that INRIX provided for the I-95 coalition ([I-95 Corridor Coalition, 2010](#)). Initiated in July 2008, this vehicle probe project sought to collect, archive, use, and evaluate speed data in several states. The University of Maryland was responsible for evaluating the quality of data from 5100 miles of freeway in NJ, DE, PA, MD, DC, VA, NC, SC, and FL. The evaluation involved a comparison of INRIX data and Bluetooth data. The data accuracy specification for the project stipulated that the average absolute speed error should be less than 10 mph and the speed error bias should be less than 5 mph. The study found that, in general, the data satisfied the contract specifications; and as the amount of data that INRIX acquired increased, the data quality also improved.

A 2005 study explored the capabilities and limitations of wireless location technology (WLT) (both point location systems and handoffs-based systems) to collect high-quality vehicle probe data by reviewing completed and ongoing deployments in the United States and abroad ([Smith, et al., 2005](#)). The study also included a survey of transportation professionals and a legal analysis. [Table 2](#) and [Table 3](#) summarize the findings from completed deployments, ongoing deployments, and WLT simulation studies.

Table 2. Summary of Independent Evaluations of the Completed Deployments.

Location	Year	Vendor	Performance Requirements	Type of Technology	Results
Washington, DC	1994–1997	Raytheon, Farradyne, Bell Atlantic.	No	WLT signal analysis using triangulation.	Only 20% of probes generated speeds. Could not consistently monitor traffic.
San Francisco and Oakland	2000	U.S. Wireless	No	WLT signal analysis using pattern matching.	60-meter mean location accuracy. 60% of location could not be matched to road. No usable traffic data generated.
Washington, DC	2000–2001	U.S. Wireless	No	WLT signal analysis using pattern matching.	5% of 10-min. intervals had no data. 6 to 8 mph mean speed estimation error. Some intervals had errors > 20 mph.
Lyons, France	2001	Abis/A	No	Unclear	Good agreement at one site, speed overestimated by 24%–32% at another.
Munich, Germany	2003	Vodafone	No	Handoff-based analysis.	Errors between 20 and 30 km/h common.
Hampton Roads, VA	2003–2005	AirSage	No	Handoff-based analysis.	68% of speed estimates had errors > 20 mph. No reliability measures could be generated.
Tel Aviv, Israel	2005	IT IS	No	Handoff-based analysis.	Limited data during off-peak hours. WLT estimates different from floating car and loop data by 10–30% during congested conditions.

Source: (I-95 Corridor Coalition, 2010)

Table 3. Description of Ongoing Deployments.

Location (Vendor)	Atlanta, Georgia (Airsage)	Missouri (Delcan)	Baltimore, Maryland (Delcan)	Kansas City, Kansas (Cellint)	Atlanta, Georgia (Cellint)
Start Year	Summer 2005	Dec 2005	Sept 2004	Feb 2006	
Budget	\$ 750,000	\$3.075 million/yr. for 2 years	\$1.9 million (from a federal earmark, matched with funds from private sector partners).	(A no-cost agreement between the two groups: KC provided personnel to independently evaluate the system and Cellint performing a no-fee deployment of the system).	<\$50,000
Objective	To evaluate the feasibility of using cellular probe to derive traffic data (an R&D effort).	To provide link-based travel time and speed data on a (color coded) condition map.	An operational test of public-private partnership (PPP) to deploy and operate regional traffic monitoring system in the Baltimore Metropolitan area.	To perform a test deployment of Cellint's Traffic Sense system on a portion of I-45 with active construction project (R&D effort).	Short-term source for a 10-mile section of the GA 400 freeway that serves the suburbs of Atlanta (where the GDOT surveillance infrastructure was made temporarily unavailable due to construction).
Type of Technology	Handoffs WLT	(None specified) Delcan used a system which fused data from WLT and GPS vehicle probe data.	GPS-equipped and WLT-based anonymous probes (same as that used in Missouri).	The system monitors the control channel of cell phones and uses pattern matching algorithms to correlate a vehicle's estimated location to the roadway network.	Same as that used in Kansas.
Performance requirement	None	None	None	None	-Provide trip times and or average speeds along the predetermined segment. -Identify a potential incident within the coverage area. -Isolate the source of incident within 250 m of the location of the incident. Starting from 75 days of NTP, the system will be benchmarked against available video detection sensors.

Location (Vendor)	Atlanta, Georgia (Airsage)	Missouri (Delcan)	Baltimore, Maryland (Delcan)	Kansas City, Kansas (Cellint)	Atlanta, Georgia (Cellint)
Data to be provided	On a 24/7 basis with a 5-minute update rate.	As link-based data with a 1-minute update rate.	At 5-minute intervals on links about 0.5–1.5 mi long.	By link at 2.5-minute intervals.	At a rate of 2.5 minutes and links of 1/3 mile (same as GDOT’s detector system).
Preliminary Results	To be evaluated by a subcontractor to Airsage.	Declan performed test drives to assess the accuracy of the path matching methods they used. According to them, 95% of test drives on the interstates and 85% of the test drives on urban roads were correctly matched. This information is not validated by any third party.	(Conducted by the University of MD) On limited access facilities, link speeds were estimated with 20% error or less, 75% of the time. Path travel time measures had 20% error or less about 90% of the time.	(‘Blind’ comparison between the vendor speeds and KC loop detectors) An average latency in the data of about 4 minutes and average difference in speeds was less than 5 mph and the system was estimated to be functional approximately 99% of the time. Travel time data evaluation was ongoing at the time of this report.	No plans to conduct independent evaluation.

Source: (I-95 Corridor Coalition, 2010)

To evaluate the potential performance of a handoffs-based monitoring system in contrasting situations, one study discussed in the report simulated two different roadway networks in VISSIM. It evaluated the effect of individual roadways, cell size, and probe penetration (i.e., number of cell phones that can be tracked). The results from the general linear model (GLM) analysis indicated that specific roadway, cell size and roadway, and cell size interaction showed a statistically significant impact on speed estimation accuracy. Neither the main effect of probe penetration nor any of its interactions was significant, which means having even one major wireless carrier participate in the project would produce sufficient samples to generate speed estimates.

This study included an Internet survey of transportation professionals (a mix of state transportation agencies, local agencies, the private sector, and researchers), which was designed to assess how agencies might potentially utilize WLT-based traffic monitoring systems and also the desired levels of accuracy and availability of such systems ([I-95 Corridor Coalition, 2010](#)). The less than 20 survey respondents who completed the survey said they would utilize WLT-based systems as follows:

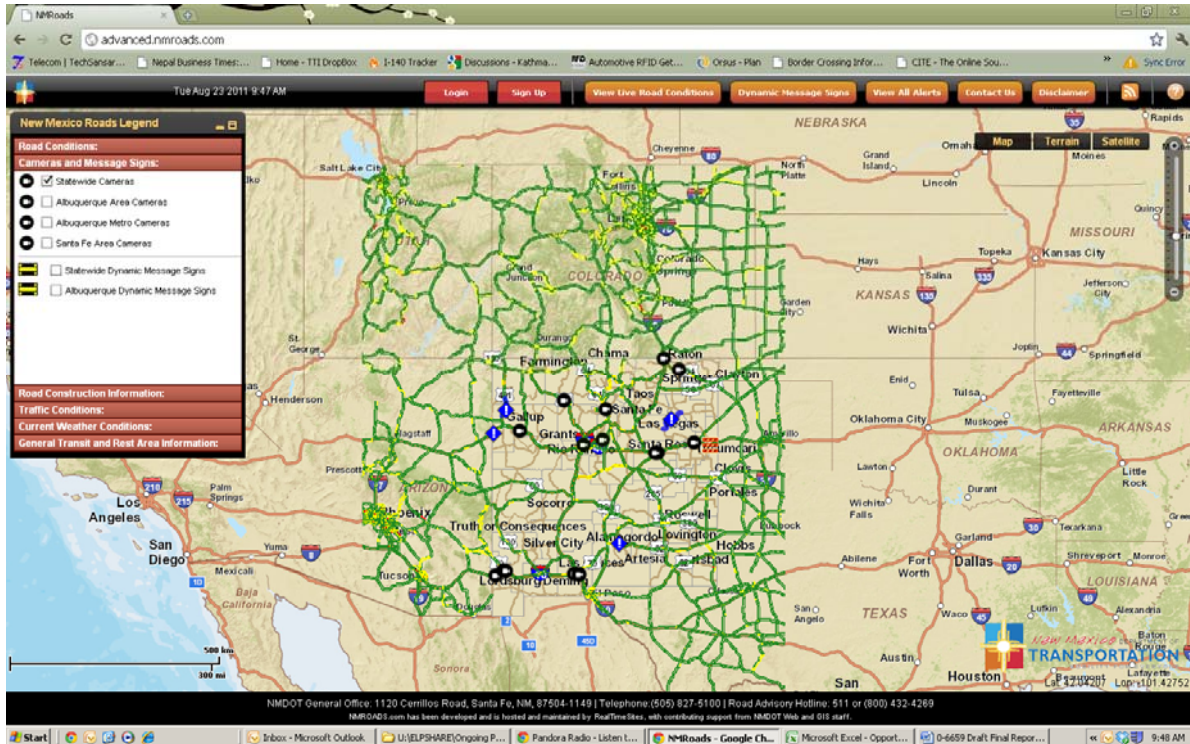
- Performance monitoring and measurement—100 percent.
- Traveler information—70 percent.
- Incident detection, real time control, and planning model—50–65 percent.

Accuracy of freeway traffic condition estimates (speed or travel time) was the preferred measure of effectiveness for evaluating a private sector data system. Respondents preferred data regarding incident detection and real time operations and control is available about 95 percent of the time, whereas traveler information and performance measurements could be available for about 85 percent of the time. The desired speed estimation accuracy that most professionals preferred was 5 mph for most applications.

The legal analysis included in the study found that if the vendor sanitizes individually identifiable Customer Proprietary Network Information (CPNI) in such a way as to transform it into aggregate information, then there would be no legal concerns with the transportation agencies obtaining or using that information ([I-95 Corridor Coalition, 2010](#)). However, individual level data (such as individual origin-destination information) is prohibited without the customer's consent.

Overall, the study concluded that it is feasible to collect freeway data of reasonable accuracy and availability using WLT-based systems (the report documented many deployments that were successful). Also, there were no obvious legal barriers for using this technology for link condition estimates. However, there were risks involved with the lack of maturity of the technology (at the time of the report) and involvement of multiple companies in the data stream. Transportation agencies should have a clear understanding of their needs before entering into such agreements and clearly define them in the service agreements. The research team suggested that a 'competitive demonstration' procurement approach for travel time data service agreements encourages continuing competition and opportunity in the industry.

The New Mexico Department of Transportation (NMDOT) recently deployed a statewide 511 system that includes traffic conditions on state highways and ITS data from urban areas (snapshots from video cameras, dynamic message signs, etc.). Figure 1 indicates the 511 system using INRIX data to show statewide highway traffic conditions..



Source: www.nmroads.com

Figure 1. Snapshot of New Mexico’s 511 System Website Showing Integration of Private Sector Data.

3.2.2. Private Data Providers

This document provides a review of private data providers and their available services. These data providers essentially provide some type of location-based services (LBS). The information service is the type of LBS applications that the researchers specifically examine in this document. The researchers reviewed 10 companies that provide some forms of LBS and traffic data service.

The information in this section comes from both data provider websites as well as publicly available literature. Most data providers rely on either fixed sensors or probe-based vehicles or both to provide real-time traffic information. Several providers also enhance the data by performing quality assurance checks before distributing the data to subscribers.

The purchasers of these data can be direct (such as automakers and media) or indirect (such as portable navigation device [PND] owners). A few providers also can provide both historical and predictive traffic information, taking into account impact factors such as work zones and weather events. In this document, the researchers highlight the findings from the review of the information gathered from various sources. These sources include web articles and published and unpublished documents to provide the most up-to-date assessment of private data providers. The information is divided into the following four categories:

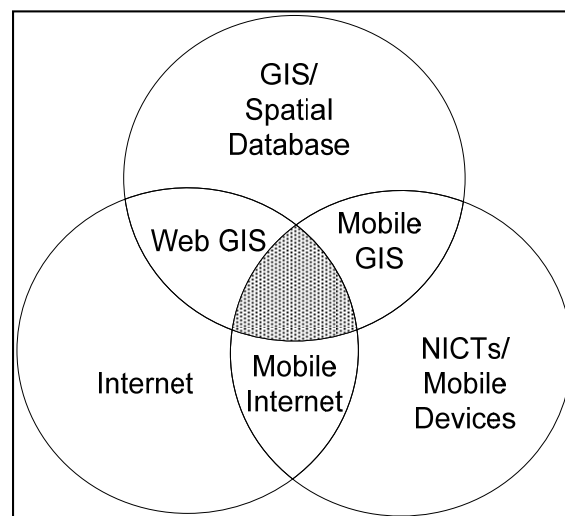
- Business models.
- Data sources.
- Coverage.
- Available services.

The end of this section contains a methodology that BMW developed for evaluating the traffic information quality based on the accuracy and relevancy of the data.

3.2.3. Location-Based Services

Location-based services are information and entertainment services, accessible with mobile devices through the mobile network and utilizing the ability to make use of the geographical position of the mobile device. LBS can be described as an intersection of the following three technologies (see [Figure 2](#)):

- New Information and Communication Technologies (NICTS, e.g., cell phones and handheld devices).
- The Internet.
- Geographic Information Systems (GIS) with spatial databases ([Shiode, et al., 2004](#)).

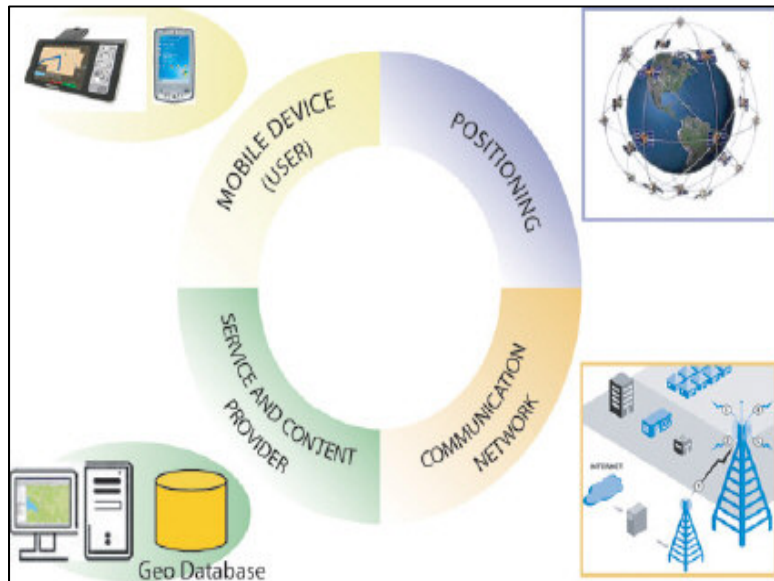


Source: ([Shiode, et al., 2004](#))

Figure 2. LBS as an Intersection of Technologies.

Figure 3 shows the basic components of LBS. The major LBS categories are:

- Emergency services.
- Navigation services.
- Information services.
- Tracking and management services.



Source: (Shiode, et al., 2004)

Figure 3. Basic Components of an LBS.

Emergency Services

One of the most evident applications of LBS is the ability to locate an individual who is either unaware of his/her exact location or is not able to reveal it because of an emergency situation (e.g., injury, hostage situation).

Navigation Services

Navigation services are based on mobile user needs for directions within their current geographical location. The ability of a mobile network to locate the exact position of a mobile user can be manifested in a series of navigation-based services.

Information Services

Location-sensitive information services refer primarily to digital distribution of information based on device location, time, and user behavior. They include finding the nearest service, accessing traffic conditions, checking weather conditions, and so forth.

Tracking and Management Services

Tracking services can be equally applicable both to the consumer and to corporate markets. One popular example involves tracking postal packages so that companies know the status of their shipment. Vehicle tracking can also be applied to locating and dispatching an ambulance that is nearest to a given call. A similar application allows companies to locate its field personnel (for example, salespeople and repair engineers) so that it is able to dispatch the nearest resources and provide its customers with accurate personnel arrival times.

3.2.4. Traffic Message Channel (TMC)

Traffic Message Channel is a means for disseminating information by some of the private data providers. This section contains a summary of expected locational accuracy of TMC data as they relate to the Texas Reference Marker (TRM) system that TxDOT currently uses. TMC is a technology to transmit real-time local traffic and weather data to portable devices, radios, in-car navigation systems, and Internet sites ([Traveler Information Services Association \(TISA\), 2011](#)). TMC can be broadcast via Radio Data System (RDS) or Radio Broadcast Data System (RBDS) technology using frequency modulation (FM) radio broadcasting. Other options for transmission are satellite radio and digital radio broadcasting technologies such as hybrid-digital (HD) radio and digital audio broadcasting (DAB). TMC messages typically include:

- An event description.
- A location.
- The direction and extent.
- The expected duration.
- Diversion advice ([Traveler Information Services Association \(TISA\), 2011](#)).

Once a message is received, the TMC decoder within the device decodes the message and, depending on the device, displays it in the form of text, graphical format, or spoken message. Depending on the TMC device, messages can be filtered to include only information for the route and immediate vicinity of the vehicle using global positioning system (GPS) data.

TMC is primarily a service for freeways and limited access highways. Secondary highways may be included in the coverage for incident reporting but are typically excluded from congestion information.

3.2.5. Private Providers Using RDS-TMC

There are two broadcasters in the United States that deliver traffic information via RDS-TMC technology: Clear Channel Communications and Sirius XM Satellite Radio. Clear Channel Communications started broadcasting over RDS-TMC in 2006, serving numerous metropolitan markets using its own Total Traffic Network that aggregates traffic data from coverage partners. The main coverage partners for the Total Traffic Network are INRIX and SpeedInfo ([Clear Channel Communications, 2011](#)). Clear Channel Communications also began broadcasting TMC via HD Radio in July 2007 ([Clear Channel Communications, 2007](#)). Other

service providers for TMC data in the United States are Navteq and Tele Atlas. Some of the radio programs that carry TMC service do not provide that service for all of its reception areas, which is sometimes referred to as white spots. Private providers use multiple sources to gather traffic data, including:

- State DOT traffic camera, volume, and speed sensor data.
- Proprietary traffic camera data and GPS sensors.
- Airborne or mobile spotter vehicles.
- Digital scanners of emergency services and police callouts.
- Floating car data ([Clear Channel Communications, 2011](#)).

TMC traffic information services are typically available to end-users via subscriptions. Although TMC supports very basic encryption for commercial services as described in ISO 14819-6, encryption is limited to the TMC location code and does not prevent injection ([International Organization for Standardization, 2006](#)).

An alternative to RDS-TMC is a wireless data network based on an FM subcarrier channel called Microsoft DirectBand™ that is used for MSN Direct. DirectBand is a proprietary standard that Microsoft owns and operates. This channel provides a data rate about 15 times higher than RDS and provides full data encryption using the RSA algorithm ([Microsoft Corporation, 2008](#)). Microsoft has announced the discontinuation of the service on January 1, 2012. However, the technology may be sold or could be put into the public domain as an open source technology.

3.2.6. TMC versus TxDOT's Current Process

RDS-TMC uses a very low-bandwidth system that was primarily designed for FM radio tuning and station name identification. Each message consists of only 37 data bits, of which 16 bits are allocated to describe the location of an incident. Because of this limitation, RDS-TMC cannot provide latitude and longitude information, but instead relies on location tables with up to 65536 (16 bit) separate locations to describe an area, state, or county. Depending on the size of the area, location information may be somewhat imprecise and limited to providing the next major intersections or freeway exits. However, this is typically adequate for navigation systems to find an alternate route.

3.2.7. Texas Reference Marker System

The Texas Reference Marker System is a mainframe-based system that documents physical and performance characteristics of the state-maintained highway network using the statewide reference marker network as a geo-referencing tool ([Figure 4](#)). With TRM, the location of features on the ground is defined in terms of mileage displacement from the nearest marker. TRM is centerline based, although it does provide for the identification of features on either side of the centerline. Even though TRM relies on displacement from markers as the mechanism to reference features to the highway network, the system also enables the calculation

of cumulative distances by using the relative location of the markers along the highway network. This conversion enables the production of maps documenting feature locations and characteristics in a GIS environment.

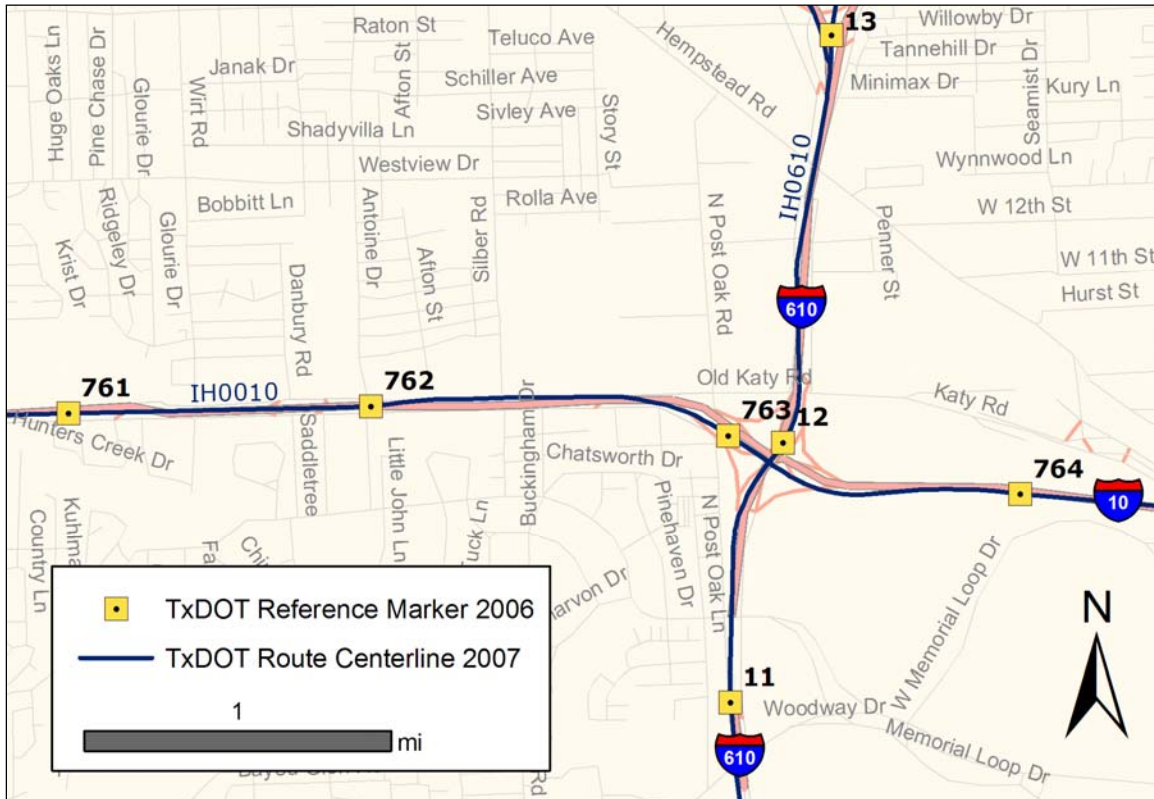


Figure 4. Reference Markers on the State Highway Network at the Intersection of I-10 and I-610 in Houston.

TRM is currently the major repository of state highway network and associated data. Examples of roadway attribute data include annual average daily traffic (AADT), classification, surface type, location of features (e.g., culverts, signs, and streams), and administrative data (e.g., county and district). TPP produces a variety of data files based on TRM. Examples are:

- The Roadway/Highway Network Inventory (RHiNo) file.
- The Point file.
- The GEO-Point file.
- The GEO2-HINI file.
- The TRM end-of-year (TRMEOY) file.

Likewise, several TxDOT asset management systems rely on TRM data. Examples are:

- The Pavement Management Information System (PMIS).
- The Highway Performance Management System (HPMS).
- The Bridge Information System.

While TRM provides data for a wide range of reporting options, the structure and characteristics of the data have shortcomings that limit the usability of the system, particularly during the project development process. For example, TRM is centerline based, which means the positional accuracy of any feature or measure (such as beginning and ending of a utility line or project limits) cannot be better than the positional accuracy of the underlying centerline map. TRM is also cumulative distance dependent, which means the positional accuracy of any feature or measure cannot be better than the longitudinal positional accuracy of *both* reference markers and the underlying centerline map. As a result, it is difficult to determine the actual location of features using cumulative distances alone.

3.2.8. Traffic Data Providers

The following private data providers have useful website information:

- Airsage.
- Cellint.
- Delcan.
- INRIX.
- NAVTEQ.
- OnStar.
- TomTom.
- Total Traffic Network (TTN).
- TrafficCast.

AirSage

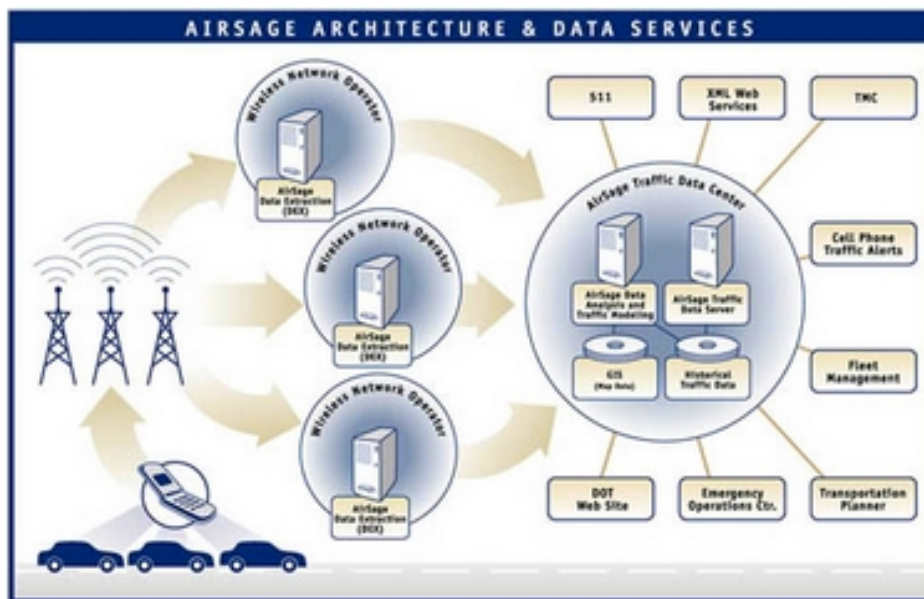
AirSage anonymously collects and analyzes wireless signaling data and then converts the data into meaningful information such as traffic condition and cellular signal quality (<http://www.airsage.com>). Consumers, businesses, government agencies, and other organizations can use this aggregated information to model and analyze the location and movement of people and assets. AirSage product categories include traffic information as well as performance monitoring and analysis of cellular network.

AirSage's patent-protected Wireless Signal Extraction (WiSE™) technology aggregates, anonymizes, and analyzes signaling data from individual handsets using the cellular network, determines accurate location information and converts it into real-time anonymous location data. In essence, each individual handset becomes a mobile location sensor, allowing AirSage to determine how phones move over time. The AirSage technology works by mining handoff data

that cellular service providers collect. AirSage processes data on cellular handoffs, as well as transitions between sectors of a cell, to estimate a vehicle's location on the roadway network. It then uses these locations to determine speed and travel time information on the network (Smith, et al., 2005).

Several subsystems work together to create the WiSE™ system architecture and data services (see Figure 5). The AirSage Data Extraction Subsystem (DEX) resides within the wireless carrier environment and is responsible for aggregating and anonymizing network data. Patented technology and multiple layers of privacy protection ensure that there is no release of proprietary, customer-identifying data from the secure environment of the wireless carrier. Inside the AirSage Data Center, the AirSage Data Analysis Subsystem receives data from the DEX(s) and uses proprietary algorithms to convert it into location and movement information. The AirSage Content Delivery Platform packages the information and delivers it to the customer.

AirSage currently provides real-time, historical, and predictive traffic information for 127 cities, which cover approximately 85 percent of the U.S. population. Table 4 indicates the data service that AirSage offers.



Source: www.airsage.com
Figure 5. AirSage WiSE.

Cellint

Cellint is a global provider of traffic information systems based on mobile networks (<http://www.cellint.com>). Cellint's proprietary patented technology utilizes pattern matching analysis on anonymous, real-time data extracted from the signaling links of mobile networks for all active mobile phones. This platform is used by cellular operators to provide road traffic

information services and optimize their RF performance. The high resolution traffic information service is available to multiple user groups such as:

- Mobile phone subscribers.
- Navigation providers.
- Government agencies.
- Road operators.
- Mapping and media portals.

Table 4. AirSage Services.

Services	Descriptions
Real-Time Traffic	AirSage real-time traffic is an XML-based data feed showing traffic flow data and several other outputs. Specifically, the traffic data feed contains: date and timestamp, market, TMC code, actual speed, and a historical mode value. Customers have the option to purchase a real-time visualization of their traffic market to avoid the need to integrate the data feeds into the visualization page themselves.
Movement Analytics	This service provides anonymous and aggregated cellular movement analysis, which clients can use to generate new insights for various business needs such as time-sensitive population and market segmentation, population movement patterns for real estate site selection, and travel time trends for transportation planning.
Network Analytics	AirSage can analyze signal data for carrier networks that, in turn, can provide information such as dropped/lost call density mapping, signal quality density mapping, and traffic density mapping.
Consulting Services	In addition to providing traffic and cellular signal quality data, AirSage also provides various types of consulting services. These services pertain to integrating real-time data or understanding the impact of aggregate movements of people such as traffic counts, travel time forecasts, speed surveys, incident reports, and origin-destination analysis.

Cellint’s coverage is limited to the agencies/regions that deploy Cellint’s system. Atlanta, Georgia, is one of the cities that deployed Cellint’s system. The website with the display of traffic data is located at <http://www.georgia-navigator.com/maps/atlanta>. Figure 6 shows the web-based interface of traffic data collected using Cellint’s system in Atlanta.

For traffic monitoring service, Cellint refers to its product as TrafficSense. TrafficSense connects to cellular network switching centers and incorporates anonymous signaling data to provide traffic information for all highways, arterials, and surface streets. The main raw deliveries of TrafficSense are:

- Incident alert—similar to road sensors.
- Travel time.
- High-resolution local speed—similar to road sensor.

The TrafficSense platform includes the following service modules:

- Integration with external data sources, such as police reports and GPS probe data, with both manual and digital inputs.
- Incident report database that analyzes each event in the flow data, and reports all relevant details, including related links and junctions, start point, end point, delay, and level of severity of each slowdown/incident.
- Reporting mechanism through TMC format, text, and voice for navigation systems, mobile phones, web, and media.
- High-Definition traffic flow data interface for integration with dynamic map data.
- Synthetic map reports of the traffic flow for web/mobile phones.
- Interactive Voice Response (IVR) interface for mobile phone.

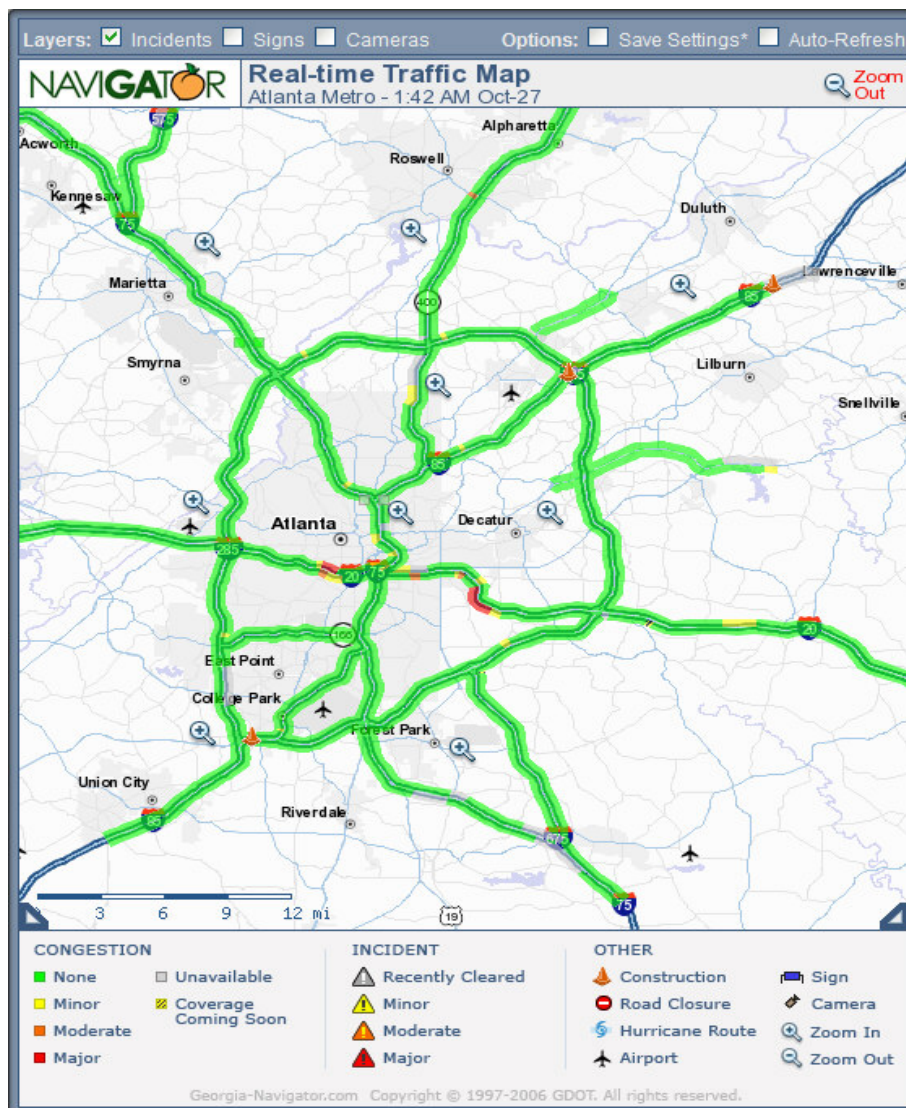


Figure 6. Cellint's Deployment in Atlanta, Georgia.



Source: www.cellint.com
Figure 8. Example of Cellint's User Interface.

Delcan

Delcan provides a broad range of integrated systems and infrastructure solutions. Delcan's expertise areas focus on three groups: transportation, water, and information technology (<http://www.delcan.com>). On the transportation side, Delcan has ventured into the traffic data service. The Delcan system uses partner ITIS Holdings' Wireless Location Technology (WLT), which fuses data from WLT-based monitoring and GPS vehicle probes. The WLT system relies on monitoring cellular hand-offs, then uses pattern recognition software and an accompanying traffic model to estimate speeds. The system also uses a number of GPS-equipped vehicle fleets to augment the cellular phone data. These include national fleets, parcel delivery companies, and taxi companies (Bogenberger, 2003).

INRIX

INRIX provides traffic information by fusing the data from GPS-enabled vehicles and mobile devices, traditional road sensors, and other sources (<http://www.inrix.com>). The majority of INRIX data come from crowd-sourced GPS-enabled vehicles (currently more than 2 million vehicles). The crowd-sourced traffic network is built on a foundation of commercial fleets—taxi cabs, delivery vans, and long-haul trucks and a growing number of consumer vehicles and mobile devices. INRIX products are classified into five different groups:

- Automotive solutions.
- Mobile solutions.
- Public sector.
- Fleet solutions.
- Internet and media.

INRIX provides coverage in countries across North America and Europe:

- Two hundred sixty thousand miles of real-time traffic flow on highways and arterial roads in the United States and Canada.
- One million kilometers of real-time traffic on primary and secondary roads across 18 countries in Europe.

INRIX currently offers the following traffic information services:

- Real-time traffic flow.
- Predictive traffic flow.
- Total fusion traffic flow.
- Historical traffic flow.
- Congestion alerts (available for 20 countries across North America and Europe).
- Journalistic incident data (based on strategic partnerships with regional incident providers including automobile clubs across Europe and Clear Channel Total Traffic Network in North America).
- Traffic maps—images and vectors representing color-coded traffic conditions on road segments.
- Traffic cameras—images from traffic cameras can be sent to mobile phones, displayed on a web page, or viewed on a connected navigation device. This service is provided via exclusive partnership with Vizzion (aggregator of traffic camera images).

NAVTEQ

NAVTEQ is a provider of maps, traffic and location data (digital location content) enabling navigation, location-based services and mobile advertising around the world. The company is a wholly-owned subsidiary of Nokia, but operates independently (<http://www.navteq.com>). NAVTEQ digital maps provide a representation of the detailed road network including up to 260 attributes such as turn restrictions, physical barriers and gates, one-way streets, restricted access, relative road heights, addresses, signage, and speed restrictions. NAVTEQ database also contains millions of Points of Interests (POIs) for routing applications.

NAVTEQ Traffic is the company's traffic solution product. NAVTEQ started delivering real-time, personalized traffic linked to the map in a navigation system in 2004 (<http://www.traffic.com>). In North America, NAVTEQ Traffic business models are flexible to support different products and pricing scenarios as application requires. NAVTEQ traffic provides real-time traffic services that include:

- Comprehensive nationwide coverage.
- Traffic flow conditions with speed values.
- Unplanned incidents such as accidents and stalled vehicles.
- Planned incidents such as road construction and closures.
- Traffic data designed for seamless integration with digital map content.

NAVTEQ maintains its own map database, street level imagery, and utilizes multiple sources of data to provide traffic information including:

- Proprietary sensors: NAVTEQ operates the world's largest proprietary sensor network, and its highly accurate sensor network covers 35 percent more roadway than the nearest competitor.
- Probe: Robust commercial and consumer GPS and cellular probes improve coverage and accuracy.
- Data validation: Proprietary data corroboration and verification methods check and recheck data accuracy.
- Data processing: The NAVTEQ Smart Traffic Processor™ blends and optimizes the widest array of traffic information to provide comprehensive and reliable real-time traffic information. Unique processing capabilities combine and prioritize multiple data sources to provide the most accurate speed values possible.

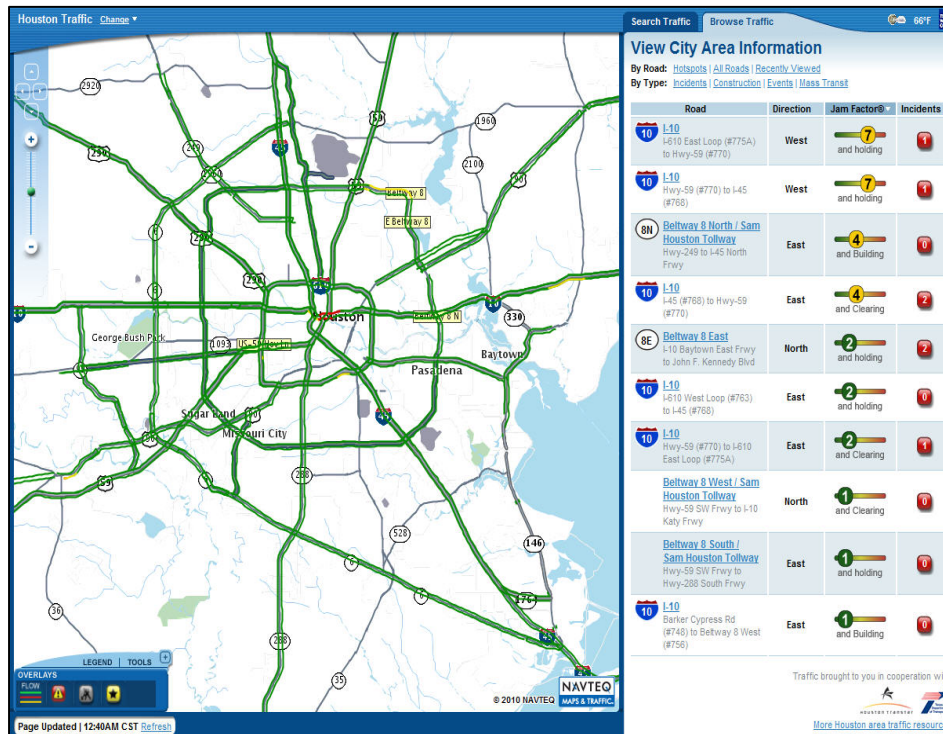
Street-level imagery refers to panoramic images enabling users to look around a 360 degree photo. These images are typically used to supplement maps, directions, and local search, and are collected using specially equipped vehicles outfitted with cameras. Street-level imagery is generally served from a website and uses a proprietary viewer embedded in a standard internet browser.

The primary focus of street-level images is on public roads of commercial interest and areas with high concentration of Points-of-Interest (POIs). Other than efforts to support this POI collection, driving in residential and suburban areas is limited.

NAVTEQ uses proprietary algorithms to detect and blur portions of images that may raise privacy concerns, including faces and license plates. The automated software and advanced algorithms are made available from the partnership with Microsoft in computer vision techniques. As with all automated approaches, there is a chance that some images may come through that should have been blurred. In those instances, NAVTEQ is providing internet-based services for individuals to identify and report concerns regarding any published images.

NAVTEQ accepts requests to blur or remove images of faces or persons, homes, license plates, acts of violence, nudity, and unlawful material. Depending on the nature of the reported image, the turnaround time to change the database is generally a few days. The original unblurred street-level imagery is retained for one year or in accordance with local laws in order to assist with continued development of street-level imagery technology and privacy algorithms.

NAVTEQ provides real-time traffic information in many major metropolitan areas through <http://www.traffic.com>. This service is currently available for North America. NAVTEQ Traffic is currently providing real-time traffic for 52 city areas in North America. [Figure 9](#) shows an example of Houston traffic conditions.



Source: www.traffic.com
Figure 9. Houston Traffic Conditions.

NAVTEQ Traffic Product Portfolio includes an array of traffic distribution channels, providing traffic information in various formats depending on customers' needs:

- NAVTEQ Traffic Satellite™: NAVTEQ is the exclusive provider of real-time traffic services for all satellite radio providers who offer traffic information via satellite radio in North America.
- NAVTEQ Traffic RDS™: Real-time traffic delivered over FM radio using a radio data system (RDS) sub-carrier channel. RDS is well-suited for auto companies and PND manufacturers.
- NAVTEQ Traffic ML™: Real-time traffic designed for mobile and server-based navigation, as well as mapping applications.
- NAVTEQ Traffic Online™: Real-time traffic available via consumer traffic websites.
- NAVTEQ Traffic Digital™: Real-time traffic delivered over digital radio's high bandwidth will mark a major leap forward as additional data services beyond traffic become available.

OnStar

OnStar Corporation (<http://www.onstar.com>) is a subsidiary of General Motors that provides subscription-based communications, in-vehicle security, hands free calling, turn-by-turn navigation, and remote diagnostics systems throughout the United States, Canada, and China.

OnStar services are only available currently on vehicles manufactured by General Motors and Saab Automobile. The service is available for all vehicles that have the factory-installed OnStar[®] hardware. A similar service, known as ChevyStar[®] in Latin American markets, is also available. The service currently has more than 5 million subscribers (<http://en.wikipedia.org/wiki/OnStar>).

The OnStar service relies on Code Division Multiple Access (CDMA) mobile phone voice and data communication, primarily via Verizon Wireless in the United States and Bell Mobility in Canada, as well as location information using GPS technology. Drivers and passengers can use its audio interface to contact OnStar representatives for emergency services, vehicle diagnostics, and directions. OnStar-equipped vehicles with an active subscription will also contact representatives based in Pontiac, Michigan; Charlotte, North Carolina; Makati, Philippines; and Oshawa, Ontario, in the event of a collision that deploys the airbags. Newer models will contact OnStar in any type of collision whether airbags deploy or not. This new service is called Advanced Automatic Crash Response (AACR) and is designed to assist emergency response efforts.

When a driver presses the Red OnStar Emergency button or the Blue OnStar button, current vehicle data and the user's GPS location are immediately gathered. This information is then sent to OnStar. OnStar Emergency calls are routed to the OnStar Center with highest priority. Three centers receive emergency calls: Pontiac, Michigan; Charlotte, North Carolina; and Ontario, Canada. All centers are open 24 hours a day. OnStar-equipped vehicles have Stolen Vehicle Tracking, which can provide police with the vehicle's exact location, speed, and direction of movement.

SpeedInfo

SpeedInfo provides traffic data using its solar-powered Doppler radar combined with its wireless network design expertise (<http://www.speedinfo.com>). The autonomous speed sensors are attached to existing infrastructure such as light or sign poles, and real-time traffic flow data are then sent via the AT&T[®] Wireless network. The company headquarters is located in San Jose, CA.

SpeedInfo customizes and disseminates traffic data to SpeedInfo partners. It can customize the data each partner receives to include text, graphics, or both, each available in a variety of formats to fit a partner's business model. The company provides average speeds for important travel corridors where no data currently exists and guarantees performance and data availability levels. Partners have the flexibility to communicate SpeedInfo data as a complete solution or to integrate it with other traffic-related data to create distinct, private-branded services for drivers.

Solar-powered Doppler radar (see [Figure 10](#)) attached to existing infrastructure measures the speed of vehicles on both sides of the highway from a single device. The DVSS-100 sensor is fully self-contained and roadside-mounted. Because it is solar powered, it does not require any wiring or special hook-ups to install it. The installer can mount the DVSS-100 on existing poles, road signs, or overpasses. The sensors directly measure traffic flow at programmable rates. The

default rate is every 30 seconds during daylight hours with the average speed transmitted once per minute over the AT&T Wireless[®] network. SpeedInfo claims that the data transmission is usually reliable with virtually no downtime because the sensors are mounted 8 to 20 feet above the ground. Table 5 lists the cities where SpeedInfo currently provides data coverage.

Table 5. Cities Served by SpeedInfo.

Dayton, OH	Oakland, CA	San Jose, CA
Denver, CO	Olympia, WA	Santa Barbara, CA
Everett, WA	Omaha, NE	Ventura, CA
Lincoln, NE	Raleigh-Durham, NC	Washington, DC
New York, NY	San Francisco, CA	

Source: www.speedinfo.com



Source: www.speedinfo.com

Figure 10. SpeedInfo Doppler Radar Sensor.

TomTom and Tele Atlas

TomTom NV is a Dutch manufacturer of automotive navigation systems, including both stand-alone units and software for personal digital assistants and mobile telephones (<http://www.tomtom.com>). TomTom claims to be the leading manufacturer of navigation systems in Europe.

Tele Atlas is a Netherlands-based company founded in 1984 that delivers digital maps and other dynamic content for navigation and location-based services, including personal and in-car navigation systems. It also provides data used in a wide range of mobile and internet map applications (<http://www.teleatlas.com>). The company competes on a global basis with

companies like NAVTEQ, as well as with local map suppliers in individual countries. Since July 30, 2008, the company has been a wholly-owned subsidiary of automotive navigation system manufacturer TomTom. Table 6 shows the location-based services that TomTom offers.

Table 6. TomTom Services.

Services	Descriptions
HD Traffic	<p>HD Traffic is a traffic monitoring service that uses multiple sources to provide traffic information including:</p> <ol style="list-style-type: none"> 1. Traditional sources: Governmental/third party data such as inductive loops in the pavement, cameras, and traffic surveillance. 2. New sources: Traffic flow of 16.7 million anonymous mobile phone users. <p>TomTom merges the information and uses algorithms to improve the data and filter out anomalous readings. The system sends updates to all HD Traffic users every three minutes. Users can receive the service through a connected navigation device or through a specially designed antenna. Most current devices receive the updated road congestion conditions automatically. Rerouting can be set to be transparent to the user with the only sign that the route has been changed due to a traffic jam being a sound indication from the device and a changed ETA.</p> <p>The first launch of the system was in the Netherlands in 2007, and subsequent launches occurred in the UK, France, Germany, and Switzerland in 2008.</p>
IQ Routes	<p>TomTom developed IQ Routes and made it available in the spring of 2008 on the TomTom GO 730 & 930. It uses anonymous travel time data accumulated by users of TomTom satnav devices. Newer TomTom devices use this data to take into account the time and day when determining the fastest route. Travel time data are stored in Historical Speed Profiles, one for each road segment, covering large motorways, main roads, and also small local roads. Historic Speed Profiles are part of the digital map and are updated with every new map release. They give insight into real-world traffic patterns. This is a fact-based routing system based on measured travel times, compared to most other methods, which use speed limits or 'assumed' speeds.</p>
Enhanced Positioning Technology	<p>This service offers continuous navigation even when a navigation device cannot receive GPS satellite signals (e.g., in tunnels or among tall buildings).</p>

Total Traffic Network (TTN)

TTN sells raw traffic data to two categories of customers: personal navigation device manufacturers (e.g., Garmin, TomTom) and automakers (e.g., BMW, Mazda), which, in turn, sell subscriptions to their customers. Currently, customers with valid subscriptions can access the data anywhere in the U.S. or Canada with TTN's data service (<http://www.totaltraffic.com>). TTN delivers real-time traffic data via in-car and portable navigation systems, broadcast media, wireless, and Internet-based services using a key journalistic approach to verify data sources. Through its network of publicly available and privately held assets, TTN provides incident,

event, construction, congestion, and impact information for customers, and claims to be the leading company in North America providing a traffic subscription service to consumers.

The data channel runs concurrently with the audio channel, and customers do not have to tune the radio to any particular station for the TMC receiver to pick up a signal. The GPS unit tunes using its own FM receiver completely in the background and automatically selects the best station in the area.

TTN owner Clear Channel Radio is the broadcaster that delivers real-time traffic data directly to vehicles using its own network of reporters, traffic cameras, aircraft, broadcast towers, and strategic partners to verify and produce traffic information. TTN uses many methods for gathering traffic information in the area, including (but not limited to) DOT and proprietary traffic cameras, DOT speed sensors, INRIX GPS sensors, airborne/mobile spotter vehicles, digital scanners that cover many local emergency services, police callouts, and traffic “Tip Lines.” The local operations center staff use their resources to locate and verify the traffic data before entering these into the TTN system.

TTN claims that its data are tested for accuracy against a scientific ground-truth testing methodology called Floating Car Data Quality. TTN drives almost 200,000 miles per year in North America to benchmark, validate, test, and improve the information it provides to customers. TTN serves more than 100 metropolitan markets in:

- United States.
- Canada.
- Mexico.
- New Zealand.

The RDS-TMC service is designed to be primarily a freeway service, used on limited access highways in the local region. Secondary roads may have coverage for some incidents, but not with flow or congestion information. [Table 7](#) shows a full list of the navigation data coverage areas; the [totaltraffic.com](#) website has embedded hyperlinks that include details such as maps, hours of operations, and a list of the FM-TMC radio stations. TTN provides coverage for the following cities in Texas:

- Austin, TX.
- Dallas, TX.
- El Paso, TX.
- Houston/Beaumont, TX.
- McAllen, TX.
- San Antonio, TX.

Table 7. List of TTN Coverage Cities.

Albany, NY	Detroit, MI	Minneapolis, MN	San Diego, CA
Albuquerque, NM	El Paso, TX	Mobile, AL	San Francisco, CA
Allentown, PA	Fort Myers, FL	Montreal, Canada	Sarasota, FL
Atlanta, GA	Fresno, CA	Nashville, TN	Savannah, GA
Augusta, GA	Grand Rapids, MI	New Orleans, LA	Seattle, WA
Austin, TX	Greensboro/Winston -	New York, NY	Spokane, WA
Bakersfield, CA	Salem, NC	Norfolk, VA	Springfield, MA
Baltimore/Frederick,	Greenville/Spartanburg,	Oklahoma, City, OK	St. Louis, MO
MD	SC	Omaha, NE	Stockton/Modesto,
Baton Rouge, LA	Harrisburg, PA	Orlando/Daytona Beach,	CA
Binghamton, NY	Hartford/Bridgeport/New	FL	Syracuse, NY
Birmingham, AL	Haven, CT	Ottawa, Canada	Tallahassee, FL
Boston, MA	Honolulu, HI	Panama City, FL	Tampa, FL
Charleston, SC	Houston/Beaumont, TX	Philadelphia, PA	Toledo, OH
Charlotte, NC	Huntsville, AL	Phoenix, AZ	Toronto, Canada
Chattanooga, TN	Indianapolis, IN	Pittsburgh, PA	Tucson, AZ
Chicago, IL	Jacksonville, FL	Portland, OR	Tulsa, OK
Cincinnati, OH	Kansas City, MO	Portsmouth/Manchester,	Tuscaloosa, AL
Cleveland, OH	Las Vegas, NV	NH	Vancouver, Canada
Colorado Springs,	Lexington, KY	Providence, RI	Washington, DC
CO	Los Angeles, CA	Raleigh-Durham, NC	West Palm Beach,
Columbia, SC	Louisville, KY	Richmond, VA	FL
Columbus, OH	McAllen, TX	Rochester, NY	Worcester, MA
Dallas, TX	Melbourne, FL	Sacramento, CA	
Denver, CO	Memphis, TN	Salt Lake City, UT	
Des Moines, IA	Miami, FL	San Antonio, TX	
	Milwaukee, WI		

Source: www.totaltraffic.com

TrafficCast

TrafficCast is developing technology, applications, and content based on advanced digital traffic data. It provides technology and data analysis for real-time and predictive traffic information that enhances and enables location-based and dynamic navigation services. TrafficCast serves the interactive, mobile, enterprise, and public sector markets (<http://www.trafficcast.com>). Table 8 shows the services that TrafficCast currently offers.

Table 8. TrafficCast Services.

Services	Descriptions
BlueTOAD™	<p>Bluetooth Travel-Time Origin And Destination traffic monitoring service. Traces anonymous Bluetooth signals to derive travel time, road speeds and vehicle patterns.</p> <p>BlueToad devices receive signals emitted by Bluetooth-equipped electronics such as mobile phones, car radios, navigation devices, and computers. Only the internal identifier, known as the MAC address, is transmitted to the BlueToad receiver; these are not associated with individuals, so personal information is unavailable and the data is anonymous.</p> <p>TrafficCast’s adoption of Bluetooth involves mobile-device hardware and software platforms. It incorporates ZigBee wireless mesh networking with cellular packet backhaul for real-time traffic monitoring.</p>
Dynaflow 1.0	Use patented models to estimate travel speeds for missing links. The data gaps can be either from sensor or probe data.
Dynaflow 2.0	Predictive traffic information product. Use historical speed trends and anticipated traffic impacts such as construction, weather, and upcoming events to model and forecast road speeds up to 48 hours in advance.
TrafficSuite™	Traffic data management system. Manage the collection and integration of traffic information. Speed, flow, incident, weather, and construction data can be integrated from a number of data sources. The system utilizes proprietary normalization technology and techniques to deliver a complete, robust view of traffic information.

Summary of Traffic Data Providers

This section provides a comparison of traffic data providers described above. The comparison categories are:

- Business model.
- Data sources.
- Coverage.
- Services provided.

Table 9 summarizes the business model of each company reviewed. INRIX sells the data service directly to both partners and end users. TTN provides traffic data through the Traffic Message Channel (TMC) and therefore primarily focuses on selling its traffic products to automakers and PND manufacturers. OnStar provides various types of location-based services through partnership with specific automakers and relies on subscription-based revenues. NAVTEQ and TomTom started out as a mapping service company and later ventured into traffic data products by gathering from both traditional and probe-based data sources. TrafficCast does not maintain its own sensors, but relies on available data sources to provide historical and predictive traffic information capability. SpeedInfo uniquely relies on its solar-powered Doppler

radar sensor for speed data collection. Delcan, AirSage, and Cellint rely on anonymous wireless data matching technology (cellular probes) to provide traffic data.

Table 9 provides a brief description of the data sources for each provider, and Table 10 is a matrix of the primary data sources for each provider. Primary data sources are the sources used to provide traffic data by the company. Some providers may purchase the data from others to fuse with the data from their own sensors.

Table 9. Traffic Data Provider Comparison—Business Model.

Provider	Business Model
INRIX	INRIX is a traffic data service company providing both historical and real-time traffic information to public and private sectors in North America and Europe.
Total Traffic Network	TTN sells raw traffic data to PND manufacturers and automakers, which in turn sell subscriptions to their customers. TTN is a subsidiary of Clear Channel Radio.
SpeedInfo	SpeedInfo collects speed data using solar-powered Doppler radar attached to an existing structure such as light or sign pole. SpeedInfo creates, enhances, and then distributes the data to its partners. The data package can be customized for each partner to include text, graphics, or both.
NAVTEQ	NAVTEQ is a provider of maps, traffic and location data (digital location content) enabling navigation, location-based services, and mobile advertising around the world. NAVTEQ Traffic is a traffic solution product that delivers real-time, personalized traffic linked to the map in a navigation system. In North America, NAVTEQ Traffic business models are flexible to support different products and pricing scenarios as application requires.
OnStar	OnStar Corporation is a subsidiary of General Motors, providing: subscription-based communications, in-vehicle security, hands free calling, turn-by-turn navigation, and remote diagnostics systems throughout the United States, Canada, and China.
TrafficCast	TrafficCast is developing technology, applications, and content based on advanced digital traffic data. It provides technology and data analysis for real-time and predictive traffic information that enhances and enables location-based and dynamic navigation services. TrafficCast serves the interactive, mobile, enterprise, and public sector markets.
TomTom	TomTom is a Netherlands-based manufacturer of automotive navigation systems. Tele Atlas, a wholly-owned subsidiary of TomTom, delivers digital maps and dynamic content for navigation and location-based services. The company competes on a global basis like NAVTEQ, as well as with local map suppliers in individual countries.
AirSage	AirSage collects and analyzes anonymous wireless signaling data from cell phones to provide aggregated information on the location, movement, and flow of people and assets. AirSage technology is software-based and thus less intrusive for carrier and faster to deploy than hardware-based technology.
Delcan	Delcan provides traffic information using a proprietary algorithm to fuse data from cellular probes and GPS-enabled vehicle probes.
Cellint	Cellint is a global provider of traffic information systems based on mobile networks. Cellint's coverage is limited to the agencies/regions that chose to deploy Cellint's system.

Table 10. Provider Primary Data Sources.

Provider	GPS-enabled Vehicles	Cellular Probes	Fixed Point Sensors	Others
AirSage		x		
Cellint		x		
Delcan		x		
INRIX	x	x	x	
NAVTEQ	x	x	x	
OnStar	x			
SpeedInfo			x (radar)	
TomTom	x	x	x	
Total Traffic Network	x	x	x	Airborne/Mobile Spotters, Cameras
TrafficCast	x		x	Bluetooth

Table 11 shows the coverage comparison of the data providers. INRIX, TTN, NAVTEQ, and TomTom provide their service for many metropolitan areas nationwide and some also compete in the international arena. AirSage, Delcan, and Cellint all rely on cellular probes but there are no independent data available on the true market penetration of their services. SpeedInfo and TrafficCast service coverages are relatively localized compared to the rest of the providers.

Table 11. Traffic Data Provider Comparison—Coverage.

Provider	Coverage
AirSage	AirSage is partnering with one wireless carrier in the United States. AirSage’s nationwide coverage includes 85% of the U.S. population and currently provides real-time, historical, and predictive traffic information for 127 U.S. cities.
Cellint	Cellint’s coverage is limited to the cities that deploy Cellint’s system.
Delcan	Delcan’s coverage is limited to the cities that deploy Delcan’s system.
INRIX	20 countries across North America and Europe.
NAVTEQ	NAVTEQ is currently providing real-time traffic for 52 city areas in North America and is expanding to Europe.
OnStar	OnStar services are only available currently on vehicles manufactured by General Motors and Saab Automobile, so the coverage includes the operational area of these vehicles. The service is only available for vehicles with the factory-installed OnStar hardware. A similar service is known as ChevyStar in Latin American markets. The service currently has more than 5 million subscribers.
SpeedInfo	Over 14 metropolitan areas in the United States.
TomTom	TomTom provides digital map service to many countries around the world.
Total Traffic Network	Over 100 metropolitan markets in four countries—the United States, Canada, Mexico, and New Zealand.
TrafficCast	TrafficCast service is currently available in at least four cities in North America as well as in Shanghai, China.

Table 12 compares the services offered by the providers. INRIX offers a wide range of data service and has strategic partnerships with several providers such as NAVTEQ and TomTom. NAVTEQ and TomTom compete on a global basis in both mapping and traffic data services. AirSage, Delcan, and Cellint offer comparable services as they are based on cellular probes. SpeedInfo is unique in that its service relies on its proprietary Doppler radar sensor and wireless communication capability.

Table 12. Traffic Data Provider Comparison – Services.

Provider	Services
AirSage	AirSage’s products include: <ul style="list-style-type: none"> • Real-Time Traffic (XML-based data feed); • Movement Analytics (aggregated cellular movement analytics for urban planning and business needs); • Network Analytics (e.g., traffic density mapping, dropped call density mapping, signal quality density mapping); and • Consulting Services (e.g., traffic counts, travel time forecasts, incident reports, speed surveys, and origin-destination analysis).
CellInt	CellInt’s TrafficSense is a traffic monitoring service. TrafficSense can provide travel time, speed, and incident alerts.
Delcan	From the literature, Delcan can provide travel time and speed data service. The exact information that the Delcan system can provide was unavailable.
INRIX	INRIX classifies its production solutions into five different groups – automotive solutions, mobile solutions, public sector, fleet solution, and internet and media.
NAVTEQ	NAVTEQ Traffic product portfolio includes various traffic distribution channels and different formats depending on customers’ needs. The distribution channels include satellite, radio data system (RDS) for auto companies and PND manufacturers, mobile and server-based navigation, consumer traffic websites, and high bandwidth digital radio.
OnStar	Vehicle owners can choose between two service plans: <ul style="list-style-type: none"> • Safe & Sound (\$18.95/mo), which includes Automatic Crash Response, Stolen Vehicle Assistance, Roadside Assistance, Remote Door Unlock, Remote Horn and Light Flashing, Red Button Emergency Services and OnStar Remote Vehicle Diagnostics; and • Connections & Directions (\$28.90/mo), which includes all services in the Safe & Sound plan, plus turn-by-turn navigation.
SpeedInfo	SpeedInfo provides data for broadcast media and government planning, as well as mobile applications (e.g. wireless devices, navigation systems, HD radio, satellite radio)
TomTom	TomTom’s HD Traffic is a traffic monitoring service which utilizes both traditional data sources (i.e. public-sector loops, cameras, surveillance) and cellular probes. TomTom’s IQ Routes is a navigation service that is available on select TomTom devices. IQ Routes utilizes measured travel times to provide fact-based routing system.
Total Traffic Network	TTN sells its products to two groups of customers – automotive and personal device. Customers may need to pay subscription charges when purchasing these products.
TrafficCast	TrafficCast offers the following services: <ul style="list-style-type: none"> • BlueTOAD™, a travel-time origin and destination traffic monitoring service using Bluetooth technology; • Dynaflow 1.0, a patented model for estimating travel speeds on missing links; • Dynaflow 2.0, predictive traffic information by utilizing historical speed trends and anticipated traffic impacts such as construction, weather, and scheduled events; and • TrafficSuite, a traffic information management tool for integrating traffic, incident, weather, and construction data from various sources.

3.3 ASSESSING TRAFFIC INFORMATION QUALITY

Bogenberger of BMW developed a method for assessing the quality of traffic information reporting, which emerged from the area of signal detection theory (Bogenberger, 2003). The method proposes two indices (QKZ_1 and QKZ_2) to describe the quality of traffic information. The first index describes the detection rate and the second index describes the false alarm rate or the relevance of the traffic information. Figure 11 shows how to calculate these two indices. Quality index one (QKZ_1), the detection rate, describes the degree to which the described area of congestion in the traffic message coincides with the actual area of congestion. The second Quality index (QKZ_2), the false alarm rate, describes the proportion of irrelevant traffic messages (i.e., non-congested areas that received a message of traffic congestion). These two indices range from 0 to 1.

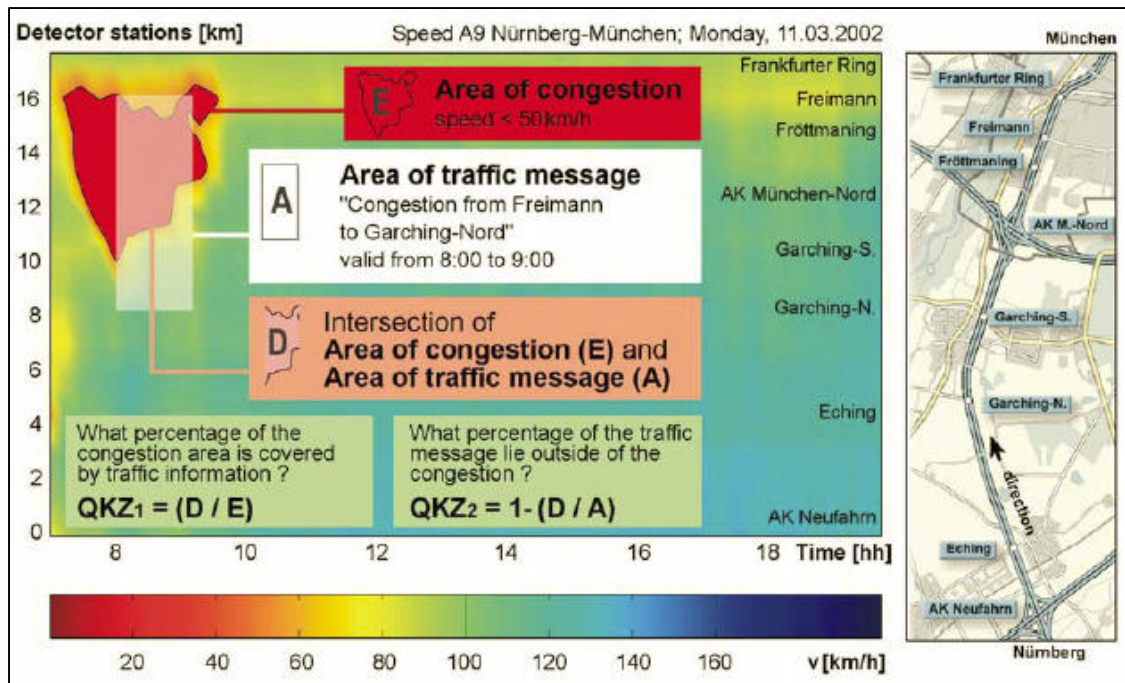


Figure 11. Quality Indices QKZ_1 and QKZ_2 (18).

Bogenberger (2003) provided a quality diagram based on the values of QKZ indices. The quality of traffic information is graded with letters from A (best) through F (worst) as shown in Figure 12. Under this grading scheme, the information quality is considered best when $QKZ_1 = 1$ and $QKZ_2 = 0$ (i.e., when the congestion is detected at 100 percent and the percentage of irrelevant traffic message is 0). The field study found that travelers want to be informed of almost all of the congestion events ($QKZ_1 = 0.90$). However, travelers also accepted the fact that some of these reported events may be irrelevant to them ($QKZ_2 = 0.45$) such as congestion events reported from other locations. In general, a poor detection rate is considered to be the more critical issue as it can lead to degradation of consumers' acceptance. Irrelevant traffic

messages can cause a problem if the information is used in real-time applications such as dynamic route guidance (resulting in suboptimal routing).

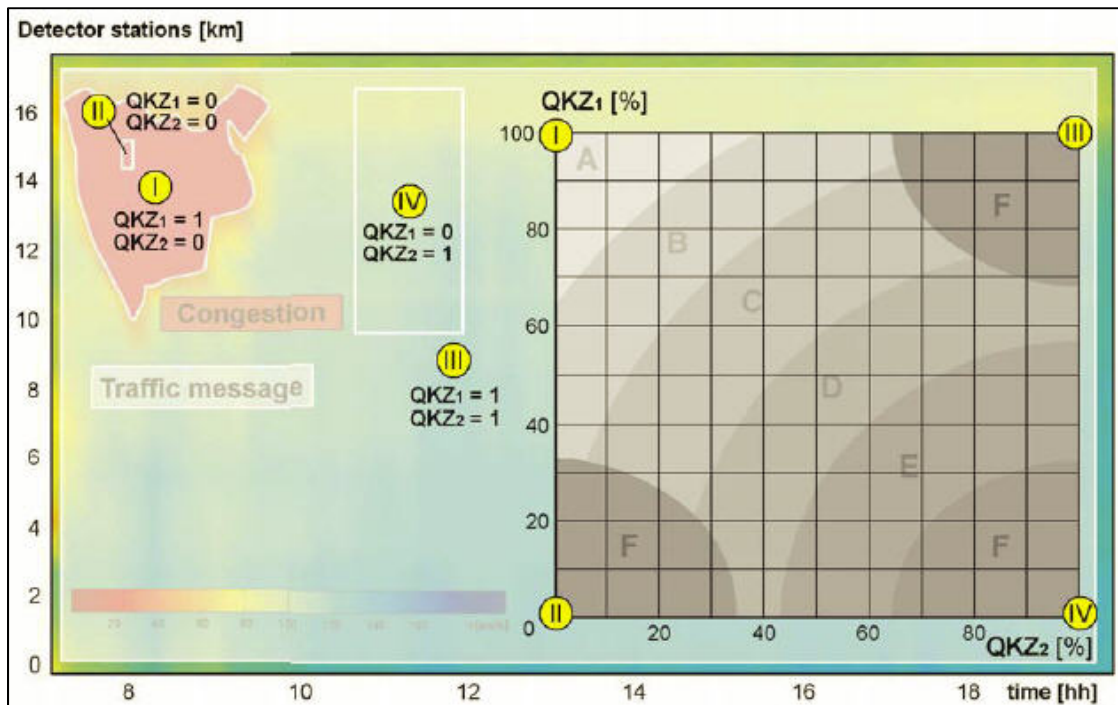


Figure 12. Quality Diagram (I8).

3.4 SURVEY OF PRIVATE DATA PROVIDERS AND CONSUMERS

The information provided in the next two major sections of this report are intended to serve the needs of two separate projects. There was a similar Federal Highway Administration (FHWA) project that ran parallel to this one with a similar task, so research personnel conducted only one set of surveys to serve the needs of both projects. The FHWA's Office of Operations sponsored the other project. The reader will see that some of the wording in survey forms and results indicate that the survey was disseminated under the auspices of the Federal Highway Administration.

3.4.1. Interview Methodology

The work efforts related to the survey included the following steps:

- Development of the survey mechanism.
- Procurement of approval through the Institutional Review Board process.
- Administration of surveys.
- Analysis and documentation of results.

3.4.2. Development of Survey Mechanism

The first aspect of developing the survey was determining the most appropriate survey mechanism. The most common mechanisms are pre-arranged telephone surveys, web-based surveys, and paper surveys using traditional postal mail or fax. Other more specialized methods have also been utilized, such as focus groups, text-message surveys, and random phone calls. None of these methodologies is applicable to the information-gathering needs of this work effort.

The type of survey mechanism most applicable to the detailed information desired from this group of participants was a telephone interview. The personal contact increases the urgency and commitment of the respondent to the successful conclusion of the survey in conjunction with the interviewer. The challenge of a telephone interview is to keep the interview on track and focused on the base survey questions while allowing and capturing the additional detail that could be offered. Overall, the benefits of a telephone survey mechanism far outweigh the challenges.

Once the decision to use a telephone survey mechanism was made, the research team crafted the actual survey instrument. Initially, they developed a set of survey goals and general questions to scope the entire survey. Efforts then focused on fleshing out additional details and question points, and developing a completed draft.

Initially, the Texas Transportation Institute component of the research team reviewed the draft survey to allow for internal refinement. After completing that step, TTI sent the improved draft to external team members for their review and input. The final survey versions for both the data provider and data consumer are included in [Appendices A, B, and C](#).

3.4.3. The Institutional Review Board (IRB) Process

According to Federal regulations, the protection of human subjects of research is required to assure both the safety and privacy of individuals who participate in any type of study. As codified in the Code of Federal Regulations (CFR), 45 CFR 46, this policy applies to all research involving human subjects conducted, supported, or otherwise subject to regulation by any Federal department or agency. In general, simple surveys are exempt from the full implementation of the protection guidelines, but a review process by an authorized review board must be performed to ascertain a finding of Exempt.

The Texas A&M University System maintains a fully authorized IRB charged with the review of human subject research to ensure that ethical and safe research practices are used when research is conducted with human subjects. IRB approval is required before anyone can begin research involving human subjects, even those studies where the presumption is exemption from IRB protocols. Additionally, all interviewers or persons having direct contact with survey respondents are required to be certified in IRB protocols and must produce a compliance certificate upon request. The Texas A&M testing and certification program is administered through the Collaborative Institutional Training Initiative.

In accordance with these requirements, the research group submitted supporting documentation for this telephone survey to the IRB review panel for oversight and a confirmation of the Exempt classification. The documents include a project information form, a disclosure of potential conflict of interest form, an information sheet for potential respondents, email text for contacts regarding surveys (both initial and follow-up), and the full text of the survey documents for both the provider and the consumer. The TAMU IRB panel found that the survey was exempt and provided authorization to begin all work efforts associated with the actual recruitment of respondents and conducting the survey.

3.4.4. Administration of Surveys

While identifying potential participants was fairly simple given the project team's experience and contacts, contacting potential participants and engaging their support for the survey was a time-consuming process spanning several months. In all, the project team received surveys and conducted interviews with six private-sector data providers. The private-sector data providers interviewed included:

- Air Sage.
- ATRI.
- INRIX.
- NAVTEQ.
- TomTom.
- TrafficCast.com.

The project team also either interviewed or developed responses from previous interactions for the following private sector data consumers:

- Houston-Galveston Area Council.
- Maricopa Association of Governments.
- Michigan Department of Transportation.
- San Francisco Bay Area 511 Program.
- Texas Department of Transportation.
- Wisconsin Department of Transportation.

The San Francisco Bay Area 511 program is the only one of this list that only purchases real-time data. The rest purchase historical data. The steps to perform the actual data collection via the telephone survey mechanism involved:

1. Identification of target providers.
2. Recruitment of potential respondents.
3. Scheduling of interview times.
4. Conducting the survey.

The work plan identified the target companies and personal contacts within those companies based primarily on the experience of the research team. The researcher responsible for the interviews then made an initial contact via email to schedule the actual interview.

Pursuant to IRB guidelines, initial electronic mail contact with potential respondents must follow a prescribed protocol. Figure 13 shows the initial contact template included in the IRB submission package for this work effort. As noted elsewhere, this template served the needs of both FHWA and this TxDOT project.

Additionally, the protocol requirements include an electronic mail template for the actual delivery of the survey document and scheduling of the telephone interview time. Figure 14 shows the required protocol template for this work effort. The brackets within each template indicate that researchers developed a single template that could be used for both providers and consumers by simply switching key phrases. This plan ensures consistency in the application of the IRB protocols and the overall protection of human subjects.

Hello, *Mr./Ms.* <>,
I am currently assisting the Federal Highway Administration with an investigation to determine the capability and the advantages and disadvantages of using private-sector data for national mobility performance management. As part of this effort, we are interviewing both producers and consumers of private-sector data to determine a typical scope of services, technical issues, and typical uses and pricing.

I was hoping that you could direct me to the best point(s) of contact within your agency—including yourself—who can provide information about *<being a producer of private-sector data><being a consumer of private-sector data>*.

Once the appropriate points of contact are identified, I will send along additional guidance regarding the specific types of information we are hoping to obtain. Thank you in advance for any assistance you can provide.

Figure 13. Electronic Mail Template for Initial Contact.

Overall, consumers cited a number of advantages to the use of private-sector data, including:

- Cost-effectiveness.
- Consistency of the data collection approach.
- Comprehensive coverage.
- Frequency of updates.
- Environmentally sensitive data collection (as compared to in-house efforts such as probes).

Hello, *Mr./Ms.* <>, *Mr./Ms.* <> identified you as the appropriate point of contact regarding a survey on the use of private-sector data and the advantages/disadvantages of that data source for national mobility performance management. We'd like to ask for your assistance by completing the attached survey related to private-sector data as a <*consumer of the data*><*provider of the data*>. We will also be performing surveys with <*producers of private-sector data*><*consumers of private-sector data*> to understand both aspects of the data. You are being provided the survey questions in advance of the scheduled telephone interview time. Although your participation is voluntary, your assistance in supporting this effort is vital to adequately document the capability of using private-sector data to support national mobility performance management practices. Your participation and the information you provide will not be kept confidential and will instead be documented and disseminated in a final report intended to encourage widespread improvements to TIM performance measurement practices. We'd appreciate your response no later than <*two weeks out*>. If you have any questions, please don't hesitate to contact me. Thank you in advance for your participation.

Figure 14. Electronic Mail Template for Follow-Up Contact.

However, respondents also identified some concerns with private-sector data use, including:

- No influence on the data collection methodology.
- Lack of per-lane data.
- Lack of full ownership of the data due to licensing and cost considerations.

3.4.5. Analysis and Documentation of Results

Table 13 summarizes the responses received from the consumer survey. To date, data providers appear to have made the bulk of their inroads with customers such as car companies, cell phone companies, and consumer GPS devices. Responses from the data providers indicate that the market sector for agencies as data consumers is small, but growing. Providers are responding with aggressive growth plans to service those areas. The responses obtained from the consumer side of the equation include three state DOTs, two regional association of governments, and a regional 511 service. This cross-section may serve as something of a validation of the data providers' expansion plans. Respondents to the data consumer survey indicated two status levels—either seeking information via a Request for Information (RFI), or already having purchased data.

Primarily, consumers responded that they purchased private sector data because of:

- Cost-effectiveness.
- Faster turnaround.
- Accuracy.
- Availability.

Table 13. Summary of Historical Data Consumer Survey Results.

Factors	Wisconsin DOT	HGAC	Michigan DOT	Texas DOT ^d	Phoenix MPO (MAG)
Status	RFI	Purchased	Purchased	Purchased	Purchased
Service Purchased ^a	H	H H H H			
Aggregation Level	Hourly day-of-week averages	15 min.	5 min.	Hourly day-of-week averages	Weekday
Data Purchased ^b	S/TT, PM	S/TT	S/TT S/TT,	PM	PM
Applications ^c	PM, TM	PM, TM, OD	PM	PM	PM
Coverage	All arterials	Houston region	MI Freeways	Statewide TMC network	Region
Timeframe	1–2 years	1 year	5 years	2009	1 year
Validation Criteria	Not yet established	Not yet established	Avail >99.5% Accuracy less than ±10 mph	None	Not yet established
Validation techniques	N/A	N/A	Probe, fixed point, re-ID	None	Probe, fixed point.
Pricing (in thousands)	\$80K (Est.)	\$77K	\$200K per year	\$28K	Negotiating
Licensing	Multiple Use	Multiple Use	Single Use	Single Use	Multiple Use

^a Service Purchased: H = Historical, RT = Real-time

^b Data Purchased: S/TT = Speed or Travel Time, PM = Performance Measures

^c Applications: PM = Performance or Congestion Monitoring, TM = Traffic Model Validation or Calibration, OD = Origin-Destination Studies

^d See <http://apps.dot.state.tx.us/apps/rider56/list.htm> for actual study results.

Service Purchased

All respondents mentioned in [Table 13](#) indicated that they purchased historical data and not a real-time data feed. Some providers offer two types of historical data. The first allows for the purchase of discrete data, which is each individually recorded data point. An example would be all vehicles and their speed within a segment and time of day. The other, more commonly used method, is the purchase of aggregate data, which would provide one data point (such as a speed or travel time) for each section and time interval. Respondents uniformly indicated that they purchase aggregate data.

Of the agencies contacted in this research project, the San Francisco Bay Area 511 program (SF program) is the only program that purchases real-time data. In this case, the provider, SpeedInfo, provides only speed data to the Metropolitan Transportation Commission (MTC).

Data Purchase

Respondents were also fairly evenly split on the particular attributes of the data purchase. Both travel times and speeds were identified as a purchase, interest, or as performance measures. The finding of performance measures was somewhat interesting as during the provider survey, most providers indicated that they felt the marketplace would want to calculate its own measures from the purchased data. The SF agency purchases speed data covering 24 hours a day, seven days a week, aggregated every minute. The agency chose private data due to its cost effectiveness and its accuracy and availability. There was no latency specified in contract agreements.

Applications

Responses from data consumers indicated that the purchased data were used primarily for planning purposes, such as performance monitoring/congestion mitigation, origin and destination studies, or traffic modeling validation/calibration. Most consumers were purchasing a one-year data set, although the Michigan DOT response was striking in that it was for a five-year period. The desired aggregation level varied widely, from 5 minutes to 1 hour. Consumers were not asked what referencing system they were using, although some responses indicated the use of Traffic Message Channel or larger, corridor scale segmentation.

The SF program uses the real-time data purchased from SpeedInfo for motorist information and for 511 services. In addition, the Metropolitan Transportation Commission (MTC) uses fused data in its planning department. The MTC plans to initiate a major analysis of available data sources in the near future. The final deliverable will be a recommendation of which data source(s) appears to best fit their needs. They are also separately including the needs of their highway and arterial operations (HAO) section and the planning section at MTC. The report will include separate recommendations for 511, HAO, and Planning.

Validation

By far the most interesting area of the consumer responses related to validation of the data. Only Michigan DOT and Maricopa Association of Governments identified techniques used to validate the purchased data; only Michigan DOT had specific criteria in place to measure those validation activities. In terms of availability, those criteria mirrored the FHWA requirements for real-time applications, while the accuracy requirements seemed significantly more lenient than what other studies (I-95 Corridor Coalition) have reported was possible.

The SF program has validated the data using fixed point data collection and re-identification studies using Bluetooth and license plate readers. SpeedInfo's data feed includes a confidence factor for each data point. The SF program disregards data falling below a confidence factor value that is selected based on their understanding of how SpeedInfo's data collection system functions. The program recently discovered that there are certain situations when the confidence factor does not properly reflect inaccurate data. They are working internally to develop an algorithm to detect this situation. If there is a drop in data accuracy, the SF program

uses Caltrans data if it is available for that roadway segment, or they extrapolate from an adjoining segment if it has live data.

Pricing

Not all consumers responded with pricing information. Understanding the pricing models in play are difficult as each provider has different pricing models based on the application, type of data, and coverage area. This was confirmed in the responses from the data providers as none except INRIX would detail their pricing structure. There is little capability to infer standard pricing information that would be applicable over a large (multi-state or national) region from the responses received.

Respondents did indicate, however, that future purchases are under consideration. This supports observations from the data provider responses indicating a growing market in this arena and a desire to capture that market. Providers also identified ongoing collection and analysis research to fulfill additional market desires, such as volumes, easier access to per-lane data, and more. Most responses from data consumers stated that validation plans and requirements are not yet developed, indicating that this could change in the future as agencies gain more experience with this type of data.

The SF program purchases speed data for \$110 per month for each bi-directional sensor station. The agreement between the program and the provider stipulates that the program can reduce its payment to the provider if either the data availability or the data accuracy falls below predetermined levels. The data availability must remain above 90 percent; otherwise, the provider's invoice must be reduced by a commensurate amount. For accuracy, the contract stipulates that at least 85 percent of speed data readings are accurate to within 7.5 percent or 5 mph, whichever is greater during all environmental conditions.

Further, it states that 95 percent of all speed data readings must be accurate to within 15 percent or 10 mph, whichever is greater, during all environmental conditions. Finally, if there is a problem with a sensor that cannot be fixed within 30 days, the invoiced amount charged for the service shall be reduced by \$100 for each inaccurate sensor. These data must also be removed from the data feed.

3.4.6. Private Data Providers

[Table 14](#) summarizes the information across the data providers. Overall, perhaps the most interesting finding concerning the private-sector data providers was the diversity of data sources in use. Providers are using a combination of GPS data from fleet vehicles, consumer devices, and cell phone applications, as well as data from fixed sensors installed and maintained by other agencies, and fixed sensors installed and maintained by the data provider.

Among the providers, there is no single data source model in use. Correspondingly, there does not appear to be any single business model in use. Each provider seems to have developed a

somewhat well-defined niche or area, although many providers spoke about a desire to break out of that niche and expand their potential market, perhaps with new data offerings.

Table 14. Summary of Historical Data Available by Provider.

Factors	Airsage	ATRI	INRIX	NAVTEQ	TomTom	TrafficCast
Data Available ^(a)	S, TT, I, Q, V	S, TT, Q	S, TT, I, Q, V	S, TT, I, Q, V (portion of network)	S, TT, I, Q	S, TT, I, Q
Services Available ^(b)	D, A, PM	D, A, PM	D, A	D, A	D, A, PM	A, PM
Data Source ^(c)	Cell phone, 911, traffic counts	GPS on commercial truck-only fleets	State-installed sensors, commercial fleets, consumer GPS	State-installed sensors, commercial fleets, consumer GPS.	Consumer GPS, Fleet GPS	State-installed sensors, commercial fleets, consumer GPS, Bluetooth systems.
Aggregation Levels for Historical Usage	None; as captured	1 mile, 1 minute	15 – 60 minutes	15 minutes	1 hour	15 minutes
Accuracy Checks Performed	Visual camera count, Probe vehicles.	Anomaly checking done, routines not disclosed.	Independently verified in large-scale testing.	Data checks prior to map matching. Comprehensive drive testing.	Data checks prior to map matching.	Simple adjacent pts compared, some clients doing accuracy checks.
Documented Quality Levels	None provided. Stated they meet Section 511 requirements.	None; burden is on receiver of data.	Accuracy above 95% Availability above 99.9%	None provided.	None provided. Stated they can meet Section 511 requirements.	None provided. Stated they can meet Section 511 requirements.
Pricing	Specific pricing information not provided.	Specific pricing information not provided. Not for profit.	Full use open licensing is \$800/ mi/yr plus \$200/mi one-time setup fee. 25% discount on other roads purchased in conjunction.	Specific pricing information not provided.	Specific pricing information not provided.	Specific pricing information not provided.

^a Data Available: S = Speed, TT = Travel Time, I = Incidents, Q = Quality, V = Volumes, GPS = GPS fleet

^b Services Available: D = Discrete Data (individual data points), A = Aggregate Data, PM = Performance Measures

^c National Coverage: Not listed in table. All providers indicated national coverage, except TrafficCast which is currently in urban areas.

^d Map Matching: Not listed in table. All providers except ATRI indicated a minimum use of TMC. ATRI uses mileposts. INRIX, NAVTEQ and TomTom also use proprietary segmentation smaller than TMC.

Even the fleet-equipped GPS data sources show a wide range of diversity. While no provider would detail their fleet arrangements for protection of their business practices, several spoke in general about the range of fleet types. From long-haul trucking, to delivery vehicles, to taxicabs, providers have actively sought data from whatever fleets are available. Many spoke about continuing to expand their fleet coverage as the best method of accessing additional data points.

Several providers spoke about the changing marketplace in terms of the amount of data now available. While low availability of data used to be the paradigm a few years ago, the new paradigm is the vast availability of data and the comparative richness of the sources. Some providers spoke about past moves to change their models and business practices to actually reduce the number of individual data sources, primarily migrating to consumer GPS information. More than one provider spoke of receiving millions, if not billions, of individual data points per day.

Coverage

[Table 14](#) does not include a listing for national coverage as part of the summary information. With the exception of TrafficCast, all providers indicated national coverage capability on main roadways, typically down to primary arterial level. This would correspond to Functional Class (FC) 3 roadways in the Traffic Message Channel (TMC) mapping system.

Map Matching

All of the private sector data providers provide their data mapped to some system that allows for the geographic identification of the roadway segment to which it applies. With the exception of ATRI, which uses mileposts, all providers utilize TMC as a minimum. INRIX, NAVTEQ, and TomTom also have proprietary mapping, which allows data to be mapped to segments at a finer (smaller) resolution than TMC.

Data Available

[Table 15](#) lists the types of data that are available across the different providers. The interviewer asked providers to identify the types of data they supplied differentiated by the historical and real-time markets. Providers chose from the following options:

- Speeds.
- Travel times.
- Incident/event data.
- GPS fleet data.
- Volumes.
- Arterial data.
- Per-lane availability.
- Other.

Table 15. Summary of Types of Data Offered.

Data Type	Historical (Percent)	Real Time (Percent)
Speeds	100	83
Travel times	100	83
Incidents/events	67	67
Quality/metadata	100	67
GPS fleet information	34	17
Volumes	50	17
Arterial	100	67
Per-lane availability	17	17
Other	50	34

Speeds (S) and travel times (TT) were the prevalent data provided and were available from all the vendors as historical data. All vendors stated the provision of some type of quality (Q) data, although the specific information provided varied by provider. Also, all vendors except ATRI stated the availability of Incident data. AirSage and INRIX stated the availability of volume data across the network, while NAVTEQ stated the capability for a portion of the network. Volume data comes from a variety of data sources including fixed sensor data sources installed and maintained by public agencies, camera counts, and probe vehicles. Data availability on a per-lane basis is still in its infancy as a provider offering, although a number of respondents stated offerings under research and development.

As indicated in the responses summarized in [Table 14](#), the availability of data changes between the real-time and historical markets. Some of the variability in the real-time responses is due to ATRI not being in the real-time market.

As expected, the “bread and butter” of the providers is speed and travel-time data, in both historical and real-time contexts. Associated with those data is the provision of quality or metadata expressing items such as confidence intervals, sample sizes, or other quality indicators. It should be noted, however, that there is little consistency in terms of what is actually provided as a quality indicator. This appears to be a negotiable item in contracts.

A number of the data providers use consumer GPS devices to some degree. These data may not arrive in sufficient quantities to include in real-time information, but can be added to the existing data sets once consumers upload these at a later date. Providers spoke of receiving data in this manner, ranging in age from a few days to several months.

A similar situation exists pertaining to arterial coverage. Several data models use data from consumer GPS devices to some degree. These data may not arrive in sufficient quantities to include in real-time information but can be added to the existing data sets once consumers upload these at a later date. Providers spoke of receiving data in this manner, ranging in age from a few days to several months.

While only one vendor claimed to have per-lane information available, additional respondents indicated they were actively working to provide this level of data service. The responses to “Other” types of data included the use of Bluetooth data and fleet diagnostics, such as engine parameters. These would primarily be of use only to the original GPS-equipped fleets.

Services Available

The interviewer asked providers if they offered the following types of data products, differentiated as historical or real-time data:

- Raw data for purchase.
- Refined/aggregate data for purchase.
- Data warehousing.
- On-demand data access.
- Performance measures.

Raw or discrete data is defined as the individual or discrete data elements or points. While providers said they would not sell fleet GPS data, many providers do sell the complete data stream on individual points, stripped of any identifying information. When purchasing discrete data, a consumer would get all of the individual speed or travel time points within a section, within a timeframe, whereas they would only get one value under the purchase of aggregate data.

Aggregation Level

The principal service offering, however, is refined or aggregated data. Aggregate data are available from all of the responding providers on a historical basis. What is different across the providers is the level of aggregation. Some providers use 5 minutes, others use 15 minutes, and still others use 60 minutes. ATRI provides the lowest level of aggregation, at 1 mile or 1 minute. Other providers vary from 15 to 60 minutes. In part, the differences are due to the wide variety of data sources. On any given device, GPS data are typically recorded at 1-second intervals but that can be altered. Data from fixed-point sensors are typically recorded at 20-, 30-, or 60-second intervals. Cellular data might be recorded at sub-second levels.

Data Sources

Each provider essentially had a unique (to some degree) set of data sources. While there was some overlap, no responding provider utilized exactly the same data model as another provider. Providers are using an expansive range of data sources including global positioning system (GPS) data from fleet vehicles, commercial devices, cell phone applications, fixed sensors installed and maintained by other agencies, fixed sensors installed and maintained by the data provider, and cell phone location.

Data Filtering

Although not specified in [Table 14](#) the interviewer also asked providers to detail the manner in which their data could be analyzed. All respondents indicated the ability to do data filtering or sorting based on typical parameters such as date, time, roadway, region, state, or data source. The provision of these capabilities stands to reason because they are somewhat inherent in any database or archive, although the extent or level of discreteness can vary greatly.

Accuracy Checks

The data providers were cautious about discussing any accuracy checks they perform to validate their data offerings. With the exception of ATRI, which stated that none are performed, most providers did not disclose specific checks or algorithms. TrafficCast did state that a part of its general methodology included simple adjacent point comparison routines but also stated that it employed more sophisticated methods. INRIX, in part due to the comparisons performed by the I-95 corridor coalition, stated that large-scale client testing has verified its data. NAVTEQ claimed that it does extensive drive testing across all types of roadways in all markets at all times of the day and days of the week. With the exception of ATRI, all providers stated they have an extensive data-checking process in place to ensure overall data quality. A number of providers also have integration routines employed to merge data from disparate sources into a seamless coverage of their network. However, they did not provide descriptions of these routines.

Quality Levels

With the exception of INRIX, data providers were also cautious about the quality levels they meet. INRIX explicitly claimed an availability of more than 99.9 percent and an accuracy of greater than 95 percent. AirSage also claimed to achieve at least a certain level of quality.

The interviewer asked providers if they were aware of, and were capable of meeting, the requirements in the FHWA Final Rule on the Real-Time System Management Information Program, which took effect December 23, 2010. The purpose of this program is to implement subsections 1201(a)(1), (a)(2), and (c)(1) of the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU) (Pub. L. 109-59; 119 Stat. 1144), pertaining to congestion relief. [Table 16](#) provides an overview of the information delivery requirements from the program. The full provisions of the program can be found in 23 CFR Part 511.

While the providers were aware of the ruling, there was no concern associated with either the time frame for implementation or the requirements. In general, based on the information provided during the survey as far as data latency and availability, the existing data parameters would appear to exceed the FHWA rule-making requirements. Only one provider (INRIX) had a specific comparison (available on their website) of information regarding the FHWA requirements and their standard numbers for reporting time frames, accuracy, and availability. Providers were aware of the requirements and expressed no concern over meeting the real-time requirements and by extension, accuracy and availability levels for historical data.

Table 16. Information Delivery Requirements for Real-Time System Management Information Program.

Information Type	Metropolitan Area (Minutes)	Non-Metropolitan Area (Minutes)	Availability (Percent)	Accuracy (Percent)
Implementation or removal of lane closure	10	20	90	85
Roadway- or lane-blocking traffic incident information	10	20	90	85
Roadway weather observation updates	20	20	90	85
Travel time along highway segments	10	N/A	90	85

Pricing

In general, the availability of pricing information was minimal. Most providers appear to negotiate each purchase individually. Pricing is tied to the usage of the data. Data that are used for a single application employs one price point; those used for multiple applications require a different price point. Providers also make a distinction between uses, such as modeling or O-D studies, and derivative products, such as summaries distributed to external sources. While providers did not disclose the various price points, all stated that they exist. The INRIX pricing provided in [Table 14](#) is the complete package pricing for an all data, all access license. Reduced requirements and uses would result in lower price points.

Data Imputation

This aspect of data primarily applies to real-time information and is not detailed as a line item in [Table 14](#). Providers responded in one of two ways when asked about data imputation or filling in the gaps in real-time data. A number of providers stated they have the ability to impute based on their historical data archives and data-checking routines. Providers also stated that they flag such data as being all or partially composed of historical versus real-time data. INRIX explained in detail how the quality measures associated with any particular data point would change, based on the amount of historical data being used. Essentially, the confidence interval expressed for the data point, such as a speed or travel time, would range from very high with no historical data in use to very low with significant historical data in use. ATRI does no data imputation at all.

Data Provision

Providers were asked to detail the ways in which they provided data to their customers. For real-time usage, the universal answer was some type of data feed, typically Extensible Markup Language (XML) updated on a 1-minute interval. Providers also stated that they could provide map outputs, but those processes are still fed in the background by a data feed.

Smartphone displays were also a standard answer, but they are also powered by a background data feed.

For the historical context, a wider variety of data provision mechanisms is possible. Some providers utilize an Internet-based portal access to the database, and customers can perform and save their own query results. Other providers execute the query for the customer and ship the resulting data file via electronic mail or CD-ROM. Typically, they provide the file in either XML or Comma Separated Variable (CSV) format.

3.4.7. Survey Forms

The Appendix has the survey forms that the research team used to contact data providers and consumers. This was a telephone interview in all cases, which began by initial contacts with each company or agency to determine the appropriate person to contact.

4. PROVIDE FEEDBACK TO TXDOT

4.1 INTRODUCTION

The objective of this activity was to convey the results of Task 2 to interested Texas Department of Transportation officials so they could offer meaningful feedback on the preliminary findings of Task 2, Gather Relevant Information. In this task, the research team identified:

- TxDOT staff to contact.
- A suitable means of conveying and discussing the findings.
- One or more appropriate locations for the exchange of ideas.

The research team provided an opportunity for selected TxDOT decision makers to comment on the literature review findings, the results of interviews with state officials experienced with private data providers, and the results of interviews with private data providers. The research team had solicited guidance from members of the Project Monitoring Committee on individuals who should be contacted to include in the discussions, the discussion format, and the discussion location.

4.2 METHODOLOGY

The following two subtasks were included:

- Subtask 3a. Determine Most Appropriate Methodology.
- Subtask 3b. Document TxDOT Input.

4.2.1. Determining the Survey Methodology

Researchers proposed to the PMC to use a webinar to provide findings from Task 2 and to solicit feedback from TxDOT staff, and the panel agreed with this approach. PMC members provided an initial list of 60 TxDOT names to contact to gather the desired information. Researchers developed the discussion format and determined that a webinar would best serve the needs of the project. Once the survey instrument was ready, researchers had to submit the instrument for approval to the Texas A&M University Institutional Review Board. This step is required any time human subjects participate in research. IRB review slowed the progress on this task, but the board's ruling of Exempt meant that the research project would not have to undergo further intense scrutiny.

4.2.2. Conducting the Webinar

The research team used the webinar services that are available through the Texas Transportation Institute. The Research Supervisor scheduled the webinar after an initial contact with prospective participants and determining the date that would be most suitable for participants. TTI sent out the survey instrument and a summary of Task 2 findings about one week in advance of the scheduled date. Survey participants received the web address and phone number for participation in the webinar at this time as well.

The TTI Research Supervisor conducted a webinar to solicit information from TxDOT employees. One other TTI engineer was involved in the setup of the webinar and a total of 20 TxDOT employees or offices were involved (one or more offices provided more than one person). According to information available to research personnel, the following districts or entities were involved:

- Abilene District.
- Amarillo District.
- Austin District (2).
- Beaumont District.
- Brownwood District.
- Bryan District.
- Corpus Christi District.
- Dallas District (2).
- El Paso District.
- Fort Worth District (no survey response).
- Laredo District.
- Lubbock District.
- Lufkin District.
- Odessa District.
- Paris District.
- Pharr District.
- Traffic Operations Division (no survey response).
- TransGuide San Antonio.
- TranStar Houston.
- Wichita Falls District.
- Yoakum District.
- Strategic Policy and Performance Department.

The webinar began at 2:00 p.m. on July 18, 2011, and lasted just over an hour. The Research Supervisor used a PowerPoint presentation to supplement the two documents used as part of a mail-out prior to the webinar. The TxDOT Project Director provided an initial list of 60 names, all of which were sent the mail-out prior to the web seminar. Based on the list of names that the webinar system had documented, the participation rate was 33 percent. The two documents sent to the list of 60 email addresses are included elsewhere in this report, one of which contained the questions used in the webinar.

The webinar included explanations of each of the 11 questions to make sure that the respondents understood the intent. Other content involved information on the following subjects:

- Accuracy of private sector (PS) data.
- Sources of the data.

- Data cost variables.
- Cost comparison of PS data with fixed sensor costs.
- Geographic coverage by the various PS providers.

The webinar involved about 15 minutes at the end for questions from participants. No comments were posted in the chat box either during the webinar or as the webinar concluded, although there were a few questions and comments over the phone during the final few minutes. One person asked about the sample size used for preparing the cost graphic. This cost graphic was prepared using some default cost information, but was intended for users to input their own cost data.

There was a comment from the Fort Worth District about already having the detection infrastructure in place to do what commercial providers could offer, with a concluding comment that TxDOT has more to offer these providers than they have to offer TxDOT. Two issues with probe data from commercial providers are: 1) they generally do not offer vehicle count data—they typically offer link travel times or speeds; and 2) they do not normally offer per-lane data. The local agency could negotiate the cost paid for data from a commercial provider by offering the private company its own fixed sensors as a data source. The cost of private sector data is dependent on many factors such as:

- The number of miles covered.
- The number of ways the agency will use the data.
- Whether the data are real-time or historical.

The agency considering PS data should consider that the cost of these data is likely to be less than its own data using fixed sensors, but agency sensors also provide more types of data than just speed and travel times. An example cited for the cost of real-time data from INRIX is \$800 per mile plus an additional \$200 per mile for a one-time setup charge. However, the agency should realize that this is a starting point for negotiations. One of the stipulations might be that PS data must meet certain accuracy and availability criteria, or penalties can be assessed against the provider. For example, if PS speed data fall below a prescribed threshold of, say, 95 percent accuracy on speed, the agency would pay less than if the accuracy met the contracted conditions.

4.3 DOCUMENTATION OF TXDOT INPUT

The discussion during the webinar focused on the state-of-the practice in the use of real-time commercial traffic routing data and related specific issues for an implementation of such data use at TxDOT. Potential topics for this discussion may include the following:

- How applicable are the findings of real-time data use in other states to the TxDOT business process, and what changes would be required to the process to implement these practices at TxDOT?
- What near-term and long-term benefits can TxDOT expect from successful implementation of a private data use program, and what are the anticipated costs?

- How should issues such as data usability, accuracy, and reliability be addressed in a potential program implementation?
- What are legal, technical, and economic requirements for implementation of such a program at TxDOT, and what would be a reasonable time horizon until TxDOT can realize benefits from any such program?

4.3.1. Survey Responses

The research team documented the responses from TxDOT staff. An initial cursory analysis of the TxDOT feedback indicated that there is a case to be made for separating results by more urban districts versus more rural districts. Of course, this is a subjective exercise with some districts with a relatively equal split between urban and rural.

The discussion below includes a summary of responses, including the average values (where appropriate) to indicate the general attitude of TxDOT toward the question. Complete responses can be found at the end of the document. The survey included 12 questions, but the last one on data fusion was somehow omitted on all but one survey form.

Question 1 asked the percentage of real-time and historical data collected within the TxDOT respondent’s district by TxDOT forces. Just over half of the respondents answered the question indicating that, on average, TxDOT collects 76 percent of its own historical data and 80 percent of its historical data. The relatively high standard deviation in [Table 17](#) indicates that there is a sizeable range in the answers. Over half of the respondents answered the question.

Table 17. Q1: Percent of Data Collected In-House.

Components of Question	Mean	Std Dev	n
a. Real-Time	76%	39.5	13
b. Historical	80%	36.2	14

Question 2 (see [Table 18](#)) asked about the value placed by each respondent on data accuracy, availability, cost-effectiveness, and turnaround time (time from data collection to being able to use the data). The numbering system allowed respondents to score each criterion with a score from 0 to 10, with 10 representing the highest rank (most important criterion). Most of the individual scores were between 5 and 10. On average, accuracy ranked highest, followed by cost-effectiveness, although the resulting scores are not statistically different. All but one respondent answered this question.

Table 18. Q2: Value Placed by Districts.

Components of Question	Mean	Std Dev	n
a. Accuracy	8.58	1.71	19
b. Availability	7.53	1.65	19
c. Cost-Effectiveness	8.21	1.96	19
d. Quick Turnaround	6.95	1.99	19

Question 3 (see [Table 19](#)) asked which of the four options shown would most enhance traveler information. Speed measurement was rated, on average, slightly higher than travel time. As in Question 2, the respondent was asked to assign a value from 0 to 10, with 10 being the highest value or rank. Comparing Question 3 answers with Question 2 answers indicates lower numbers overall in Question 3 responses and higher standard deviations. The reason for this difference is unclear. All respondents answered parts c and d, but one respondent did not answer parts a and b.

Table 19. Q3: Which of the Following Would Most Enhance Traveler Information?

Components of Question	Mean	Std Dev	n
a. Travel time information	6.37	3.02	19
b. Levels of congestion	5.47	2.65	19
c. Speed measurement	6.65	3.07	20
d. Alternate route information	6.00	2.60	20
e. Comments			

Question 4 (see [Table 20](#)) also used a 10-point ranking scheme, with 10 being the highest score. The strongest support was for use of private sector data to help “fill in gaps” in coverage. Use of private sector data to reduce TxDOT’s reliance on its own fixed sensors was second, but barely ahead of using PS data where (other) ITS deployment was not cost-effective. Standard deviations in Question 4 responses were similar to those resulting from Question 3. All but one person answered this question.

Table 20. Q4: Rank of Statements for Districts.

Components of Question	Mean	Std Dev	n
a. PS data where ITS deployment is not cost-effective	5.72	3.37	19
b. PS data could help create uniform coverage across jurisdictions	6.33	2.68	19
c. PS data could reduce TxDOT’s reliance on fixed sensors	5.83	2.60	19
d. Comments			

Question 5 (see [Table 21](#)) asked TxDOT respondents how they would fill the void in traffic counts not provided by private sector providers. Respondents only used the two options provided—either continue to rely on TxDOT forces or contract it out. The answers were mixed in that all respondents said that TxDOT would continue to collect all or some of the data. In about one-third of the cases, the answers indicate that “others” would also be involved, in addition to TxDOT.

Table 21. Q5: Traffic Counts Not Available from Private Sector.

Components of Question	Yes/No
a. Continue to use TxDOT forces	Y=20
b. Hire others	Y=7
c. Other (please specify)	

Question 6 (see [Table 22](#)) asked about the importance of per-lane data since most private sector data providers do not offer data on a per-lane basis. Seventeen of the 20 respondents answered this question, with exactly half (10 of 20) indicating that it would not be a limitation. Seven of 20 (about a third) indicated that it would be a problem. The PS providers that offer per-lane data use the local agency’s fixed-site detector data and fuse with other data (e.g., probe data). Districts that indicated that lack of per-lane data would be a limitation include two predominantly rural districts—Brownwood and Yoakum.

Another rural district (Lufkin, responding “no”) simply indicated that private sector data would not be justified in rural districts. The Dallas District response indicates that per-lane data would be important for comparing the utilization of general purpose lanes with managed lanes. The El Paso District comment applied to volumes on a per-lane basis, perhaps not understanding that PS data would not typically include count data. The Laredo District response seems to emphasize that this new source of data would be considered even if it had limitations.

Table 22. Q6: Is Lack of Per-Lane Data a Limitation.

Components of Question	Yes/No
a. Yes	Y=7
b. No	N=10
c. Explain Comments for “No” responses: ELP – Do not use per-lane volumes LUF – Not justified in rural districts PHR – Need all the help we can get TransGuide – Travel time not done on a per lane basis Comments for “Yes” responses: BWD – Lane utilization critical in design DAL – Need to compare general purpose lanes with managed lanes YKM – Mostly need turning movement counts so need by lane (for traffic signal studies)	

Question 7 (see [Table 23](#)) asked TxDOT about the conditions that might cause them to consider purchasing private sector real-time data. The thought was that TxDOT fixed sensor failure rate and PS coverage, accuracy, and cost, might reach some acceptable value (assuming they are not already) and that the addition of count data might cause districts to reevaluate PS data.

Respondents who answered said they would consider purchasing PS data if TxDOT fixed sensor failures reach 56 percent in the next five years and 61 percent beyond the 5-year mark. PS coverage would need to reach 80 percent in both the near-term and long-term, on average, before the eight responding TxDOT personnel would consider PS data. Data accuracy would need to reach just over 80 percent in the next five years and over 90 percent thereafter, again for eight respondents. Six respondents said that when the cost of PS data is about the same as fixed sensors, they would consider purchasing PS data. Finally, six respondents said they would consider PS data if it included count data.

Less than half of the participants responded to this question, and some who responded did not answer all sub-parts of the question. Two less respondents answered a, d, and e sub-parts compared to others.

Table 23. Q7: Conditions for District to Consider Real-Time PS Data.

Components of Question	Ave (next 5)	Ave (>5)	n
a. % failures in TxDOT fixed sensors reach this value	56%	61%	6
b. PS coverage in your district above this value	80%	80%	8
c. Speed/TT accuracy reaches this value	81%	93%	8
d. Cost of PS data about the same as fixed sensors	Yes=6	Yes=6	6
e. Count data becomes available from PS providers	Yes=6	Yes=6	6
f. Comments			

Question 8 (see [Table 24](#)) asked TxDOT about the conditions that might cause them to consider purchasing private sector historical data. As in Question 7, the thought was that TxDOT fixed sensor failure rate and PS coverage, accuracy, and cost, might reach some acceptable value (assuming they are not already) and that the addition of count data might cause districts to reevaluate PS data.

Comparing these results with Question 7 results indicates that the average failure rates in TxDOT fixed sensors would need to be larger for historical data than for real-time data to justify consideration of PS data. At face value, this comparison indicates that there would need to be greater failures in fixed sensors used for historical data than for those used for real-time data. The average coverage is about the same for historical compared to real-time. Survey results indicate that the accuracy of speed and/or travel time measurements from PS data could reach much lesser values compared to real-time data. Only four respondents answered parts d and e, indicating they would consider purchasing PS data if the costs were about the same as fixed sensors and if PS providers include traffic counts as part of their data offerings.

Less than one-third of the participants answered this question. The reason(s) for the low response rate is unclear.

Table 24. Q8: Conditions to Consider Purchase of Historical PS Data.

Components of Question	Ave (next 5)	Ave (>5)	n
a. % failures in TxDOT fixed sensors reach this value	65%	70%	6
b. PS coverage in your district above this value	82%	82%	6
c. Speed/TT accuracy reaches this value	50%	60%	6
d. Cost of PS data about the same as fixed sensors	Yes=4	Yes=4	6
e. Count data becomes available from PS providers	Yes=4	Yes=4	6
f. Comments			

Question 9 (see [Table 25](#)) asked about the use of historical PS data and offered two answers plus an “other” option. There were nine respondents who said they would use PS historical data for origin-destination studies and nine who said they would use it for model calibration.

Table 25. Q9: Use of Historical PS Data if Purchased.

Components of Question	Responses
a. Origin-destination studies	O-D: 9
b. Calibrate simulation models	Calibrate model: 9
c. Other (please specify)	None

Question 10 (see [Table 26](#)) asked whether the use of the Traffic Message Channel mapping would be an impediment to the district using PS data. Four respondents indicated that it would clearly be a problem, but a higher number (7) indicated that it would not be an issue. There was a low response rate on this question.

Table 26. Q10: Would Use of TMC be an Impediment to Using PS Data.

Components of Question	Yes/No
a. Yes	Yes=4
b. No	No=7
c. Explain: 4 responded “uncertain”	

Question 11 (see [Table 27](#)) asked respondents to volunteer other ideas that PS data might be used for. One example that the survey cited was “hurricane evacuation” and some respondents agreed with that idea. Most of the examples were general in nature, but two were specific—compare travel time along a segment of I-35 and compare with SH 130, which is a toll facility; and use of PS data to meet the USDOT mandate for real-time monitoring systems (Sec 1201). Two districts along the border between Texas and Mexico suggested uses related to international ports of entry (e.g., travel time to cross the border). Nine respondents offered suggestions, although several of them offered more than one suggestion.

Table 27. Q11: Other Examples of Long-Term Opportunities for PS Data.

<ul style="list-style-type: none"> • Tolling • Operational validation • Hurricane evacuation • Other evacuations (non-hurricane) • Flooding • International POEs • Border violence (causing traffic anomalies) 	<ul style="list-style-type: none"> • Work zones (2), • O-D freight • Real-time system management • USDOT mandate for real-time monitoring systems (Sec 1201) • Incident avoidance • Special events (2) • Travel time comparison I-35/SH 130
---	--

Question 12: Are there data fusion issues that might be a problem for TxDOT? This question was inadvertently omitted from all but one of the surveys. That response indicated that there would be no issues anticipated with data fusion based on private sector data.

4.3.2. Overall Comments on the Survey

The intent of this survey was to gather TxDOT input—mostly from district personnel. The sample included 17 districts of the total 25 TxDOT districts; therefore, the sample size is adequate for district representation. The Institutional Review Board required that researchers include the following statement in the mail-out to participants, “You may elect to not answer any question and you may terminate the interview at any time.” The fact that some respondents did not choose to answer all questions (per IRB policy) and the fact that two districts had more than one person responding is a potential source of bias in the results.

5. DEVELOP OPPORTUNITY MATRIX

5.1 INTRODUCTION

Based on input from the TxDOT survey, a review of other state DOTs, and researcher understanding of ITS needs, the TTI team developed a comprehensive list of opportunities for TxDOT to consider pertaining to future use of private sector data. Specific opportunities for applying private data were reviewed in light of accuracy of the data, coverage areas, data availability, cost, and control of the data stream. The list of opportunities considered included:

- Enhance traveler information in urban areas such as:
 - Travel time information.
 - Levels of congestion.
 - Speed measurement.
 - Alternate routes.
- Introduce traveler information in areas where ITS deployment is not cost-effective.
- Improve continuity of data based on existing ITS coverage across jurisdictions.
- Develop statewide 511 system.
- Reduce ITS deployment costs by limiting deployment of fixed data collection devices.

TxDOT has deployed a variety of field devices to relay traveler information to motorists and other users. These devices include dynamic message signs, highway advisory radio, and others. In many urban areas, Transportation Management Centers receive data from vehicle sensors and cameras, and the data are processed and converted to useful information to be disseminated to the traveling public. However, there are situations where gaps in coverage exist and where private sector data could fill the gaps. For example, when TMCs detour traffic from freeways to surface streets, there might be no means of monitoring the congestion levels on the streets without private data. The same could be true of rural areas where the deployment of ITS is minimal.

If private agencies have coverage on these roadways and have sufficient data, TxDOT could purchase the data and provide traveler information without making huge investments to deploy ITS. Even with much of the desired coverage in place through past TxDOT efforts, it is conceivable that data from private providers could fill in gaps that would be difficult or unfeasible using traditional methods.

Should TxDOT decide to move forward with purchasing data from a private source, an important consideration in the licensing process is building in qualifiers that specify penalties if the private data fall below preset performance thresholds. The San Francisco 511 program gives one example. As [Table 28](#) indicates, the speed data from the PS provider must be within 5 mph or 7.5 percent of true speed (whichever is greater) 85 percent of the time, or within 10 mph or 15 percent of true speed (whichever is greater) 95 percent of the time. The data must be available

90 percent of the time. If these requirements are not met, the cost owed to the provider drops by the same amount as the accuracy or availability.

Table 28. San Francisco 511 Real-Time Pricing Stipulations.

Value	% of Data	Within
Speed data ^a	85% 95%	7.5% or 5 mph ^b 15% or 10 mph ^b
Availability ^c	90%	N/A

^a In all environmental conditions.

^b Whichever is greater.

^c Data availability or accuracy < amts shown reduces cost by same amount.

5.2 SAFETEA-LU SECTION 1201

Section 1201 of the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU), published on November 8, 2010, establishes the provisions and minimum parameters for the Real-Time System Management Information Program to be established by State DOTs, other responsible agencies, and partnerships with other commercial entities. SAFETEA-LU mandates that the program be established on all Interstate routes within four years (November 8, 2014) and on other significant roadways as identified by State and local agencies within six years (November 8, 2016). [Table 29](#) identifies the key requirements of the information delivery timeframes.

Table 29. Information Delivery Requirements of Section 1201.

Information Type	Metropolitan Area (Minutes)	Non-metropolitan Area (Minutes)	Availability (Percent)	Accuracy (Percent)
Implementation or removal of lane closure	10	20	90	85
Roadway- or lane-blocking traffic incident information	10	20	90	85
Roadway weather observation updates	20	20	90	85
Travel time along highway segments	10	N/A	90	85

The research team understands that the Traffic Operations Division is currently considering ways how to meet the requirements of Section 1201. While many urban areas as well as some freeway corridors between major cities, have transportation management center coverage and infrastructure, the vast majority of the areas do not have coverage. This will require either the installation of additional infrastructure, or the purchase of private sector data to fill the

gaps. Also, information on road weather conditions is not a typical component of current TMC activities and information delivery. The key issues are therefore:

- How to comply with the requirements of the final rule in a cost-effective manner.
- How to integrate the disparate sources of data to meet the requirements.
- How to assemble the information and present it in a concise but comprehensive way to the traveling public.

5.3 STRENGTHS AND WEAKNESSES OF EACH METHOD

The options being considered in this analysis are:

- Use of fixed sensors (e.g., inductive loops or non-intrusive technologies).
- Use of private sector data.
- A combination of the two.

TxDOT uses the following fixed sensors for collecting real-time data include the following primary technologies:

- Inductive loops.
- Video imaging.
- Radar detectors.
- Magnetometers.

It is appropriate to consider the strengths and weaknesses of fixed sensors versus private sector data in order to maximize the use of known information about each approach. Each source of data has its own inherent strengths and weaknesses, so TxDOT should weigh each of the metrics in terms of its importance in TxDOT practice. The proposed metrics for making this comparison are:

- Detection accuracy.
- Data source (origin and quality control).
- Control of data stream.
- Desired uses of data.
- Data coverage.
- Life-cycle cost.

5.3.1. TxDOT-Maintained Fixed Sensors

Detection Accuracy

Inductive loops are the most mature of the technologies listed, so installers know much about how to install them. The best count accuracy for vehicle detection assuming proper installation and maintenance of inductive loops indicates ± 2 percent error. A more realistic range for count accuracy is ± 5 percent. Speed errors are often in the ± 5 to 10 percent range.

Video imaging accuracy is a function of lighting and weather conditions, and their position beside and above the roadway. Occlusion is a function of the mounting height and lateral distance from lanes being monitored, and it compromises accuracy in most situations.

The best count accuracy for vehicle detection using video assuming perfect weather and daylight conditions is about ± 5 percent error. Count accuracy for nighttime conditions and/or poor weather and with a high percentage of tall vehicles falls within ± 10 to 20 percent. Speed errors are usually in the ± 5 to 10 percent range.

Radar detectors (typically mounted side-fire) are not affected significantly by weather or light conditions, but are affected by occlusion, which (like video) is a function of the mounting height and lateral offset from detected lanes. The best count accuracy for vehicle detection using radar is in the ± 2 to 5 percent error range, but can be as high as ± 5 to 10 percent with high truck percentages. Speed errors are usually in the ± 5 to 8 percent range.

Magnetometers mounted in the pavement are becoming more prevalent as loop replacements and are about as consistent as loops for detection of most vehicles. Problematic vehicles include motorcycles and large trucks. Of course, no weather or light conditions affect their performance and occlusion is not an issue. The best count accuracy for vehicle detection with Sensys Networks magnetometers is ± 2 percent error. A more realistic range for count accuracy is ± 5 percent. Using single magnetometers (two stations per lane spaced a known distance apart longitudinally) speed errors are usually in the ± 2 to 10 percent range. Performance improves (e.g., motorcycle detection) by using multiple magnetometers instead of just one. Software enhancements improve truck detection.

Data Source

With fixed sensors, TxDOT usually has full control of the data source and determines the quality of the data and whether the data are useful. TMC control in larger urban areas usually means that the data coming into the center goes through a Q/C algorithm. Out-of-bounds data usually result in the sensor being flagged and perhaps taken off line and eventually replaced. However, limited resources result in some of the field devices running for long periods of time, especially in smaller urban areas, without adequate Q/C checks. Some problems are intermittent and difficult to diagnose. One of the downsides to any problems or failures is that TxDOT is responsible for remedying the problem.

Control of Data Stream

TxDOT control means that there is less doubt about the data source and how the data might have been “filtered” before final use. Having full control involves a higher comfort level than having partial or no control, at least until TxDOT uses a low-control data source long enough to build confidence in the data.

Desired Uses of Data

With full TxDOT control, TxDOT can use the data in any way it wants. In most cases, fixed sensors generate speeds, counts, occupancies, and length-based classification, and the data are available on a per-lane basis. TxDOT has the option of using any or all of the available data with little or no difference in cost. TxDOT can archive the data for subsequent retrieval, again for little or no additional cost.

Data Coverage

With fixed sensors, the data coverage is whatever TxDOT considers feasible within the limited resources available. Sensor spacing and the parameters defining the data stream are based on TxDOT design, although again, based on limited resources. Resulting coverage is typically limited to the most congested portions of urban systems, with outlying areas not covered as well. Reaching these lesser congested areas is often desirable, but limited resources do not allow or delay the expansion until the problem worsens.

Life-Cycle Cost

Determining the life-cycle cost of fixed sensors is challenging at best. Most agencies do not maintain the foundational cost data to be able to calculate life-cycle costs. TTI has developed guidance based on previous research and calculations from the Utah DOT. UDOT costs might be different from TxDOT costs, at least in terms of the replacement cycle of some in-pavement sensors or due to differences in weather patterns. For detectors not affected by weather, this factor is not usually an issue. For purposes of this analysis, these differences will be considered minimal. TTI used the UDOT data and other sources to develop a life-cycle cost comparison. A later section in this chapter provides this comparison.

5.3.2. Private Sector Data

The data provided by private sector providers are generally limited to speeds and travel times. From these values, one can identify incidents and bottlenecks. The data do not usually contain vehicle counts, but private sector providers sometimes enter into arrangements with public sector agencies to access count data from the public sector’s fixed sensors. These shared arrangements have implications on the price negotiated with private sector providers.

Detection Accuracy

For private sector data, the accuracy is a function of the number of probes in the traffic stream. Data from the largest PS providers have multiple sources, but the primary source is based on GPS devices. These devices are known to generate accurate speeds under almost all conditions. Based on this research, the speed accuracy of PS data is usually within the bounds of ± 5 to 10 percent and is expected to improve with time since additional probes are being added daily through voluntary incentive programs. Private providers have algorithms that provide the necessary Q/C, so the result is an accuracy level with such a modest error that the average driver will not be affected.

Data Source

Table 30 indicates the source of data for various providers. GPS use has grown substantially in recent years due to improving accuracy and reasonable cost. The use of cellular probes alone is not viewed as having the same accuracy as GPS, assuming the PS provider determines speed based on cell tower “hand-offs.” This process would generate no location information between towers; for roadways with adjacent frontage roads, there would be no way to distinguish between vehicles on the main line and those on the frontage roads (which usually have different speeds). SpeedInfo uses Doppler radar, which is a reliable speed detection device. Bluetooth is also known to generate accurate speeds.

Table 30. Provider Primary Data Sources.

Provider	GPS-enabled Vehicles	Cellular Probes	Fixed Point Sensors	Others
AirSage		Yes		
Cellint		Yes		
Delcan		Yes		
INRIX	Yes	Yes	Yes	
NAVTEQ	Yes	Yes	Yes	
OnStar	Yes			
SpeedInfo			Yes (radar)	
TomTom	Yes	Yes	Yes	
Total Traffic Network	Yes	Yes	Yes	Airborne/Mobile Spotters, Cameras
TrafficCast	Yes		Yes	Bluetooth

In considering the accuracy desired for the various sources of data (fixed or private), one should consider the differences in travel times (prediction errors) that would result from speed errors. The first consideration is that the errors could be random, resulting in errors that tend to cancel each other when considered on an aggregate basis. For example, a speed reading that is 5 percent high would be canceled by another speed reading that is 5 percent low. Obviously, the magnitudes of the errors are not necessarily equal; if these are not, there would be a resulting error based on the difference. There could be an overall bias in speed readings, which means that all speed detections would be either high or low and there would be no cancelling as in the first example.

Figure 15 shows some of the errors in travel times that would result from the typical errors in both fixed and private data if the prevailing speed is 50 mph. In this case, the graphic shows errors in travel time based on speed measurements that range from 5 percent below the actual speed to 5 percent above the actual speed. For example, a trip of 30 miles would range from a low of 33.5 minutes to a high of 37.8 minutes. The actual travel time if the speed measurement had no error would be 36.0 minutes. The average motorist would not notice the difference.

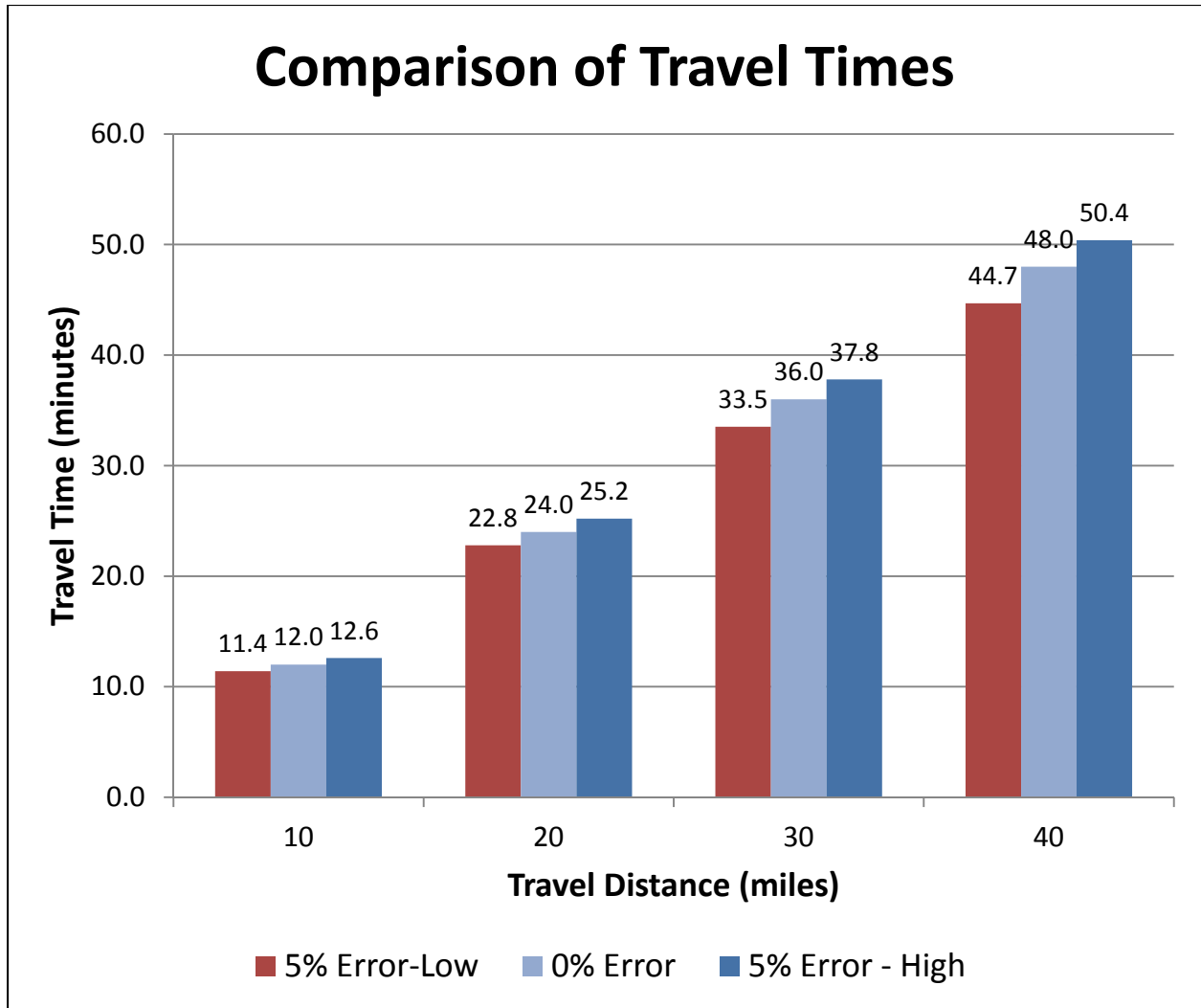


Figure 15. Travel Time Errors Based on ±5 Percent Speed Error.

Control of Data Stream

With the use of PS data, TxDOT has little or no control over the data stream. While this might appear to be an issue at the beginning of some future contract period, TxDOT will need to weigh the pros and cons and decide whether the merits are worth the risk. Since TxDOT has the denser urban areas covered with fixed sensors, the best approach might be to test PS data in urban fringe or rural areas to see how any apprehensions might play out. One precedent in this decision has been TxDOT's use of toll tag systems in Houston and other urban areas where there are sufficient vehicles with tags to serve as probes. In some cases, others have provided the data stream.

Desired Uses of Data

With private data, the more ways TxDOT uses the data the higher the cost will be. Real-time data cost more than historical data, and faster updates will likely increase the cost. The only way to accurately determine the cost of some of the PS data is for TxDOT to discuss the desired uses, coverage area, update frequency, and other details with one or more providers. A few have established a starting point in the negotiation process. For example, INRIX and SpeedInfo have both set costs of data based on some set conditions. Again, changing the uses of the data from the preset conditions will change the cost.

Data Coverage

The data coverage that TxDOT could expect would include the Traffic Message Channel network throughout the state. This would involve all major freeways and other major roadways throughout the state and most urban arterials. Coverage on lower volume roadways is a function of the number of probes that are generating data. Probes include fleet vehicles such as trucks and taxi cabs, so areas with a sufficient number of trucks such as commercial zones and industrial areas should have sufficient coverage. Based on the survey of TxDOT personnel, the TMC network is not necessarily a hindrance to using private sector data, but TxDOT must realize that the segments in rural areas could be longer than the spacing between fixed sensors such as Bluetooth. [Chapter 2](#) of this report offers more details on the TMC network.

Life-Cycle Cost

As noted elsewhere, the cost of some private sector data will not be known to a prospective DOT until the DOT negotiates a price with a provider. One exception is SpeedInfo. This company installs and maintains autonomous Doppler radar units alongside the roadway and uses its own solar power and wireless communications to generate data for the operating agency. The cost of this service is \$110 per month per bi-directional station.

The other advertised cost is from INRIX. It amounts to \$800 per mile per year with an additional first-year cost of \$200 per mile. There are also discounts available for some of the network, but few details are available.

An additional up-front cost that TxDOT must consider is the cost of its own independent verification of PS data. One low-cost option would be the use of Bluetooth systems interspersed along major routes with update frequencies similar to that of PS providers.

5.3.3. Summary

Table 31 provides a summary of the factors cited above, with the exception of life-cycle cost. The cost discussion follows. The comparison includes two different types and orientations of radar detectors: side-fire and parallel to the traffic stream. TxDOT uses products from two manufacturers in side-fire to cover freeways as a fixed sensor. Doppler radar is oriented parallel (or approximately parallel) to traffic and is a proven technology for accurate speed detection.

As noted elsewhere, Bluetooth readers detect devices passing in vehicles that generate a sufficiently strong signal. Each device (e.g., cell phones) generates a unique MAC address that can be read at two points with known separation distance. The link travel time is the difference in the timestamps at the two detection points.

Table 31. Summary Comparison of Data Sources.

Measure of Performance	Private Sector Data	Bluetooth	Loops	Video	Side Fire Radar	Magnetometers
Speed Accuracy (%)	±5–10	±5–10	±5–10	±5–20	±5–10	±2–10
Count Accuracy (%)	N/A (w/o TxDOT sensors)	N/A	±2–5	±5–20	±2–5	±2–5
Data Source	GPS: High Doppler Radar: High Bluetooth: High	High	High	Medium	High	High
Control of Data Stream	Low	Low	High	High	High	High
Uses of data						
- Speed/Travel Time	Yes	Yes	Yes	Yes	Yes	Yes
- Counts	No	No	Yes	Yes	Yes	Yes
- Occupancy	No	No	Yes	Yes	Yes	Yes
Coverage	TMC Network	As determined by TxDOT				

Note: N/A = Not Applicable

Figure 16 summarizes the cost of private sector data compared to fixed sensors estimated on a per-mile basis for 1-, 5-, and 10-year costs—low to high. Of course, such comparisons require many assumptions. TTI developed this comparison in Excel format to allow users to

input location-specific values to make it more meaningful. In the figure, “BT” is Bluetooth, “Pvt Sec” is private sector data, and “Mag” is magnetometers.

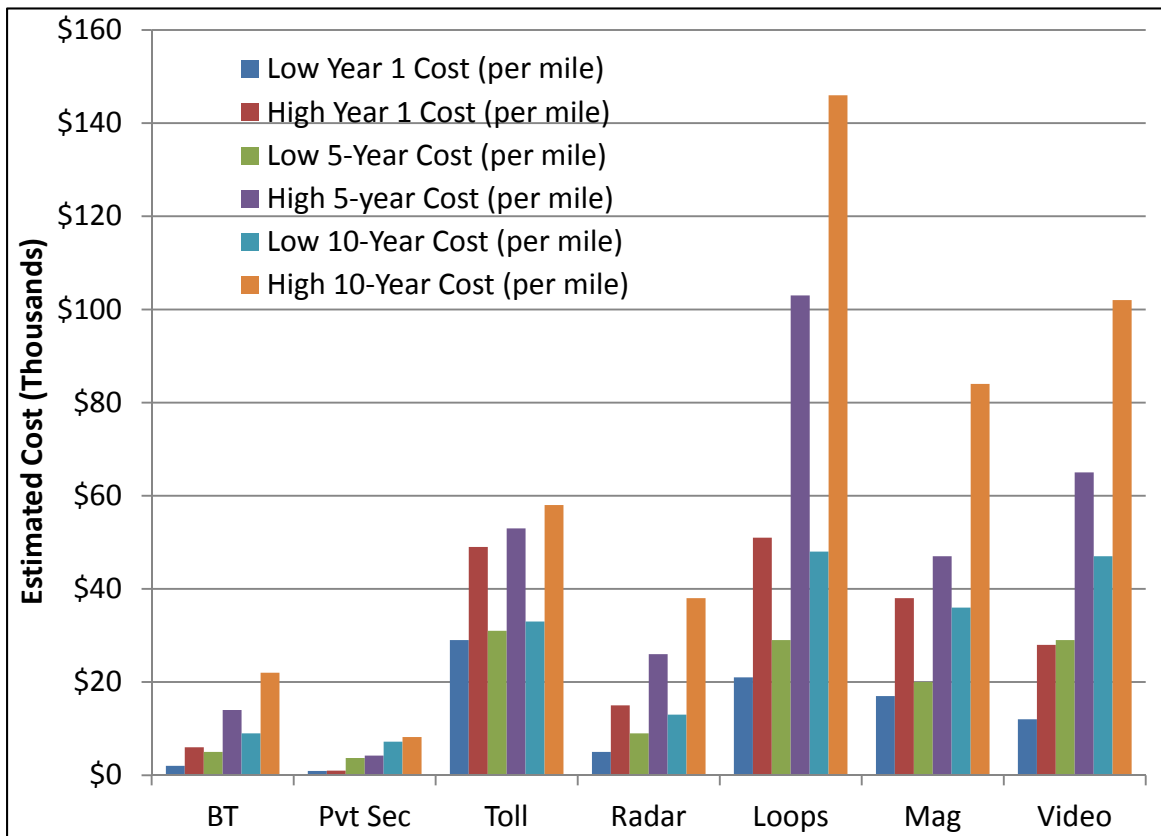


Figure 16. Comparative Costs of Detection.

5.4 COMPREHENSIVE OPPORTUNITY MATRIX

Based on input from the TxDOT survey, a review of other state DOTs, and researcher understanding of ITS needs, the TTI team developed a comprehensive list of opportunities for TxDOT to consider pertaining to future use of private sector data. The researchers hope that opportunity matrices presented in this research in subsequent sections would provide TxDOT with subjective and qualitative tools to determine appropriateness of implementing private sector data to achieve its intended goals and objectives.

They reviewed specific opportunities for applying private data in light of accuracy of the data, coverage areas, data availability, cost, and control of the data stream. The list of ITS application areas considered includes the following as shown in [Table 32](#). Application areas are by no means exhaustive; rather, these areas originated from the survey of TxDOT staff.

While considering the use of private sector data, its strengths and weaknesses as well as opportunities such data provide in relation to the specific ITS application area should be carefully evaluated. [Table 33](#) provides an evaluation of strengths, weaknesses, and opportunities when using private sector data in different ITS application areas. Even though the evaluation is entirely subjective, it provides an excellent starting point for TxDOT to build an understanding of the private sector data. Researchers not only used results from the survey performed among TxDOT staff, but also considered scope and cost of private sector data in developing such evaluations. On a cautionary note, researchers believe that the evaluations provided in [Table 33](#) could change over time as private sector data improves along with changes in the needs of TxDOT.

Table 32. List of Opportunities Considered in the Study.

ITS Application Group	ITS Application Area
Traveler information	Enhance coverage of traveler information in urban areas
	Enhance traveler information in rural areas
	Statewide 511 system
	Emergency evacuation
	Work zone information
System planning	Performance measurement
	Model input and calibration
System operation	Faster identification of congested areas
	Predictive information

The research also identified six governing factors that would come into play while making decisions to use private sector data. These governing factors do not exert equal influence in making the decision to utilize private sector data and vary depending on the application area as well as the urgency of implementing them. Hence, the relative importance of these governing factors may also vary between districts due to the district's regional needs, funding availability, and so forth. [Table 34](#) presents the relative importance of six governing factors in relation to specific application areas based on the survey and researchers' knowledge of TxDOT's needs. The importance is presented on a scale of 1 through 3, 1 being less important (less concerning) and 3 being of highest importance (most concerning).

Table 33. Strength, Weaknesses, and Opportunities of the Private Sector Data in Relation with ITS Application Areas.

Application Area	Strength	Weakness	Opportunities
Enhance coverage of traveler information in urban areas	Where TxDOT has ITS deployments, private sector data can be used to improve the information provided to travelers.	Close coordination of city and the state is required regarding sustained funding for procurement of data on arterials maintained by the city but are in TxDOT's ROW—who will pay for what and where?	Provide traveler information on arterials and state highways where there is limited ITS deployment and be able to meet SEC 1201 requirements.
Enhance traveler information in rural areas	Acquiring private sector data could be more cost-effective than TxDOT deploying and maintaining fix point sensors.	Complex procurement language may be necessary to cover for data gaps and availability in case enough probe vehicles are not available.	Enhance coverage of rural areas where ITS is not available and not cost-effective to deploy fixed-point sensors and also meet SEC 1201 requirements.
Statewide 511 system	TxDOT can quickly deploy the 511 system using the private sector data and show traffic conditions on rural as well as urban roadways throughout the state.	Covering the entire state could be expensive. Complex procurement language may be necessary to cover for data gaps, data availability.	The statewide 511 system will show traffic conditions at rural areas where ITS is not available and not cost-effective to deploy and also merge traffic conditions from existing ITS in urban areas.
Emergency evacuation	Private sector data will act as an additional source of traffic information.	The private sector may not be able to report traffic conditions on roadways due to absence of probe vehicles.	Determine alternate routes in dynamic environment and provide that information to the traveling public.
Work zone information	Identify alternate routes, proliferation of congested links around work zones.	Per lane information is not available; hence, private sector data may not be effective for traffic routing within the work zone in a smaller area.	Traffic management operators can monitor how and where the congestion is moving and expanding around at and around the work zone.
Performance measurement	Private sector data can be both cost-effective and efficient while reporting performance on a continuous basis (year after year).	Since volume is typically not available from private sectors, MPOs have to fuse private sector data with volume data from fixed sensors if performance measures use combination of volume and speed.	Many districts (and even MPOs) in Texas have not been able to establish congestion-related performance measures mainly due to lack of continuous data source to measure performance. There is a growing trend among states and MPOs to use private sector data to fill that void.

Application Area	Strength	Weakness	Opportunities
Model input and calibration	Private sector data such as historic snapshots of traffic conditions can be helpful when simulating and modeling wider areas (regional or corridor wide).	Mesoscopic and microscopic modeling needs highly granular traffic conditions data, which may not be available from private sector data.	Mostly beneficial while comparing results of the macroscopic modeling with historic traffic conditions provided by the private sector data.
Faster identification of congested areas	Private sector data already cover most major urban areas in Texas and provide snapshots of traffic conditions in user-selected intervals.	In urban areas where there is already wide deployment of ITS to measure traffic conditions, private sector data may be viewed as just another “redundant” data source.	Traffic management operators can be provided with a regional view of traffic conditions that will allow them to quickly identify where the congestion is building and point surveillance cameras to the area.
Predictive information	Many private sector agencies already provide short-term prediction of travel time and speed on roadway segments.	The prediction models of private sector agencies are not transparent to data subscribers; hence, their performance is difficult to ascertain.	Traffic management operators can monitor, proactively, where congestion will build up and focus ITS resources in that area.

Note: ROW = Right of Way, MPO = Metropolitan Planning Organization

Table 34. Relative Importance of Various Governing Factors.

Application Area	Spatial Coverage	Cost Effectiveness	Information Accuracy	Data Reliability	Control of Data Stream	Quick Procurement
Enhance coverage of traveler information in urban areas	1	2	2	1	2	3
Enhance traveler information in rural areas	3	2	1	1	3	3
Statewide 511 system	3	3	2	1	2	2
Emergency evacuation	3	1	3	3	3	2
Work zone information	1	2	3	3	2	3
Performance measurement	3	2	3	2	1	2
Model and calibration	3	1	3	2	1	2
Faster identification of congested areas	3	1	3	2	3	1
Predictive information	1	1	3	2	3	1

6. INVESTIGATE DATA FUSION

6.1 INTRODUCTION

Traditionally, TxDOT and its districts have collected most of the necessary traffic data by using fixed location sensors with supplementary data collected through probe vehicles. Private traffic data providers offer data such as travel speeds, travel time, and delay. This research project addresses the challenges of:

- Evaluating the benefits of utilizing private data in conjunction with TxDOT data.
- Determining how to merge TxDOT data with private data to enrich traffic information provision to TxDOT and its districts.

Here is the outline of the data fusion task:

- Investigate previous studies conducted on techniques for merging different data sources.
- Identify possible data attributes and propose prospective data fusion architecture to fuse TxDOT and INRIX databases to obtain an enriched traffic database. This process demonstrates a pictorial and conceptual framework for data fusion architecture by taking an example case using the El Paso Reference Mile Marker with INRIX data files.
- Summarize findings and results.

Figure 17 presents the work-flow of assigned tasks to UTEP on the data fusion process.

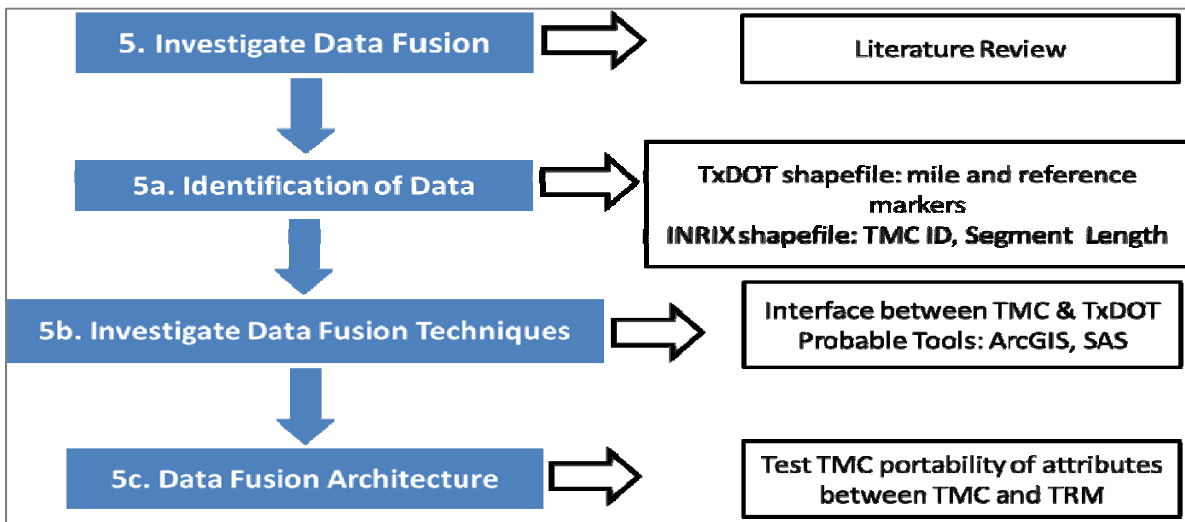


Figure 17. Framework for Data Fusion.

6.2 PAST STUDIES ON DATA FUSION

The application of data fusion techniques consists of merging multiple data sources to obtain a single data stream that can be utilized for a specific purpose. This concept is relatively new, especially for real-time transportation data. Currently, data fusion techniques are being utilized for:

- Research related to robotics.
- General image processing.
- Non-government projects such as weather surveillance.
- NASA missions (Dailey, et al., 1996).

From a transportation engineering perspective, the application of data fusion techniques would entail the combination of data from different transportation data collection sources such as radar, infrared, loop detectors, and video (visual). Most DOTs (including TxDOT) utilize one or more of the abovementioned data collection methods for Advanced Traveler Information System (ATIS) applications. As an example of what data fusion would require, Figure 18 shows the sequencing of major functions of the data fusion process with respect to ATIS applications.

Raw Data Collection	<ul style="list-style-type: none"> • Transmitting and receiving error-free data from field sensors
Data Identification	<ul style="list-style-type: none"> • Matching the sensed data with the source for missing data values
Data Alignment	<ul style="list-style-type: none"> • Configuring identified sensed data to a common spatial and temporal reference/origin, as well as transforming compatible representations and/or languages
Data Combination	<ul style="list-style-type: none"> • Performing various association analyses (e.g., statistical correlations, pattern recognition) to improve detection, classification, and tracking of entities of interest
State Estimation	<ul style="list-style-type: none"> • Predicting the kinematic (time and/or spatial) performance of an entity of interest
Performance Assessment	<ul style="list-style-type: none"> • Applying techniques to assess fused data quality and fusion processes

Figure 18. Functional Steps in the Data Fusion Process.

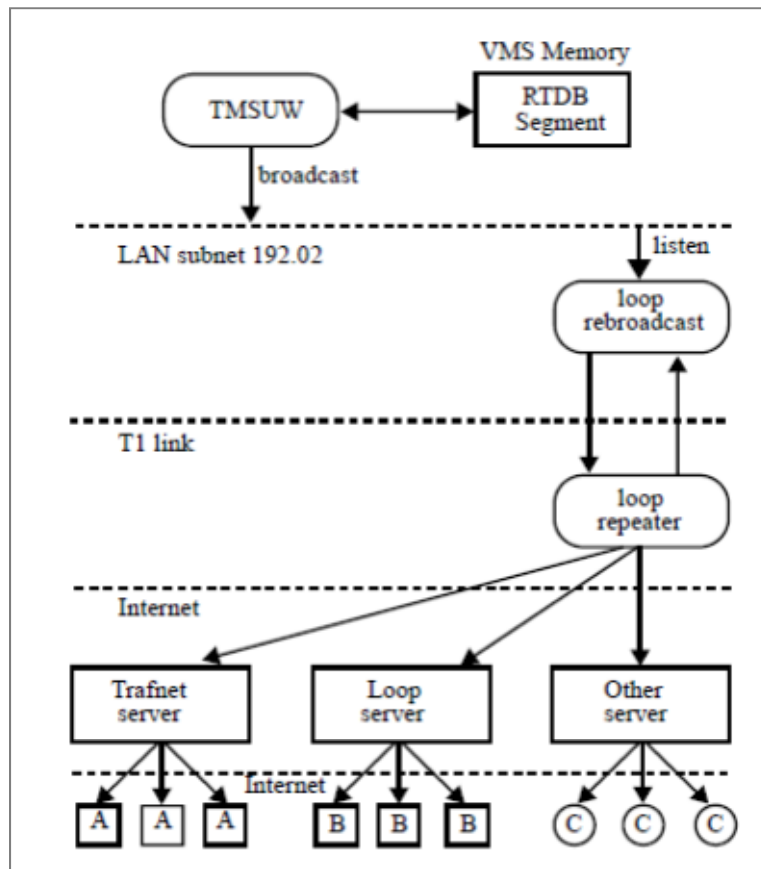
The available literature on data fusion applications to transportation data has primarily dealt with traffic information provision related to traffic conditions or incidents and is sparse.

Therefore, the intent of this literature review is to synthesize data fusion techniques within a transportation context.

Dailey, et al. (1996) focused on the data fusion application known as Traffic Systems Management Center (TSMC) traffic reporting system. The authors set up two goals for the data fusion project:

- Gather traffic congestion information from all available sources to make reliable traffic predictions.
- Support travelers by providing them with up-to-the-minute information on highway congestion to help their transit decision.

Figure 19 illustrates the architecture of data flows under the data fusion technique presented in the study. The process provides information to users through loop servers by transmitting estimates of occupancy, volume, speed, and length data for each loop and station.



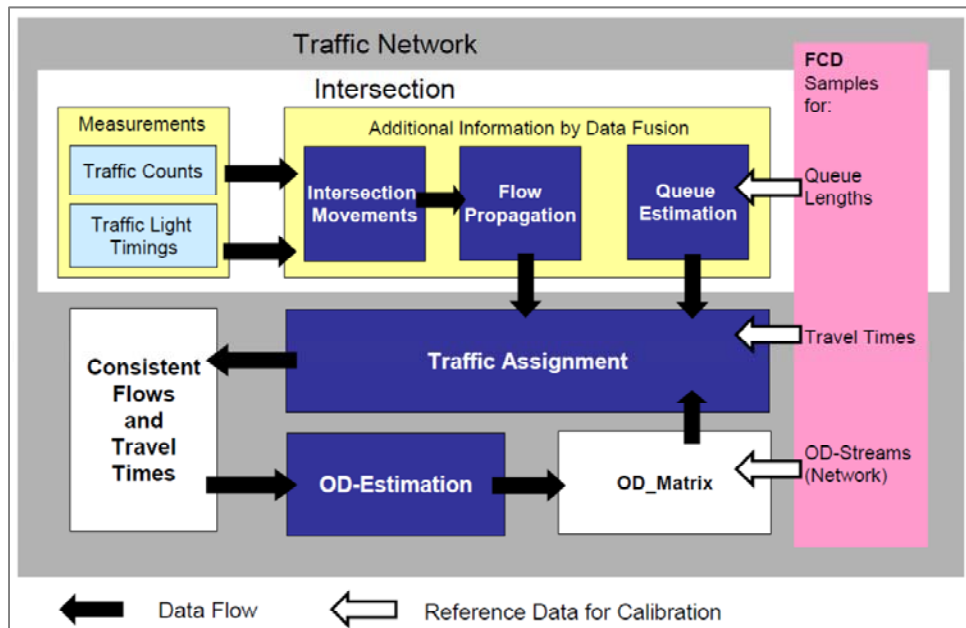
(Dailey, et al., 1996)

Figure 19. Architecture of the TSMC Traffic Reporting System.

Figure 20 shows how (Matschke, et al., 2004) applied a data fusion technique to traffic state estimation. Their procedure consisted of:

- Fusing traffic counts and traffic light timings for the estimation of volumes for all movements within an intersection.
- Propagating the estimated volumes to the adjacent sectors, offering the opportunity to obtain more exact data at each detected intersection and also to fill data gaps on links where no detector was available.
- Estimating queue lengths using a data fusion technique by combining traffic counts and traffic signal timing (Floating car data provided calibration estimate).
- Estimating the first OD matrix by applying the information on link flows (and turning movements) and estimating the route choice.
- Performing a traffic assignment based on the processed data (e.g., volumes, queue lengths, and OD matrix).

The research team then performed a new iteration of OD estimation based on these data. Floating car data again served the purposes of comparing and calibrating travel times as well as weight for the OD estimation.

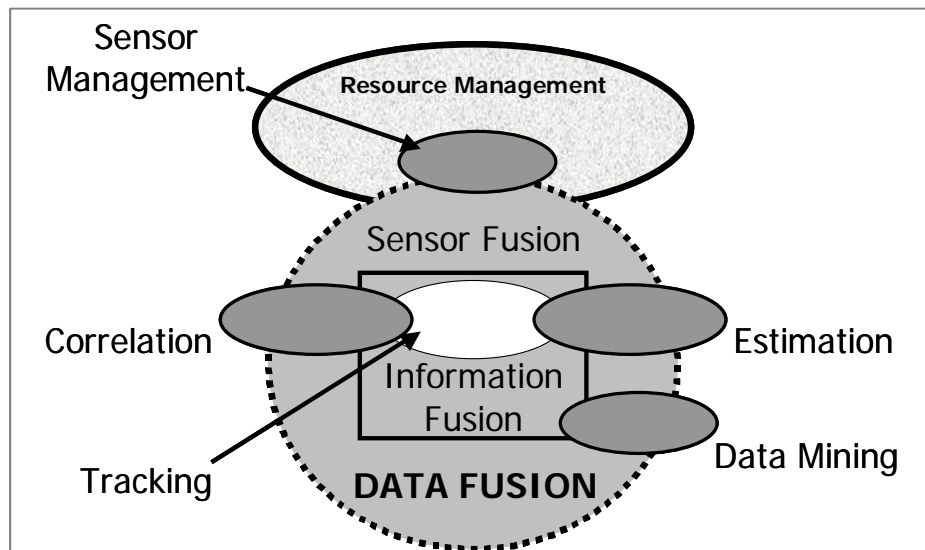


Source: (Matschke, et al., 2004)

Figure 20. Data Fusion Technique in Traffic State Estimation.

In more general terms, a report from the (U.S. Department of Transportation ITS Joint Program Office, 2003) described a data fusion methodology for delivering advanced traveler information services. As depicted in Figure 21, the methodology is described as:

- **Sensor Management:** This refers to the range of activities to ensure sensor data are accurately formatted and processed in a timely manner as the data fusion subsystems require. It also covers possible control of the sensor operations and adjustment of the data processing in order to improve estimation and prediction of selected objects.
- **Data Mining:** It is the nontrivial discovery of meaningful new correlations, patterns, and trends and the extraction of implicit, previously unknown, and potentially useful information from large amounts of data. By applying sophisticated software tools, an analyst is able to infer rules from among data objects that can be used to predict future states or guide decision making.
- **Estimation:** Refers to the use of methods to infer information or define parameters about a general population based on a limited set of observations about the population.
- **Correlation:** Refers to the degree of relationship among two or more variables.
- **Tracking:** Denotes the temporal and geospatial location of objects.



Source: (U.S. Department of Transportation ITS Joint Program Office, 2003)

Figure 21. Data Fusion in Overlapping Disciplines.

The tracking of objects involves a number of factors including uncertainty in position location, management of multiple measurements for the same object, database organization, and scalability. In short, the report offers a methodological approach to perform data fusion for delivering relevant traffic information to roadway users. Figure 22 shows that data fusion can be separated broadly into two categories, namely data centric and model centric.

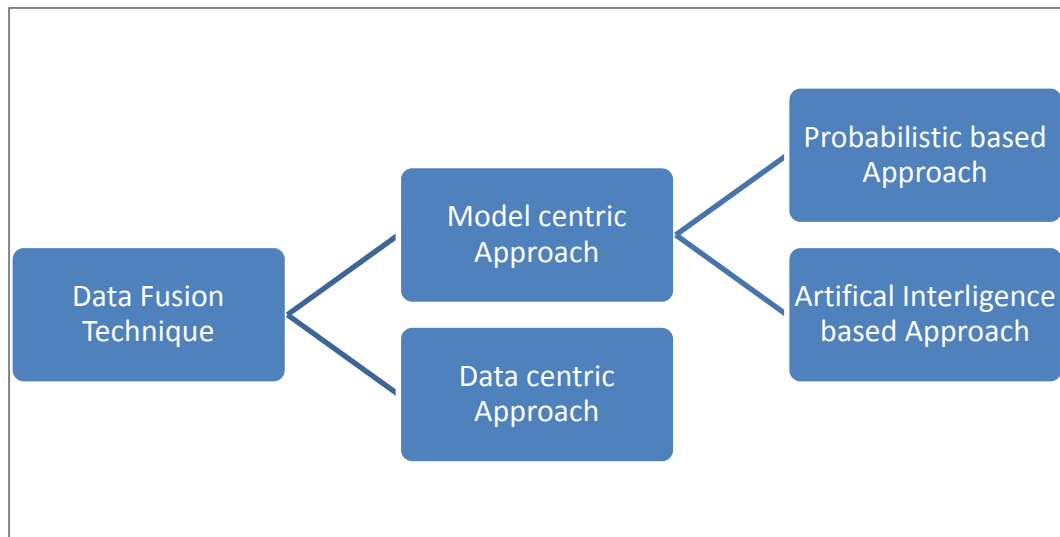


Figure 22. Data Fusion Technique Classification.

A data centric system usually focuses on analyzing large pools of data obtained from a host agency database and merging them with external data sources (e.g., private vendor data, or outside data source). The analysis and fusing of the data is typically a well-defined and properly-structured problem involving timely collection, processing, storing of information at proper locations (e.g., data warehouses), and disseminating the information based on the ATIS system design or subscriber services (U.S. Department of Transportation ITS Joint Program Office, 2003). This research project utilizes a data-centric approach.

A model centric system may draw on the data centric information to make informed estimates of patterns, trends, and “what if” scenarios. In case of ATIS applications, these estimates include a variety of topics—for example, forecast of network clearance times given the current depiction of an event or the estimated changes in historic mean speeds on highways when it rains at a certain pace and time of day (U.S. Department of Transportation ITS Joint Program Office, 2003). El Faouzi (2004) further classified the model centric approach into one or more Probabilistic-based approaches. This would require the application of probabilistic approaches such as Bayesian models (Okutani, 1987), Possibility theory (Dubois, et al., 1988), Evidential Reasoning and Evidence theory (Dempster, 1967) and (Shafer, 1976).

Evidence theory technique could be viewed as a generalization of the Bayesian approach (Dempster, 1967); and Artificial Intelligence-based approaches: the examples of such approaches could be Neuromimetic networks and artificial cognition including artificial intelligence, genetic algorithms, and neural networks. In many applications, this approach serves both as a tool to derive classifiers or estimators and as a fusion framework of classifiers/estimators.

In summary, the literature review demonstrates the use of data fusion as a valid technique to provide a more comprehensive source of data for information provision. Although the literature regarding transportation applications is sparse, data fusion provides a method to merge large sets of data from multiple sources. From a methodological perspective, this research

utilizes a data centric data fusion approach. However, if given more time, a model centric approach would allow the research team to study and minimize any uncertainties present in traffic datasets.

6.3 DATA FUSION PROCESS AND CASE STUDY

This project proposes a data fusion approach to illustrate the viability of merging TxDOT traffic data with private vendor data in a single database. More specifically, this fusion would involve combining TxDOT Reference Mile Marker ([Texas Department of Transportation, 2005](#)) data with INRIX TMC data. Fusing the data following a data centric approach utilized the ESRI ArcGIS program.

[Figure 23](#) shows the first phase of processing consists of identifying GIS shape files for INRIX TMC data and the El Paso Reference Mile Marker data. Once these files were identified, researchers loaded shape files for both data sources in ArcGIS. The coordinate system commonly known as State-Plane system is the default system and the locations of both files are in units of feet. An intermediate step consists of exporting the INRIX TMC data shape file as a text file and opening it in a spreadsheet (MS Excel[®]). The rationale for this step is to determine the centroids (mid-points) of the TMC segments for both traveling directions.

Once the analysts determined the centroids for the INRIX Traffic Message Channel segments, they imported them back into ArcGIS and set a 0.25-mile radius for the TMC segment centroids to link TMC segments with reference markers. [Tarko, et al. \(2009\)](#) used a similar value when analyzing roadway segments for improving safety in high-speed work zones. This step creates an imaginary buffer zone for associating El Paso Reference Mile Marker data to the INRIX TMC data. A Spatial Join (a tool readily available in ArcGIS) is then involved to merge the INRIX buffer Zone and El Paso Reference Mile Marker shape files. This stage of Spatial Join sets the target file to the INRIX TMC buffer shape file and the joined file is the El Paso Reference Mile Marker. Any El Paso Mile Markers falling within the 0.25 mile radius of a centroid gets attached to the INRIX buffer shape file which contains INRIX TMC IDs.

In this Spatial Join process, any El Paso Mile Markers falling outside the 0.25-mile buffer zone will not be attached to the INRIX buffer file. A secondary process associates those mile markers falling outside of 0.25-mile radius buffer with the INRIX buffer shape file. The procedure of incorporating these outlying mile markers consists of attaching the mile markers to the closest INRIX TMC centroid. This procedure is represented by the loop in [Figure 23](#).

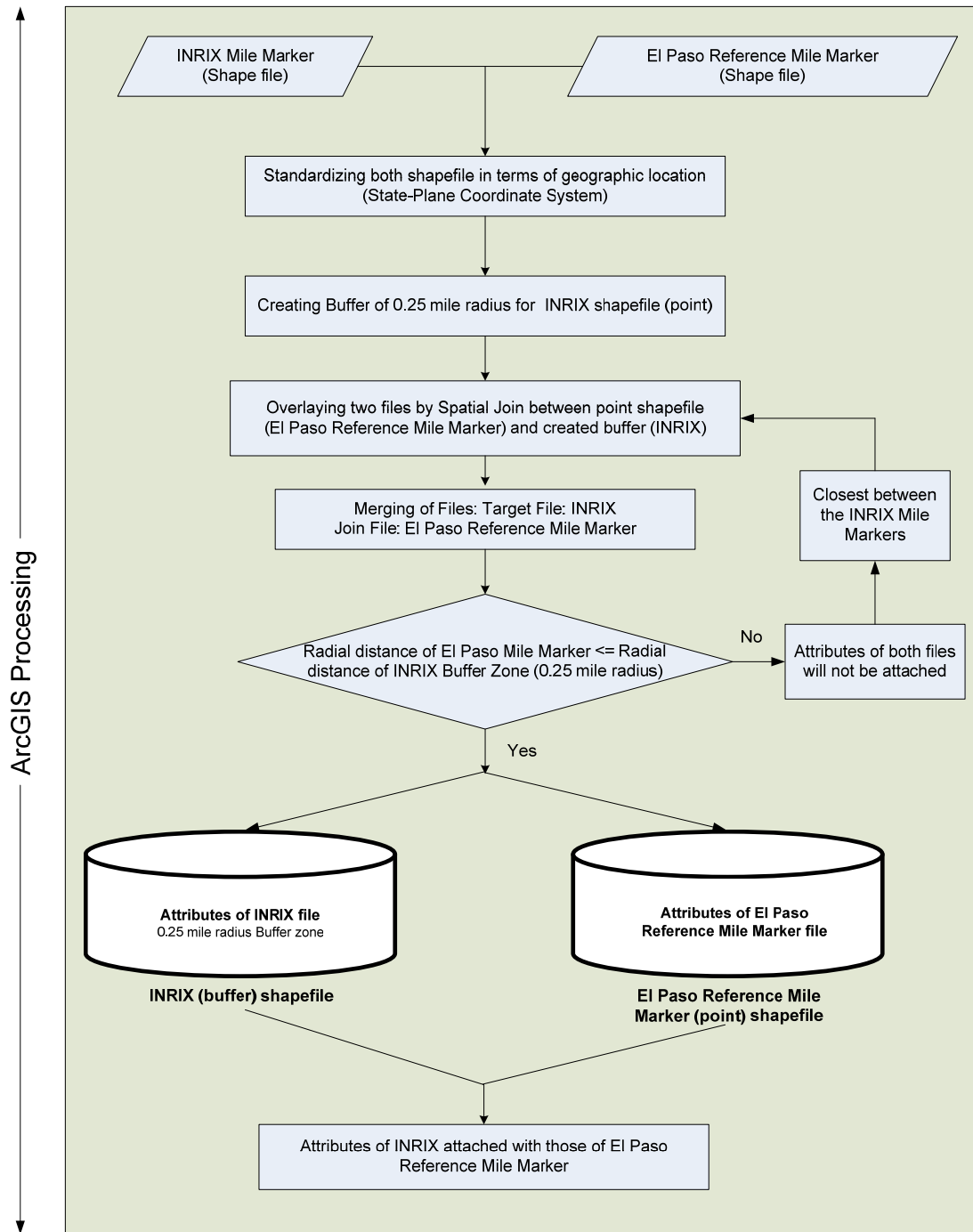


Figure 23. Conceptual Framework for Integrating INRIX TMC Data with TxDOT Reference Mile Marker Data (El Paso as Case Study).

Figure 24 shows a microscopic view of the fusion process for a highway segment in El Paso, Texas, for TMC ID numbers 115+01237, 115+01236, and 115+01235. The black line segments represent the INRIX TMCs while the blue and red circles represent the El Paso Reference Mile Markers. The vertical red lines represent the boundary of the buffer zone. The blue circles fall within the buffer zone and are associated with the specific TMC ID, in this case 115+01236. The red circles fall outside the buffer zone, so the data fusion framework will associate them with the nearest TMC centroid. This process is repeated for the entire study area (in this case, El Paso).

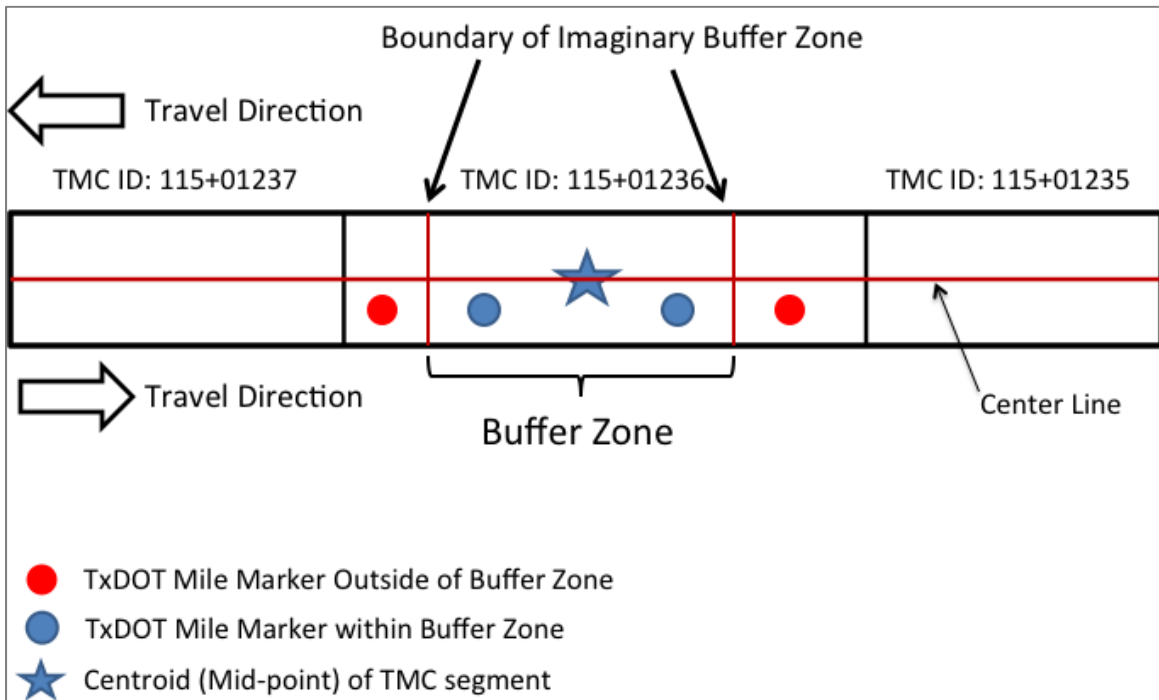


Figure 24. Data Fusion of INRIX and TxDOT Mile Marker—A Microscopic View.

For illustrative purposes, researchers constructed a small case study from the El Paso TxDOT Reference Mile Marker and INRIX TMC data sets. The case study involved data fusion for I-10 between Loop 375 and Zaragoza Road in El Paso (a total of 180 segments in both directions). Assuming a 0.25-mile buffer radius as noted earlier, the results of the data fusion indicated that 72 percent of the El Paso Reference Mile Markers were within the INRIX TMC buffer zones. The remaining 28 percent were associated with the nearest TMC centroid. Changing the buffer zone radius will either increase or decrease the percentage of El Paso Reference Mile markers within the buffer zone. The determination of what constitutes an appropriate buffer radius can be further evaluated as needed in an implementation phase of this project. Additionally, the data fusion took approximately 3 hours to perform; however in an implementation phase, the data fusion conceptual framework could be fully automated, increasing the time efficiency of the data fusion process.

6.4 SUMMARY

This section presented a data fusion conceptual framework to merge TxDOT Reference Mile Marker data with data from private vendors (e.g., INRIX Traffic Message Channel segments). It developed a case study to test the data fusion framework. The outcome of the fusion resulted in a more comprehensive dataset that can provide more information to TxDOT regarding various aspects of traffic. The data fusion framework presented in this section can be modified to incorporate various data sets (e.g., crash) for increased levels of information density. Future research is needed to:

- Test the viability and quality of other data sources available to TxDOT (e.g., crash data, weather data, etc.).
- Study the effect of buffer zone radii.
- Fully automate the data fusion process.

7. SUMMARY OF FINDINGS AND RECOMMENDATIONS

7.1 KEY FINDINGS

To gather information on providers and consumers of private sector data, the research team conducted a survey of providers and consumers as listed in Table 35. Table 36 summarizes the results of the consumer survey, and Table 37 summarizes the provider survey.

Table 35. Providers and Consumers Providing Input.

Private Data Providers	Consumers of Private Data
Air Sage ATRI INRIX NAVTEQ TomTom TrafficCast.com	Houston – Galveston Area Council Maricopa Association of Governments Michigan Department of Transportation San Francisco Bay Area 511 Program Texas Department of Transportation Wisconsin Department of Transportation

Table 36. Summary of Historical Data Consumer Survey Results.

	Wisconsin DOT	HGAC	Michigan DOT	Texas DOT ^d	Phoenix MPO (MAG)
Status	RFI	Purchased	Purchased	Purchased	Purchased
Service Purchased ^a	H	H	H H H		
Aggregation Level	Hourly day-of-week averages	15 min.	5 min.	Hourly day-of-week averages	Weekday
Data Purchased ^b	S/TT, PM	S/TT	S/TT S/TT,	PM	PM
Applications ^c	PM, TM	PM, TM, OD	PM	PM	PM
Coverage	All arterials	Houston region	MI Freeways	Statewide TMC network	Region
Timeframe	1–2 years	1 year	5 years	2009	1 year
Validation Criteria	Not yet established	Not yet established	Avail >99.5% Accuracy < ±10 mph	None	Not yet established
Validation Techniques	N/A	N/A	Probe, fixed point; re-ID	None	Probe, fixed point.
Pricing (in thousands)	\$80K (Est.)	\$77K	\$200K per year	\$28K	Negotiating
Licensing	Multiple Use	Multiple Use	Single Use	Single Use	Multiple Use

^a Service Purchased: H = Historical, RT = Real-time

^b Data Purchased: S/TT = Speed or Travel Time, PM = Performance Measures

^c Applications: PM -Performance or Congestion Monitoring, TM = Traffic Model Validation or Calibration,

OD = Origin-Destination Studies

^d See <http://apps.dot.state.tx.us/apps/rider56/list.htm> for actual study results.

Table 37. Summary of Historical Data Available by Provider.

Factor	Airsage	ATRI	INRIX	NAVTEQ	TomTom	TrafficCast
Data Available ^(a)	S, TT, I, Q, V	S, TT, Q	S, TT, I, Q, V	S, TT, I, Q, V (portion of network)	S, TT, I, Q	S, TT, I, Q
Services Available ^(b)	D, A, PM	D, A, PM	D, A	D, A	D, A, PM	A, PM
Data Source ^(c)	Cell phone, 911, traffic counts	GPS on commercial truck-only fleets	State installed sensors, commercial fleets, consumer GPS	State installed sensors, commercial fleets, consumer GPS.	Consumer GPS, Fleet GPS	State installed sensors, commercial fleets, consumer GPS, Bluetooth systems.
Aggregation Levels for Historical Usage	None; as captured	1 mile, 1 minute	15–60 min.	15 min.	1 hour	15 min.
Accuracy Checks Performed	Visual camera count, Probe vehicles.	Anomaly checking done, routines not disclosed.	Independently verified in large-scale testing.	Data checks prior to map matching. Comprehensive drive testing.	Data checks prior to map matching.	Simple-adjacent points compared, some clients doing accuracy checks.
Documented Quality Levels	None provided. Stated they meet Section 511 requirements.	None-burden is on receiver of data.	Accuracy above 95%. Availability above 99.9%.	None provided.	None provided. Stated they can meet Section 511 requirements.	None provided. Stated they can meet Section 511 requirements.
Pricing	Specific pricing information not provided.	Specific pricing information not provided. Not for profit.	Full use open licensing is \$800 per mile per year plus \$200 per mile one-time setup fee. 25% discount on other roads purchased in conjunction.	Specific pricing information not provided.	Specific pricing information not provided.	Specific pricing information not provided.

^a Data Available: S = Speed, TT = Travel Time, I = Incidents, Q = Quality, V = Volumes, GPS = GPS fleet

^b Services Available: D = Discrete Data (individual data points), A = Aggregate Data, PM = Performance Measures

^c National Coverage: Not listed in table. All providers indicated national coverage, except TrafficCast which is currently in urban areas.

^d Map Matching: Not listed in table. All providers except ATRI indicated a minimum use of TMC. ATRI uses mileposts. INRIX, NAVTEQ and TomTom also use proprietary segmentation smaller than TMC.

7.2 CRITICAL FACTORS FOR TXDOT

The key factors that appear to be most important to TxDOT in deciding whether to purchase private sector data are:

- Data accuracy and availability (includes consideration of the data source).
- Cost.
- Network coverage.
- Control of the data stream.

Section 1201 of SAFETEA-LU mandates that state DOTs and other operating agencies establish a monitoring program to be established on all Interstate routes by November 8, 2014, and on other significant roadways as identified by the States and local agencies by November 8, 2016. [Table 38](#) identifies the key requirements of the information delivery timeframes.

Table 38. Information Delivery Requirements of Section 1201.

Information Type	Metropolitan Area (Minutes)	Non-metropolitan Area (Minutes)	Availability (Percent)	Accuracy (Percent)
Implementation or removal of lane closure	10	20	90	85
Roadway- or lane-blocking traffic incident information	10	20	90	85
Roadway weather observation updates	20	20	90	85
Travel time along highway segments	10	N/A	90	85

7.2.1. Data Accuracy and Availability

Sources of the data are important in the resultant accuracy. GPS and fixed sensors are generally reliable sensors that provide the level of accuracy that TxDOT needs. [Table 39](#) summarizes the sources of data by provider. As a general rule on speed accuracy, most providers are claiming speed accuracy levels around 90 to 95 percent and availability levels above 95 percent.

7.2.2. Cost

In most cases, TxDOT will not know the exact cost of private sector data without entering into a negotiation phase. However, specific cost information from INRIX indicates a first-year cost of \$800/mi, plus a one-time setup fee of \$200/mi. SpeedInfo provides a self-contained Doppler radar system at a cost of \$110/bi-directional station. [Figure 25](#) shows a comparison for one year of data for a 20-mile segment and sensors at 3- to 5-mile spacings. The

first three sensors represent PS providers and the fourth is a side-fire radar fixed sensor. Of course, the radar offers per-lane data including speeds, counts, and vehicle lengths across at least 8 lanes. The cost of a verification mechanism is an additional cost that TxDOT must consider for PS data at the beginning and periodically.

Table 39. Provider Primary Data Sources.

Provider	GPS-enabled Vehicles	Cellular Probes	Fixed Point Sensors	Others
AirSage		x		
Cellint		x		
Delcan		x		
INRIX	x	x	x	
NAVTEQ	x	x	x	
OnStar	x			
SpeedInfo			x (radar)	
TomTom	x	x	x	
Total Traffic Network	x	x	x	Airborne/Mobile Spotters, Cameras
TrafficCast	x		x	Bluetooth

7.2.3. Network Coverage

The coverage of PS data in Texas is the entire Traffic Message Channel network. It includes all interstate routes both urban and rural and many non-interstate routes. Coverage is a function of the number of probe vehicles in the traffic stream. For example, in urban areas, many arterial streets would have sufficient fleet vehicles (e.g., trucks) to provide adequate coverage.

7.2.4. Control of the Data Stream

With PS data, TxDOT has little control of the data stream. It is advisable that TxDOT establish minimum thresholds of acceptability in both accuracy and availability. If the provider does not meet the established limits, TxDOT costs would decrease by some agreed-upon amount. Of course, checking the accuracy and availability requires an ongoing verification methodology.

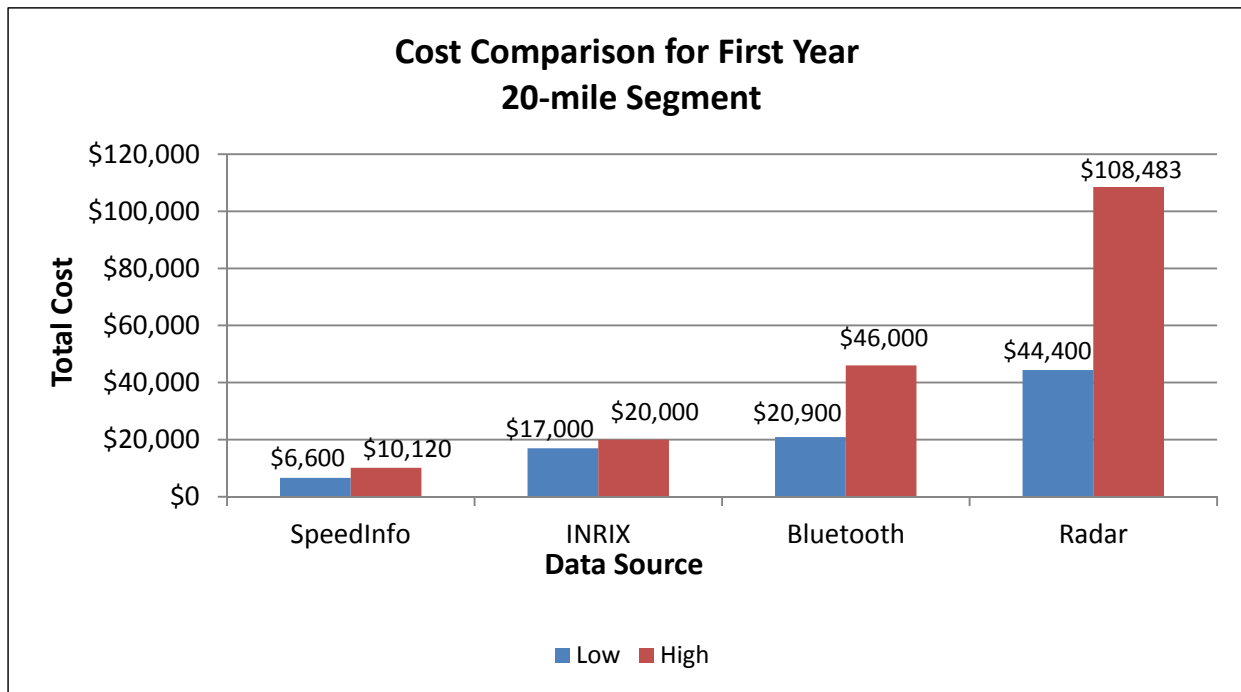


Figure 25. Cost Comparison for PS versus Radar Fixed Sensor.

7.2.5. Summary Comparison

Table 40 summarizes the major MOEs for comparing PS data with other alternatives. SF is side-fire radar and Mag is magnetometers.

Table 40. Comparison of Performance of Various Data Sources.

Performance	PS Data	BT	Loops	Video	SF Radar	Mag
Speed Accuracy(%)	±5-10	±5-10	±5-10	±5-20	±5-10	±2-10
Count Accuracy(%)	N/A (w/o TxDOT sensors)	N/A	±2-5	±5-20	±2-5	±2-5
Data Source	GPS: High Doppler Radar: High BT: High	High	High	Medium	High	High
TxDOT Control	Low	Low	High	High	High	High
Uses of data						
Speed/TT	Yes	Yes	Yes	Yes	Yes	Yes
Counts	No	No	Yes	Yes	Yes	Yes
Occupancy	No	No	Yes	Yes	Yes	Yes
Coverage	The TMC Network	As determined by TxDOT				

7.2.6. Provide Feedback to TxDOT

The research team conducted a webinar to provide information on private sector data providers and consumers, then ask for their feedback by having them fill out a survey form. Feedback from 20 TxDOT engineers and planners (mostly districts) indicated the following:

- TxDOT responders on average ranked accuracy and cost effectiveness higher than availability and quick turnaround.
- For enhancement of traveler information, speed/travel time measurement ranked slightly higher on average than alternate route information or levels of congestion.
- On average, creating uniform coverage rated higher than cost effectiveness and reduction of TxDOT's reliance on fixed sensors.
- Assuming data purchased from PS providers, all 20 responders said TxDOT forces would continue to provide count data since PS providers do not typically do this.
- Per-lane data were not critical to 10 responders, but it was to 7.
- On average, TxDOT responders said that if fixed sensors reach a 60 percent failure rate they would purchase real-time data from the PS.
- If responders purchased PS historical data, they would use it for origin-destination studies and for model calibration.
- Use of the Traffic Message Channel was not a deterrent to PS data for 7 responders, but it was to 4.
- Responders suggested the following examples of long-term opportunities for PS data:
 - Tolling.
 - Operational validation.
 - Hurricane evacuation.
 - Other evacuations (non-hurricane).
 - Flooding.
 - International POEs.
 - Border violence (causing traffic anomalies).
 - Work zones (2).
 - O-D freight (re: rail).
 - Real-time system management.
 - USDOT mandate for real-time monitoring systems (Sec 1201).
 - Incident avoidance.
 - Special events (2).
 - Travel time comparison I-35/SH 130.

7.2.7. Opportunity Matrix

Other than Section 1201, the PS data could also meet the following needs, some of which could be longer term in nature:

- Enhance traveler information in urban areas such as:
 - Travel time information.
 - Levels of congestion.
 - Speed measurement.
 - Alternate routes.
- Introduce traveler information in areas where ITS deployment is not cost-effective.
- Improve continuity of data based on existing ITS coverage across jurisdictions.

- Develop statewide 511 system.
- Reduce ITS deployment costs by limiting deployment of fixed data collection devices.

Based on this list, the research team developed an opportunity matrix, with implementation timing to depend on how quickly TxDOT initiates the process. Table 41 shows a subjective ranking of various governing factors that should guide the initial steps of decision making. Researchers recommend a one-year implementation of a trial PS network to include workshops with key TxDOT personnel around the state to more thoroughly cover the implications of PS data.

Table 41. Relative Importance of Various Governing Factors.

Application Group	Application Area	Spatial Coverage	Cost Effectiveness	Information Accuracy	Data Reliability	Control of Data Stream	Quick Procurement
Traveler information	Enhance coverage of traveler info. in urban areas	1	2	2	1	2	3
	Enhance traveler info. in rural areas	3	2	1	1	3	3
	Statewide 511 system	3	3	2	1	2	2
	Emergency evacuation	3	1	3	3	3	2
	Work zone information	1	2	3	3	2	3
System planning	Performance measurement	3	2	3	2	1	2
	Model and calibration	3	1	3	2	1	2
System operation	Faster identification of congested areas	3	1	3	3	3	1
	Predictive information	1	1	3	3	3	1

Note: 1: Less concerned with, 2: Neutral, 3: More concerned with.

7.3 RECOMMENDATIONS/IMPLEMENTATION STRATEGY

The initial steps to implement the findings of this research should provide TxDOT with further guidance on meeting the SAFETEA-LU Section 1201 requirements. During the initial portion of that period, say one year, researchers recommend that TxDOT select two or more providers of PS data and select a trial network that already has a means of verification or could easily be modified to serve that purpose. This might be a corridor with sufficient fixed sensors, probe vehicles, toll readers, or Bluetooth devices. The authors believe that this effort could be conducted as an implementation project since it could be initiated immediately and provide the timely results TxDOT needs to continue planning for meeting the Section 1201 requirements.

If TxDOT considers the results of this proposed evaluation as acceptable, the research team recommends moving forward with a more significant purchase of private sector data to fill gaps in the TxDOT network.

Recommended key tasks in the pilot project include the following:

- Assessment of needs and requirements of districts and identify the role of private sector data to meet those needs—tie with regional ITS architecture and ITS strategic plans of the districts.
- Conduct a statewide workshop and provide vendors with opportunities to demonstrate their capabilities.
- Identify case study sites and/or corridors for the pilot test. The sites should include regions/corridors with varying degrees of ITS deployment and traffic conditions—rural areas with limited ITS deployment, urban areas with increased ITS deployment, specific urban and rural corridors as well as different application areas—travel time measurements, incident detection, emergency evacuation, and so forth.
- At the case study sites, procure real-time as well as archived data from more than one private sector agencies.
- At the case study sites, implement one or more ITS applications (e.g., displaying travel).
- Perform detailed evaluation of procurement issues, quality, accuracy, and reliability issues pertaining to application of private sector data at case study sites to implement specific ITS application areas.
- Perform detailed evaluation of life cycle costs (deployment, installation, license costs, evaluation, maintenance, etc.) to use private sector data at case study sites to implement specific ITS application areas.
- Develop a guidebook for districts and TxDOT partner agencies to perform pre-procurement planning, procurement, and deployment of private sector data.

8. REFERENCES

- Bogenberger, K. 2003.** The Quality of Traffic Information. [Online] 2003. [Cited: April 11, 2011.]
http://www.trafficforum.ethz.ch/vwt_2003/beitraege/VWT19proceedings_contribution_28.1-28.17.pdf.
- Clear Channel Communications. 2007.** Kenwood KNA G510 In-Car Navigation System Will Offer Real-Time Traffic Data from Clear Channel's Total Traffic Network. *Clear Channel Radio Press Release*. [Online] 2007. [Cited: 12 2011, April.]
<http://www.clearchannel.com/Radio/PressRelease.aspx?PressReleaseID=1486>.
- . **2011.** Total Traffic Network. [Online] 2011. [Cited: April 12, 2011.]
<http://totaltraffic.com/Partners/Inrix.aspx>.
- Dailey, D. J., Harn, P. and Lin, P. 1996.** *ITS Data Fusion, Final Research Report, Research Project T9903, Task 9*. 1996.
- Dempster, A. P. 1967.** Upper and Lower probabilities induced by Multivalued Mapping. *Annals of Mathematical Statistics*. 1967, Vol. 38, pp. 325–339.
- Dubois, D. and Prade, H. 1988.** *Possibility Theory*. New York : Plenum Press, 1988.
- El Faouzi, N-E. 2004.** *Data Fusion in Road Traffic Engineering: An Overview*. s.l. : Proceedings of SPIE, Vol. 5434, 2004.
- Hallenbeck, Mark E. and Yao, Wu. 2007.** *Review of Private Sector Data for Roadway Monitoring*. s.l. : Washington State Transportation Center, University of Washington, 2007.
- I-95 Corridor Coalition. 2010.** *Validation of INRIX Data: Two-Year Summary Report*. 2010.
- International Organization for Standardization. 2006.** Traffic and Traveler Information (TTI) - TTI Messages via Traffic Message Coding - Part 6: Encryption and Conditional Access for the Radio Data System - Traffic Message Channel ALERT C Coding. *International Standard ISO/DIS 14819-6:2006*. [Online] 2006. [Cited: April 12, 2011.]
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=37267.
- Matschke, I., Friedrich, B. and Heinig, K. 2004.** *Data Fusion Technique in the Context of Traffic State Estimation*. s.l. : Proceedings of Triennial Symposium on Transportation Analysis, 2004.
- Microsoft Corporation. 2008.** A Comparison of RDS and MSN Direct. [Online] 2008. [Cited: April 12, 2011.]
http://www.msndirect.com/RDS_MSNDirect_Comparison.pdf.
- Okutani, I. 1987.** *The Kalman Filtering Approaches in some Transportation and Road Traffic Problems*. s.l. : Proceedings of the 10th ISTTT, Gartner and Wilson Eds, 1987.
- Shafer, G. 1976.** *Mathematical Theory of Evidence*. Princeton : Princeton University Press, 1976.
- Shiode, N., et al. 2004.** *The Impact of Penetration of Location Based Services*. 2004.
- Smalley, D. G. and McCasland, W. R. 1993.** *Public/Private Sector Partnership to Provide Traffic Information by Cellular Telephone Reports, Research Report 1941-1*. College Station, Texas : Texas Transportation Institute, 1993.

Smith, B. L. and Fontaine, M. D. 2005. *Private-Sector Provision of Congestion Data. Final Report NCHRP 70-01.* s.l. : University of Virginia – Center for Transportation Studies, 2005.

Smith, B. L., et al. 2005. *Private Sector Provision of Congestion Data.* s.l. : University of Virginia/Old Dominion University, 2005.

Tarko, A.P., Islam, M.B. and Thomaz, J. 2009. *Improving Safety In High Speed Work Zones: A Super 70 Study, Joint Transportation Research Program, Project No. C-36-59FFF.* s.l. : Federal Highway Administration, 2009.

Texas Department of Transportation. 2005. *Texas Reference Marker (TRM) System User's Manual.* Austin, Texas : s.n., 2005.

Traveler Information Services Association (TISA). 2011. *The Traffic Message Channel.* [Online] 2011. [Cited: April 12, 2011.] <http://www.tmcforum.com/technologies/tmc/>.

U.S. Department of Transportation ITS Joint Program Office. 2003. *Data Fusion for Delivering Advanced Traveler Information Services.* Washington, DC : s.n., 2003.

Young, S. 2007. *Real-Time Traffic Operations Data Using Vehicle Probe Technology.* Ames, Iowa : s.n., 2007.

APPENDIX A. DATA PROVIDER SURVEY FORM

Data Provider Survey

Technical Support and Assistance for the FHWA's Office of Transportation Operations

Under the "Private Sector Data for Performance Management" contract



Private Sector Data Marketplace Review - Providers

To develop a comprehensive understanding of transportation mobility performance management the Federal Highway Administration (FHWA) is sponsoring a research project to determine both the capability and advantages/disadvantages of using private sector data for national mobility performance management.

This survey consists of a series of questions categorized into five sections: (1) CONTACT INFORMATION, (2) SCOPE OF SERVICES, (3) TECHNICAL ASPECTS OF PROVIDING PRIVATE SECTOR DATA, (4) PRICING OF PRIVATE SECTOR DATA, and (5) CLIENT INFORMATION. Please respond to these questions regarding the provision of private sector data to the best of your ability. If you feel that someone else within or outside your company is better suited to complete this survey, please forward the survey and associated email participation request to them or provide their contact information so that we may contact them directly.

This study is anonymous and no participant names will be included anywhere in the research deliverables. The records of this study will be kept private. No identifiers linking you to this study will be included in any sort of report that might be published. Research records will be stored securely and only research staff will have access to the records.

Surveys will be conducted via telephone at a time of mutual convenience. This survey questionnaire will be provided to the respondents in advance in order to help them prepare for the questions. If desired, a written completed survey can be returned via fax (979-845-9873), email (r-brydia@tamu.edu) or mail (Texas Transportation Institute, Texas A&M University System, 3135 TAMU, College Station, TX 77843, attention Mr. Robert E. Brydia).

If you have specific questions as you complete the survey, please contact Mr. Brydia at r-brydia@tamu.edu or (979) 845-8140. Thank you in advance for your participation.

PART 1. CONTACT INFORMATION

1. As an initial step, please provide your contact information below to facilitate any follow up that may be required.
 - a. Name
 - b. Company
 - c. Phone Number
 - d. Email Address

PART 2. SCOPE OF SERVICES

2. What types of services do you offer?
 - a. Raw data for purchase? No Historical Real-Time
 - b. Refined / aggregate data for purchase? No Historical Real-Time
 - c. Data warehousing? No Historical Real-Time
 - d. On-demand data access? No Historical Real-Time
 - e. Performance measures? No Historical Real-Time
3. What types of data do you offer?
 - a. Speeds? No Historical Real-Time
 - b. Travel times? No Historical Real-Time
 - c. Incident / Event data? No Historical Real-Time
 - d. Metadata? No Historical Real-Time
 - e. Quality data? (sample sizes, quality indicators, confidence levels, blending ratios, standard deviations, etc.) No Historical Real-Time
 - f. GPS fleet data? No Historical Real-Time
 - g. Other? No Historical Real-Time
 - h. Volumes? No Historical Real-Time
 - i. Arterial data? No Historical Real-Time
 - j. Per lane availability? No Historical Real-Time
4. What data management services do you offer?
 - a. Metadata? No Historical Real-Time
 - b. Data sorting by _____ (region / state / roadway / date / time / source) No Historical Real-Time
 - c. Data checking (e.g., range checks)? No Historical Real-Time
 - d. Custom filtering / queries? No Historical Real-Time
 - e. Other No Historical Real-Time
5. How do you sell your data?
 - a. Direct?
 - b. Distributor / Reseller?
 - c. Other?
6. What types of licensing do you use for your data?
 - a. Single-use?
 - b. Open
 - c. Depends on type? (historical vs. real-time)

PART 3. TECHNICAL ASPECTS OF PROVIDING PRIVATE SECTOR DATA

7. Can you describe your overall data collection process?
 - a. Generation / Sources of Data
 - i. Timeframes of Data Collection?
 - ii. Market Penetration?
 - iii. Number of Probe Vehicles?
 - b. Data Aggregation?
 - c. Data Fusion?
 - d. Data Dissemination?
8. What is your coverage? If not national, extent by:
 - a. States?
 - b. Regions?
 - c. Roadway Corridors?
9. What are your typical data aggregation routines?
 - a. What temporal levels do you aggregate to?
 - b. What spatial levels do you aggregate to?
 - c. Are custom levels available?
10. What data fusion techniques do you use to get the final product?
 - a. Other sources of data?
 - b. Matching techniques?
11. What is the accuracy and availability information you provide with your data?
 - a. How do you verify accuracy?
 - b. Does the accuracy or availability ever drop too low for the data to be useful?
 - i. In what typical type of situations would this occur?
 - ii. How often would this occur?
12. Do you have procedures to compensate for the drop in availability and/or accuracy?
13. How do you / will you, provide proof of compliance with Section 1201 requirements?
14. How do you disseminate the information to users?
 - a. Data feeds (XML, etc.)
 - b. Map display over the Internet
 - c. Smartphone displays
 - d. In-vehicle displays
 - e. Other _____
15. How do you blend historical and real-time data?
16. What is the typical latency in reporting real-time data?

PART 4. PRICING OF PRIVATE SECTOR DATA

17. Can you provide typical pricing information for acquiring your real-time data?
 - a. Are costs primarily based on
 - i. Days of data
 - ii. Coverage?
 - iii. Aggregation level required?
 - iv. Usage of the data?
 - v. Other
18. Can you provide typical pricing information for acquiring your historical data?

- b. Are costs primarily based on
 - i. Days of data
 - ii. Coverage?
 - iii. Aggregation level required?
 - iv. Usage of the data?
 - v. Other?

PART 5. CLIENT INFORMATION

- 19. What categories do your clients fall into? (Percentage to make up 100)
 - a. State
 - b. Local region / cities
 - c. MPOs
 - d. Private companies
 - e. Other
- 20. Could we obtain a client list of agencies that might be willing to talk with us relating to their experiences in the use of private sector data?

APPENDIX B. DATA CONSUMER SURVEY FORM

Data Consumer Survey

Technical Support and Assistance for the FHWA's Office of Transportation Operations

Under the "Private Sector Data for Performance Management" contract



Private Sector Data Marketplace Review - Consumers

To develop a comprehensive understanding of transportation mobility performance management the Federal Highway Administration (FHWA) is sponsoring a research project to determine both the capability and advantages/disadvantages of using private sector data for national mobility performance management.

This survey consists of a series of questions categorized into five sections: (1) CONTACT INFORMATION, (2) SCOPE OF SERVICES, (3) TECHNICAL ASPECTS OF UTILIZING PRIVATE SECTOR DATA, (4) CLIENT USES OF PRIVATE SECTOR DATA, and (5) PROCUREMENT AND PRICING OF PRIVATE SECTOR DATA. Please respond to these questions regarding private sector data usage in your agency to the best of your ability. If you feel that someone else within or outside your agency is better suited to complete this survey, please forward the survey and associated participation request to them or provide their contact information so that we may contact them directly.

This study is anonymous and no participant names will be included anywhere in the research deliverables. The records of this study will be kept private. No identifiers linking you to this study will be included in any sort of report that might be published. Research records will be stored securely and only research staff will have access to the records.

Surveys will be conducted via telephone at a time of mutual convenience. This survey questionnaire will be provided to the respondents in advance in order to help them prepare for the questions. If desired, a written completed survey can be returned via fax (979-845-9873), email (r-brydia@tamu.edu) or mail (Texas Transportation Institute, Texas A&M University System, 3135 TAMU, College Station, TX 77843, attention Mr. Robert E. Brydia).

If you have specific questions as you complete the survey, please contact Mr. Brydia at r-brydia@tamu.edu or (979) 845-8140. Thank you in advance for your participation.

PART 1. CONTACT INFORMATION

1. As an initial step, please provide your contact information below to facilitate any follow up that may be required.
 - a. Name
 - b. Company
 - c. Phone Number
 - d. Email Address

PART 2. SCOPE OF SERVICES

2. What types of services did you investigate?
 - a. Raw data for purchase?
 - b. Refined / aggregate data for purchase?
 - c. Data warehousing?
 - d. On-demand data access?
 - e. Performance measures?
3. What types of data did you wind up purchasing?
 - a. Speeds?
 - b. Travel times?
 - c. Congestion indicators?
 - d. Incidents / Event data?
 - e. Quality indicators?
 - i. Please describe
 - f. Metadata?
 - g. Performance measures
 - h. Fleet data?
 - i. Other?
4. Did you purchase data for:
 - a. State?
 - b. Region?
 - c. Corridor?
 - d. Other?
5. What are the temporal characteristics of the data you purchased?
 - a. Number of days?
6. What were the driving factors in deciding to purchase commercial traffic data?
 - a. Cost effectiveness?
 - b. Quicker turnaround?
 - c. Accuracy and Availability?
 - d. Others _____

PART 3. TECHNICAL ASPECTS OF UTILIZING PRIVATE SECTOR DATA

7. What is the update period (refresh rate) for the real-time data your agency recently bought?
 - a. ___ Minutes

8. Who defined the refresh rate for the real-time data?
 - a. The commercial provider had already pre-set a refresh rate, which could not be changed.
 - b. The agency defined the refresh rate and the vendor was able to provide it.
9. What was the latency of the recently purchased speed data for real-time operation?
 - a. _____ Minutes
10. Did you define minimum and maximum latency prior to the purchase of the real-time speed data?
 - a. Minimum = _____ Minutes (e.g., 5 minutes)
 - b. Maximum = _____ Minutes (e.g., 10 minutes)
11. What techniques have you used to validate the data?
 - a. Probe vehicle method
 - b. Fixed point data collection
 - c. Re-identification studies (Bluetooth, License Plate, etc.)
12. Describe the MOEs used to validate the real-time and archived data. *(For example, speed data shall have a maximum average absolute error of 10 mph for all speed ranges.)*
13. Do you aggregate the data beyond what you purchased from the private sector provider?
 - a. What temporal levels do you aggregate to?
 - b. What spatial levels do you aggregate to?
14. What data fusion techniques do you use to get the final product?
 - a. Other sources of data?
 - b. Matching techniques?
15. What is the accuracy and availability information you required for your data?
 - a. How do you verify accuracy?
 - b. Does the accuracy or availability ever drop too low for the data to be useful?
 - i. In what typical type of situations would this occur?
 - ii. How often would this occur?
16. Do you have any procedures to compensate for the drop in availability and/or accuracy?
17. How do you get the data from the private sector provider?
 - a. Data feeds (XML, etc.)
 - b. Map display over the Internet
 - c. Aggregated numerical reports, such as spreadsheets
 - d. Other _____

PART 4. CLIENT USES OF PRIVATE SECTOR DATA

18. How are you using the private sector data?
 - a. Traffic operations
 - i. Motorist information:
 1. Map based displays of Link travel times / average speeds?
 2. Trip planning
 3. 511 services
 4. Highway Advisory Radio

5. Dynamic Message Signs

- ii. Work zone information
 - iii. Incident management?
 - iv. Performance measurement
 - v. Other (please specify) _____
- b. Transportation planning
- i. General traffic O&D studies
 - ii. Freight O&D studies
 - iii. Performance measurement
 - iv. Other (specify) _____ ?
- c. Other (specify) _____ ?
19. From your perspective, what are the advantages of purchasing private sector data?
20. From your perspective, what are the disadvantages of purchasing private sector data?
21. How do you compensate for the lack of traffic volumes in private sector data,
- a. Our agency continues to collect volume counts
 - b. We contract it out
 - c. Both

PART 5. PROCUMENT AND PRICING OF PRIVATE SECTOR DATA

22. Can you provide us with cost information for the recently purchased data?
- a. Real-time data
 - i. \$ ___ per mile per year
 - ii. _____ Total miles purchased
 - iii. For ___ number of years
 - b. Archived data
 - i. \$ ___ per mile per year
 - ii. _____ Total miles purchased
 - iii. For ___ number of years
 - c. Based on _____ days of data
23. Do you use a payment plan tied to availability and accuracy of the data?
24. Do you think your agency will continue purchasing commercial data in the future?
25. Has your agency allocated budget for regular purchase of commercial traffic data?
26. Has your agency released guidelines for procurement of commercial traffic data?
- a. If yes, can you provide us copy?
 - b. If no, are there plans to prepare one?
27. Can you provide us a copy of an RFP recently issued by your State DOT to purchase commercial data?
28. Are you procuring private sector data in coordination or combination with other agencies in your area?

APPENDIX C. TXDOT SURVEY FORM



Survey Instrument
Research Project 0-6659
Synthesis of TxDOT Uses of Real-Time Commercial Traffic Routing Data

Note: this email is being distributed to a list of knowledgeable Texas Department of Transportation (TxDOT) representatives. These representatives were determined based on guidance from the TxDOT Project Director.

Instructions to TxDOT District Representative:

The Texas Transportation Institute is conducting a study for TxDOT to determine the applicability of private sector data for TxDOT use. You were selected to participate in this study because you are knowledgeable about TxDOT's current data collection and use. Through this study, we are hoping to determine whether TxDOT should utilize the data offerings by private sector providers to supplement its own data collection efforts and, if so, how.

As part of this study, the research team would like for you to participate in a web seminar on Monday, July 18 beginning at 2:00 p.m. Your participation in this webinar is completely voluntary and if you agree to participate, you will be asked to respond to a series of questions on how private sector data might apply to TxDOT practice. It will take approximately 30 minutes to complete the webinar; you are being provided information about private sector data in this mail-out, although some of this will be repeated during the webinar. Please read ahead of time if possible. The purpose of the webinar is to provide information to participants and to clarify any questions you might have. Therefore, you are encouraged to read the materials in advance to ensure understanding. Complete the questions at a convenient time after the webinar.

All responses to questions will be anonymous and no participant's name or contact information will be included in any report. You may elect to not answer any question and you may terminate the interview at any time. You should not expect any direct benefit or compensation for participating in the survey, and the risks associated with your completing the questions are minimal—no greater than risks ordinarily encountered in daily life. Your responses to the survey will be stored securely and only the TTI research team will have access to the responses. If you have any questions regarding this study, you may contact Dan Middleton at 979-845-7196 or d-middleton@tamu.edu.

This research study has been reviewed by the Human Subjects' Protection Program and/or the Institutional Review Board at Texas A&M University. For research-related

problems or questions regarding your rights as a participant in this study, please feel free to contact their offices at 979-458-4067.

You should have received an initial mail-out for this survey to schedule the date for the webinar. We hope to include a significant number of the 25 TxDOT districts to discuss this project. You are being provided some of the research findings in this mail-out and a researcher will summarize these findings at the beginning of the webinar. Some of the questions we plan to ask pertain to this material so reading the material in advance is encouraged.

Thank you in advance for your cooperation and support of this research project.

PROJECT SURVEY FINDINGS (See attachment A – please read in advance if possible)
PARTICIPANT INFORMATION AND SURVEY QUESTIONS

Name: (only if you wish to be contacted): _____

TxDOT District: _____

Position: (optional): _____

Telephone: (optional): _____

Email: (optional): _____

(Introduction by research staff to include brief overview of the survey information of private data providers and private data users.)

Private sector data are available from a variety of sources, primarily generated by probe vehicles, collected anonymously in most cases to be delivered by third party companies to DOTs as either real-time data or historical data. The data consist primarily of speed and travel time, which can be used to derive incident information. Real-time data usually involve latencies of a few seconds to a few minutes, and can be delivered in various aggregation levels down to as short as 1 minute. Historical data are archived and made available to DOTs in a variety of formats depending on DOT needs. Much of the historical data is aggregated in 15-minute or 1-hour intervals.

This research also needs to investigate the timing of future TxDOT decisions. In this discussion, near-term is the next 5 years and long-term is more than 5 years. Near-term opportunities will include application of private data to enhance existing service and information provided by TxDOT based on the current coverage and data quality of private data and the need to expand the services. Long-term opportunities will include applications that have potential to be implemented by TxDOT because of availability of private data.

Questions for Interviews

1. About what percent of your district traffic data (historical and real-time) is collected in-house (by TxDOT personnel)?
 - a. Historical: ___%
 - b. Real-time: ___%
 - c. Comments: _____

2. Previous discussions with consumers of private sector data indicated that they purchased private data primarily for the following reasons. What value from 1 to 10 would you place on each? (1 is lowest, 10 is highest)?
 - a. Accuracy _____
 - b. Availability _____
 - c. Cost-effectiveness _____
 - d. Faster turnaround _____

3. Private sector data could potentially enhance traveler information in urban/rural areas. Which of the following areas would be of the greatest benefit in your district? (Rank: 1 is lowest, 10 is highest)?
 - a. Travel time information _____
 - b. Levels of congestion _____
 - c. Speed measurement _____
 - d. Alternate route information _____
 - e. Comments: _____

4. Please rank the following statements in terms of importance to your district:
 - a. Private sector data could be used to introduce traveler information in areas where ITS deployment is not cost-effective. Rank (1 low to 10 high): _____
 - b. Private sector data could be used to create uniform coverage across jurisdictions (e.g., expand to arterials). Rank (1 low to 10 high): _____
 - c. Private sector data could reduce TxDOT's reliance on fixed sensors maintained and paid for by TxDOT. Rank (1 low to 10 high): _____
 - d. Comments: _____

5. Traffic volume data are generally not available from private data providers, so how would your district acquire this data?
 - a. Continue to collect by TxDOT forces _____
 - b. Hire others to collect the data _____
 - c. Other (please specify) _____

6. Most of the providers of private sector data do not currently offer per-lane data although one or more say they plan on providing it soon. Do you believe this is a limitation for your district?
 - a. y, n
 - b. Please explain: _____

7. Under what conditions would you consider purchasing real-time private sector data for your district? (enter information for all that apply)

Criteria for Selecting Private Data	Within next 5 yrs	>5yrs
% failures in fixed TxDOT sensors reaches this value	%	%
Private sector coverage of the TxDOT road network more than this value	_%	_%
Speed/travel time accuracy of private data within this value	%	%
Cost of private data about the same as point sensors	y/n	y/n
Count data become available from private providers	y/n	y/n

Comments: _____

8. Under what conditions would you consider purchasing historical private data for your district? (enter all that apply)

Criteria for Selecting Private Data	Within next 5 yrs	>5yrs
% failures in fixed TxDOT sensors reaches this value	%	%
Private sector coverage of the TxDOT road network more than this value	_%	_%
Speed/travel time accuracy of private data within this value	%	%
Cost of private data about the same as point sensors	y/n	y/n
Count data become available from private providers	y/n	y/n

Comments: _____

9. What would you use historical private data for if purchased?
- Origin-destination studies ___
 - Calibrate simulation models ___
 - Other (please specify) ___

10. The Traffic Message Channel (TMC) mapping system does not have the same resolution as the TxDOT network (see Figure 1). Except in urban areas, the TMC system offers less coverage. Do you believe that the use of TMC will be an impediment to using private provider data?

- Yes ___
- No ___
- Explain ___

11. Can you think of other examples of long-term opportunities for the use of private sector data (e.g., hurricane evacuations)?

- Idea 1: _____
- Idea 2: _____
- Idea 3: _____
- Comments: _____

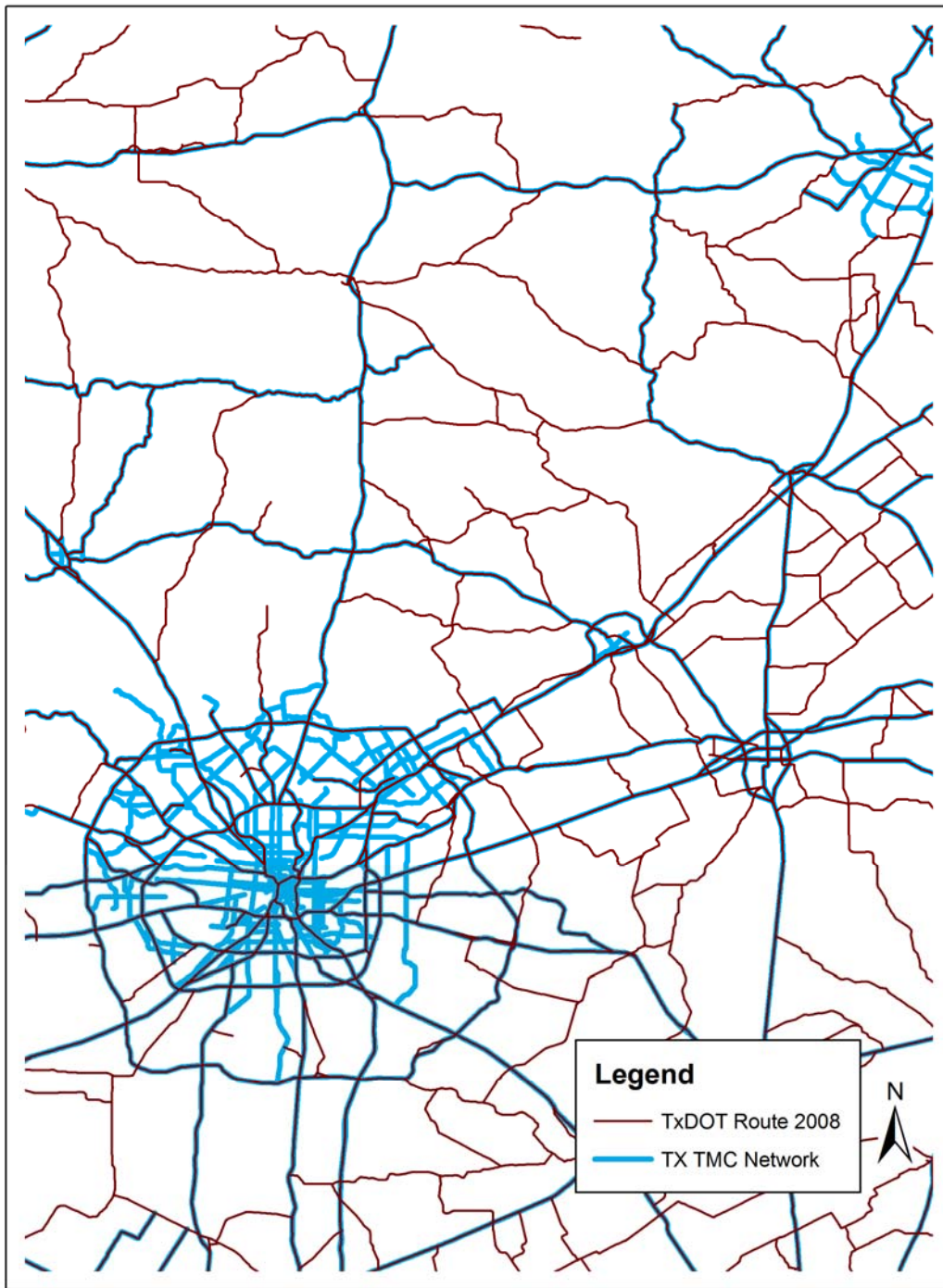


Figure C-1. Comparison of TMC Network with TxDOT 2008 Routes near San Antonio.

