Technical Report Documentation Page

1. Report No. FHWA/TX-14/0-6697-CTR-1	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Integrating Public and Private Data Sources for Freight Transportation Planning		5. Report Date December 2013; Published November 2014
		6. Performing Organization Code
7. Author(s) Dan Seedah, Alejandra Cruz-Ross, Bharathwaj Sankaran, Peter La Fountain, Prateek Agarwal, Haegon Kim, Meredith Cebelak, Sarah Overmyer, Jolanda Prozzi, Dr. William J. O'Brien, Dr. C. Michael Walton		8. Performing Organization Report No. 0-6697-CTR-1
9. Performing Organization Name and Address Center for Transportation Research The University of Texas at Austin 1616 Guadalupe Street, Suite 4.202 Austin, TX 78701		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. 0-6697-CTR
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Implementation Office P.O. Box 5080 Austin, TX 78763-5080		13. Type of Report and Period Covered Technical Report December 2011 to December 2013
		14. Sponsoring Agency Code

15. Supplementary Notes

Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.

16. Abstract

The Moving Ahead for Progress in the 21st Century Act (MAP-21) stipulates that state transportation agencies expand their interest in freight initiatives and modeling to support planning efforts, particularly the evaluation of current and future freight transportation capacity necessary to ensure freight mobility. However, the understanding of freight demand and the evaluation of current and future freight transportation capacity are not only determined by robust models, but are critically contingent on the availability of accurate data. Effective partnerships are clearly needed between the public and private sectors to ensure adequate freight planning and funding of transportation infrastructure at the state and local levels. However, establishing partnerships with firms who are both busy and suspicious of data-sharing, remains a challenge. This study was commissioned by the Texas Department of Transportation (TxDOT) to explore the feasibility of TxDOT entering into a data-sharing partnership with representatives of the private sector to obtain sample data for use in formulating a strategy for integrating public and private sector data sources. This report summarizes the findings, lessons learned, and recommendations formed from the outreach effort, and provides a prototype freight data architecture that will facilitate the storage, exchange, and integration of freight data through a data-sharing partnerships.

17. Key Words	18. Distribution Statement
Freight data, data integration, private sector data	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161; www.ntis.gov.

19. Security Classif. (of report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	160	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

THE UNIVERSITY OF TEXAS AT AUSTIN
CENTER FOR TRANSPORTATION RESEARCH

Integrating Public and Private Data Sources for Freight Transportation Planning

Dan Seedah
Alejandra Cruz-Ross
Bharathwaj Sankaran
Peter La Fountain
Prateek Agarwal
Haegon Kim
Meredith Cebelak
Sarah Overmyer
Jolanda Prozzi
Dr. William J. O'Brien
Dr. C. Michael Walton

CTR Technical Report:	0-6697-CTR-1
Report Date:	December 2013
Project:	0-6697-CTR
Project Title:	Integration of Data Sources to Optimize Freight Transportation in Texas
Sponsoring Agency:	Texas Department of Transportation
Performing Agency:	Center for Transportation Research at The University of Texas at Austin

Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.

Center for Transportation Research
The University of Texas at Austin
1616 Guadalupe, Suite 4.202
Austin, TX 78701

http://ctr.utexas.edu/

Disclaimers

Author's Disclaimer: The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the Federal Highway Administration or the Texas Department of Transportation (TxDOT). This report does not constitute a standard, specification, or regulation.

Patent Disclaimer: There was no invention or discovery conceived or first actually reduced to practice in the course of or under this contract, including any art, method, process, machine manufacture, design or composition of matter, or any new useful improvement thereof, or any variety of plant, which is or may be patentable under the patent laws of the United States of America or any foreign country.

Notice: The United States Government and the State of Texas do not endorse products or manufacturers. If trade or manufacturers' names appear herein, it is solely because they are considered essential to the object of this report.

Engineering Disclaimer

NOT INTENDED FOR CONSTRUCTION, BIDDING, OR PERMIT PURPOSES.

Project Engineer: Dr. C. Michael Walton
Professional Engineer License State and Number: Texas No. 46293
P. E. Designation: Research Supervisor

Acknowledgments

The authors express their sincere appreciation to the Project Director, Kevin Pete, and following members of the Technical Advisory Panel: Caroline Mays, Orlando Jamandre, Jennifer Moczygemba, Brian Burk, Cary Karnstadt, Dieter Billek, Frank Phillips, Shawn Bichsel, and Stacey Worsham. Finally the authors thank John Esparza (Texas Trucking Association) and Richard Sherman (Council of Supply Chain Management Professionals) for their insight and recommendations to start approaching freight stakeholders.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

"Better and more timely data are essential, not only to make policies, programs, and implementing agencies more accountable, but to shift to a more outcome-oriented system" (New Transportation Agenda Conference Report).

The passage of the Moving Ahead for Progress in the 21st Century Act (MAP-21), enacted in 2012, provides guidance and regulation and creates *momentum* to improve the condition and performance of the national freight network. The Act also supports investment in freight-related transportation projects. MAP-21's stipulations expand agencies' interest in freight initiatives and modeling within statewide planning efforts, particularly the evaluation of the current and future freight transportation capacity necessary to ensure freight mobility. For example, in the case of project authorization, MAP-21 specifically states that a description of how the project will improve the efficient movement of freight can be created through data and information that support quantitative analysis. However, the understanding of freight demand and the evaluation of current and future freight transportation capacity are not only determined by robust models, but are critically contingent on the availability of accurate data. In this regard, insufficient and inferior quality data is the most commonly cited challenge in the development of freight models (TRB 2003, Mani and Prozzi 2004, Cambridge Systematics and GeoStats 2010, Tok et al. 2011, Prozzi et al. 2011, Quiroga et al. 2011, Rhodes et al. 2012, Chase et al. 2013).

State departments of transport (DOTs) typically (a) rely on the data compiled and published by federal agencies, such as the Freight Analysis Framework, (b) obtain one of the private commercial sources of data related to freight movements, such as the IHS Global Insight TRANSEARCH database, or (c) collect primary data through interview and surveys of freight stakeholders. Few states, however, collect primary freight data, partly because this can be a costly and time-consuming process. Also, currently few, if any, procedures at the federal, state, or local levels provide specific guidance on the collection of freight data (Cambridge Systematics 1997, Prozzi 2004).

As part of TxDOT Project 0-6297: *Freight Planning Factors Impacting Texas Commodity Flows* (2011), the CTR research team developed Texas Department of Transportation (TxDOT) Product 0-6297-P3: *Relational Multimodal Freight Database* using publicly available databases. This database product was created with the intention of providing TxDOT with a tool to easily access the publicly available freight data, and to identify any variables not captured by any publicly available database that TxDOT should consider acquiring. After evaluating these public databases, the research team found that no single publicly available data source contained all the data variables necessary to populate a *Multimodal Freight Database* for Texas that satisfies all TxDOT's freight data needs. Additional limitations identified in currently existing public and commercial databases include the following:

- reporting variations that complicate data interpretation, comparison, and combination from various sources,
- insufficient coverage of movements,
- uneven coverage of different sectors and transportation modes,
- insufficient geographic detail at urban and local levels, and

- insufficient coverage of international shipments.

To address the limitations of public and commercial sector databases identified in TxDOT Project 0-6297 and other related studies, this study team was commissioned to explore the feasibility of entering into a data-sharing partnership with representatives of the private sector (i.e., shippers, receivers, trucking companies, forwarders, etc.), and obtain sample data that can be used in formulating a strategy for integrating both public and private sector data sources.

This research study

- developed a strategy for collecting and integrating available freight data;
- explored the feasibility of entering into a data sharing partnership with the freight community for the collection of detailed and robust freight data that will satisfy the needs of transportation planning agencies;
- developed a prototype Integrated Freight Database, and
- provides recommendations to TxDOT on the cost-effectiveness of acquiring and maintaining a freight data sharing partnership to populate an integrated Freight Database.

This document is structured as follows: Chapter 2 summarizes the outcome of five workshops hosted in TxDOT's District and regional offices to determine the freight data needs of agency staff and obtain insight into how an Integrated Freight Database should be developed to adequately address statewide planning efforts. In Chapter 3, the study team reviews the state of the practice in freight data collection in terms of data variables captured, technology used, the cost of the collection method, and the reliability of the collected information to identify the data gaps that remain given the freight data needs identified by planning agencies. In Chapter 4, the study team identifies the data gaps that remain between the needed freight data variables, data variables captured in current databases, and data that can be obtained using primary data collection methods. Chapter 5 summarizes information garnered by the study team after conducting a series of key person interviews with representatives of Texas's freight community to determine their willingness to entering into a data sharing partnership with TxDOT and provide the study team with sample data. In Chapter 6, the study team reviews and assesses information about existing data architecture initiatives to gain insight into whether some of existing data architectures can be tailored or built upon for the development of an Integrated Freight Database. Chapter 7 provides a strategy for integrating multiple data sources and for collecting and compiling primary data from Texas's freight stakeholders. Recommendations on how TxDOT can populate and maintain an Integrated Freight Database including cost estimates for populating the database is provided in Chapter 8.

Chapter 2. Identifying Texas Freight Data Needs

In April 2012, the study team hosted five workshops and invited potential users of an Integrated Freight Database in TxDOT's District and Regional Offices, MPOs, and city offices to attend. Appendix B provides a sample workshop agenda. During these workshops, the research team

i) reviewed the freight data variables captured in available public and commercial databases (see Appendix A),

ii) explored how freight data is currently obtained, the issues and constraints in obtaining freight data, the data variables needed by modelers and planners, and the anticipated freight data needs given the models currently under development,

iii) solicited insight on how a proposed Statewide Freight Database will be used, and

iv) asked workshop participants to provide input in terms of the data architecture that will be developed and the types of architecture that can be supported by the different agencies.

This chapter summarizes the outcome of these workshops, including information obtained on how an Integrated Freight Database should be developed to adequately address statewide planning efforts.

2.1 Participants

Working from a previously compiled contacts list from TxDOT Research Project 5-6297-01: *"Relational Multimodal Freight Database,"* the research team contacted TxDOT districts, metropolitan planning organizations (MPOs), and city personnel in each of the areas where the workshops were conducted. Additionally, TxDOT district contacts were asked to provide names of individuals potentially interested in the workshops. Table 2.1 provides the number of participants by agency that attended each workshop.

Table 2.1: Workshop Participants by Agency

	Dallas	Amarillo	Midland	Beaumont	Pharr
TxDOT Officials	5	2	4	9	3
MPO Officials	3	2	1	2	3
City or other entity officials	-	1	1	-	3

In total, 39 individuals participated in the five workshops that were hosted as part of this project. The attendance of the workshops was mainly composed of TxDOT (59%) and MPO (28%) officials.

2.2 Freight Data Usage and Needs

After the research team presented the freight data variables captured in available public and commercial databases, participants were given the opportunity to discuss the following questions:

i) What freight data do you use?

ii) What data sources do you currently use?

iii) For what purpose(s) do you use the freight data?

iv) Have you experienced any issues in obtaining reliable freight data?

v) What freight data variables do you need?

vi) What level of detail do you require?

vii) How would you use the proposed Statewide Freight Database (what queries will you run)?

The following subsections summarize information obtained at the workshops regarding data usage and needs in each region.

2.2.1 Dallas Workshop

At the Dallas workshop, participants indicated that freight data was attained from a variety of sources: Freight Analysis Framework (FAF), Commodity Flow Survey (CFS), Carload Waybill Sample, and TxDOT Statewide Analysis Model (SAM). These sources include the following relevant data:

- truck counts and growth and
- percentage of truck traffic.

However, the following limitations were also acknowledged about current data sources:

- data was found to be outdated and aggregated;
- commercial data is costly;
- the reliability of “free public data” was questionable, and
- county-level data is based on assumptions used in disaggregating the data.

According to the participants, freight data is used to determine current roadway conditions, provide public information, and develop environmental documents for indirect and cumulative impacts. Additionally, participants stated that freight data is used for truck traffic projections, time-of-day and route information, as well as determine truck traffic generators.

Concerning variables for the database, participants expressed an interest in the following:

- *Regarding demand and capacity*: Data should be broken down by modal demand and current capacity of the system. This information is desired for pavement design and system life cycle cost analysis.

- *Regarding the level of disaggregation*: Origin-destination (O-D) of goods by commodity classification, tonnage, and mode.
- *Regarding vehicle sizes:* Lengths of vehicles to determine vehicle turning radii.
- *Regarding oversize/overweight (OS/OW) permits:* The number of OS/OW permits on corridors as well as the average weight of OS/OW trucks.
- *Regarding air quality:* Vehicle fleet age, engine type, vehicle type, speeds, and EPA Smartway partnership for air quality reports.
- *Regarding roadway use*: Traffic lanes used by trucks during trips.

According to the participants, the ideal database should provide as much data as possible and at a minimum, county- or corridor-level data. The database should be simple to use, reliable, and frequently updated, with the capability of showing when the data was collected. Participants indicated the database should be web-based, showing information via a graphic interface that uses the transportation network, and be able to aid potential users in determining congestion and continuity issues. The database should have the capability of determining traffic flow and the percent of truck traffic.

2.2.2 Amarillo Workshop

At the Amarillo workshop, participants indicated that they planned solely with data provided by TxDOT. This includes the following data:

- truck volumes, percentages of trucks on highways, 20- or 30-year equivalent single axle load (ESAL) counts;
- counts and classification counts—TxDOT collects traffic data through permanent or movable counters; based on these numbers they were able to estimate annual average daily traffic (AADT);
- commercial vehicle surveys performed by TxDOT, and
- rail counts at intersections.

The following concerns were also acknowledged about current data sources:

- no O-D data is available, and
- counting does not take place at the local level.

The study team suggested that potential Ports-to-Plains economic viability studies could be helpful. However, participants indicated that the Ports-to-Plains data provides more of a vision for an economic development of a corridor. Currently, project planning in the Amarillo region does not take into these account freight movements:

- there is no level of detail regarding the type of goods moved (i.e., by commodity);

- more data collection is needed for pavement design, construction, turning radii, hazmat cargo, among others,
- the MPO indicated that they do not have a means of obtaining specific data or conducting any classification counts. Their planning mostly relied on any freight data they could be obtained from TxDOT, and
- the last time the MPO updated their freight databases was when the interchange was under construction in 2001.

Participants also expressed interest in the following areas:

- The City representative indicated they would be interested in attaining whatever freight data might be made available for planning purposes.
- Participants mentioned that the City Commission would like some maps of hazmat routes that go through town as they only have maps that show the materials and hazmat that travel via the interstate highways (but not through "the Loop," for example).
- Outside of the MPO area, it would be useful to be better aware of the energy industry developments; oil field equipment and wind generation structures have created significant damage to pavements and roads.
- Data could be used for project prioritization. Participants indicated that Texas's rural plan had a ranking methodology that could provide better results if more data were available.
- Performance measures would benefit from freight data since very little information is available for quantifying at this time.

Participants identified the following items as desired freight data variables:

- *Regarding the number and type of load*: Number of trucks, type of loads (i.e., hazmat), and value of goods. Commodity data (product coming in or out) at the county level was specifically mentioned. Freight can be considered according (although not limited) to truck weight (over 80,000 pounds plus the extra 5%).
- *Regarding frequency*: Times and season variations, weekday variations for traffic control changes, and frequency.
- *Regarding classification information*: Roadway, Interstate, and US highway "pass through" collectors (FM roads).
- *Regarding the level of disaggregation*: County-level data was specifically mentioned as a desired level of disaggregation.
- *Regarding air and intermodal freight*: Not considered crucial at the moment but could potentially become relevant in the future. For example, Amarillo is considered to be well positioned for a containerized yard, and with the interstates' intersection, international airport, and rail facilities, it could become relevant quite soon.

According to the participants, the ideal database should be web-based with a graphic user interface, geographical information system (GIS), and depiction of highways and rail line locations. Participants suggested that a Google Earth interface would be more user-friendly, and expressed interest for an interactive database with maps. Being able to determine the best port of entry, shortest route, and the most effective way to transport goods into the state was also part of the ideal database. In addition, road construction maps with needs and level of service indicated were noted as ideal. It was mentioned that some parts of the database should be free and other components accessible through a fee.

2.2.3 Midland/Odessa Workshop

At the Midland/Odessa workshop, participants indicated that transportation planning was done solely with ESALs data used for design. Other external factors influence their planning, such as the development of the South Orient Railroad and other negotiations with Ferromex.

Currently, project planning in the Midland/Odessa region does not take into account certain freight movements:

- Not all products that are brought into the Midland/Odessa or even up to the Fort Stockton region are being captured, and
- there is cut-throat competition among all industries, so information about new developments is never shared.

Participants in this region are concerned with the following issues:

- Union Pacific Railroad is solely interested in long haul trips and business. However, short-haul service by this carrier should be promoted in the area—the more loads moved by rail, the better.
- Roads are being destroyed. Oil pipes, sand, wind mills, cement, and other heavy seasonal deliveries are destroying the pavements in the region. It was noted that the oil industry caused 30% or more reduction in service life.
- Seasonal movements (brought in by train but needing truck delivery for the final leg of the trip) create unpredictable traffic spikes.
- Because of cut-throat competition, there is little opportunity to plan appropriately for these new developments. Road construction or improvements are more responsive in nature to what is happening.
- Local infrastructure was often left out of traffic counts—this would be a desired addition.

Participants identified the following items as desired freight data variables:

- Participants expressed much interest for *any* kind of freight data that might be made available to them.

- *Regarding the traffic generators*: Knowing the main traffic generators for current use and what potential driver generators will come into the area was noted as important for planning purposes.
- *Regarding traffic diversion*: Knowing how much traffic is going to be diverted would be a good measure for the area. Route change, overloading, and the type of mixture of traffic were identified issues.
- *Regarding the type of loads*: Knowing which type of heavy loads (5 to 10% excess in weight) are destroying the roadways would be helpful. Commodity data was also requested.
- *Regarding the level of disaggregation*: Having route and distribution patterns at the best level of detail was desired.
- *Regarding the government and private party participation*: It would be desirable to work with the government (i.e., maybe an agreement with the Railroad Commission) to better understand the location and impacts of new energy developments. The latter would allow local agencies to better plan for road damage and improvements. In addition, better communication with private stakeholders through data sharing would enable planners to better prepare road infrastructure for potential new developments.

According to the participants, the ideal database would show an aerial view of their infrastructure and how it is being used. Additional features desired for the database include the ability to show traffic generators (current and future), changes in modes, commodity information (weigh out, cube out), loads locations, OS/OW loads, and networks used.

2.2.4 Beaumont Workshop

At the Beaumont workshop, participants indicated that they used data provided by TxDOT's Transportation Planning and Programming (TP&P), simple traffic counts, and vehicle registration data. This includes the following data:

- ESALs, AADT, and traffic trends;
- the number of rail conflicts;
- percentage of trucks, and
- the number of heavy vehicles registered in each county.

Limitations of the current data identified include the following:

- While it was easy to simply count trucks traveling through the area to obtain data, it would be beneficial to know routes and where the vehicles are registered.
- TxDOT representatives indicated that there was difficulty keeping up with the rapidly changing business models for planning purposes.

The data collected was used to create, maintain, and improve the efficiency of the roadways. TxDOT participants stated that freight data was used for pavement design, and highway capacity analysis. The MPO also uses the data for air quality modeling.

Participants identified the following items as desirable freight data variables:

- truck volumes and percentages;
- trip origins and destinations;
- types of vehicles;
- registration of vehicles;
- number of daily rail traffic, and
- effect of train crossings on air quality

According to the participants, the ideal database should have route level detail. However, participants were unsure of their use of the database, stating that TP&P would be the most likely determinants for TxDOT.

2.2.5 Pharr Workshop

At the Pharr workshop, participants indicated that transportation planning in the region was done using the following methods:

- TxDOT officials indicated that traffic counts and truck percentages were used and that pavement designs were based only on the percentage of trucks with no definition of whether the truck may be full or not.
- Brownsville MPO officials indicated that they used the TxDOT map with truck values for the region.
- It was noted that algorithms are used on traffic counts to estimate traffic and new information could improve the current process.
- The Port of Brownsville has confidential information about the freight that they share with the MPO and with the airport for planning purposes.

Currently, project planning in the Lower Rio Grande Valley region does not take into account freight movements, as no disaggregated data is available. In addition, participants stated that the following reasons were partly to blame for the latter:

- The MPO stated they used private studies shared with them to obtain information about truck routes in Cameron County. This official also noted it is very difficult to obtain O-D data for truck routes. Generally, the consultants hired to obtain the O-D information visited companies as well as sent a letter requesting the information.
- Participants highlighted that many times industry decisions are made very quickly with no transportation planning taken into account. Economic development divisions promoted development without taking into account other variables such as providing

transportation infrastructure. For example, participants mentioned that United Parcel Service (UPS) moved their air freight cargo operations to the McAllen airport.

- It was also noted that from an economic development perspective, capacity is the only concern for forecast models and this might not be enough.
- It was also brought up that new roads in Mexico are shifting produce to the Pharr area and that this information needs to be captured.
- Participants noted that from a statewide perspective, it is more difficult to see which variables are useful parameters to be used at the local level in each region.

Participants identified the following items as desired freight data variables:

- *Regarding port-of-entry data:* Land ports-of-entry data in this region is important to adequate transportation network planning for any proposed port-of-entry modifications. The U.S. Customs and Border Protection has data from an archived history on international crossings that might be of use.
- *Regarding competitiveness*: Opportunity costs of the system were also mentioned. The need to be competitive and to reduce time and cost is important for companies and municipalities. Congestion pricing was also mentioned.
- *Regarding O-D data:* The MPO specifically desired O-D data for planning. Freight data was not currently integrated into the travel demand models used, but efforts were being made to do so. Flow and count data on at least major arterials, O-D, and counts were identified as needed variables. City of McAllen officials indicated that they are more interested in truck traffic, O-D, and future needs so that proper planning and development can be done.
- *Regarding commodity type data*: It was noted during the meeting that commodity information was needed for environmental justice planning. Also, truck routes were mentioned as an appropriate parameter; truck routes by commodity would be even better.

According to the participants, the ideal database should be a web-based or CD database. A feature that enables users to find a particular area with a query function or via click and zoom was also requested. Different database variables at a microscopic level (i.e., data disaggregated at the county level) would best interact with current regional models currently used in Hidalgo and Cameron counties. Simplicity and user-friendliness was also mentioned—it is important that the database is not limited to only running on high performance computers.

2.3 Summary of Workshops

Despite the varied locations of the workshops, some common themes were noted. These themes are summarized in the following tables. Table 2.2 shows data currently used by participants; Table 2.3 summarizes all general concerns related to the current available data freight; Tables 2.4 and 2.5 review the freight variables and type of database requested by participants.

Table 2.2: Data Currently Used

	Dallas	Amarillo	Midland	Beaumont	Pharr
Truck volumes, percentages of trucks, 20- or 30-year ESAL counts	X	X	X	X	X
Traffic counts and classifications	X	X		X	X
Rail counts at intersections		X		X	
The number of heavy vehicles registered in each county				X	
Confidential information from marine port for MPO and airport planning purposes					X

As observed in Table 2.2, traffic indicators such as truck volumes, percentages of trucks, and ESAL counts, as well as traffic and rail counts, are the most relevant data variables currently used by the workshop participants to plan transportation infrastructure in their region. It can also be observed that in comparison to the other areas, participants from Beaumont cited the highest number of available data sources.

Table 2.3: General Concerns of Currently Available Freight Data

	Dallas	Amarillo	Midland	Beaumont	Pharr
Assumptions used to disaggregate the data	X				X
Lack of O-D data		X			X
Lack of transportation planning involvement in industry decision-making				X	X
High costs of acquiring data	X				
Outdated and aggregated freight data	X				
Reliability of "free public data"	X				
Lack of traffic counts on local infrastructure		X			
There is no level of detail regarding the type of goods moved (i.e., by commodity)		X			
More data collection is needed for pavement design, construction, turning radii, hazmat cargo, etc.		X			
Lack of information about seasonal movements			X		
Unknown routes and unknown vehicle registration data				X	

As observed in Table 2.3, participants generally had three major concerns: (*i*) the type of assumptions used to disaggregate data, (*ii*) the lack of O-D data, and (*iii*) the lack of transportation planning consideration in the industry decision-making process. It can also be observed that Dallas and Amarillo expressed more concerns than any of the other areas.

Table 2.4: Freight Data Variables Desired

	Dallas	**Amarillo**	**Midland**	**Beaumont**	**Pharr**
Modal demand and current capacity of the system	X	X		X	
O-D of goods by commodity classification, tonnage, and mode	X	X		X	X
Types and number of OS/OW vehicles	X		X		
Seasonal variations, weekday variations, time of day and frequency of trips		X	X		
Vehicle turning radii	X				
Air quality variables: Vehicle fleet age, engine type, vehicle type, speeds and EPA Smartway partnerships	X				
Traffic lanes used by trucks during trips	X				
Inclusion of local infrastructure			X		
Traffic generators			X		
Vehicle registration				X	
Number of daily rail traffic				X	
Effect of train crossings on air quality				X	

As observed in Table 2.4, the most popular freight data variables requested by participants were modal demand and current capacity of the system, and O-D of goods by commodity, classification, tonnage, and mode. Other variables mentioned as relevant were variations (seasonal, weekday, time of day) and frequency of trips, as well as types and numbers of OS/OW vehicles. It can also be observed that Dallas and Beaumont participants requested more data variables than any of the other regions. Participants from Midland generally expressed an interest for *any* kind of freight data that might be made available to them.

Table 2.5: Freight Database Features

	Dallas	Amarillo	Midland	Beaumont	Pharr
Web-based with graphical user interface and GIS of the transportation network	X	X	X		X
Ability to determine the best port of entry, shortest route, and the most effective way to transport goods into the State		X	X		X
County and corridor level data		X	X	X	
Simple and reliable					X
Frequent updates	X				
Show traffic generators (current and future), changes in modes, commodity information (weigh out, cube out), loads locations, OS/OW loads, and networks used			X		
Ability to find a particular area with a query function or via click and zoom					X
Ability to determine traffic flow and percentage of truck traffic	X				
Road construction maps with needs and LOS indicated		X			

As observed in Table 2.5, the most popular freight database features requested by participants were (*i*) a web-based platform, with graphical user interface and GIS, (*ii*) the ability to determine the most effective or shortest routes to transport goods within Texas, and (*iii*) a disaggregation of the data at the county or corridor level. It can also be observed that participants from the Amarillo, Midland, and Pharr areas expressed interest for the highest number of features for the potential freight database.

Participants of the workshops indicated two main uses for freight data: planning and the development of environmental documents. The following are the planning-related uses specifically mentioned for the freight data:

- Pavement design;
- Highway capacity analysis;
- Route information;
- Time-of-day information;
- Determination of current roadway conditions;
- Determination of truck traffic generators;
- Providing public information;
- Project prioritization, and
- Travel demand modeling

The most commonly mentioned planning uses for freight were (i) pavement design and (ii) highway capacity analysis.

2.4 Conclusion

This chapter summarizes the outcome of workshops with agency staff. Five one-day workshops (held in Dallas, Amarillo, Midland/Odessa, Beaumont, and Pharr) helped the study team to determine that there is a statewide need for additional freight data. Thirty-nine participants attended the workshops and were given a presentation regarding the project and existing freight databases. In addition, the study team took the opportunity to seek input from the participants regarding their desired variables, as well as their "must haves" for the statewide freight database.

In summary, participants indicated the need to have efficient data sources to enable them to appropriately plan transportation infrastructure for their regions. Specifically, the data requested included quantity of goods, commodity types, and mode of transport disaggregated at the county level. Additionally, they expressed interest for data to be collected related to the capacity of the system, the O-D of freight (organized by their classification), tonnage, type, and port-of-entry information.

Finally, in the case of the creation of a potential statewide freight database, participants indicated as "must haves" that county- and corridor-level of disaggregation would be desirable. Also, participants expressed that having a user-friendly web-based database with a graphical user interface of the transportation network and with GIS capabilities would be helpful for their planning activities. The ability to determine the most practical port of entry, shortest routes, and most effective way to transport goods, were also requested functionalities of the potential database.

In Chapter 3, the study team reviews the state of the practice in freight data collection in terms of data variables captured, technology used, the cost of the collection method, and the reliability of the collected information, and uses this information to identify the data gaps (Chapter 4) that remain given the freight data needs identified by planning agencies.

Chapter 3. Freight Data Collection Methods

Findings from the workshops determined that state, city, and MPO planners and modelers in Texas typically rely on TxDOT's traffic counts and percentage of truck traffic for freight-related planning. TxDOT's Statewide Analysis Model (SAM)[1], which is the state's multimodal travel demand model that provides highway traffic forecasts for highway passenger travel and freight transport, uses the TRANSEARCH database as its main source of freight data. Ideally, collecting primary freight data is the best source of data for planning purposes. However, the collection of primary freight data can be a costly and time-consuming process at the state level. This chapter presents a review of the state of the practice in freight data collection in terms of data variables that can be captured, technology required, the cost of the collection methods, and their ability to close the gap that remains given the freight data needs identified by planning agencies and currently available data sources. In addition, data collection methods used in existing freight databases were compiled to provide an overview of how agencies are currently collecting freight data.

3.1 Primary Freight Data Collection Methods

A number of primary freight data collection efforts are available. The more traditional approaches involve collecting freight data directly from freight stakeholders (e.g., shippers, carriers, receivers, and freight forwarders) through surveys such as roadside intercept surveys, mailout/mailback questionnaires, a combination of telephone and mailout/mailback questionnaires, driver trip diaries, and telephone interviews, and Internet surveys. Done correctly, these survey methods, in general, are regarded as very reliable for obtaining freight data. However, these methods are costly and time consuming—especially when detailed disaggregated freight data is required. Progress is being made in some states and regions to cost-effectively procure more accurate truck travel data using technological applications instead of the traditional survey methods, which are reliable but very expensive.

3.1.1 Survey Data Collection Methods

- **Telephone interviews**—In general these yield a high response rate and facilitate follow-up, but depending on the sample size, timeframe, and nature of the survey, the number and skills of the interviewers, this method can become too costly.
- **Mailout/mailback surveys**—In general these are less costly, but have a lower response rate. Reliability and completeness depend upon finding the appropriate individual within an organization or company.
- **Combination of telephone with mailout/mailback**—In general this approach yields a higher response than mailout/mailback surveys, but it is likely to be more expensive.

[1] SAM is TxDOT's multimodal travel demand model that provides highway traffic forecasts for highway passenger travel and freight transport, intercity and high speed passenger rail ridership, freight rail tonnage and train forecasts, and forecasts of air passenger travel to and from Texas airports.

- **Commercial vehicle trip diaries—**Often used to calibrate urban models, requesting diaries may result in low response rates at times because the person agreeing to the survey is not the truck driver.
- **Roadside/intercept interviews**—Often used for truck surveys, these in general yield a high response rate. The disadvantages of this method include potential disruption to traffic flow, safety hazards for the interviewers, less ability to follow up with respondents, the effect of factors such as weather, time of day, and lighting on implementation, and restricting the sample to a particular location rather than including an entire region.
- **Personal interviews**—The most costly method of conducting surveys, personal interview generally involve a smaller, more select, or targeted sample. Not appropriate for the population of statewide truck travel databases.

3.1.2 Technological Approaches[2]

In the case of truck data, a number of ITS technologies are able to collect various truck travel attributes; including routing, time, carrier, and O-D information (see Table 3.1). These technologies are described in this chapter.

Table 3.1: Attributes Captured by Current ITS Technologies (CTR, 2012)

Attribute	Inductive Loop Detectors	WIM Sensor Systems	Sensors	Video Image Detection	AVC	AVI Toll and CVO Tags	AVL GPS	License Plate Matching
Vehicle classification	X	X	X	X	X	X		
Vehicle weight		X						
Vehicle speed	X	X	X	X		X	X	X
Vehicle delay data	X		X	X				
Vehicle incident data	X		X	X				
Traffic volume data (classification)	X	X	X	X	X	X		
Commodity/cargo type				X	X	X	X	
Payload (cargo) weight								
Truck O-D patterns						X	X	X
Trip O-D patterns						X	X	X
Average tour length						X	X	X
Number of stops per tour						X	X	X
Number of truck stops for LTL shipments						X	X	X
Travel time	X		X			X	X	X
Transit time						X	X	
Travel time reliability						X	X	

[2] Most of the information for these technologies was gathered from a recently completed CTR study on data collection for the Dallas/Fort Work Region titled *Framework for Collecting Truck Travel Data for the Dallas/Fort Worth Region (2012)* and the FHWA study *Vehicle Detection and Surveillance Technologies (2007).*

Within the last decade, the use of global positioning systems (GPS) has also been extended to fleet management for a number of purposes, such as driver notification, logging of hours, in-cab scanning, navigation, Wi-Fi capability at truck stop locations, and maintenance alerts. Several companies, such as Qualcomm and Rand McNally, market these types of devices to commercial carriers. These devices are relatively inexpensive, while reducing operational costs and improving the productivity of the carriers.

There are three main categories of ITS technologies for collecting freight data. The first category, *in-roadway sensors*, includes devices such as inductive loop detectors, weigh-in-motion sensor systems, piezoelectric detectors, bending plate and magnetic detectors. These technologies are either embedded in roadway pavements, roadway subgrades, or otherwise attached to the surface of the roadway. The second category, *over-roadway sensors*, includes devices such as passive and active infrared detection, microwave detection, ultrasonic detection, passive acoustic detection, and video image detection. These technologies are mounted above the surface of the roadway either above the roadway itself or alongside the roadway, offset from the nearest traffic lane by some distance. The last category, *vehicle-based technologies*, include devices such as automated vehicle classification (AVC), automated vehicle identification (AVI), commercial vehicle operations tag (CVO), GPS, and license plate matching. These devices are installed on vehicles and data is transmitted either through wireless readers or satellite.

In-Roadway Sensors

Inductive Loop Detectors—Inductive loop detectors are the most prevalent detection system in use today. This detector type operates such that when a vehicle crosses a loop or stays within a loop area, it causes a reduction in loop inductance and an increase in oscillator frequency (CTR 2012, Martin et al., 2003). The vehicle's presence is detected by frequency changes exceeding the threshold determined by sensitivity settings. Inductive loop detectors may be saw-cut, trenched-in, or preformed and are used frequently in intersection control, traffic recording, and traffic monitoring (CTR, 2012). Costs for installation range from $8,000 for a corridor to $15,000 for an intersection. Operations and management costs range from $600 to $1,400 per year based on installation type (ITSJPO 2012). Figure 3.1 illustrates how inductive loops detect a vehicle.

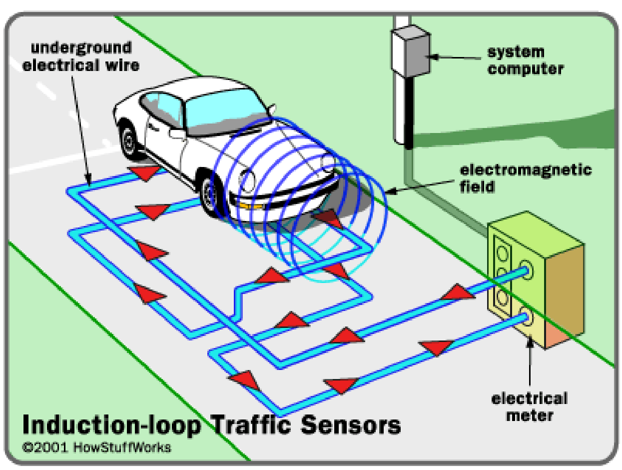


Figure 3.1: Inductive Loop Traffic Sensor (CTR, 2012; HowStuffWorks, 2012)

Weigh-in-motion (WIM) Sensor Systems—WIM technologies capture axle weight spacing, vehicle length, traffic volume, speed, invalid measurements codes, weight violations, and classification based on the number and spacing of axles (FHWA, 2007). The technologies use components such as scales, cables, loop detectors, controllers, cabinets, modems, and cabling (see Figure 3.2). According to the Federal Highway Administration (FHWA)(2007), WIM systems increase the capacity of weigh stations and are often used when heavy truck traffic volumes cannot otherwise be accommodated. Categories of WIM systems are listed in Table 3.2 along with the corresponding data each provide (FHWA, 2007; ASTM E1318-02, 2002).

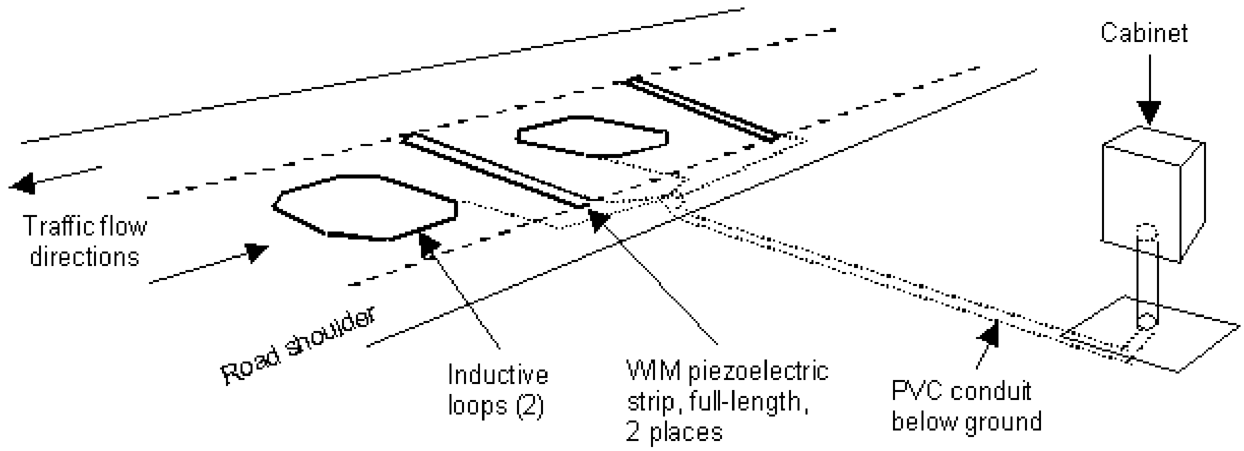


Figure 3.2: WIM installation with full-length piezoelectric sensors (CTR, 2012; FHWA, 2007)

The accuracy of WIM systems is a function of the vehicle dynamics, pavement integrity, composition, and design; variance inherent in the WIM system; and calibration (FHWA, 2012).

It is recommended that WIM systems be expressed as life cycle costs (FHWA, 2007) consisting of initial capital cost and maintenance costs. Capital costs for the various technologies (such as piezoelectric sensors, bending plates, and load cells) include in-road WIM equipment, installation labor and materials, initial calibration, and traffic control (Table 3.3). Maintenance costs—arising from differences in traffic volumes, truck weights, weather, original installation procedures, roadbed condition, and onsite quality control, among others—consist of labor and materials, traffic control, and system recalibration (Table 3.4). The costs are based on performing annual routine maintenance (e.g., road inspection and crack filling) on the roadbed surrounding the WIM system (FHWA, 2007). Piezoelectric sensors are assumed to require replacing every 3 years, bending plates refurbishing every 5 years, and single load cells every 5 years (FHWA, 2007). Note that these costs may vary by manufacturer and sensor model.

.

Table 3.2: WIM System Categories, Applications, and Data Items (FHWA, 2007)

Data Item	Category			
	Type I	**Type II**	**Type III**	**Type IV**
Speed range	16 to 130 km/h (10 to 80 mi/h)	16 to 130 km/h (10 to 80 mi/h)	16 to 130 km/h (10 to 80 mi/h)	3 to 16 km/h (2 to 10 mi/h)
Application	Traffic data collection	Traffic data collection	Weight enforcement	Weight enforcement
Number of lanes	1 or more	1 or more	1 or more	1 or more
Bending plate	●	●	●	●
Piezoelectric	●	●		
Load cell	●	●	●	●
Wheel load	●		●	●
Axle load	●	●	●	●
Axle group load	●	●	●	●
Gross vehicle weight	●	●	●	●
Speed	●	●	●	●
Center-to-center axle spacing	●	●	●	●
Vehicle class (via axle configuration)	●	●		
Site identification code	●	●	●	●
Lane and direction of travel	●	●	●	
Date and time of passage	●	●	●	●
Sequential vehicle record number	●	●	●	●
Wheelbase (front-to-rear axle)	●	●		
Equivalent single-axle load	●	●		
Violation code	●	●	●	●
Acceleration estimate			●	●

Table 3.3: Budgetary initial capital costs of WIM systems (FHWA, 2007)

Capital Cost Component (Costs shown are in 2007 dollars)	**Piezoelectric**	**Bending Plate**	**Single Load Cell**
In-road equipment	$4,500	$13,000	$34,000
Installation labor and materials	$3,500	$6,500	$10,500
Traffic control	$1,000 (1 day)	$2,000 (2 days)	$4,000 (4 days)
Total capital cost	$9,000	$21,500	$48,000

Table 3.4: Life-cycle maintenance costs of WIM systems (FHWA, 2007)

Cost Component (Costs shown are in 2007 dollars)	Piezoelectric (3 years)	Bending Plate (5 years)	Single Load Cell (5 years)
In-road equipment	$4,000	$6,000	$1,000
Labor and materials	$4,000	$5,500	$500
Traffic control	$1,500 (1 day)	$1,500 (1 day)	$750 (1/2 day)
Total life-cycle cost	$9,500	$13,000	$2,250

Piezoelectric Detectors—Piezoelectric detectors are electromechanical systems that react to changes in compression. Composed of metal strips placed on or near the road surface, these detectors are not sensitive to electromagnetic fields or radiation and can operate in a variety of temperatures. Detection occurs when a vehicle crosses over the device, generating a voltage from the piezoelectric material. This voltage is proportional to the weight of the detected vehicle; by measuring the change in voltage, an axle weight can be measured. Piezoelectric detectors are used for counts, detection, and classification, and used in WIM applications (CTR, 2012). Costs for these detectors range from $150,000 for a single lane to $300,000 for four lanes (FHWA 2000). Figure 3.3 illustrates the composition of a piezoelectric tube.

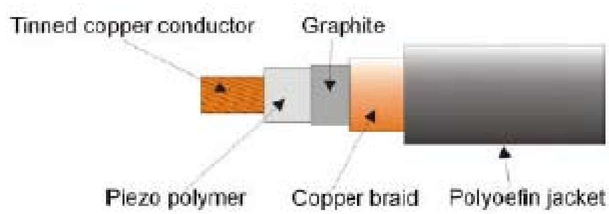
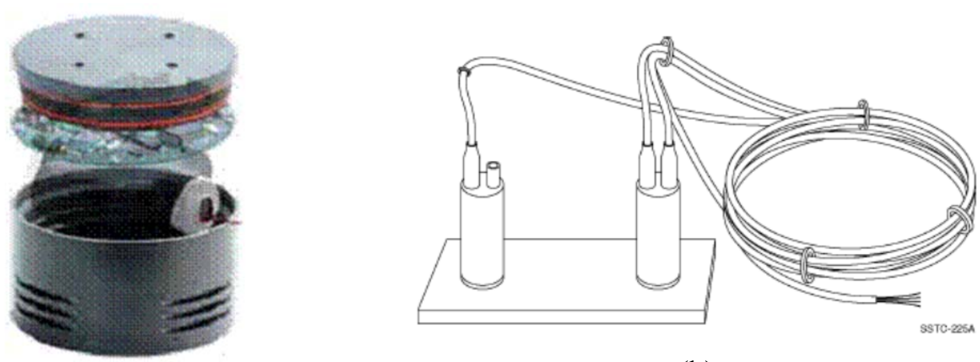


Figure 3.3: Composite materials of a piezoelectric tube (CTR, 2012; Martin, 2003)

Bending Plate—Bending plate WIM systems use plates with strain gauges bonded to the underside (see Figure 3.4). As vehicles pass over the bending plate, the system records the strain measured by the strain gauges and calculates the dynamic load. The static load is estimated using the measured dynamic load and calibration parameters. The calibration parameters account for factors that influence estimates of the static weight, such as vehicle speed and pavement/suspension dynamics. The accuracy of bending plate WIM systems can be expressed as a function of the vehicle speed reached while over the plates, assuming the system is installed in a sound road structure and subject to normal traffic conditions (FHWA, 2007).

Figure 3.4: Bending Plates for WIM systems (Florida DOT, 2007)

Magnetic Detectors—Magnetometers detect magnetic disturbances in the earth's magnetic field when a vehicle passes over the detector. These detectors can often be found on bridges where inductive loops cannot easily be installed. This technology has been proven to be more accurate for counts, easier to install, and less intrusive than inductive loops. The self-powered vehicle detector can collect vehicle counts, classifications, occupancy, presence, and speed. These detectors are approximately $600 per lane in addition to receiver costs (FHWA 2000). See Figure 3.5 for examples.



(a) (b)

Figure 3.5: Magnetic detectors: a) Groundhog® G-8 magnetometer and road-weather sensor (Nu-Metrics, 2011), and b) Model 701 microloop probe (3M Company, 2011)

Over-Roadway Sensors

Passive Infrared Detection—Passive infrared detectors recognize emitted energy from vehicles, road surface, and other objects (Figure 3.6). Detectors are typically mounted overhead and can capture volume, speed, class, occupancy, and presence data. Per installation, this detector type costs around $1,200 (ITSJPO).

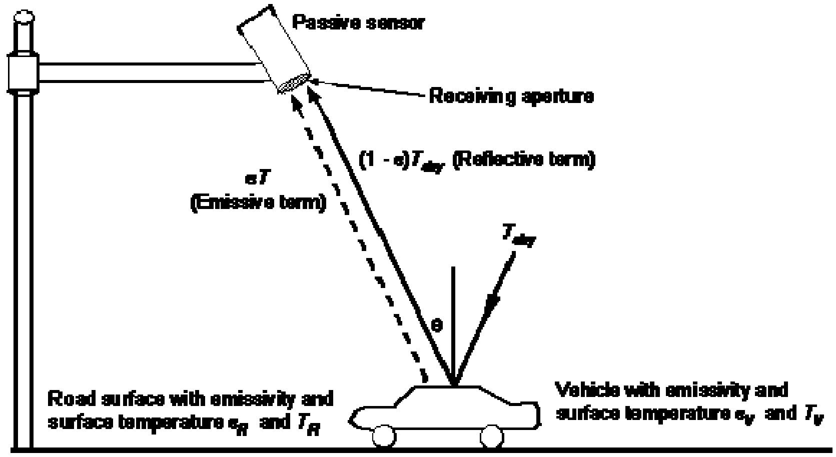


Figure 3.6: Emission and Reflection of Energy by Vehicle and Road Surface (Martin,2003)

Active Infrared Detection—Active infrared detectors use infrared energy to detect vehicles in a similar manner to microwave radar detectors. The detector emits multiple radiation beams and converts the energy into electric signals to detect vehicles (Figure 3.7). Vehicle temperature, size, and structure are determined by the amount of radiation emitted. The two- and three-dimensional images created allow for the classification of vehicles into 11 different classes. Per installation, this detector type costs around $7,500 (ITSJPO).

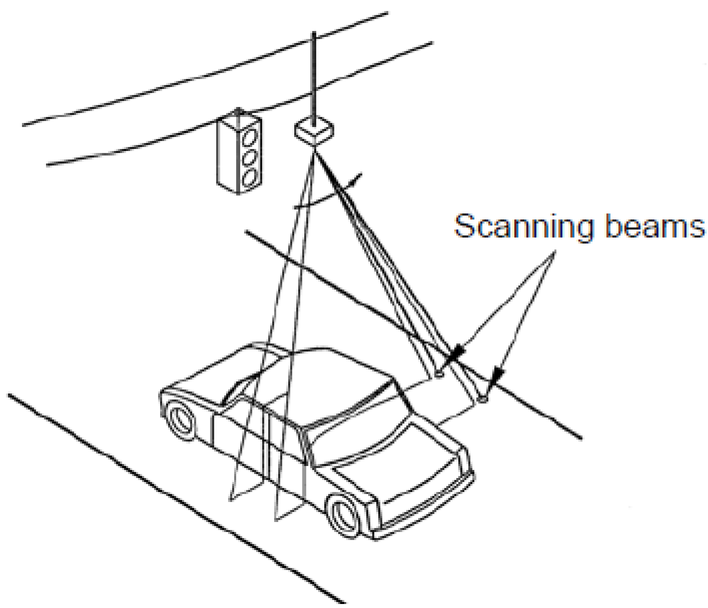


Figure 3.7: Active Infrared Detector (Martin 2003)

Microwave Detection—Microwave detectors can be mounted overhead or on the side of the road and use either Doppler microwave detectors or frequency-modulated continuous wave (FMCW) detectors. Detection occurs as an object moving through the sensor's field returns the microwaves to the sensor at a different frequency than was transmitted. Doppler detection is effective only for vehicles moving at a minimum speed, and only for the determination of speed.

FMCW devices can detect the presence of motionless vehicles as well as the speed of moving vehicles. Additionally, FMCW can provide real-time volume and occupancy, detect queues, and classify vehicles based upon length. Per corridor, installation costs approximately $11,000 with operation and maintenance costs of $500 per year. For intersections, installations cost approximately $14,000 with $100 of operation and maintenance costs per year (ITSJPO).

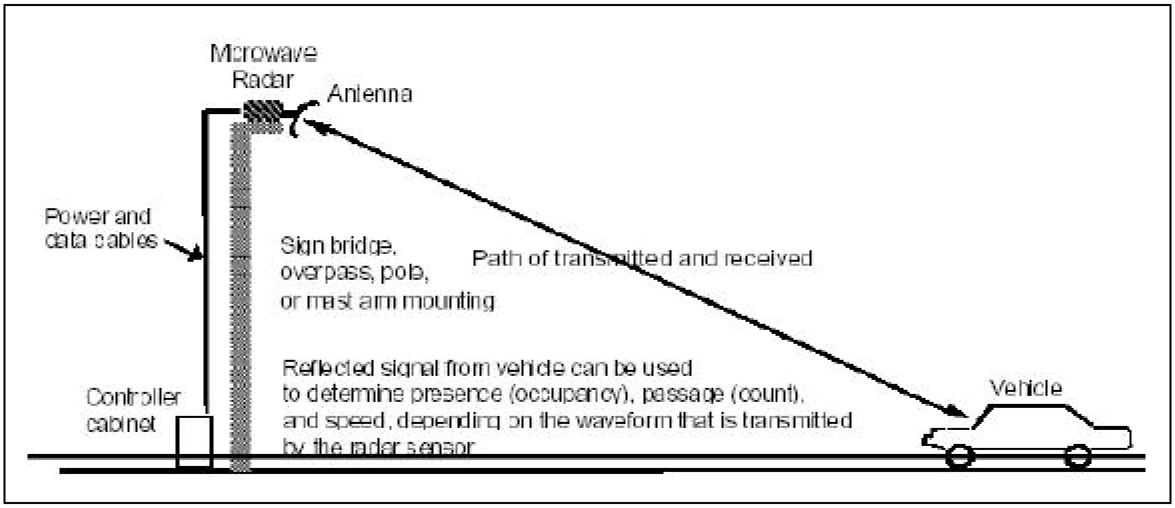


Figure 3.8: Microwave Detector (Martin 2003)

Ultrasonic Detection—Using an active acoustic detector, an ultrasonic detector transmits sound waves towards the detection zone and senses the waves returned by vehicle reflection. Two types are available: pulsed ultrasonic and continuous wave (CW) detectors. They can detect volume and presence and cost approximately $2,000 per lane (FHWA 2000).

Passive Acoustic Detection—Passive acoustic detectors measure the sounds produced from passing vehicles and are typically mounted to existing road structures (see Figure 3.9). The primary uses for this detector type are vehicle sensing and intersection control. Passive acoustic detection is capable of capturing count, classification, occupancy, presence, and speed data. Corridor installations cost around $8,000 with $400 in operation and maintenance costs per year, while intersection installations cost approximately $15,000 with the same operation and maintenance costs as the corridor application (ITSJPO).

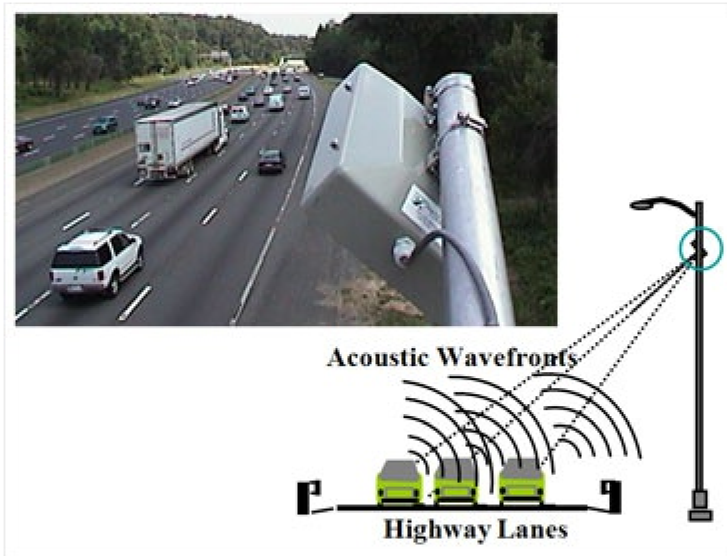


Figure 3.9: Passive Acoustic Detection (SmarTek 2012)

Video Image Detection—Video image detection uses multiple cameras, computers for digitizing and processing images, and software for the interpretation and conversion of images. These devices have the ability to capture count, classification, density, occupancy, presence, speed data, and incident detection. Limitations of this technology include the reduction of accuracy due to occlusions, lighting conditions, and high-density, slow-moving traffic. Costs per installation type range from $5,000 per approach to approximately $18,000 per intersection (FHWA 2000).

Commercial Vehicle Information System and Network (CVISN)—CVISNs employ radio frequency transponders to gather time-stamped information on location, speed, and direction of travel, as well as vehicle identification. The limited numbers of reader locations that are often on primarily major rural routes have lessened the technology's popularity in recent years. CVISN costs are estimated to be $48,000 (ITSJPO). See Figure 3.10 for CVISN uses.

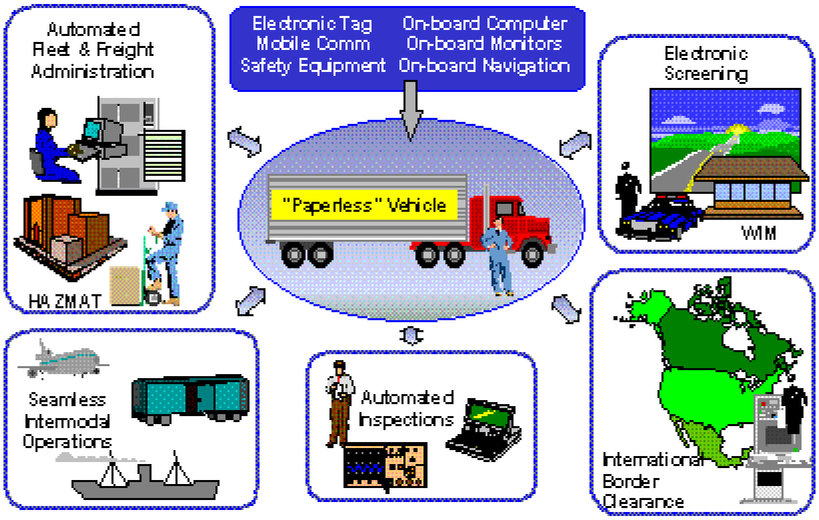


Figure 3.10: User services provided by CVISN (FHWA 1997)

Vehicle-Based Technologies

Automated Vehicle Classification (AVC)—AVC devices are vehicle-based and are often deployed with electronic credentials (EC), which produce an electronic record of the vehicle type. These devices allow the motor carriers to file, obtain, and pay for required licenses, registrations, and permits. Additionally, the technology uses WIM, closed circuit television (CCTV), and roadside detection (RS-D) for vehicle classification. Projected costs for EC are the average initial cost of $500,000 and $190,000 per year for operation and maintenance (FHWA 2002).

Automated Vehicle Identification (AVI)—AVI devices allow for the identification of vehicles using DSRC devices or electronic tags. These devices comprise three pieces: a vehicle-mounted transponder, roadside communication unit, and a storage and data processing unit. AVI systems can be optical, infrared, inductive loop, radio frequency, or microwave. AVI toll tags have a unique ID embedded in the tag and transmit this unique ID when requested by a roadside toll tag reader. Most large urban areas in Texas now have toll facilities and thus toll tag readers used by TxDOT to measure travel time and volume. These toll tags are relatively inexpensive. An example of an AVI application is the PrePass system that enables participating transponder-equipped commercial vehicles to be pre-screened throughout the nation at designated weigh stations, port-of-entry facilities, and agricultural interdiction facilities (see Figure 3.11). Costs for AVI are approximated at $84,000 (ITSJPO).

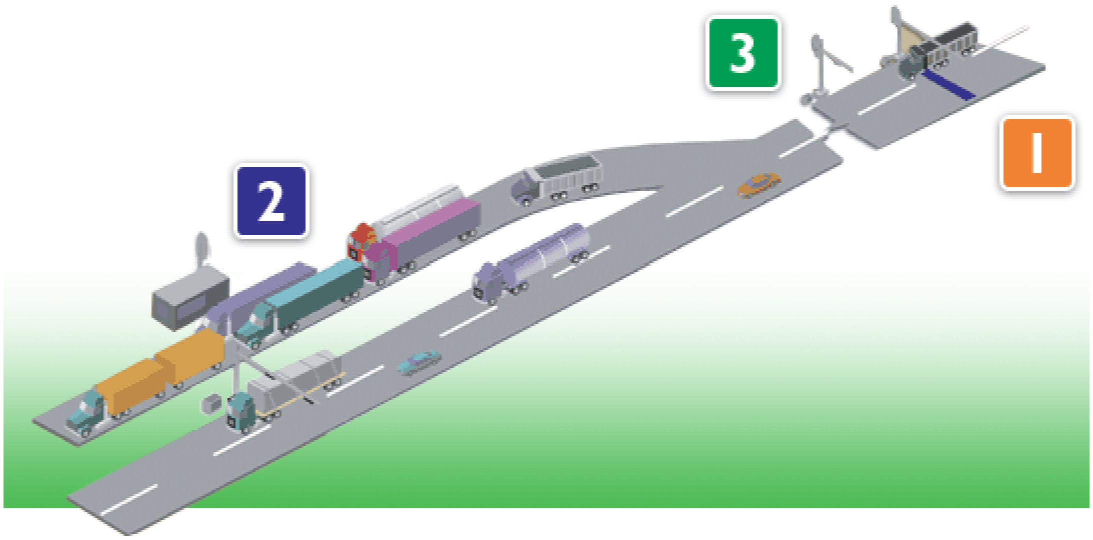


Figure 3.11: AVI PrePass System (PrePass)

> 1) Trucks enrolled in PrePass are assigned small, wireless transponders that easily mount onto the vehicles' windshields…an electronic reader on a boom over the road automatically scans the transponder and identifies the vehicle.
>
> 2) A secure PrePass computer located inside the scale house accesses the vehicle information associated with the transponder, and validates it to ensure compliance with state requirements…
>
> 3) Finally, as the truck passes beneath a second boom, a signal indicating whether the vehicle may bypass is transmitted back to the transponder. If the vehicle's information cannot be validated, or if it is selected for a random manual inspection, a red light on the transponder alerts the driver to stop. If the vehicle's credentials, safety, and weight are all in order however, a green light tells the driver to go ahead and bypass the facility.
>
> Source: PrePass 2012

<u>Commercial Vehicle Operations Tag (CVO)</u>—A CVO tag has a unique ID embedded. This tag stores data sent by a roadside CVO reader, and transmits the unique ID and the stored data when requested by a roadside CVO tag reader. Many large trucking firms now use CVO tags in their normal operations; the PrePass system, which allows certain trucks to bypass truck weight stations, also uses CVO tags. CVO tags are more expensive than toll tags but are still relatively inexpensive.

<u>AVL/GPS</u>—AVL/GPS systems determine the approximate location of vehicles and track vehicles while on the transportation network. Real-time data provided via GPS is used to determine travel times, traffic flows, and patterns on a network. The technology has the potential to provide congestion information and has been applied to routing and scheduling optimization as well as asset management. Additionally, HAZMAT Response Systems use AVL technologies to provide necessary data for the enforcement and incident management. These devices are moderately expensive and cost approximately $2,000 per vehicle installation (ITSJPO). See Figure 3.12.

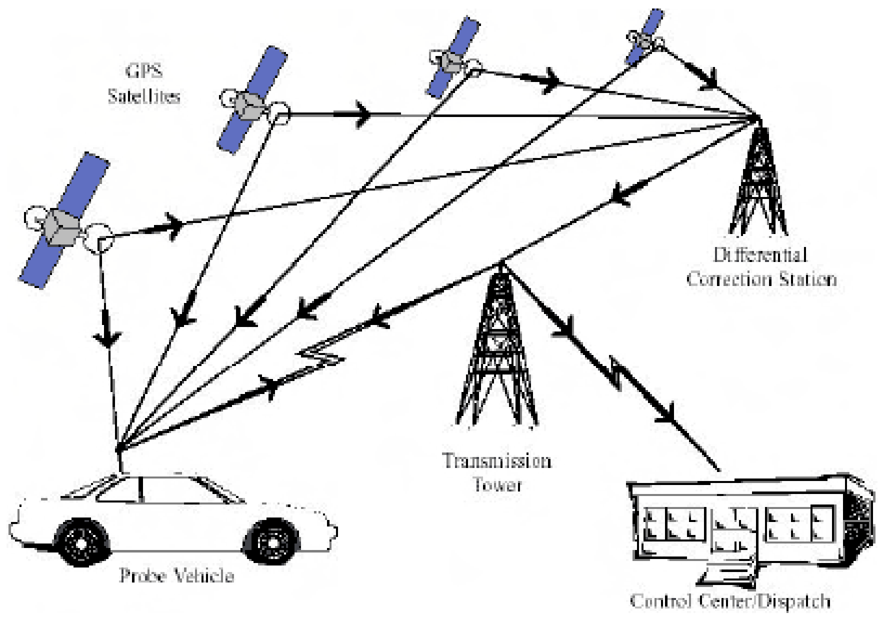


Figure 3.12: Typical Configuration for Satellite-based Probe Vehicle System (Martin 2003)

License Plate Matching—License plate matching consists of collecting license plate numbers and arrival times at checkpoints, and matching license plate numbers and arrival times between checkpoints, thus computing travel times. License plate recognition systems are composed of an imaging attainment processor, a license plate detection system, a character segmentation and recognition engine, and a computer to store the data (Figure 3.13).

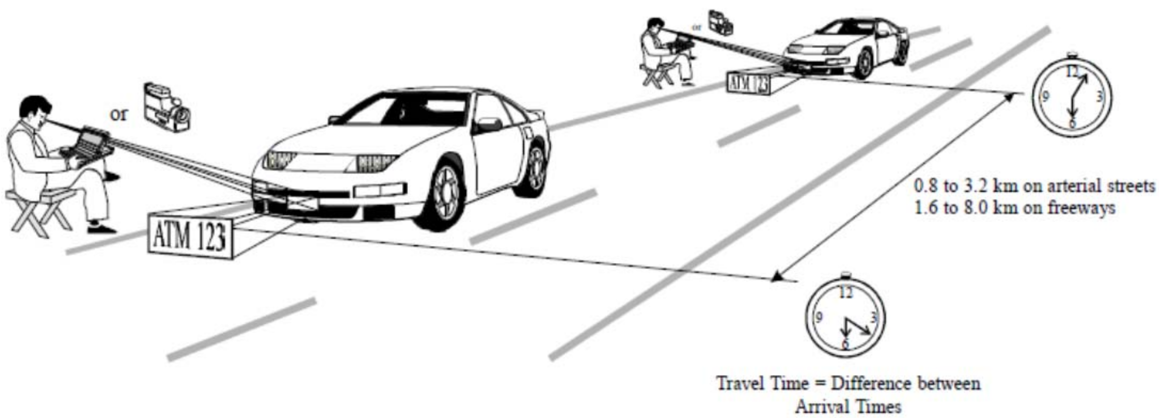


Figure 3.13: Typical Configuration for License Plate Matching (Martin 2003)

Dedicated Short-Range Communications (DSRC)—DSRC devices are being developed by the FHWA, the American Association of State Highway and Transportation Officials (AASHTO), a number of state DOTs, some of the world's largest electronic companies, toll authorities, car and truck manufacturers, and the American Trucking Association (ATA). DSRCs are medium-range communication service intended to support both public safety (e.g., collision avoidance) and licensed private operations over roadside-to-vehicle and vehicle-to-vehicle communication channels. It is anticipated that these devices will eventually be installed in all vehicles—cars and trucks—during the manufacturing process. As each DSRC device will contain a device identification number, they have been designed to replace toll tags, CVO tags, and PrePass tags. The potential exists to use these devices for collecting primary data on truck movements. For example, these devices can be used to identify the most frequently traveled truck routes in a state.

Examples of These Technologies in Use for Truck Travel Data Collection

1) McCormack and Hallenbeck (2005) of the University of Washington tested and compared the data collected with CVISN tags against GPS devices on selected routes in the Seattle/Tacoma metropolitan area. The authors concluded that both devices could be used for truck travel data collection, but also noted the advantages and disadvantages of both devices that have to be considered.
2) Greaves and Figliozzi (2008) collected one week of data from 30 trucks that had GPS devices and that operated in the Melbourne, Australia metropolitan area. Several challenges were encountered in data processing, with correctly locating trip ends, and with missing and inaccurate data. The authors, however, highlighted several uses of GPS data, such as the construction of O-D matrices by time of day, and developing trip length distributions and speed-time profiles.
3) The University of Minnesota (Liao, 2009) procured data from the Australian Transport Index database and obtained traffic data along the corridor acquired from the state DOT in an effort to assess travel characteristics along the I-94/I-90 corridor from the Twin Cities to Chicago. The study team integrated the data into a route geo-spatial database in ArcGIS, which allowed the study of truck speed variation by location and hour of day, and the identification of bottlenecks.
4) In 2010, the University of Washington (McCormack, 2010) conducted another GPS truck travel data collection study to evaluate the possibility of supporting a statewide freight network performance monitoring system that will monitor travel times and system reliability, as well as guide freight investment decisions. The research team contracted with three GPS service providers to evaluate data acquisition methods instead of partnering with trucking companies. The study team developed a data feed framework to efficiently retrieve and store a large data system. The effort involved writing software that identifies origins and destinations and creates an O-D matrix; flags errors and incomplete and external trips; categorizes trips; creates a range of trip-to-trip performance measures; and calculates sample size confidence statistics.
5) The University of Toronto (Fok and Yan, 2007) tested the use of data from a fleet-management application to develop an agent-based model for simulating goods movement throughout an urban region. A custom program was created for data processing, consisting of data cleaning, tracking vehicle GPS points, clustering trip ends into destinations, depot identification, and tour creation.

3.2 Review of Data Collection Methods used in Currently Available Freight Data Sources

In 2004, CTR published TxDOT Research Product 0-4713-P2: *State-of-the-Practice in Freight Data: A Review of Available Freight Data in the U.S.* (authored by Mani and Prozzi). This document was updated in 2010 to reflect recent changes in the data sources. The document presents a detailed evaluation of 31 publicly available and commercial databases. The survey methods, technology used, the cost of the collection method, and the reliability of the collected information for 14 of the databases are available in Appendix C. Table 3.5 provides a summary of the data collection methods of the various databases, and shows that reliable data is usually acquired when agencies are federally mandated to provide the data or some kind of formal agreement between the agencies is involved. The quality of data is, however, dependent on the accuracy of the data source and discrepancies can be minimized through adequate quality control measures.

Table 3.5: Summary of Data Collection Methods for Existing Databases

Database	Data Collection Methodology	Collecting Agency	Agencies Involved	Federally Required	Quality Control Procedure
Airport Activity Statistics of Certificated Route Air Carriers	Monthly filing of activity data	USDOT Office of Airline Statistics	Large Certificated Air Carriers	Yes	Fines
Carload Waybill Sample	Stratified sample of carload waybills reported annually by railroads	Association of American Railroads	Rail Carriers	Yes	Machine Readable Input (MRI) format reduces the non-sampling errors in data reporting.
Commodity Flow Survey	Quarterly survey a questionnaire of sampled establishments	Census Bureau	Sample of establishments from the Business Register maintained by the U.S. Census Bureau	No	To reduce the variance of estimation, establishments are sampled on the basis of a *certainty/non-certainty* design approach.
Freight Analysis Framework	Combination of establishments survey data, traffic count data, and virtual FAF O-D disaggregation	Federal Highway Administration	Multiple s sources	No	Visual inspection of data to identify discrepancies between adjacent links
Fresh Fruit and Vegetable Shipments	Combination of survey data and reported data	U.S. Department of Agriculture	Multiple sources	Both	Quality is dependent on accuracy of data reported
Grain Transportation Report	Predominately reported data	U.S. Department of Agriculture	Multiple sources	Both	Quality is dependent on accuracy of data reported

Database	Data Collection Methodology	Collecting Agency	Agencies Involved	Federally Required	Quality Control Procedure
Less Than Truckload (LTL) Commodity and Market Flow Database	Subscribing LTL carriers provide information to the database. The sample size is equal to the number of LTL carriers subscribing to the database.	American Trucking Association	LTL carriers	No—database is available only to the subscribing carriers	Not available
National Transportation Statistics	Prepared from standard data sources developed by federal government agencies, private industries and associations for each mode of transport	Bureau of Transportation Statistics	Multiple sources	Both	Quality is dependent on accuracy of data sources but annual variations in statistics for each mode is determined
Port Import/Export Reporting Service	Electronically filed or hard-copy freight vessel manifests submitted to the U.S. Customs Service	United Business Media, Inc.	U.S. Customs Service	Yes	Quality is dependent on accuracy of data reported
Quarterly Coal Report	Survey (mail surveys)	U.S. Department of Energy	Multiple sources	No	Multiple quality control procedures
Ship Movements Database	Surveys of vessels at each major port location worldwide by Lloyd's agents	Lloyd's Maritime Information Services, Inc.	Merchant vessels worldwide	No	Efficiency in data collection is ensured by the Lloyd's Agency Department
State Estimates of Truck Traffic	Vehicle counts	State Highway Agencies	Count sites	No	Vehicle classification counts from permanent count sites are collected daily and do not require any time-of-day sampling.

Database	Data Collection Methodology	Collecting Agency	Agencies Involved	Federally Required	Quality Control Procedure
Transborder Surface Freight Database	Derived from the U.S. import and export trade data developed by the U.S. Census Bureau as part of its FTS program.	U.S. Bureau of the Census *Foreign Trade Statistics*	U.S. Census Bureau for its Foreign Trade Statistics	No	Frequent analytical reviews are conducted by the BTS with the help of the U.S. Census Bureau to check for inconsistencies
TRANSEARCH	Constructed from a sample of *commercial, public,* and *proprietary* freight data sources	IHS Global Insight	Multiple	No	Employment and population statistics
Waterborne Commerce and Vessel Statistics	*Vessel Operations Reports*	U.S. Army Corps of Engineers	More than 1,500 domestic vessel operating companies	Yes	Vessel Operation Reports from all the domestic carriers ensures that the domestic waterborne shipment statistics reported in the database are accurate estimates without any sampling errors (***bias***) in the annual estimates.

3.3 Data Exchange and Data-Sharing Partnerships

Effective partnerships are needed between the government and the freight community to ensure adequate planning and funding of transportation infrastructure at the state and local levels. However, the transportation planning community struggles to understand the needs of the freight community, partly owing to the inferior freight data that are available to freight planners (Freight Stakeholders National Network, n.d.). Enhanced freight mobility through infrastructure improvements is in the interest of both the private sector and the transportation planning community. In addition, reliable freight data can be valuable to the private sector in informing investment decisions relating to equipment utilization, new markets, and business opportunities (Transportation Research Board, 2003). One untested approach is to convince a statistically representative sample of freight stakeholders in Texas to enter into a data-sharing partnership with TxDOT.

Although the option of entering into a data-sharing partnership has never been implemented, Prozzi et al. (2006) did approach a number of trucking companies that were exposed to transportation planning through their involvement with the North Central Texas Council of Government's 2004 Intermodal Freight and Safety Committee (IFS) to explore the idea. Eight trucking members of the IFS were interviewed to determine (a) whether the company would consider participating in a truck data-sharing initiative with TxDOT, and (b) what their conditions for participation would be. All the trucking companies interviewed indicated a willingness to participate in a data-sharing arrangement with TxDOT provided that certain conditions were met. The following list reiterates the conditions for participation in descending order of the number of times mentioned.

- No information about the company will be included in the aggregate database that is compiled and used by TxDOT;
- The data will not be used for law enforcement or litigation against the company;
- The Texas Motor Transportation Association will be involved to protect the interests of those that participate;
- No severe cost burden will be imposed on the trucking company in compiling the data;
- TxDOT will demonstrate to the trucking companies that the data will be used for a worthwhile purpose;
- No shipper details will be requested; and
- The trucking company will have access to the aggregated database compiled by TxDOT.

Only one trucking company indicated that TxDOT would have to compensate the company for the costs of extracting the data and providing it to the agency as a condition of participation. Given this positive response, Prozzi et al. (2006) recommended and provided guidance as to how TxDOT could proceed in establishing a data-sharing partnership with the freight community in Texas.

3.4 Conclusion

The use of intelligent transportation system (ITS) technology provides a non-intrusive approach to collecting freight data. Data that can collected include vehicle classification, weight, speed, delay data, incident data, traffic volume data, commodity/cargo type, payload (cargo) weight, truck O-D patterns, trip O-D patterns, average tour length, number of stops per tour, number of truck stops for LTL shipments, travel time, transit time, and travel time reliability.

Primary technologies can be grouped into in-roadway sensors, over-roadway sensors, and vehicle-based technologies. In-roadway sensors involve inserting the devices into the pavement and include inductive loop detectors, WIM sensor systems, piezoelectric detectors, bending plate WIM systems, and magnetic detectors. Costs for these systems vary between $600 and $300,000. Over-roadway sensors include passive infrared-detection, active infrared detection, microwave detection, ultrasonic detection, passive acoustic detection, video image detection, and commercial vehicle information systems and networks. Costs for these systems vary between $1,200 and $48,000. Use of vehicle-based technologies such as AVC and AVI devices, CVO tags, AVL and GPS systems, license plate matching, and DSRC devices is also increasing. These devices, which generally take the form of electronic tags, transmit data for multiple purposes such as toll payment, vehicle compliance checks, identification systems, and fleet tracking, and can be used as sources of freight data collection.

This study recognizes that effective partnerships are needed between government and the freight community to ensure adequate planning and funding of transportation infrastructure at the state and local levels. A review of the data collection methods of the various databases shows that reliable data is usually acquired when some form of formal agreement exists between the agencies involved.

In Chapter 4, the study team identifies the data gaps that remain between the needed freight data variables, data variables captured in current databases, and data that can be obtained using primary data collection methods.

Chapter 4. Identification of Freight Data Gaps

By comparing the freight data needs expressed during the agency workshops with the available freight data captured in current databases and data that can be obtained using primary data collection methods, data gaps can be identified. The chapter provides a detailed analysis of the identified data gaps in public, commercial, and TxDOT databases, and the potential data that can be acquired in a cost-effective manner through primary data collection methods.

4.1 Identified Freight Data Needs

The outcome of the workshops yielded this list of identified freight data needs:

- Trip O-D,
- Commodity classification,
- Shipment data,
- Mode,
- Vehicle classification and registration,
- Trip frequency and traffic count,
- Routing information and trip types,
- Accident data,
- Inventory of area businesses and trip generators,
- System infrastructure information,
- Variables relating to air quality—vehicle fleet age, engine type, vehicle type, travel speeds, and Environmental Protection Agency (EPA) Smartway partnerships,
- Traffic lanes used by trucks during trips,
- Market segmentation information,
- Service types—truckload, less-than-truckload, and just-in-time delivery,
- Trip purpose, and
- Production and attraction rates.

4.1.1 Trip Origin and Destination

A review of the existing freight databases, summarized in Table 4.1, shows that 17 databases provide some level of O-D data, as requested by the participants in the previously held workshops. All of the databases provide O-D data at the state level and 11 provide data at a more disaggregate level (i.e., county, metropolitan region, or city). Of the 11 databases that provide information at a more disaggregate level, 4 are regional-level data (Commodity Flow Survey [CFS], Freight Analysis Framework [FAF], Transearch, Carload Waybill Sample) and only 1 provides county-level information. Eight of the databases also provide information for the various ports of entry/exit. However, some of the databases provide more specific information than others. For example, the FAF database excludes smaller ports like the Port of Brownsville and Port of Orange as they may not be in the defined FAF zones.

An exception to the listed databases is the Carload Waybill Sample. This database is available in two versions—the Public Use Waybill Sample (PUWS) and the confidential Carload Waybill Sample (STB, 2012). The PUWS, which is publicly available, reports origin and destination points by Business Economic Area (BEA) and junction points, is reported by state or province. The confidential Carload Waybill sample reports data by origin and termination freight station, junction points, and rail carrier identification.

Identified Gaps

None of the databases identified provides information on in-city or zip code O-D trips. During the workshops, many participants indicated the need to have local level data for freight trip O-D. This information is required by planners, especially at the MPOs, to identify flows within their respective cities. The privately owned Transearch database provides information at the county level but the actual source of data is unknown due to proprietary reasons.

City and zip code level data can however be acquired through 1) AVI devices (includes a vehicle-mounted transponder, a roadside communication unit, and a storage and data processing unit) similar to those used in toll facilities, 2) AVL/GPS, and 3) license plate matching using video image detection (as discussed in Chapter 3). Of the above listed technologies, the use of GPS data is most commonly cited as it is unobtrusive and large historic data samples can be obtained.

Table 4.1: Available O-D Data

Database	O-D							
	Country		State	NTAR/ BEA/ CSA	County	City	Zip Code	Port Name
	Import	Export						
CFS		O	O	O				
FAF	O	O	O	O				Δ
Transearch / Reebie (Private)			O	O	O			
Carload Waybill Sample (Public Use Waybill Sample)	O	O	O	O				
Waterborne Commerce Statistics	O	O	O					O
Air Carrier Statistics (Form 41 Traffic)	O	O	O					O
Border Crossing/Entry Data	O		O					O
North America Transborder Freight Data—U.S. Transshipments through Canada or Mexico	O	O	O					O
North America Transborder Freight Data—Value to Weight Ratios by Transportation Mode and Commodity	O		O					
North America Transborder Freight Data—U.S. States Trade with Canada and Mexico through a Specific Port or in a Specific Commodity	O	O	O					
Maritime Administration (MARAD) database	O	O	O					O
USA Trade	O	O	O					O
Fresh, Fruit and Vegetables	O		O					
National Transportation Statistics (NTS)	O	O						
Annual Coal Report	O	O	O					
PIERS (private)	O	O	O					O
Texas Permitting & Routing Optimization System (TxPROS)			O					

O: Data Provided; Δ: Limited Data Provided

As documented in a Cambridge Systematics study (Beagan et al., 2007), GPS receivers in trucks can trace individual truck trip activity, which can be used in determining the origin and destination of vehicles, the roadways that are being used, trip tours, trip chaining, and trip frequencies. GPS data can be acquired directly from trucking companies or through vendors who sell GPS services to trucking companies. Data variables which can be obtained from GPS data include latitude/longitude, time/date stamp, travel direction, spot speed, and truck ID (METRANS, 2011). However, most data currently collected is for trucking company needs and not for planning purposes due to cellular costs. According to the University of Washington study[4], buying truck GPS data requires contracts and privacy agreements, which may limit the ability to distribute disaggregated data.

Integration of Data Variables

Integrating the above identified databases will require that O-D be broken down by

- Country—(Import and Export purposes),
- State,
- County,
- Business Economic Area,
- County,
- City,
  - If aggregated, City as port name (airport, marine port, terminal facility)
  - If disaggregated, actual port name (airport, marine port, terminal facility)
- Zip code and street level (if available)
  - Business/Warehouse location

4.1.2 Commodity Classification

Commodity information is used in determining industry activity and economic generators in the region, and also for forecasting studies. Based on the list of reviewed databases (see Table 4.2), the most common commodity classifications used are the Harmonized System Codes (HS Code or Harmonized Tariff Schedule), the Standard Classification of Transported Goods (SCTG), the Standard Transportation Commodity Code (STCC), and the North American Industry Classification System (NAICS). Of the 21 databases analyzed, 10 contain commodity information—some on a more disaggregate level than others (i.e., five digit in comparison to two digits).

Table 4.2: Available Commodity Data

Database	Commodity Classification				Industry Classification
	SCTG	STCC	SITC	Harmonized Tariff Schedule	
CFS	2 Digit				NAICS
FAF	2 Digit				
Transearch / Reebie (Private)		4 Digit and 5 Digit (R/W)			
Carload Waybill Sample (Public Use Waybill Sample)		2 Digit			
Waterborne Commerce Statistics			4 Digit		
North America Transborder Freight Data—U.S. Transshipments through Canada or Mexico				2 Digit	
North America Transborder Freight Data—Value to Weight Ratios by Transportation Mode and Commodity				2 Digit	
North America Transborder Freight Data—U.S. States Trade with Canada and Mexico through a Specific Port or in a Specific Commodity				2 Digit	
Motor Carrier Management Information System (MCMIS)—Census File					O
USA Trade				HS	NAICS
National Transportation Statistics (NTS)	O				
Annual Coal Report					NAICS
PIERS (private)				HS	

O: Data Provided

Identified Gaps

For majority of the databases, commodity information is only available at a national, statewide or metropolitan region level. The Transearch database is said to contain county-level information for 340 commodities (IHS, 2012). Unfortunately, commodity information cannot be directly obtained through unobtrusive primary data collection methods like GPS readings, and this information will have to be provided by the shipper or carrier. Cargo type can, however, be identified using video image detection, which can be limited in its ability to detect containerized goods.

Integration of Data Variables

In order to integrate commodity information from multiple data sources, a common classification system similar to what was developed in a previous TxDOT study (Prozzi et al., 2011) and shown in Figure 4.1 will need to be used.

	Commodity Type (STCC2)	HS
1x. Agriculture	01 - Farm Products	01 Live Animals 10 Cereals 11 Milling Products; Malt; Starch; Inulin; Wht Gluten 15 Animal Or Vegetable Fats, Oils Etc. & Waxes
	08 - Forest Products	06 Live Trees, Plants, Bulbs Etc.; Cut Flowers Etc. 09 Coffee, Tea, Mate & Spices 12 Oil Seeds Etc.; Misc Grain, Seed, Fruit, Plant Etc
	09 - Fresh Fish and Marine Products	03 Fish, Crustaceans & Aquatic Invertebrates
2x. Raw Material	10 - Metalic Ores 11 - Coal 13 - Crude Petroleum or Natural Gas 14 - Nonmetallic Minerals	26 Ores, Slag And Ash
3x. Food	20 - Food or Kindred Products	02 - Meat and Edible Meat Offal 04 Dairy Prods; Birds Eggs; Honey; Ed Animal Pr Nesoi 07 Edible Vegetables & Certain Roots & Tubers 08 Edible Fruit & Nuts; Citrus Fruit Or Melon Peel 16 Edible Preparations Of Meat, Fish, Crustaceans Etc 17 Sugars And Sugar Confectionary 18 Cocoa And Cocoa Preparations 19 Prep Cereal, Flour, Starch Or Milk; Bakers Wares 20 Prep Vegetables, Fruit, Nuts Or Other Plant Parts 21 Miscellaneous Edible Preparations 22 Beverages, Spirits And Vinegar
	21 - Tabacco Products	24 Tobacco And Manufactured Tobacco Substitutes

Figure 4.1: Sample Development of a Commodity Classification System (Prozzi et al., 2011)

4.1.3 Shipment Data

Shipment weight (tonnage) and value are the most commonly reported shipment data according to Table 4.3. Of the 21 databases reviewed, 17 provide information on shipment weight and 10 databases contain information on both shipment value and weight. Other reported data include number of the means of transportation (trucks, trains, vessels) available in two databases and number of containers (in three databases). Databases such as the Carload Waybill Sample contain additional shipment information pertaining to rail movement such as number of empties, number of railcars etc. Four other databases also provide information on ton-miles moved, which is the tonnage of shipment moved divided by the distance travelled and a measure of demand of goods in a particular area.

Table 4.3: Available Shipment Data

Database	Shipment Data				
	Weight	Value	Vehicle /Mode Count	Ton-Miles	Other
CFS	O	O		O	
FAF	O	O		O	
Transearch / Reebie (Private)	O	O	O	O	
Carload Waybill Sample (Public Use Waybill Sample)	O	O	Number of Carloads	O	Number of Loads
Waterborne Commerce Statistics	O				
Air Carrier Statistics (Form 41 Traffic)	O		O		
Border Crossing/Entry Data			Truck and Rail		Loaded/Empty Containers
North America Transborder Freight Data—U.S. Transshipments through Canada or Mexico	O	O			
North America Transborder Freight Data—Value to Weight Ratios by Transportation Mode and Commodity	O	O			
North America Transborder Freight Data—U.S. States Trade with Canada and Mexico through a Specific Port or in a Specific Commodity	O	O			
Maritime Administration (MARAD) database	O		Number of Calls		Number of Containers
MCMIS—Census File					Number of Vehicles in Fleet
Texas Crash Records Information System (CRIS)	Δ				
USA Trade	O	O			
Fresh, Fruit and Vegetables	O				
National Transportation Statistics (NTS)	O	O		O	
Annual Coal Report	O				
PIERS (private)	O	O			Container size and less than number of containers
TxPROS	O				

O: Data Provided

Integration of Data Variables

Based on the analysis of the various databases, the following variables are to be considered in the development of the integrated freight database:

- Weight,
- Value,
- Number of trucks,
- Number of trains,
- Number of rail cars,
- Number of vessels (calling at a marine port),
- Number of airlines (enplaning or deplaning at an airport),
- Number of containers (can be broken down into *empty* and *loaded*), and
- Ton-miles moved

Identified Gaps

Based on the review of shipment data, it can be inferred that most databases provide information on shipment weight and value. Ton-mile information, a measure of demand of goods in a particular area, is rare. However, this information is necessary to determine the distance travelled by carriers from their origin to their final destination. Knowledge of this type of information can advise on which infrastructure systems are being used most and the location of markets within a region or corridor.

Primary data on ton-mile information can be collected through a data sharing partnership with freight stakeholders. This approach, however, raises privacy concerns relating to the shippers' customer base, demand for a particular commodity, and market power.

4.1.4 Mode

Mode of transport is critical in identifying how goods are moved from an origin to a destination. It is also an indicator of the preferred choice of movement for a particular commodity. For example, low-value/heavy-weight commodities such as coal usually move by rail and high-value/light-weight commodities such as high-priced electronics move by air. As shown in Table 4.4, all 21 databases provide information on the mode of transport which includes truck, rail, air, pipeline, water, and multimodal. Some databases such as the CFS and North American Transborder Freight databases have information on multiple modes but others, such as the Carload Waybill Sample, Waterborne Commerce Statistics, and Air Carrier Statistics, focus on a single mode such as air, rail, and vessels.

Identified Gaps

The major challenge with mode of transport is identifying multimodal movements, i.e., movements involve two or more forms of transport. For example, according to the FAF documentation (FHWA, 2010), multiple modes and mail "includes shipments by multiple modes and by parcel delivery services, U.S. Postal Service, or couriers…and is not limited to containerized or trailer-on-flatcar shipments."

Another gap identified is the classification of "Other" or "Unknown" mode (e.g., FAF) and Free Trade Zone (FTZ) modes (e.g., TransBorder Freight Data). FAF defines "Other" or "Unknown" as "flyaway aircraft, vessels, and vehicles moving under their own power from the manufacturer to a customer and not carrying any freight, unknown, and miscellaneous other modes of transport[6]. FTZ mode as defined by the TransBorder Freight Data Program Documentation is an actual mode of transport "for a specific shipment into or out of a foreign trade zone [for which the mode] is unknown because U.S. Customs and Border Protection does not collect this information" (BTS, 2012).

Integration of Data Variables

As a result of the available reported databases, the recommended variables for mode of transport are the following:

- Truck,
- Rail,
- Air,
- Pipeline,
- Water,
- Multimodal,
- Other,
- Unknown, and
- FTZ

Table 4.5: Mode Data

Database	Mode						
	Truck	Rail	Air	Pipeline	Water	Multimodal	Other
CFS	O	O	O	O	O	O	O
FAF	O	O	O	O	O	O	
Transearch / Reebie (Private)	O	O	O		O		
Carload Waybill Sample (Public Use Waybill Sample)		O					
Waterborne Commerce Statistics					O		
Air Carrier Statistics (Form 41 Traffic)			O				
Border Crossing/Entry Data	O	O					
North America Transborder Freight Data—U.S. Transshipments through Canada or Mexico	O	O	O	O	O		FTZ, Other
North America Transborder Freight Data—Value to Weight Ratios by Transportation Mode and Commodity	O	O	O	O	O	O	FTZ, Other
North America Transborder Freight Data—U.S. States Trade with Canada and Mexico through a Specific Port or in a Specific Commodity	O	O	O	O	O		FTZ, Other
Maritime Administration (MARAD) database					O		
TxDOT's TLOG	O	O					
MCMIS—Census File	O						
Texas Crash Records Information System (CRIS)	O						
USA Trade			O		O		Containers
Fresh, Fruit and Vegetables	O	O	O		O		Piggyback
National Transportation Statistics (NTS)	O	O	O	O	O		
Annual Coal Report	O	O		O	O		
PIERS (private)					O		Intermodal
TxPROS	O						
Railroad Accident/Incident Reporting		O					

O: Data Provided

4.1.5 Vehicle Classification and Registration

Vehicle classification information for freight movement such as vehicle type (e.g., light-duty truck, heavy-duty truck, delivery truck, semi-trailer, etc.) and vehicle size and weight information are difficult to come by. This information is useful in roadway geometry and pavement design. As shown in Table 4.5, the TxDOT TLOG data, Motor Carrier Management

Information System (MCMIS) Census and Crash Files, Texas Crash Records Information System (CRIS), and the and Vehicle OS/OW databases contain some form of this information.

The TLOG data contains percent AADT information for single unit and combo unit vehicles. This information is collected through traffic counts on various on-system roadway segments. This database, however, lacks the ability to differentiate between the different vehicle types as the data is collected based on the vehicles driving over a traffic count device such as a pneumatic road tubes, or piezo-electric sensors and inductive loops embedded in the roadway.

For OS/OW vehicles, one great resource is the Texas Permitting & Routing Optimization System (TxPROS), which is a GIS-based mapping system that provides routing and permitting information for OS/OW vehicles. Using the database, the number of OS/OW vehicles, routes used, trip O-D, frequency of trips, and types of loads can be queried. Vehicle classification information, such as vehicle axle weight of the vehicle being used in the movement, is also available.

Identified Gaps

As discussed earlier, the majority of the databases do not contain information on vehicle classification or registration except for the four listed in Table 4.5. However, these databases do have their limitations. For example, the TxPROS database captures only data provided by roadway users seeking routing permits for OS/OW vehicles. The CRIS and MCMIS Crash File databases include only data from accident reports, and the TLOG file contains only two vehicle classifications (Single Unit and Combo Unit vehicles) instead of the available seven FHWA classifications for trucks (FHWA, 2011a).

Through the use of technologies such as video image detection, AVC devices, and AVI devices, it is possible to collect primary vehicle classification data. However, the cost of installing some of these devices may be prohibitive and companies may not be willing to share this information because of privacy concerns.

In addition, through TxDOT's Vehicle Titles and Registration (VTR) Division, data on vehicle registration can be obtained. However, this information may have to be anonymized to address privacy concerns. The VTR data may also be limited in its scope as it will not account for out-of-state vehicles and it is impossible to determine which routes are actually being used by the vehicles.

Table 4.6: Vehicle Classification Data

Database	Vehicle Classification	
	Vehicle Type	OS/OW
TLOG	O	
MCMIS –Crash File	Δ	
Texas CRIS	O	
TxPROS		O

O: Data Provided; Δ: Limited Data Provided

Integration of Data Variables

Based on information available in the above reviewed databases, the list of variables that need to be considered in the integrated freight database are the following:

- Vehicle type (e.g.)
  - Four or more axle single-unit trucks
  - Four or fewer axle single-trailer trucks
  - Five-axle single-trailer trucks
  - Six or more axle single-trailer trucks
  - Five or fewer axle multi-trailer trucks
  - Six-axle multi-trailer trucks
  - Seven or more axle multi-trailer trucks
- Vehicle axle weight, and
- OS/OW criteria

4.1.6 Trip Frequency and Traffic Count

Trip frequency and traffic count variables seek to measure how often a facility or infrastructure is used. The goal of this data is to determine which and when facilities are most used. The common variations proposed by stakeholders include seasonal, weekday, and time of day trips. Unfortunately, this information is not reported by any of the databases (see Table 4.6).

Traffic count information, such as the number of vehicles using a roadway, is reported annually in the TLOG databases and the number of trucks is computed as a percentage of the AADT. In addition, databases such as the North America Transborder Freight Data, the Carload Waybill Sample (Public Use Waybill Sample), and the Waterborne Commerce databases report information on the number of trucks, trains, vessels and aircrafts that either arrive or depart a port of entry/exit on a monthly or annual basis.

Identified Gaps

As discussed earlier, only the TLOG database accounts for annual traffic volumes on a route. None of the other databases provide information on the frequency of trips except for traffic counts at ports of entry/exit. The TxPROS database can be used in determining the number of trips made by OS/OW vehicles on a route but this will not be representative of all freight vehicles. The confidential version of the Carload Waybill Sample contains information on the number of trains using a particular route but this information is not publicly accessible.

Integration of Data Variables

Variables that can be included in an integrated freight database to capture trip frequency and traffic count information include:

- AADT—truck and rail (route specific),
- Average annual rail cars (route specific),

- Annual port of entry/exit traffic,
- Average weekday traffic,
- Time of day traffic, and
- Monthly traffic

Table 4.7: Trip Frequency and Traffic Count Data

Database	Time				
	Seasonal	Weekday	Time of Day	Traffic Count (AADT, AADRC, etc.)	Frequency of Data Collection
CFS					1997, 2002 & 2007
FAF					2007
Carload Waybill Sample (Public Use Waybill Sample)				Number of Carloads	Yearly
Waterborne Commerce Statistics					Yearly
Air Carrier Statistics (Form 41 Traffic)				O	Monthly
Border Crossing/Entry Data				O	Monthly, since 1995
North America Transborder Freight Data—U.S. Transshipments through Canada or Mexico				O	Annual
North America Transborder Freight Data—Value to Weight Ratios by Transportation Mode and Commodity				O	Monthly, since 2004
North America Transborder Freight Data—U.S. States Trade with Canada and Mexico through a Specific Port or in a Specific Commodity				O	Monthly, since 1994
Maritime Administration (MARAD) database				O	Annual
TLOG				O	Annual
Texas CRIS			O		Daily
PIERS (private)		O			Weekly
TxPROS		O			Daily

O: Data Provided

4.1.7 Routing Information and Trip Types

Vehicle routing information includes variables such as trip length and routes used. While six of the reviewed freight databases contained this data (see Table 4.7), only two databases are able to provide the data at a county or BEA level: the confidential Carload Waybill Sample and the FAF. The confidential Carload Waybill contains information on the actual routes being used and the stations on that route, while the Public Use file does not contain any route or station information. The FAF uses traffic assignment models to estimate the flow of traffic on the national highway network (FHWA, 2010). The methodology excludes intra-zonal truck movements (local traffic) and thus cannot be used for MPO planning purposes. Data from the FAF, however, provides a general estimate of freight flows on the major highway networks such as the interstate roadways.

Trip type tries to capture whether a trip was within a defined region (internal), between a region and another region (internal/external), or through traffic. This information is difficult to obtain unless directly reported by the industry and none of the reviewed databases contained this information.

Table 4.8: Routing Information

Database	Routing Information		
	Algorithm Used	Level of Detail	Trip Length
CFS	Best/Quickest Path	Mostly Interstates	O
FAF		Highway	
Transearch / Reebie (Private)		Highway	
Carload Waybill Sample (Public Use Waybill Sample)		Interchange State	
Waterborne Commerce Statistics		Δ	
TLOG		Interstate	
Texas CRIS		Highway	

O: Data Provided; Δ: Limited Data Provided

Identified Gaps

Aside from the confidential Carload Waybill, no other database proves county- or city-level information on routes used by the freight industry. Through the use of traffic assignment models such as the one developed in the FAF, it is possible to estimate freight flows on a roadway network but this will also require accurate O-D data to be provided by shippers and carriers.

Integration of Data Variables

Variables necessary to capture routing information and trip type include the following:

- trip length,
- routes used, and
- location of stops.

4.1.8 Accident data

Accident data reports for truck and rail movements are available in the MCMIS Crash Count report, TxDOT's CRIS, and the Federal Railroad Authority's (FRA) rail accident database (see Table 4.8). The MCMIS and CRIS databases provide information such as the location of the accident and the vehicles involved. This database can be further queried to determine the frequency and fatality of accidents. The publicly available FRA Railroad Accident/Incident Reporting database also provides information on accidents by county, track class, track type, and primary cause of accident. Detailed rail line information may also be available but not publicly accessible.

Table 4.9: Accident Data

Database	Safety
MCMIS—Census File	Crash Reports
Texas CRIS	Crash Reports
Railroad Accident/Incident Reporting	Crash Reports

Integration of Data Variables

Variables that need to be considered for accident data include

- Location (region, state, county, city, street/roadway/rail line)
- Type of roadway
- Type of accident
- Primary cause of accident
- Number of fatalities, and
- Time frame (year, month, week, day)

4.1.9 Inventory of Area Businesses and Trip Generators

An inventory of area businesses can provide information on traffic generators as requested by the participants of the workshops. The publicly available SOCRATES, U.S. Census Bureaus' County Business Patterns, and MCMIS Census File databases provide information on the types and location of industries within a zip code. This information can be used in estimating traffic generators, which is useful for transportation and land use planning. Actual freight generation information is still required from the industry stakeholders to determine realistic trip generation figures.

Integration of Data Variables

Data variables that can be obtained from SOCRATES, U.S. Census Bureaus' County Business Patterns, and MCMIS Census File databases include the following:

- Industry description (e.g., agriculture, forestry, fishing and hunting, mining, quarrying, and oil and gas extraction, Transportation and warehousing, etc.)
- Number of employees,
- Number of establishments, and
- Zip code

4.1.10 System Infrastructure Information

System infrastructure information provides a general overview of the transportation system, i.e., capacity, geometry, grades, number of lanes/tracks, location of intersections/overpasses, railroad crossings, roadway intersection turning radii, etc. This information is usually available from the agencies responsible for the infrastructure, i.e., TxDOT, counties, cities, port authorities, and railroad companies. In order to meet the needs of the freight data users, a GIS database that integrates all the data from the various planning agencies will have to be developed. The database can assist in determining capacity needs and problem spots in the system. TxDOT's Statewide Planning Map (TxDOT, 2012) is an example of a GIS database that incorporates information of all roadways for which TxDOT is responsible. It contains information such as traffic flows and roadway functional class. When used with the TLOG data (using the commonly identified control section information), it is possible to determine a roadway's number of lanes, traffic count, and other traffic data. A single database that also includes information on the location of rail grade crossings and other scarcely documented data will provide a robust infrastructure catalog of freight data.

4.2 Data Needs for Which Data Was Not Found

For a few of the requested freight data needs, no existing freight database provides the requested information, and primary data collection efforts will need to be undertaken to acquire such information. Such needs include the following:

- Variables relating to air quality—vehicle fleet age, engine type, vehicle type, roadway speeds, and EPA Smartway partnerships,
- Traffic lanes used by trucks during trips,
- Market segmentation information,
- Service types—truckload, less-than-truckload, and just-in-time delivery,
- Trip purpose, and
- Production and attraction rates.

4.2.1 Sample Graph Illustrating Data Gaps

This section demonstrates the challenges users will face when running queries with the current available freight data. Figure 4.2 describes an example of data gaps identified during integration of Freight Analysis Framework 3 (FAF3), CFS, and Carload Waybill Sample (CWS) databases into the Integrated Freight Transportation Database (IFD). Because FAF3 is the most popular data source and covers relatively more O-D zones, commodity, and modes, this example defines uses FAF3 as the default data source. FAF3, however, does not include hazardous materials data that CFS has. FAF3 and CFS's hazardous materials data can be easily matched because FAF3 is developed based on CFS. Another disconnect identified between the databases is that FAF3 is updated every 5 years but CWS is updated annually. However, should the data be integrated, it is possible that the reliability of FAF3's rail data will be much more improved.

Also identified was that FAF3 uses a two-digit SCTG commodity code and that of the CWS uses a two-digit STCC system, but they can be fairly matched. However, some data gaps were found in case of O-D zones. For example, zones 487 and 489 of FAF3 (Laredo and Remainder of Texas, respectively) cannot be matched to any of the BEA zones found in the CWS (see Figure 4.2). This data gap, therefore, needs to be examined before integrating the data sources.

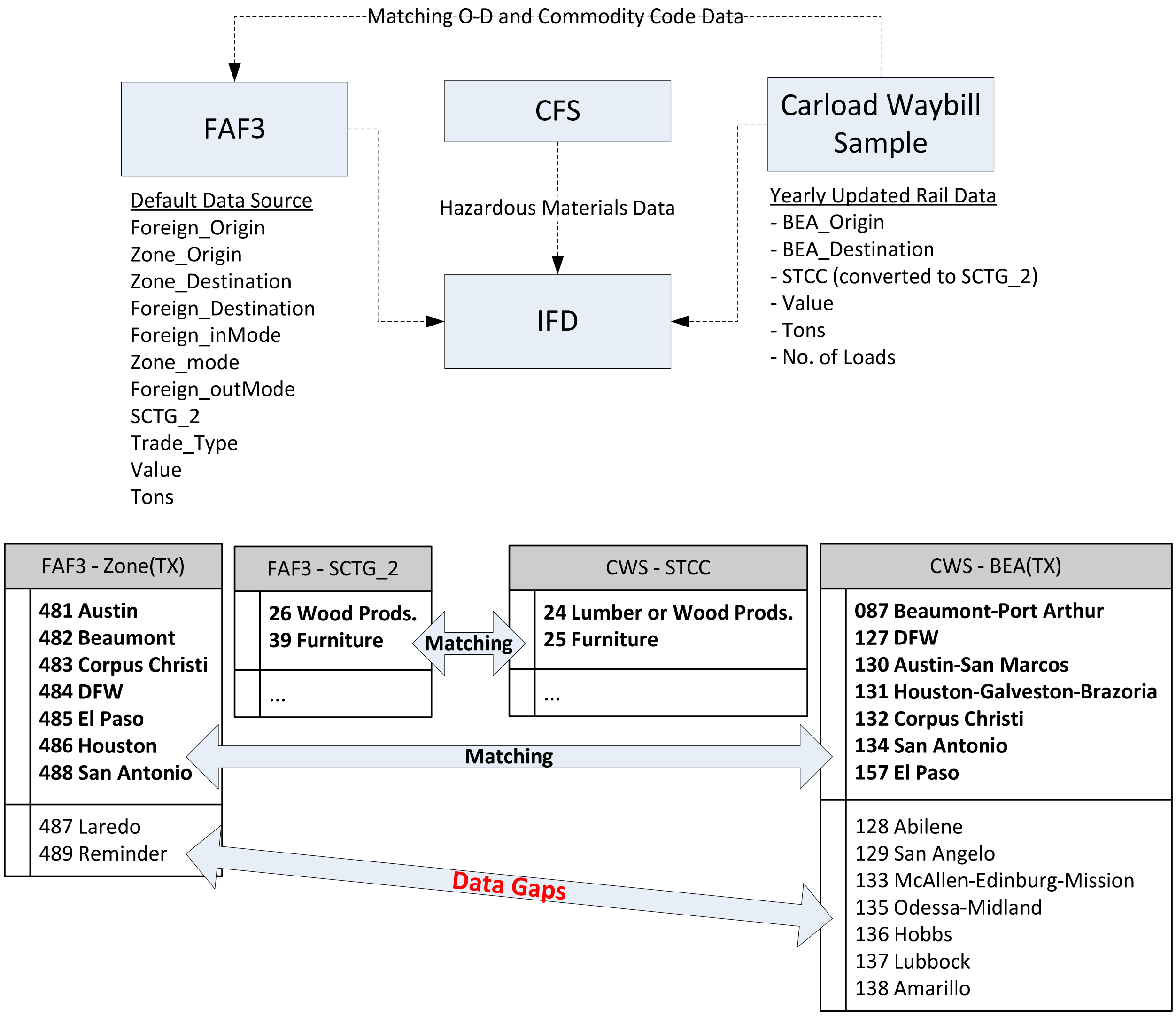


Figure 4.2: An Example of Data Gaps

An example scenario demonstrating the data gaps is illustrated in Figure 4.3, and additional scenarios can be found in Appendix D of this report.

> ***Example Scenario*:** How many tons of wood products were moved from Austin to Dallas by trucks using IH-35 in 2007?
>
> As illustrated in Figure 4.3, users will try to select "Input" from IFD, which includes several representative "Data Sources." The "mediators" then select the "Default Data Source" considering the factors from the "User Input". The "Default Data Source" here is CFS but the IFD cannot show the "System Output" result due to the "Route" specified by the user because none of data sources provide that level of information. This example scenario illustrates the data gap in the specified databases.

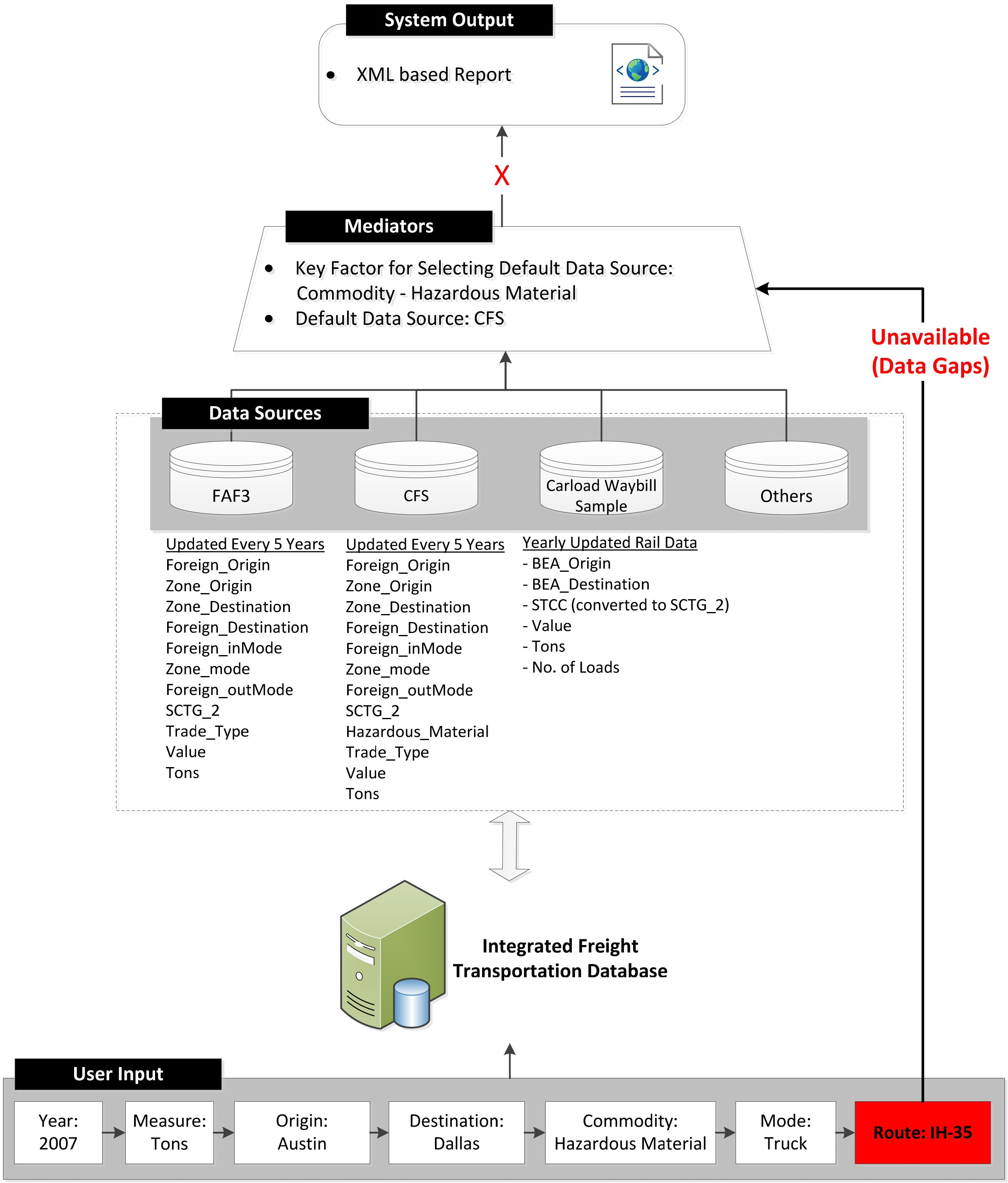


Figure 4.3: Scenario 1

4.3 Conclusion

A review of all the databases determined the gaps and the challenges of filling those gaps. Mode of transport information is the most readily available in most of the databases. However, the O-D points are insufficient to meet the needs of TxDOT. For roadway movements, an overwhelming number of databases (19) contain information at either the state or regional area level; however, none of the databases contain information at a city or zip code level—the data most desired by participants at the freight workshops. This information is required by planners, especially at the MPOs, to identify freights flows within their respective cities. Eight databases contain information on the port of entry/exits but do not provide much information about actual freight movement on the roadway infrastructure. This gap translates into scarce data for traffic generators, vehicle routing information, trip frequencies, and commodity flows at the city or zip code level.

For rail, air, and vessel movements, system-wide routing data is currently being collected but may be protected because of privacy concerns. For example, the Carload Waybill Sample comes in two versions: the Public Use Waybill Sample (PUWS) and the confidential Carload Waybill Sample. The PUWS, which is publicly available, reports O-D points by Business Economic Area but the confidential Carload Waybill Sample—which is more detailed but restricted—reports data by origin and termination freight station, junction points, and rail carrier identification.

Through the use of advanced data integration methods, it is possible to overlay some of these databases on top of each other to assist in filling some of the data gaps. As noted in a recent National Research Council report, "when separate datasets are collected and analyzed in such a manner that they may be used together, the value of the resulting information and the efficiency of obtaining it can be greatly enhanced" (NRC, 2001). For example, publicly available SOCRATES, U.S. Census Bureaus' County Business Patterns, and MCMIS Census File databases provide information on the types and location of industries within a zip code. This information can be used in estimating traffic generators. In addition, the TLOG file can be used in determining average annual daily truck traffic on various roadways, and the CRIS database can provide information such as the frequency of truck-related accidents on various roadways. This integrated approach can assist in filling some of the data gaps but not all of it. There is still a need for industry participation to provide sufficient data for needs such as variables relating to air quality (e.g., vehicle fleet age, engine type, vehicle type, roadway speeds), service types (e.g., truckload, less-than-truckload, and just-in-time delivery), trip purpose, and actual production and attraction rates.

Chapter 5 summarizes information garnered by the study team after conducting a series of key person interviews with representatives of Texas's freight community to determine their willingness to entering into a data sharing partnership with TxDOT and provide the study team with sample data.

Chapter 5. Data-Sharing Partnerships

Traditional primary data collection methods such as roadside intercept surveys, mail-out/mail-back questionnaires, combined telephone-mail-out/mail back questionnaires, driver trip diaries, personal interviews, etc., when done correctly, are regarded as very reliable for obtaining freight data. However, these methods can be costly and generally involve a smaller, more select or targeted sample that may not be appropriate for the population of statewide freight movement databases.

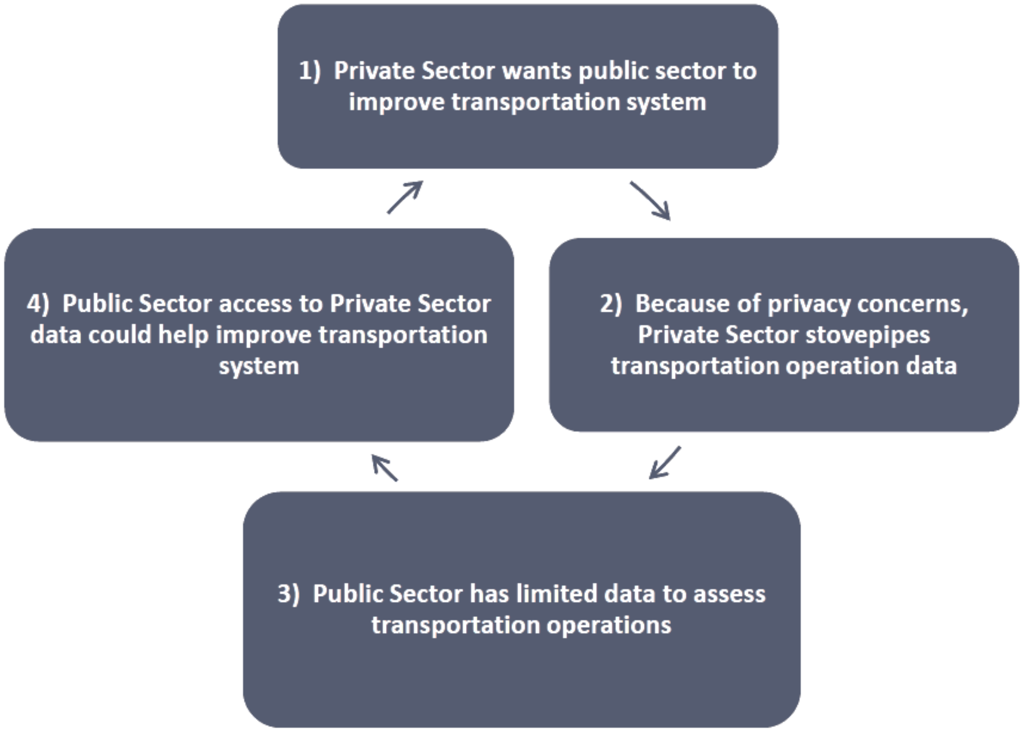
A largely untested option to fill the data gaps identified in the earlier chapter is to enter into effective partnerships with the freight community to ensure adequate and accurate freight data, as proposed and discussed by Prozzi et al. (2006). This option is based on the hypothesis that a statistically representative sample of Texas freight stakeholders (i.e., shippers, receives, trucking companies, rail carriers, ports, airports, inland port, logistics companies, freight forwarders, brokers, etc.) can be convinced to enter into a data-sharing partnership with TxDOT.

The research team developed an interview instrument and performed key person interviews with executive-level managers at trucking companies, shippers, airports, freight forwarders, and logistics companies with the objectives:

- relay the objectives of the study;
- establish the key person's interest in entering into a data-sharing partnership with TxDOT;
- establish their conditions for participation in a data-sharing partnership;
- determine which data variables will be shared with TxDOT at what level of detail;
- establish an understanding about receiving and using possibly proprietary information; and
- request a small sample of data that could be used by the study team to develop a strategy for data collection and integration.

5.1 Key Person Interviews with Representatives of Texas's Freight Community

Figure 5.1, taken from the National Cooperative Freight Research Program (NCFRP) study titled *Freight Data Sharing Guidebook,* illustrates the interaction between public and private agents and their interests for data exchange is demonstrated. In the first instance, the private sector wishes the public sector could improve the transportation system to allow them to become more competitive. In the second instance, the private sector stakeholders are reticent to share data because of privacy concerns. The third and fourth instances show that public sector planners need private sector operations data to better plan for freight infrastructure.



Source: TRB, 2013

Figure 5.1: Establishing the Need for Private Freight Stakeholder Data

Similar observations from the first and second instances in Figure 5.1 were made by the study team in an online survey (see Appendix F). Of the 32 companies that participated in the survey, when asked what benefits they will seek from a data-sharing partnership, 20 companies answered that addressing current and anticipated transportation issues such as congestion and capacity constraints will be beneficial and 19 companies cited that improving roadway safety and public education on freight will be beneficial. However, 22 cited the fear of data being mishandled or being improperly used. If guaranteed that the information will never become public or shared with others, 28 companies cited that they will be more willing to enter into a data-sharing partnership.

Based on these observations, it can be inferred that both public and private sector agencies recognize that effective partnerships are needed to ensure adequate planning and funding of transportation infrastructure at the state and local levels. Despite this recognition, the private sector is reluctant to share data because of privacy concerns, and the transportation planning community continues to struggle to understand the needs of the freight community. Enhanced freight mobility through infrastructure improvements is in the interest of both the private sector and the transportation planning community. In addition, reliable freight data can be valuable to the private sector in informing investment decisions relating to equipment utilization, new markets, and business opportunities (TRB, 2003).

With knowledge of the concerns of the private sector, the CTR study team moved forward with approaching the freight community to establish data-sharing partnerships. This option was analyzed under the hypothesis that a statistically representative sample of Texas freight stakeholders (i.e., shippers, receivers, trucking companies, rail carriers, ports, airports, inland port, logistics companies, freight forwarders, brokers, etc.) can be convinced to enter into a data-sharing partnership with TxDOT. Initial efforts for the development of this task entailed three approaches:

- *Approach #1* consisted of working with the CTR contacts and other interested parties to gather suggestions on how to best approach the freight community. The following subsections (5.1–5.4) summarize these efforts.
- *Approach #2* consisted of contacting targeted freight stakeholders in trucking or shipper companies. Section 5.5 summarizes these efforts.
- *Approach #3* consisted of cold calling to start building a network of connections within the freight community. Section 5.6 summarizes these efforts.

5.2 Development of Bilingual Brochure and Interview Instrument

Before approaching stakeholders, the study team developed a brochure relaying the objectives and benefits of the projects in English and Spanish (see Appendix E). In addition, the study team developed phone and in-person interview questions.

5.3 Development of a Survey Instrument

An initial survey was first developed and discussed with representatives of the Texas Trucking Association (TXTA) and the Council of Supply Chain Management Professionals (CSCMP). The survey included the following final questions:

- What benefits do you seek should you enter into a data-sharing partnership with TxDOT?
- What are your main concerns for entering into a data-sharing partnership?
- If guaranteed that the information will never become public or shared with others, will you be more willing to enter into a data-sharing partnership with TxDOT?
- If guaranteed that the information will never become public or shared with others, will you be more willing to enter into a data-sharing partnership with TxDOT?
- If you do agree to a data-sharing partnership, how often will you be willing to share data with TxDOT without additional effort from your end?
- What kind of data are you willing to share?
- In what format are you willing to provide the data?

5.3.1 Meetings with TXTA

TXTA is a trade organization dedicated to advocating policies for, leading training activities for, and providing information on safe, dependable, and cost-effective motor transportation systems. In July 2012, the study team initially met with John Esparza and Les Findeisen to discuss the objectives of the study and request their support and recommendations. At that time, they provided only a few recommendations on how the team should lead this effort. Subsequently, in November 2012, a second meeting took place and TXTA officials agreed to provide support to the study team in distributing information about the study, and inviting the team to relevant events to better approach trucking and freight stakeholders.

5.3.2 Meeting with CSCMP

The CSCMP is a worldwide professional association dedicated to the advancement and dissemination of research on and knowledge of supply chain management. An initial meeting with the CSCMP's Texas representative, Mr. Richard Sherman, took place in November 2012. A contact list of Texas companies and professionals in the freight industry was shared with the study team.

5.3.3 Successful Contact with Other Key Stakeholders

During the months of October/November 2012, the study team contacted the following stakeholders:

- *Dallas-Fort Worth International Airport*: A conference call with Mr. Crites and Naveen Bandla took place in November during which these stakeholders were briefed on the project objectives of the project. They showed willingness to participate and stated they have some information about freight movements going in and out the airport.
- *Carload Waybill Sample*: No further action to involve rail stakeholders was taken, as the study team secured a copy of the restricted version of the Carload Waybill Sample to be used solely for this project and other research purposes.
- *PrePass-Help INC.*: Initial contact was established with Mr. Tommy Holst to relay the study objectives and request support. A contact list of freight stakeholders and freight-related officials was shared with the study team.

5.3.4 Other Efforts

From October to December 2012, the study team also performed the following outreach efforts to gauge freight stakeholders' willingness to enter into a data partnership and obtain data samples:

- *Approach #1:* The Port of Brownsville, Ryder México, Wal-Mart, the Southwest Maquila Association, and Foreign Trade Association were contacted.
- *Approach #2:* The study team reached out to 33 "targeted" stakeholders through contact lists of Texas logistics and freight officials of trucking and shipping companies with operations in Texas. Stakeholders were sent the brochure and survey instrument; a member of the study team followed up via both telephone and a follow-up email.
- *Approach #3:* The study team started to compile information regarding trucking companies from lists such as FleetSeek, the Texas Workforce Commission, Texas Vehicle Motor Carrier Registration database, and other websites. Team members called 151 companies, seeking to get channeled to the right official. Once a logistics or transport official had been reached, a study team member sent an initial email, including the brochure and survey, and then followed up twice.

5.3.5 Initial Results

In December 2012, the study team met and reported to TxDOT its initial results:

- *Survey:* Preliminary survey results involved only nine responses. Fear of data being mishandled was highlighted as the main concern about starting a data partnership conversation. However, most respondents also mentioned that if the information was guaranteed to never become public or shared with others, they might be willing to consider entering into a data-sharing partnership.
- *Samples:* The study team was starting to build relationships with some freight stakeholders but they were unwilling to provide data samples.
- *Feedback*: Freight stakeholders indicated they could not visualize or "get a feel" of what the direct benefits would be.

The study team suggested to the project sponsor that the project required additional time and different strategies to achieve the desired results and obtain data samples. The project sponsor granted permission for the study team to continue these efforts through May 2013.

5.4 Develop Strategy for Collecting and Compiling Primary Freight Data from Texas's Freight Stakeholders

From December 2012 to May 2013, various efforts enabled the study team to enhance the promotion of Project 0-6697 and obtain four data samples from freight stakeholders, as described in this section.

5.4.1 TXTA's Support

TXTA supported the study team's outreach efforts by publishing in its weekly newsletter the survey link, and by inviting team members to TXTA events, such as Trucking Day at the Capitol and related activities.

By January 2013, the team had obtained 32 responses to the final survey (see Appendix B), summarized here:

- Most stakeholders considered that the partnership would have beneficial outcomes, including (*i*) addressing any current or anticipated transportation issues such as congestion and capacity constraints, or (*ii*) improving roadway safety and public education on freight issues.
- The major stakeholder concerns were (*i*) fear of data being mishandled or improperly used, (*ii*) burden to prepare the data, and (*iii*) fear of government involvement and law enforcement measures.
- If guaranteed that the information would never become public, 88% of respondents would be willing to participate in a data-sharing partnership.
- None of the respondents were currently participating in a data-sharing partnership.
- The data variables that could be potentially shared involve trip O-D, number of trips, vehicle type, load type, and route preference, among others.

- Most stakeholders would be willing to provide to share the data either through electronic submissions via email or via a website.

In addition, the study team participated in TXTA events, such as Trucking Day at the Capitol, where they had in-person and direct access to freight stakeholders to promote the project.

5.4.2 Further Trust-Building Efforts

Aside from survey results, the interviewees listed in Table 5.1 not only dedicated time to the study team but also provided feedback in guiding the team's second outreach effort to obtain freight data samples. These conversations and interviews provided the study team the following feedback and recommendations:

- *A clear non-disclosure contract is needed*: a written contract would reinforce a sense of trust with freight partners. A sample contract was sent to TxDOT officials for review in April and May 2013.
- *IT burden is a key disincentive for participation*: Some stakeholders' IT burden would prevent them from participating. The study team attempted to lighten the IT burden by requesting data samples in CSV formats or as Excel documents. Additionally, the study team has offered to "clean" the data, to lighten the load on freight partners. Regardless of these efforts, the firms perceive the IT component of this project as a considerable burden.
- *Certain private consulting firms are collecting similar data:* However, the data is only available at a considerable cost. In addition, in most cases the release of the data is not possible—the only way of consulting it would be through paid searches for specific issues, an approach that diverges from this project's goals.
- *Participation—secret for some, open for others:* Some stakeholders expressed preference for their participation to remain anonymous; others inquired about the possibilities of advertising their participation in such data-sharing partnership.

Table 5.1: Interviewees that Provided Feedback through Telephone or In-Person Efforts

Association /Company / Agency	First Contact's Name
Ainsworth Trucking	David Ainsworth
American Trucking Associations	Benton Landers
AFN (Illinois)	Marcus Weiss
Cannonball Trucking	Regan Eubanks
Central Freight Lines	Kris Ikijeri
Delcan	Rosie Wilson
Empire Truck Lines	Dwight Jennings
Enbridge Liquids Transportation	Michelle Lawrence
FEDEX	David Short
Fikes Truck Line	Krisha Cronin
Grocery Direct Consolidated	Kevin West
Lake Truck Lines	Douglas Cain
NFI	John Salomon
Niagara Bottling	Brian Reed
Seaboard Foods	Joe Goodwin
Stevens Transport	Todd Aaron
Texas Motor Transportation Association	John Esparza
TNT Crane Rigging	Alan Riddick
UPS	Danny Smith
Vantix Logistics	Charles Bostick
Veros Consulting	Rick Romero
Wal-Mart	Gary Brida
Western Distributing Transport	Robert Etchells
WM Dewey	Bob McDowell

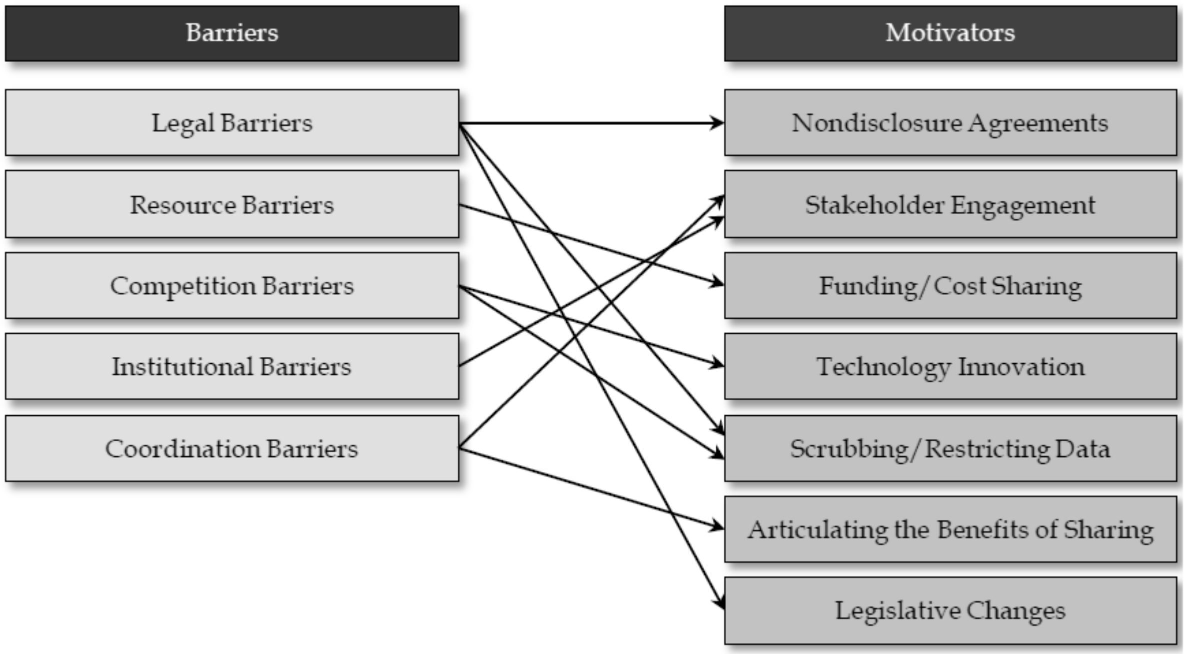
5.4.3 A Strategy for Success

The study team found that the Transportation Research Board was carrying out research through its National Cooperative Freight Research Program (NCFRP) on establishing guidelines for approaching the freight community to establish data partnerships. The study, *NCFRP Report 25*, was officially published on June 4, 2013[iii]. However, through multiple efforts, the study team was able to get hold of an unofficial draft copy in January 2013.

NCFRP Report 25 explores and addresses some of the barriers to freight data sharing between the private and public sectors (Figure 5.2), including the need to share data with academic researchers. The report establishes as a premise that "planners are only as effective as

[iii] Please see http://www.trb.org/Publications/Blurbs/169010.aspx.

the quality of their updated information on movement needs." However, data is guarded by firms in the private sector, because it affects their competitive edge in the marketplace.



Source: TRB, 2013

Figure 5.2: Barriers and Motivators to Freight Data Sharing

NCFRP Report 25 was designed to bridge the gap between these two viewpoints, and provide recommendations for effective ways to establish such partnerships. To better navigate complex interactions and hindrances to partnerships, the report reorganized these issues into "barriers" and "motivators." Barriers are the existing challenges to partnerships (i.e., legal, resource, competition, institutional, and coordination) and motivators are potential steps to alleviate those barriers (i.e., non-disclosure agreements, stakeholder engagement efforts, funding and costing sharing, technology innovation, scrubbing data, showing benefits, and legislative/policy changes). Figure 5.2 shows the interaction between both.

The *NCFRP Report 25* report provides 28 guidelines *NCFRP Report 25*, inspiring the study team to create a point-by-point skeleton of the major issues affecting CTR's outreach effort. Table 5.2 maps the NCFRP guidelines to the study team's efforts.

Table 5.2: *NCFRP Report 25* Guidelines and Summary of Related CTR Efforts

	Freight Data Sharing Guideline	**CTR Study Team Action**
1	*Identify sources of freight data via literature search and review of past research*	The study team conducted a literature review and compiled information about all free freight databases comparing their components and level of disaggregation.
	Guidelines Related to Non-Restricted Data	
2	*Use non-restricted or open source data if available*	The study team collected lists of dozens of available databases, attempting to compile all readily available, smaller, more limited public sources.
3	*Utilize non-intrusive technologies for data collection that do not require sharing agreements*	The study team examined non-intrusive data collection technologies currently available to TxDOT such as infrared detection, video image detection, passive acoustic detection, and ultrasonic detection.
	Guidelines to Address Privacy Concerns	
4	*If unrestricted data is not enough, be aware that privacy concerns must be addressed*	The study team ensured that data would remain private and showed willingness to find adaptable solutions for each freight stakeholder.
5	*A non-disclosure agreement can be a good tool to support a data-sharing arrangement*	Most stakeholders were not willing to start a conversation without legal protections.
6	*A stable contracting relationship with data providers can be very helpful in successful data sharing*	The study team established personal rapport with freight stakeholders. The team is continuing ongoing efforts to build successful relationships on a long-term basis.

	Freight Data Sharing Guideline	**CTR Study Team Action**
7	*A less formal agreement to maintain confidentiality of private sector data may be sufficient*	The study team discussed confidentiality concerns from the very beginning with potential partners; many were trusting of The University of Texas brand to keep their data secure. However, only three samples were obtained under this "less formal" agreement. Thus, the study team will start sending the non-disclosure agreement template to interested stakeholders in order to start the negotiations on "comfortable grounds."
8	*Begin negotiations of disclosure and use restrictions on freight data as early in the process as possible*	*Idem.*
9	*Public agencies desiring to obtain data from private companies may need to research Freedom of Information Act (FOIA) laws*	The study team researched applicable FOIA laws and incorporated these into the template disclosure agreement.
10	*Consider seeking enabling legislation and public agency policy approaches to support data sharing and protect the data*	N/A (Currently, the efforts are merely preliminary, and the legislative session has passed. This could be an item for the next legislative session.)
	Guidelines for Scrubbing or Restricting Access to Freight Data	
11	*Consider the use of software and database tools to protect and access freight data by removing private or competitive information*	The study team assured partner firms that data would be aggregated and their identifying information would be scrubbed from the database.
12	*Build access restrictions into the data set as an alternative to scrubbing*	In addition to data scrubbing, the study team assured partner firms that data would be restricted to users based on a hierarchal system.

	Freight Data Sharing Guideline	CTR Study Team Action
	Guidelines for Stakeholder Engagement	
13	*Place a high priority on coordination and devote the needed resources to extensive coordination with public and private stakeholders*	The study team focused on coordination when reaching out to stakeholders, relying primarily on two team members who orchestrated efforts to avoid duplication of efforts.
14	*Consider the use of trusted third parties (associations, consultants, or academics) as intermediaries or data analysts*	TxDOT, the project sponsor, chose the study team, a neutral third party, to carry out this important outreach effort. Data partners were more open to discussions when informed that CTR would be the guardian of the data.
15	*Investigate possible partnership with trade associations to facilitate data sharing*	The study team engaged not only trade associations such as TXTA and CSCMP, but also other DOTs and freight consultants to research lessons learned and gain knowledge of similar experiences or projects. The study team sought to partner specifically with TXTA, which in turn lent support to the study team.
16	*Coordinate with local or regional agencies that may have closer relationships with data providers*	The study team reached out to MPOs and other planning organizations, and learned that some MPOs manage databases that could contain useful criteria for the development of a statewide freight database.
17	*Consider gradual implementation of data acquisition coupled with coordination about successes*	N/A (This could be a future consideration)
	Guidelines for Articulating Benefits of Sharing	
18	*Define and articulate the benefits, goals, and purpose of data sharing to stakeholders*	During the outreach efforts, the study team found that promoting the project as a "pilot program" and emphasizing the benefits of participation were effective outreach tools.

	Freight Data Sharing Guideline	**CTR Study Team Action**
19	*Include a stipulation that data is for limited or one-time use and cannot be used for any other purposes such as regulation*	N/A (This could be a future consideration)
20	*Publicize the cooperation among project partners and seek to give the project visibility to stakeholders and the public*	The study team created a website providing information on the project and demonstrated to stakeholders how the data will be used and where the data will eventually reside.
21	*Explain clearly to stakeholders that sharing of data will support improved freight infrastructure decisions that will benefit those stakeholders*	The study team repeated these benefits and assurance throughout every conversation with private sector firms.
22	*Add value to the data and make it available to all stakeholders*	N/A (This could be a future consideration)
23	*Use technologies that are useful for other purposes*	N/A (This could be a future consideration)
24	*Explore new market opportunities with potential data providers*	N/A (This could be a future consideration)
	Guidelines for Funding for Data Sharing and Projects	
25	*Attempt to include funding for research and data collection in public sector contracts*	N/A (This could be a future consideration)
26	*Be sure to include funding to cover costs of data sharing and needed agreements to protect data*	N/A (This could be a future consideration)

	Freight Data Sharing Guideline	**CTR Study Team Action**
27	*Where appropriate, try to obtain join public-private funding for projects*	N/A (This could be a future consideration)
28	*Consider gathering data from volunteer stakeholder groups or round tables*	The study team focused on these efforts during the beginning phases of the project. This issue refers back to the type of non-disclosure agreement to be negotiated.

5.4.4 Samples Obtained from Confirmed Partner Firms and Their Characteristics

After following the recommendations of *NCFRP Report 25* and carrying out multiple efforts, four samples were shared with the study team. These four firms made a hard commitment to share data by sending a data sample.

Grocery Direct:

- Location: Dallas, Texas
- Point of Contact: Kevin West
- Result: Provided data
- Description: "The largest grocery-based consolidation and transportation services company in the Southern United States." The company ships 400 million pounds of freight per year.

Niagara Water

- Location: Ontario, CA
- Point of Contact: Brian Reed
- Result: Provided data
- Description (provided over the phone): #1 Bottling company in the country; $1B in revenue. Their Texas operations happen primarily out of Dallas (7 sites, 260 water trucks/day) and Houston (2 sites, 120 trucks/day). The company is focused on short-range distribution in the Texas market, with some routes into Oklahoma and Arkansas. The company does not own/run their own trucks (provided referral to a contact at an outside freight company, AFN, whom the study team later spoke with). They mostly deal with heavy trucks due to the weight of water. They ship resin for bottles, plastic for caps, and the water itself. Niagara has data on speed for transit times; knows exactly when trucks leave/arrive.
- Motivation for project participation: Previously involved with "benchmarking studies" with MIT; the "good citizen" aspects of the project.

Skinner Transportation

- Location: Austin, Texas
- Contact: Steve Skinner (survey contact)
- Result: Provided data
- Description: Skinner ships asphalt and other road building material. This all-Texas company showed distinct interest in the project from an early stage. Steve Skinner is Austin-based and has supported our project since late 2012.

5.4.5 Cooperative Research and Development Agreement Data

The study team learned that the Houston-Galveston Area Council (HGAC) has a partnership in place with the Port of Houston and selected drayage carriers. Through the latter, HGAC collects data on drayage trucks that travel between the Port of Houston and diverse surrounding warehouse and intermodal facilities. The database's main purpose is to calculate the environmental and emissions caused by drayage operations. However, the study team could also use this data for this project. The database would have to be requested through a FOIA request.

5.5 Conclusion

Effective partnerships are clearly needed between the public and private sectors to ensure adequate freight planning and funding of transportation infrastructure at the state and local levels. Despite this recognition, roadblocks continue to exist in moving these partnerships forward. The greatest challenge is getting a foot in the door with firms who are busy, preoccupied, and suspicious of the nature of a pilot project involving data-sharing. Of the 493 companies contacted, 151 expressed an interest in receiving additional information about the project. However, only 33 participated in one-on-one interviews and 3 provided sample data, notwithstanding the lack of a non-disclosure agreement. Nonetheless, through these efforts, the study team initiated productive relationships with freight stakeholders, and recommends continuing to develop these freight relationships on an ongoing, long-term basis in order to gain trust and secure partnerships. In summary, the following lessons were gleaned from the outreach efforts:

1. Most stakeholders interviewed considered that a partnership would have beneficial outcomes, including addressing any current or anticipated transportation issues such as congestion and capacity constraints, and providing recommendations in the design and development of new infrastructural projects which can impact freight operations.
2. The majority of stakeholders were concerned with the mishandling or improper use of data, the time commitment required in scrubbing and preparing data in-house, and new government regulations and law enforcement measures pertaining to data security.
3. Lightening the information technology requirements for stakeholders is highly recommended. Stakeholders were more willing to share data when the study team offered to accept data samples in either CSV or Excel formats. In addition, offering to "clean" or scrub the data for them was also welcomed.
4. If guaranteed that the information would never become public, 88% of survey respondents were willing to participate in a data-sharing partnership.
5. None of the respondents interviewed or surveyed are currently participating in a data-sharing partnership.
6. Data variables that stakeholders were willing to share include trip O-D, number of trips, vehicle type, load type (truckload, less-than-truckload), route preference, commodity being transferred, cargo weight, and mode of transport.
7. Most stakeholders were willing to share data either through an electronic submission via email or via a secured website.

8. A clear non-disclosure contract is required: a written contract would reinforce a sense of trust with freight partners and the sample provided in the *Freight Data Sharing Guidebook* is a good option.
9. Support from trade associations such as TXTA was found to be invaluable in the outreach efforts. TXTA published the project information in its weekly newsletter, including a link to the online survey, and invited team members to events, which provided an opportunity to meet potential stakeholders.
10. Participation—secret for some, open for others: Some stakeholders expressed preference for their participation to remain anonymous; others inquired about the possibility of advertising their participation in such data-sharing partnership.
11. Most stakeholders contacted asked to see something more tangible to share with company executives and decision-makers before deciding whether to enter into a data-sharing partnership. Therefore, a built-out demo website (complete with the initial architecture) that demonstrates the integration of public and private datasets could promote confidence in future data partnership efforts.

In Chapter 6, the study team reviews and assesses information about existing data architecture initiatives to gain insight into whether some of existing data architectures can be tailored or built upon for the development of an Integrated Freight Database.

Chapter 6. Data Architecture Review

As part of the study, a review and assessment of available database architectures that have been proposed in the literature was conducted. The purpose of the review was to gain insight from previous work to use in the development of the Freight Data Architecture.

In general, considerable development has gone into supporting the capture and exchange of freight data, as well as architectures to support the use of the data. These include a variety of state and federal efforts as well as the provision of commercial systems. These efforts do not yet provide the richness required of a multi-modal data set that supports inter-city, inter-state, and intra-city freight data queries as determined from earlier tasks. However, the efforts to date provide the necessary groundwork to support such specifications. As part of the literature review on freight-related data architectures, multiple reports were examined in addition to information available on federal or state websites. A brief synopsis of the reviewed systems and lessons learned is provided in this section.

6.1 Automated Commercial Environment/International Trade Data System

—Primary Reference: U.S. Customs and Border Protection. *ACE 101.* February 2012, pp. 1–7.

6.1.1 Purpose and Content

Automated Commercial Environment (ACE) is the U.S. commercial trade processing system developed to automate border processing, improve border security, and promote economic security in the U.S. through legitimate international trade and travel. The U.S. Customs and Border Protection (CBP) agency plans to expand ACE to offer cargo processing capabilities throughout all modes of transportation, replacing current systems with automated commercial systems (Quiroga, et al., 2011). ACE provides a foundation of solid technology for all border security initiatives within CBP to (1) enable trade members access to and control of their trade information via reports; (2) accelerate lawful trade by providing CBP with support tools to efficiently manage imports/exports and relocate commodities quickly across the border; (3) enhance communication, collaboration, and compliance efforts between CBP and the trade community; (4) facilitate efficient collection, processing, and analysis of commercial import and export data; and (5) offer an information-sharing platform for trade data throughout government agencies (U.S. Customs and Border Protection, 2012).

6.1.2 System Overview

The ACE Secure Data Portal is an interactive and web-based tool that provides a single, user-friendly gateway to access CBP information. The portal has three primary user parties: (1) CBP; (2) trade community; and (3) participating government agencies (PGAs). The ACE Secure Data Portal offers these features to each account user: (1) near real-time (every 2 hours) review of CBP entry, entry summary, and specific data from the reports; (2) account-based display; (3) secure account setting and ACE Portal account access based on individual companies' organizational structures; and (4) the ability to post information, tracking, and/or respond to CBP on compliance and operational issues. Figure 6.1 shows the conceptual architecture of ACE.

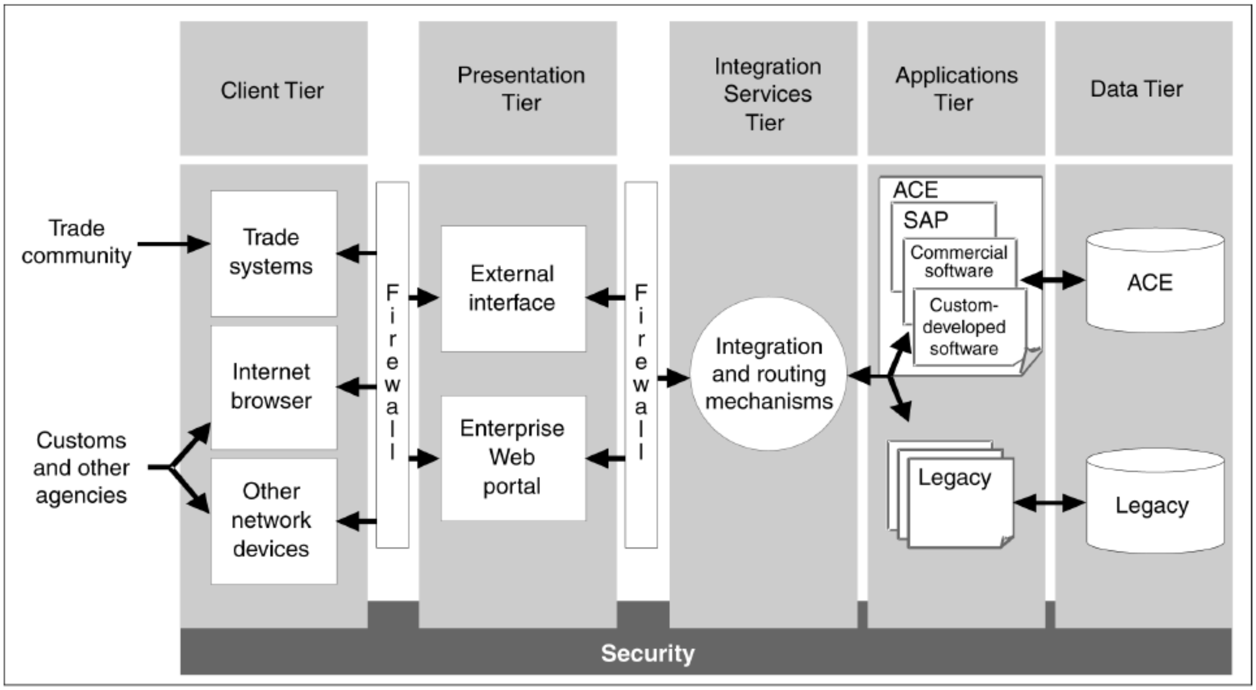


Figure 6.1: Conceptual ACE Technical Architecture (Byrd & Rogers, 2005, p. 28)

The International Trade Data System (ITDS) is a federal program that encourages PGA to join in ACE (U.S. Customs and Border Protection, 2012). The program supports plans of PGAs for identification, documentation, and execution in order to improve business process through using ACE. One of the mechanisms on the ITDS program uses to integrate PGAs into ACE is through the developing the ACE/ITDS standard dataset. This dataset is a collection of data requirements for regulatory and enforcement processes on international trade and U.S. border. Confirming data collaboration to help the entire implementation of ACE throughout all relevant federal agencies is its objective (Quiroga, et al., 2011).

6.1.3 Lessons Learned

Customization

Users should be able to customize system output by selecting input data elements.

(Near) Real-Time Based System

The ACE trade reports are refreshed every 2 hours. This capture rate is useful for providing disaggregated data to support detailed data queries for a Texas freight data system.

Security

In security, e-Manifest has been used in ACE, the electronic submission of trip, conveyance, equipment, crew, passenger, and shipment information. An e-Manifest is submitted to CBP via the ACE secure data portal, electronic data interchange (EDI), or a combination of

EDI and the ACE portal through self-filing or use of third parties (U.S. Customs and Border Protection, 2011).

6.2 Freight Analysis Framework 3 (FAF3)

—Primary reference: Southworth, Peterson, Hwang, Chin, & Davidson (2011). *The Freight Analysis Framework, Version 3 (FAF3): A Description of the FAF3 Regional Database And How It Is Constructed.* Office of Freight Management and Operations, Oak Ridge National Laboratory. Federal Highway Administration, Washington, D.C., 2011, pp. 1–15.

6.2.1 Purpose and Content

The FAF assembles data from a variety of sources to draw a comprehensive national picture of freight movements among states and major metropolitan areas, by all modes of transportation (Southworth, Peterson, Hwang, Chin, & Davidson, 2011). Operating with a national level of current freight flows to, from, and within the U.S., the FAF assigns certain flows to the transportation system, and predicts future patterns of such freight flow. FAF3 is the third version of database; FAF1 provided estimates for truck, rail, and water tonnage for calendar year 1998, and FAF2 created a more comprehensive picture through 2002 Commodity Flow Survey (CFS).

The FAF3 data products contain both a regional database and network database with highway flow assignment as described in Southworth, Peterson, Hwang, Chin, & Davidson (2011). This document specifically explains the development methodologies for the regional FAF3 data products from various data sources, using many data modeling tools.

6.2.2 System Overview

As explained in Southworth, Peterson, Hwang, Chin, & Davidson (2011), the regional FAF3 database contains a matrix of freight volumes or freight flows, reported in both annual tons and dollar values for shipments to, from, and within regions for 2007, with conditional estimates for the current year, and forecasts to 2040. The principal dimensions of this flow matrix are shipment origination region (O), shipment destination region (D), the class of commodity being transported (C), and the mode of transportation used (M).

The CFS is the foundation for the development of the FAF. It provides shipper-sampled and subsequently expanded estimates of both tons shipped and dollar value trades within and between all U.S. regions for all modes of freight transportation. However, according to Southworth, Peterson, Hwang, Chin, & Davidson (2011), the CFS has some deficiencies that require significant additional effort to address in order to create a complete accounting of freight movements within the U.S.: (1) the CFS does not provide imports, while CFS reporting of export flows is also subject to data quality issues resulting from limited sample size, and (2) the CFS either does not collect data from certain freight-generating and receiving industries, or collects insufficient data to cover the industries in a comprehensive manner (referred to in Figure 2.2 as Foreign & Domestic Commodity Flows).

In addition, the CFS table contains a significant number of missing cells. This omission may be due to a small number in the cell, which was rounded to zero (such as for less than half a kilo-ton or half a mega-dollar), or the value may have been restrained for confidentiality reasons, or suppressed due to a high coefficient variance in the flow estimation, implying low confidence in the expansion of the sample. In order to fill in the missing values in the FAF flow matrix (shown in Figure 6.2 as the Missing Flow Value Inference), a combination of log-linear modeling and iterative proportional fitting was used.

6.2.3 Lessons Learned

FAF3 is a statistical-based model using CFS. It is useful for setting transportation plan or policy or allocating road/rail construction budget by year taking into account the traffic forecast. It also allows for the development and updating of data to meet the growing demand for freight data and minimize the gap among the FHWA, state DOTs, and MPOs. See Figure 6.3 provides an input/output display sample.

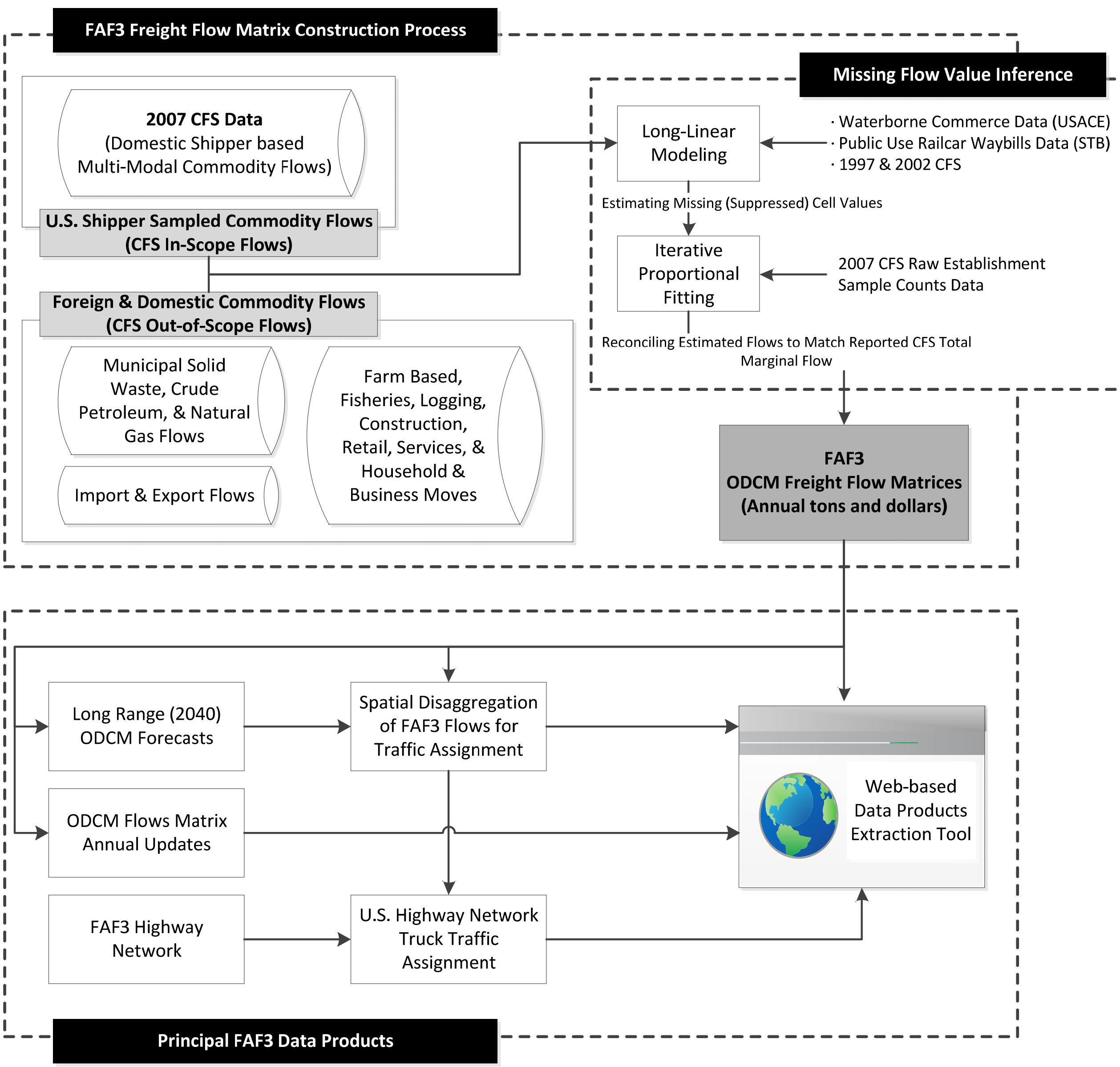


Figure 6.2: FAF3 Architecture

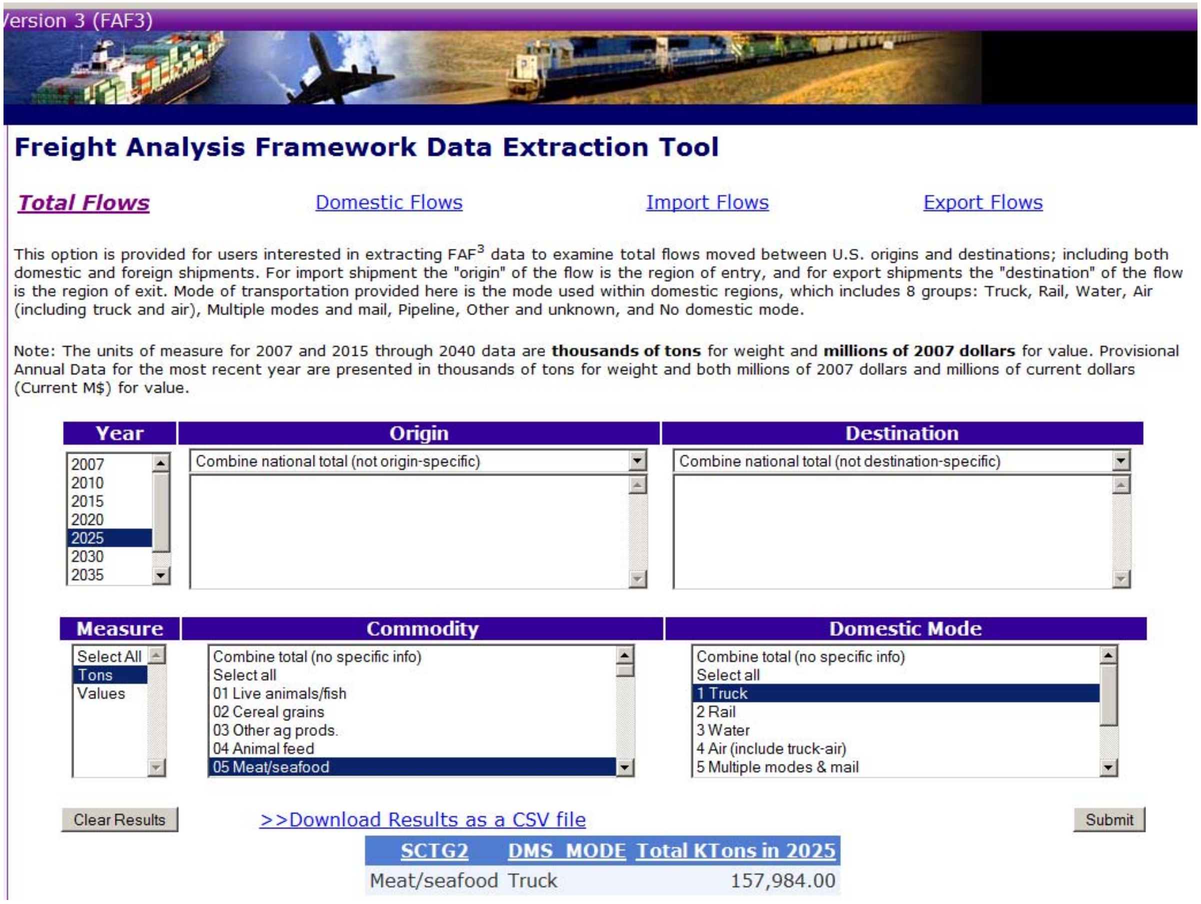


Figure 6.3: FAF3 Input and Output Display Sample (Federal Highway Administration, 2012)

6.3 Highway Performance Monitoring System (HPMS)

—Primary Reference: Federal Highway Administration. *Highway Performance Monitoring System Field Manual*. Washington, D.C., pp. 1.1–7.4. http://www.fhwa.dot.gov/policyinformation/hpms/fieldmanual/hpms_field_manual_2012.pdf, 2012.

6.3.1 Purposes and Content

The HPMS is a system that contains data about the extent, condition, performance, use, and operating characteristics of the U.S. highway network (Federal Highway Administration, 2005). HPMS data are used for a variety of applications, including the following (Quiroga, et al., 2011):

- Providing input for reports to Congress on the condition, performance, and investment needs of U.S. highways, reports that affect the scope and size of the federal-aid highway program;
- Assessing modifications in highway system performance and assigning federal-aid highway funds to each state;
- Assembling freight corridors and determining freight movement performance;

- Special policy and planning studies;
- Travel and congestion monitoring, public road usage, and fatality rate calculations;
- Investment needs and planning at the state level; and
- Air quality conformance and planning.

6.3.2 System Overview

Various agencies use HPMS data, including federal, state, and local agencies, as well as research agencies (Figure 6.4). HPMS relies on annual data from state DOTs (Federal Highway Administration, 2005). The HPMS field manual provides guidelines to state DOTs on the data obtaining and reporting procedures for HPMS, including precision levels and sample size estimation procedures. HPMS includes limited data on all public roads; detailed data on a sample of the arterial and collector functional systems; and area-wide summary information for urbanized, small urban, and rural areas; as follows (Quiroga, et al., 2011):

- "Universe" data show fundamental inventory data on all open public road systems in the HPMS database. This inventory includes 46 data items for National Highway System (NHS) sections and 28 data items for local roads.
- "Sample" data show 98 data items containing additional inventory, condition, use, pavement, operational, and improvement data for 120,000 sections of roadway selected as standard samples.
- "Summary" data show travel data for all serviceable systems in urbanized, small-sized urban, and rural areas, as well as air quality in non-attainment and maintenance areas. In addition to other HPMS data, each state has to submit linear referencing system (LRS) data in one of three categories: (1) maps and computer files, (2) files and maps for new links and nodes, or (3) geographic information system (GIS) files.

The FHWA receives, processes, analyzes, populates, and applies HPMS data provided by the states. The quality and integrity of these data depend on the state processes used to create the basic data inputs. The questions, concerns, and criticisms issuing from the various data users reflect directly on the states' ability to accurately represent the condition and performance of their highway systems. The FHWA provides guidance, educating, and technical support to states, but fundamentally providing quality data is a state responsibility (Federal Highway Administration, 2006).

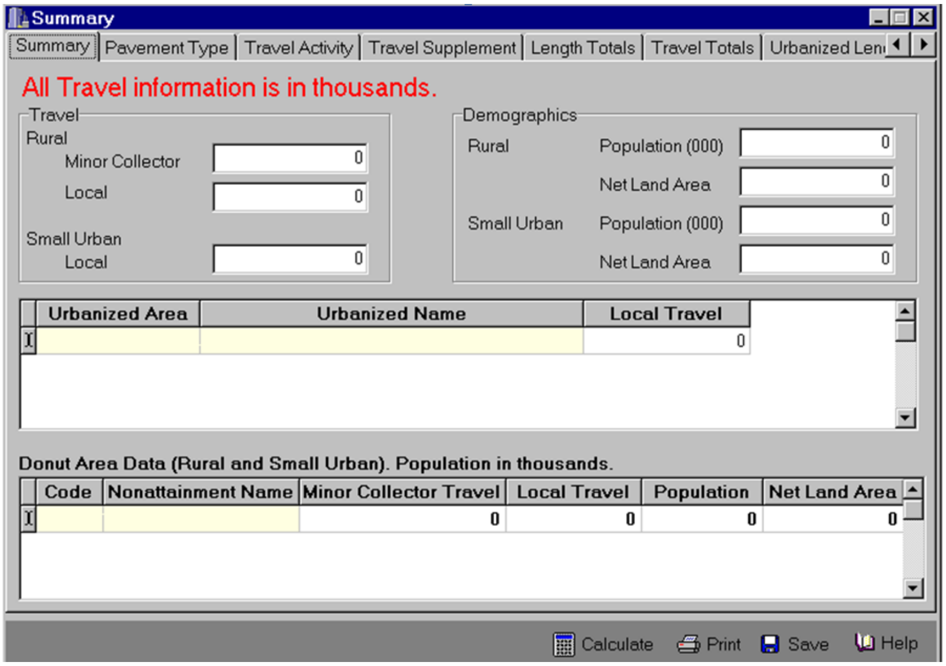


Figure 6.4: HPMS Software Summary Screen (Federal Highway Administration, 2011)

6.4 Planning, Environment, Realty Geographic Information System (HEPGIS)

—Primary Reference: Federal Highway Administration. *HEPGIS*. http://hepgis.fhwa.dot.gov/hepgismaps 11/help/About. html, 2012.

6.4.1 Purpose and Content

The HEPGIS is an interactive web-based GIS that enables users to access transportation-related geo-spatial data using only a web browser. Simple navigation tools let users locate and zoom in on an area, and to create maps providing various geographic features and themes. Users can not only display information about a specific feature, but also print or save an electronic copy of any map they create (Federal Highway Administration, 2012).

6.4.2 System Overview

Separate application tabs access maps and data that focus on different transportation areas of interest. Current applications include the following:

- Demographic Information—focusing on Census demographic data;
- Highway Information—focusing on national highway network and highway performance data integrated with FAF3;
- MPO Boundaries—focusing on urban area and metropolitan planning boundaries; and
- Federal Lands—focusing on locations of and access to various federal and tribal lands (Federal Highway Administration, 2012).

FAF3 data do not provide an estimation of the average daily truck traffic (ADTT) used to move freight between the shipping zones. The primary source of information for developing the procedures for converting commodity flows in tons to truck trips was the 2002 Vehicle Inventory and Use Survey (VIUS) database (tonnage to truck payload conversion process). The VIUS provides national and state-level estimates of the total number of trucks by truck type. These data are gathered through surveys of a sample of the motor carrier industry, and the survey is conducted every 5 years as part of the U.S. Economic Census (Battelle, 2011).

The truck flow maps and other thematic maps, for example, provide visual presentations of the volume and spatial variation of freight traffic. The outputs of the analyses can be expected to support policy makers in assessing improvement and policy options that affect freight transportation (Battelle, 2011). Figures 6.5, 6.6, and 6.7 depict the HEPGIS interface, showing 2007 ADTT and both the 2010 and 2040 FAF3 truck flow on the highway network. HEPGIS is currently partially available but considering the specific characteristics of truck traffic is important in adapting existing transport demand modeling techniques. The use of GIS-Transportation application software for the analysis facilitates communication of outputs of the analysis to policy makers.

Figure 6.5: 2007 AADT

Figure 6.6: 2010 FAF3 truck flow on highway network

Figure 6.7: 2040 FAF3 truck flow on highway network

6.5 Oregon Freight Data-Mart (OFDM)

—Primary Reference: Figliozzi, M. A. and Tufte, K. "*Challenges and Opportunities for Online Freight Data Mapping Integration and Visualization*." *Transportation Research Forum.* Fargo, North Dakota, pp. 1–15, 2009.

6.5.1 Purpose and Content

The Oregon Freight Data-Mart (OFDM) provides an online environment to integrate, visualize, and disseminate freight data in Oregon. The OFDM is a data visualization tool based on Google Maps. Google Maps was chosen for visualization because it has high compatibility, which can be used to combine different types of data and can be accessed by any user from any internet browser.

6.5.2 System Overview

The user interface designs to use integrated visualization of data sources using multiple hierarchical layers and clickable links able to explore and expand details. Using Google Maps to display the freight data has a significant advantage: the possibility of drawing on other Google services like Google Traffic, Google Street View, or satellite images.

The integration of current freight data with the Google Maps application means that as Google provides more useful services, the OFDM can take advantage of these services, most often with limited time and financial overhead. Google Earth can be applied as a backbone to create maps that can be exported to KML/KMZ files, a widely accepted format, which can be displayed later in Google Maps (Figliozzi and Tufte, 2009). At a conceptual level, the OFDM system architecture and process are described in Figure 6.8 and Figure 6.9. Table 6.1 presents the summarized data sources and their characteristics.

The OFDM has some provision for using multiple data sources, including ITS data. Figures 6.10 and 6.11 present an example of the use of the OFDM to highlight truck travel volumes collected from live data at a certain location.

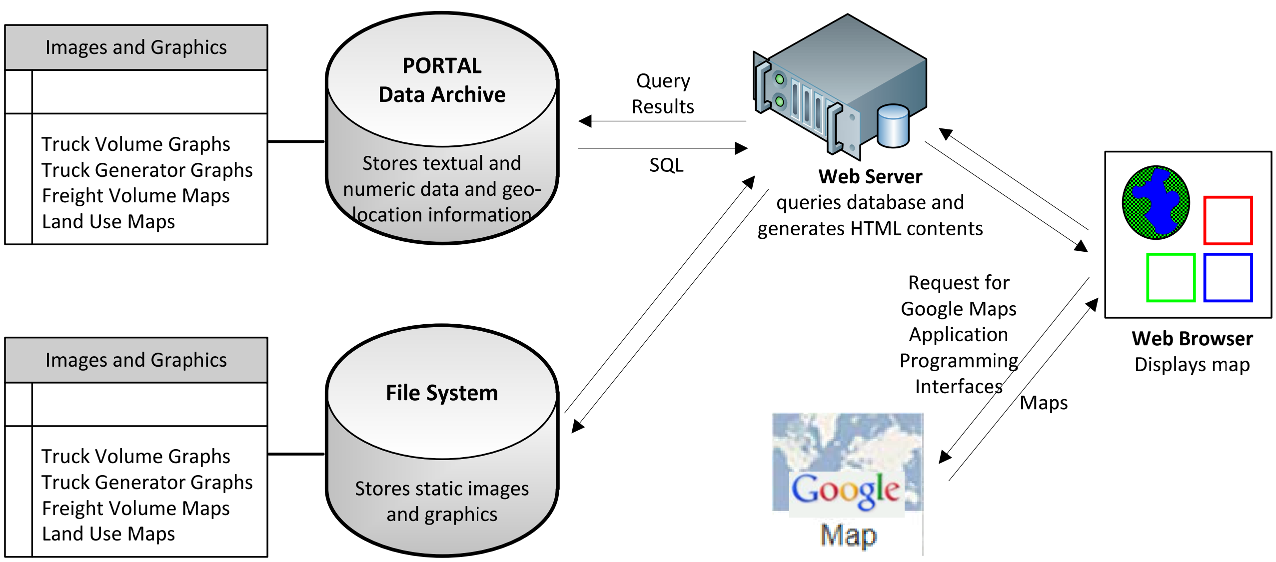


Figure 6.8: OFDM System Architecture (Modified from Figliozzi and Tufte, 2009, p. 4)

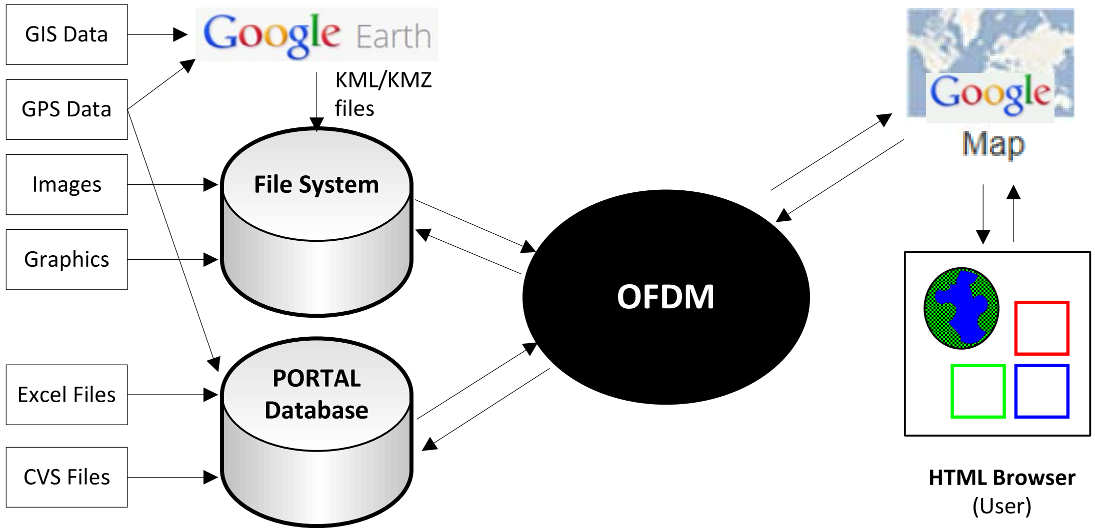


Figure 6.9: OFDM System Process (Modified from Figliozzi and Tufte, 2009, p. 4)

Table 6.1: OFDM Data Sources (Figliozzi and Tufte, 2009, p. 6)

Name	Data Provider	Type of Data	Source Instrument	Collection Metadata	Analysis Report
Truck Incidents	PORTAL/ODOT	Database	Input by ODOT ATMS Operators	Database Field Descriptions	
Truck Volumes	Port of Portland	Survey Data, Excel File	Field Collection, Consultants	Report Description	
Truck Generators	Port of Portland	Survey Data, Excel File	Field Collection, Consultants	Minimal Metadata	
Bottlenecks	ODOT/ OTREC Project	Text Data	Loop Detectors, Ground Truth GPS Data	Detailed Description and Methodology	Continuous Collection and Analysis – Reports
Weigh-In-Motion Stations	ODOT	Database	Scales, Transponder Readers	Detailed Description and Methodology	Continuous Collection
Highway Speed and Reliability	PORTAL/ODOT	Database	Loop Detectors, Cameras	Detailed Description and Methodology	Continuous Collection and Analysis – Reports
Freight Volume Maps	Metro	Maps	Variety of Truck Counts	No Metadata	
Land Use Maps	Metro	Maps	Norms and Regulations	No Metadata	

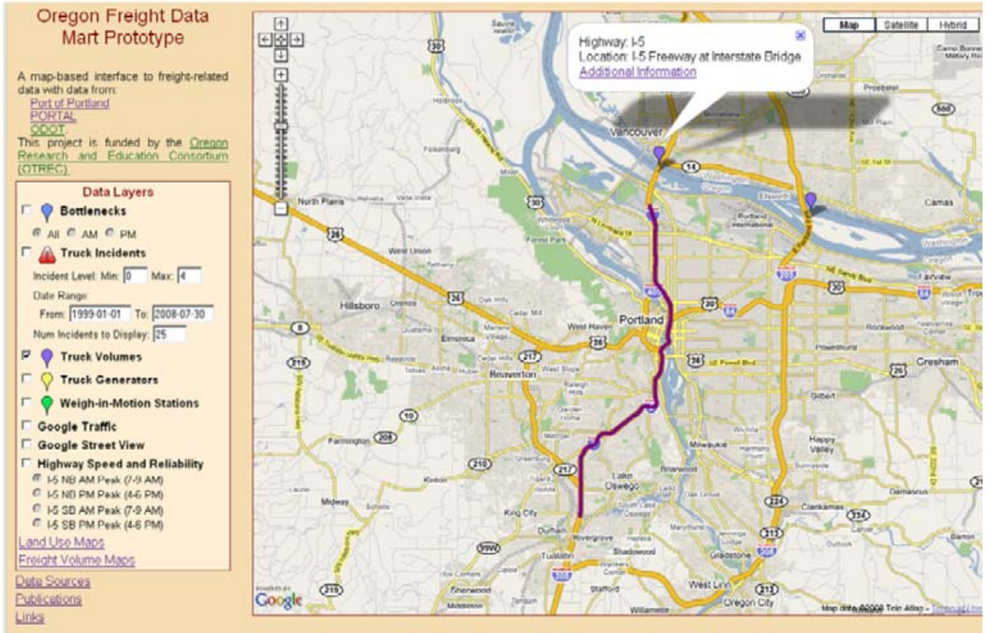


Figure 6.10: Truck Volume Display—Map View (Figliozzi and Tufte, 2009, p. 11)

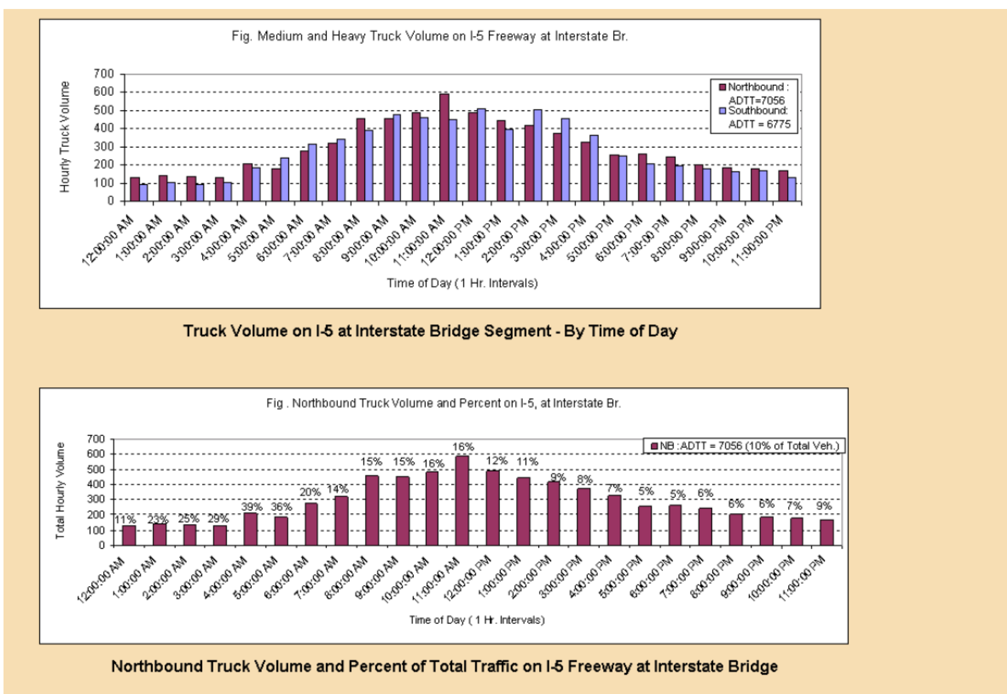


Figure 6.11: Truck Volume Display—Detail for I-5 at Interstate Bridge (Figliozzi and Tufte, 2009, p. 11)

6.6 National ITS Architecture

—Primary Reference: Mitretek Systems and TransCore, Inc. *Developing Traffic Signal Control Systems Using the National ITS Architecture*. FHWA-JPO-98-026, ITS Joint Program Office, Federal Highway Administration, Washington, D.C., pp. 2-1 – 2-71, 1998.

6.6.1 Purpose and Content

The National ITS Architecture is a framework for the inter-relation of functions, physical entities, and information flows among functions and entities (Research and Innovative Technology Administration, 2012(1)). The National ITS Architecture has been extensively implemented around the country, including traffic management centers, traffic signal systems, and tolling operations.

6.6.2 System Overview

The National ITS Architecture includes user services, a logical architecture, a physical architecture, and standards, as summarized below (Quiroga, et al., 2011).

User Services: A user might be the public or a system operator; user services refer to the functional description of what system provides to the user.

Logical architecture: This component refers to the processes, data flows among processes, terminators (i.e., entry and exit points such as sensors, computers, and human operators), and data stores required to satisfy the functional requirements of the 33 user services defined in the logical architecture (Research and Innovative Technology Administration, 2012(2)). Nested data flow diagrams (DFDs) illustrate the graphical representative processes, data flows, terminators, and data stores at varied separate levels. The highest level is a DFD called Manage ITS that has nine first-level processes, all of which are DFDs. In turn, each of these processes has subordinate processes, some of which are DFDs. Version 7.0 of the National ITS Architecture (released January 2012) includes 3,871 logical data flows, of which 365 have as a source node one of the Manage Commercial Vehicles processes or subprocesses. Freight-related data elements typically cover vehicles and their interaction with the road environment.

Physical architecture: This component provides a representative methodology of the functionality defined by the user services and the logical architecture from an integrated system. By defining subsystems based on the functional similarity of process specifications and physical locations of functions within the transportation system, this objective is satisfied. Figure 6.12 shows the four general categories of subsystems: Centers, Field, Travelers, and Vehicles. Generally, the physical architecture covers subsystems, architectural flows that connect subsystems and terminators, and equipment packages that divide subsystems into deployment-sized pieces. Basic Traffic Signal Control System Architecture is described in Figure 6.13.

Standards: The RITA (Research Innovative Technology Administration) ITS standard database has 96 ITS standards. They include document types such as guides, data dictionaries, message sets, and protocols.

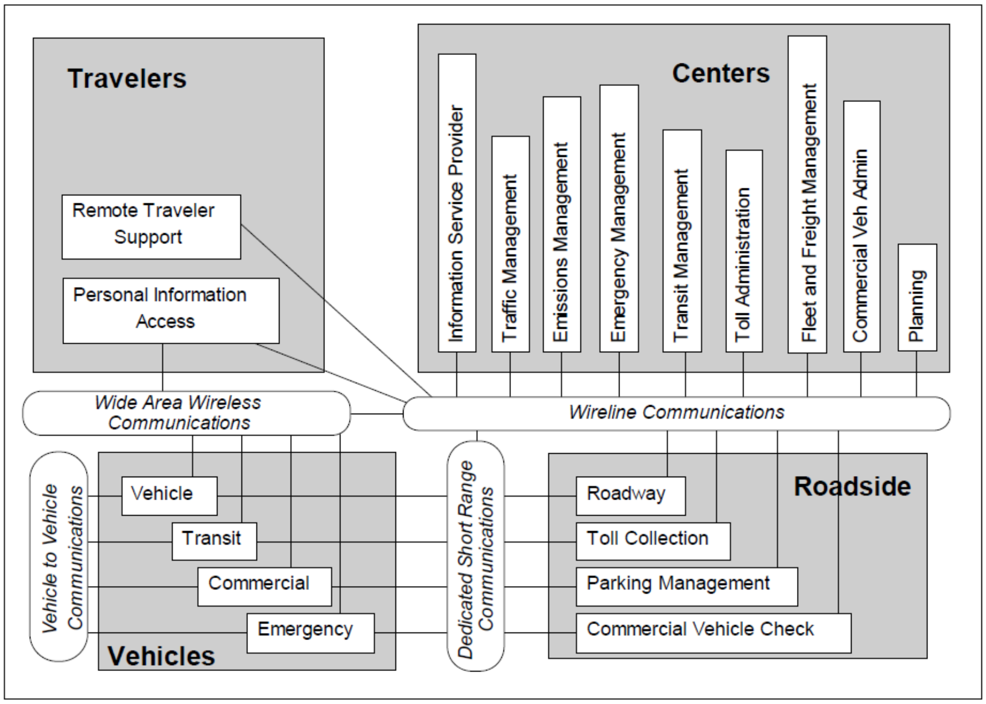


Figure 6.12: National ITS Architecture Subsystems and Communications (Mitretek Systems and TransCore, Inc., 1998, p. 2-13)

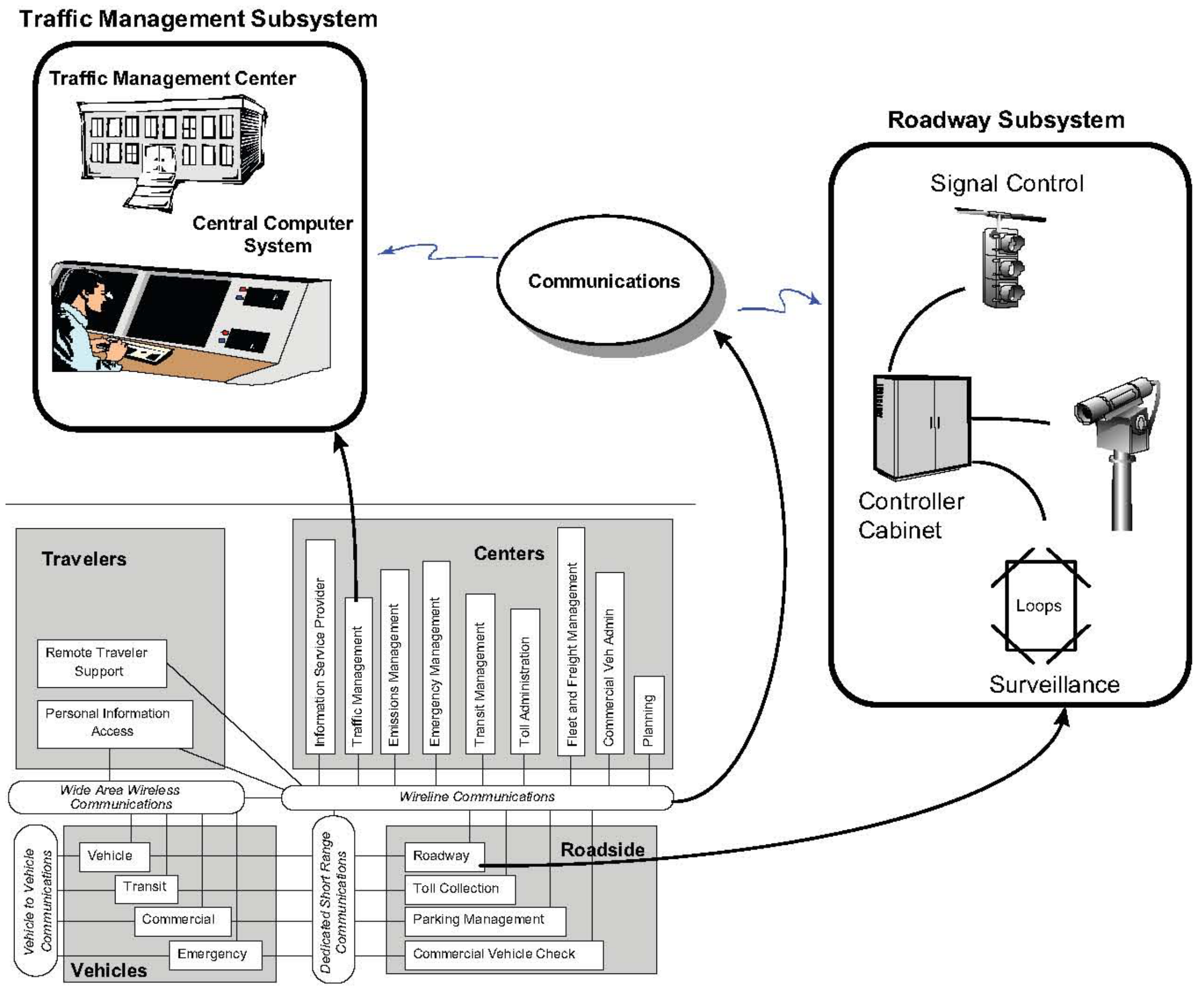


Figure 6.13: Basic Traffic Signal Control System Architecture Depiction (Mitretek Systems and TransCore, Inc., 1998, p. 2-14)

The physical architecture deals with market packages as well, which represent pieces of the physical architecture that address specific services. Typically, a market package includes several different subsystems, equipment packages, terminators, and architectural flows that provide the desired service. The physical architecture contains 13 market packages connected to commercial vehicle operation.

6.7 TransXML

—Primary Reference: Ziering, E., Harrison, F., & Scarponcini, P. *TransXML: XML Schemas for Exchange of Transportation Data.* Transportation Research Board, NCHRP Report 576, Washington, D.C., 2007.

6.7.1 Current and Potential Future Scope of TransXML

Currently, four business areas in data exchange of surface transportation have been applied for TransXML: design, construction and materials, bridge, and safety. One recommendation is to expand use of TransXML into other parts of transportation activity: geospatial, ITS/operations, travel and traffic modeling and simulation, and freight logistics. The current and potential future uses of TransXML are illustrated in Figure 6.14.

The freight logistic view involves the intermodal freight supply chain—sharing of information on shipment and equipment status and location. The Open Applications Group (OAGI), where TranXML (as opposed to TransXML) was developed, created a specification of the freight logistics view (Zeiring et al., 2007). As of this writing, the active links in TranXML are supported by OAGI and others date from 2000 to 2004. Unclear is whether TranXML is currently being used or has been replaced. At any rate, the TranXML documents refer to data interchange among shippers and these possibly support the needs of freight data in Texas, but additional study is needed.

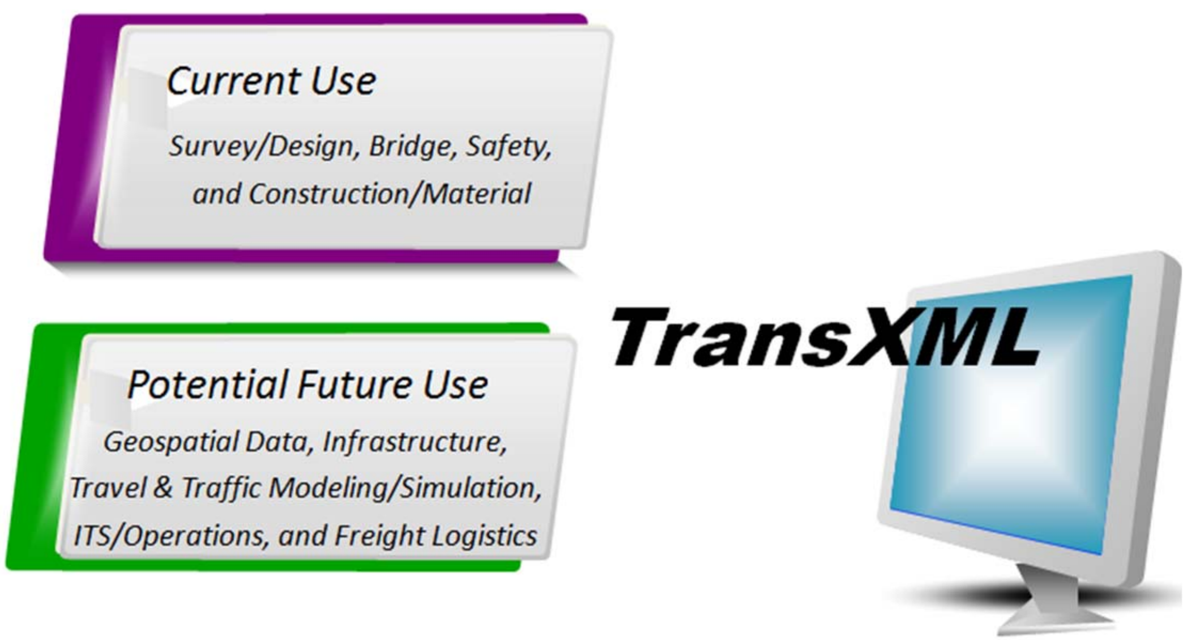


Figure 6.14: TransXML Current and Potential Future Uses

6.8 Conclusion

The findings from the stakeholder workshops suggest that freight data users in Texas desire a robust, visually orientated information system that can handle a wide range of queries of varying levels of detail. The literature was reviewed with these requirements in mind and it was determined by the study team that a wide range of applicable data standards are available for use in representing freight data. The following summarizes the study team's findings:

- *Security of data is an issue, particularly across multiple data sources:* Secure design comprises at least two aspects: provision of secure access to multiple data sources and provision of user roles and permissions. These issues are generally well understood and very project specific. Security issues also involve contractual understandings among data providers and users.
- *Source and quality of the data should be made visible to the users: T*he review of various data sources made it clear that providing users with a clear sense of where the data is being collected from is very helpful. This encourages an independent assessment of the data quality by users and prevents the misuse and misrepresentation of data.
- *The incorporation of commercial software architectures such as Google Earth and Maps as a visualization front end:* Use of such tools has benefits in terms of speed and cost of deployment and the outsourcing of significant development and maintenance tasks to other parties. Disadvantages of using such systems include reliance on a third party and subsequent lack of control and the possibility of being forced into maintenance/update cycles that do not match internal cycles.
- *Provision for decision support tools that supplement data available for user queries:* Missing or incomplete data can possibly be supplemented through statistical analysis or other modeling methods. User queries for disaggregated data along routes may also require inferences from simulation tools that build from the available data. The data architecture should be designed to allow addition of decision support modules.
- *Use of ITS (Intelligent Transportation System) data is an option*: Collection of ITS data from electronic monitoring systems provide an opportunity to collect rich data about freight movement in Texas. The quantity of available ITS data will only increase with time and the proposed data architecture should be able to leverage its benefits.

Chapter 7 provides a strategy for integrating multiple data sources and for collecting and compiling primary data from Texas's freight stakeholders.

Chapter 7. Proposed Data Architecture

Using the information gleaned during the development of this study, the study team developed prototype freight data architecture and supporting description and specifications that will facilitate the storage and exchange of data through a data sharing partnership with members of Texas's freight community. The proposed conceptual freight data architecture was developed based on interactions with potential partners (Chapter 5) and the stakeholder workshops organized in the six TxDOT districts (Chapter 2). The conceptual architecture includes mediator architecture to support translation of publicly available and partner freight data to a common/shareable representation is presented. Sample logical and data models are specified, including a data field mapping adapter that reconciles user queries with existing database fields and facilitates the preservation of data in its original form, allowing future updates of existing datasets. A list of items to be considered in estimating the cost for developing and maintaining the system is also discussed.

A limited functionality prototype of the integrated freight database was developed and presented to TxDOT. The prototype system, called Unity DB, can be accessed at http://www.unitydatabase.com.

7.1 Conceptual Freight Data Integration Architecture

It was determined from the survey of freight stakeholders that electronic file submission (i.e., either by email or directly via a website) is the preferred method for data exchange for any freight data sharing partnership (see Figure 7.1).

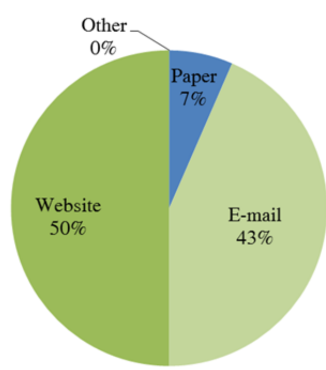


Figure 7.1: Stakeholder Preferred Data Submission Format

It is therefore imperative that any successful freight data sharing architecture have this capability available in addition to any other features such as data security and data privacy that will need to be included. Based on this knowledge and information gleaned during the study period, the study team proposes a conceptual system architecture with the following minimum capabilities as illustrated in Figure 7.2 and described in detail in subsequent subsections:

1. Integration and use of publicly available data
2. Electronic submission of data by freight data sharing partners
3. Data quality and validation

4. Automated data scrubbing and aggregation
5. Secure data storage and restricted access
6. Value added services through integration into existing TxDOT data centers
7. Data output and analysis tools

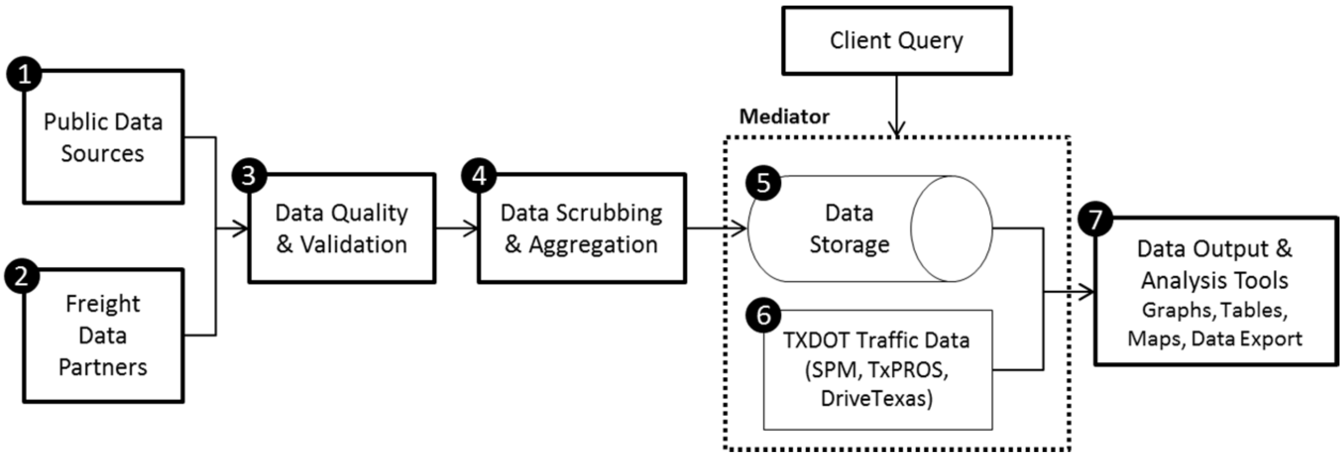


Figure 7.2: Conceptual Architecture for an Integrated Freight Database System

7.1.1 Integration and Use of Publicly Available Data

From earlier tasks, the study team determined that a total of 21 public freight databases can be readily used in the development of an integrated freight database. This data collected by government agencies and other industry associations usually do not have the level of disaggregation required for county or zip code level planning. Through the use of data fusion and mediation processes, it is possible to integrate these databases into a single system as demonstrated in the prototype system. Data integration will, however, have to be done carefully because of varying naming schemes and reporting methodologies. The mediator architecture section of this tech memo describes in detail a recommended approach for integrating these databases with private sector data.

7.1.2 Electronic Submission of Data by Data-Sharing Partners

Electronic data submission involves the submission of data either by e-mail or a web form. Despite electronic data submission being the preferred data exchange medium, phone conversations with potential data sharing partners determined that partners may not have sufficient resources to prepare, examine, or scrub the data into a particular standardized format before submission. Therefore, the receiving agency will need to have in place an infrastructure that facilitates this process. The data sharing partner may only need to submit the data in the format they feel comfortable with and the data exchange system should take care of the rest automatically. The minimal standard that needs to be adhered to by the data sharing partner is the file exchange format (e.g., CSV, Excel) and a partner-specific agreed upon and consistent data layout (e.g., number of columns, variable types).

Not all electronic data will require manual submission. Data from non-intrusive technologies such as radio frequency identification (RFID) tags and GPS devices can be set up to automatically transmit data to the integrated system. However, privacy concerns will have to be addressed, and data aggregation, scrubbing, and cleaning will still be required before final inclusion into the integrated freight database.

7.1.3 Data Quality and Validation

Ensuring data quality is primary to the success of any data sharing program. Data cleansing or data cleaning is required to ensure that data being stored is accurate, complete, relevant and consistent. Data cleansing can be performed through the provision of standard data dictionary definitions for each data source to be included in the integrated freight database. A system should be in place to detect missing or inaccurate data types and notify the system administrator of any such errors.

Data validity can be checked using the following suggested constraints (Oracle 2013, IBM 2013, Wikipedia 2013):

i) Data-Type Constraints: Values in a particular column must be of a particular data type, e.g., Boolean, numeric (integer or real), date, etc.

ii) Range Constraints: numbers or dates should fall within a certain range set by a maximum or minimum value. This can be performed using the database *check constraint* feature, e.g., setting the figure for AADT to not be negative and less than 500,000 for a specific region or area.

iii) NOT NULL Constraints: to prevent null values from being entered into a column

iv) Unique Value Constraints: A field, or a combination of fields, must be unique across a dataset, e.g., using designated roadway names to describe a particular roadway link.

v) Foreign-key Constraints: Also known as referential constraints or referential integrity constraints, these are used in defining required relationships between and within tables, e.g., referencing all NAICS commodity code classifications in multiple tables to a single table containing all NAICS commodity code definitions.

vi) Regular Expression Patterns: These are used in validating text fields if data is required to have a certain pattern, e.g., five digit zip codes (00000).

vii) Cross-field Validation: To ensure that certain conditions that use multiple fields must hold, e.g., data uploaded for 2012 cannot contain 2013 data or percentage of truck traffic cannot exceed 100.

viii) Duplicate Data Elimination: elimination of duplicate or already reported data. These will, however, need to be done carefully based on a knowledgeable set of unique value constraints.

In addition to using software to set data constraints, other data quality steps may need to be performed manually by the receiving agency. These steps include checking for data accuracy, data consistency (i.e., comparison with previously reported data), and data uniformity (i.e., units of measurement).

7.1.4 Automated Data Scrubbing and Aggregation

As identified in this study and other related studies (Cambridge Systematics et al. 2013), the private sector is reluctant to share data for fear of government regulation and industry competition, among other reasons such as resource constraints in making the data available in an acceptable format. Data scrubbing and/or aggregation steps are therefore necessary to address privacy concerns and ensure trust between parties involved in the data sharing partnership. Data scrubbing will remove sensitive data such as company name, customer address, tracking data, specific commodity identifiers, or any other variables that may endanger the competitive advantage of partners. In addition to data scrubbing, data aggregation will consolidate data from multiple sources, making it impossible to trace original sources; aggregation is useful for reporting network- or zone-level information. For data for which scrubbing and aggregation may be insufficient, noise infusion can be applied. Commonly used by the U.S. Census Bureau, noise infusion is a method of disclosure avoidance in which values are perturbed prior to reporting by applying a random noise multiplier to the originally reported magnitude data (Evans et al. 1996, U.S. Census Bureau 2013).

7.1.5 Secure Data Storage and Restricted Access

Essential to data sharing programs is the ability for the receiving agency to securely store and restrict access to data received from partners. Options include the use of data encryption technologies such Secure Sockets Layer (SSL) encryption during the data submission process and 256-bit Advanced Encryption Standard (AES) encryption for data storage.

In addition, access to raw data should be restricted to users at a level sufficient to examine data integrity but still protect the privacy of partners involved in the data partnership program. Access to output data should also be restricted at different usage levels. For example, agency staff can have access to samples of the prior scrubbed data and the general public can be restricted to summary data aggregated from multiple databases. Exact user policies and roles can be further defined in detail should TxDOT choose to implement the integrated freight database.

7.1.6 Value-Added Services through Integration into Existing TxDOT Data Centers

TxDOT currently has available a suite of tools that, when linked into the integrated freight database, will add additional value to the database. Examples of identified TxDOT data tools include the following:

i) TxDOT's Statewide Planning Map (SPM): This contains relevant information on TxDOT on-system roadways. Example data reported by the SPM include current and historic traffic counts, Texas trunk system routes, roadway speed limits, geopolitical boundaries, imagery, and other geospatial and biogeographic data such as watersheds, aquifers, and vegetation (TxDOT 2013a).

ii) DriveTexas.org Road Construction and Incident Reporting System: DriveTexas.org is another online TxDOT mapping system that can complement the integrated freight database system by providing live traffic feed data. Still in development, the system provides live information on roadway accidents, closures, construction work zones, damages, ice/snow conditions, and other condition types such as special events and parades (TxDOT 2013b).

iii) Texas Permitting & Routing Optimization System (TxPROS): TxPROS is an online permitting and GIS-based mapping system capable of producing information on the routing of OS/OW loads permitting system. TxPROS data can complement the integrated freight database system by providing routing information, number of permits, vehicle sizes and weights, and frequency of OS/OW trips on various roadways (Texas DMV 2013).

iv) TxDOT's Statewide Analysis Model (SAM): SAM is a state-of-the-practice multimodal travel model that provides highway traffic forecasts for both highway passenger travel and freight transport, intercity and high speed passenger rail ridership, freight rail tonnage and train forecasts, and forecasts of air passenger travel to and from Texas airports. The most recent version, SAM-V3, provides travel forecasts at a level of detail suitable for use in comparative analyses of large scale transportation corridor projects and other large scale investments. The model can also be used to perform analyses of the transportation outcomes and economic impacts of state-level transportation, land use, and economic policy decisions and strategies. It is the primary tool used by TxDOT for evaluating large intercity transportation projects through the state. SAM-V3 already incorporates publicly available data into its models and compatible to proprietary data sources such as Wood and Poole employment forecasts and TRANSEARCH/Global Insight data (TxDOT 2013c).

7.1.7 Data Output and Analysis Tools

During the workshops conducted as part of this project—which solicited feedback from potential users of the integrated freight database in TxDOT's district and regional offices, MPOs, and city offices—the most popular freight database features requested by participants were

i) a web-based platform with graphical user interface and GIS,

ii) the ability to determine the most effective or shortest routes to transport goods in the State of Texas,

iii) disaggregation of the data at the county- or corridor-level,

iv) show traffic generators (current and future), changes in modes, commodity information (weigh out, cube out), loads locations, OS/OW loads, and networks used, and

v) the ability to determine traffic flow and percentage of truck traffic

The final requirement of this system should incorporate the above-requested features in addition to the following:

vi) Avoid combining data from multiple databases if integration is not possible. Instead, data from each database should be shown separately[iv].

vii) Provide users with a summarized view of the data as well as the ability to download the complete queried data.

[iv] An ongoing project by the study team to identify differences in data element definitions and methods for bridging these differences is currently underway. Additional information on this project is available at http://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=3537

viii) Display information as graphs and tables where possible.

7.2 Proposed Mediator Architecture

The mediator architecture proposed by the study team and shown in Figure 7.3 provides a strategy that maintains the integrity of datasets[v], facilitates future updates, and speeds up the processing time of user queries. This architecture was used in the development of the prototype system. Datasets used in the development of the prototype system include the Freight Analysis Framework (FAF); Commodity Flow Survey (CFS); TxDOT Highway Performance Monitoring System (HPMS); three private datasets; and National Corridors Analysis and Speed Tool (NCAST).

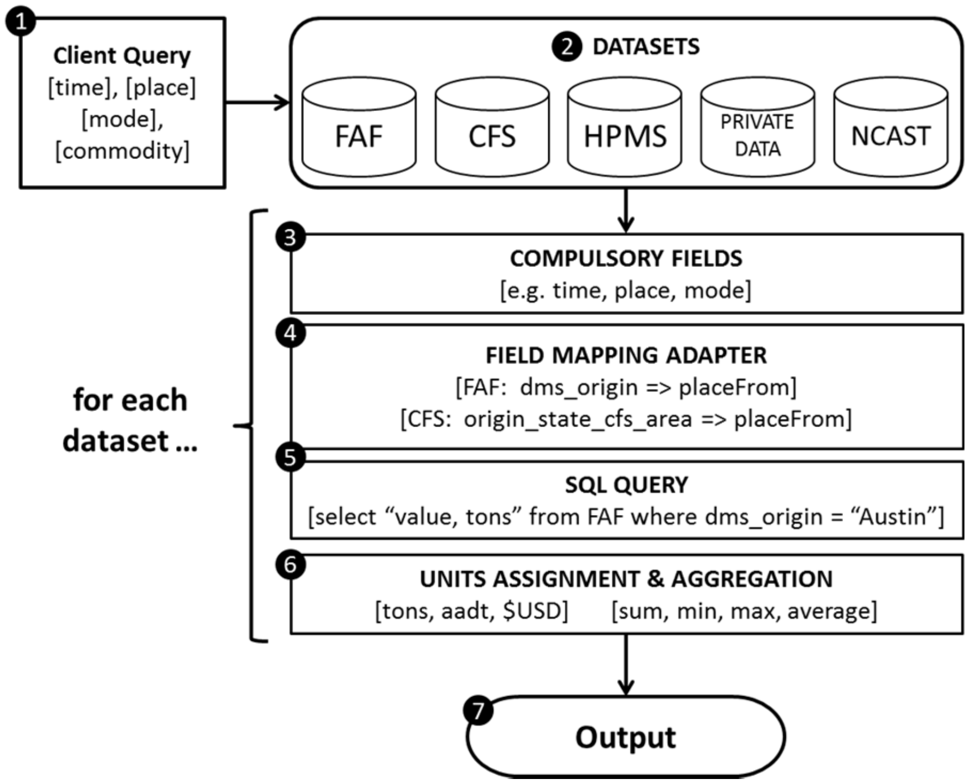


Figure 7.3: Proposed Mediator Architecture

7.2.2 Client Queries

The study team recommends that client queries should have the following minimum options available to users as tested in the prototype system (see Figure 7.4):

1. **Time**—For the prototype system, this was a date field that provided users with the option of querying data *from* a specific time frame or *limited to* a specific time period, e.g.,

---

[v] The term *dataset* is used henceforth to refer to a single freight database table used in the development of the prototype integrated system. A single dataset is made up of columns that represent variables in that dataset and rows that represent records contained in the dataset.

"show all records from January 2004" or "show all records from January 2004 to August 2012".

2. **Place**—The place option is broken down into two main options:
   a. Geographical Location—when a user chooses to search for data only for a specific geographical location, e.g., "show all inbound and outbound records relating to the city of Austin."
   b. O-D—when a user chooses to search for data relating to freight movement from an origin to a destination, e.g., "show all records for which origin equals Austin and destination equals Dallas".

3. **Mode**—Users should have the capability to filter records based on a specific mode of transport, e.g., air, marine, truck, pipeline, or rail. An additional option is the ability to select features unique to a specific mode, e.g., highway name for trucks, number of rail cars for rail, number of vessels for marine, etc.

4. **Commodity**—Users should have the option of filtering between multiple commodities.

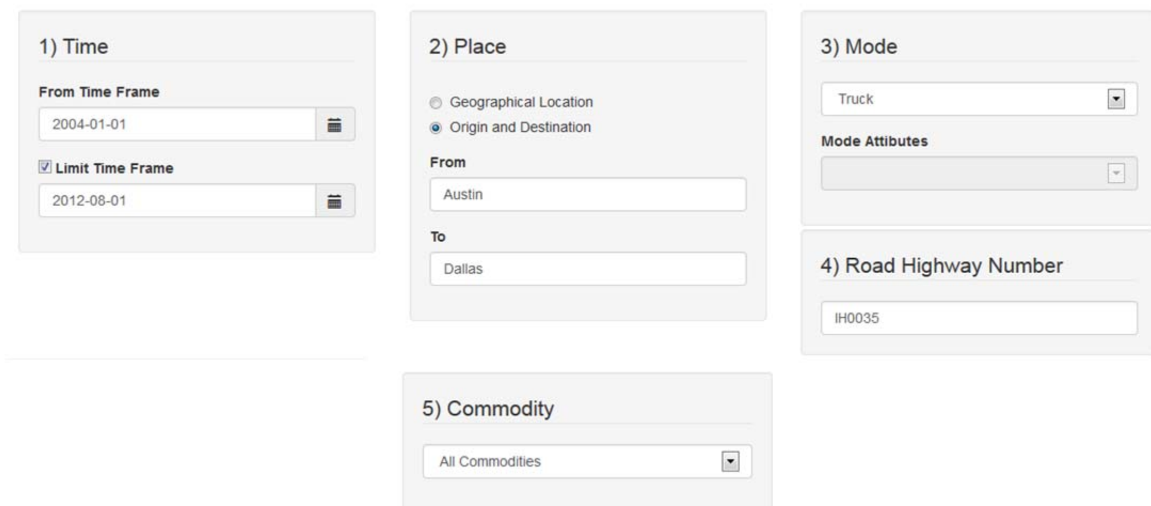


Figure 7.4: Example screenshot of minimum available query options

7.2.3 Datasets

In order to keep the integrity of the datasets and facilitate future updates of existing datasets, the study team proposes that the data dictionary structure of datasets be kept the same as the original. Field mapping adapters are then linked to these datasets when users submit queries.

For datasets referencing similar lookup tables, a single lookup table is recommended. For example, both the FAF and the CFS use the SCTG classification system. Therefore, linking both datasets to a single SCTG table is much more efficient than having duplicate tables (see Figure 7.5).

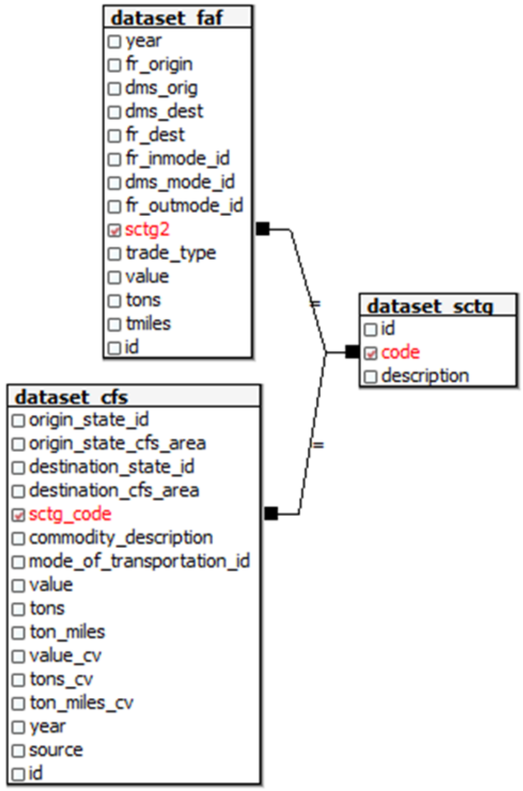


Figure 7.5: Example showing FAF and CFS datasets referencing a common SCTG lookup table

7.2.4 Compulsory Fields

The compulsory fields' layer provides an additional layer for speeding up query processing time. The purpose of this layer is to prevent the system from querying data tables for which the minimum required search options are not specified by the user. For example, for most datasets time and place are required. Therefore, if the user fails to specify a time option during the querying phase, then the system will alert the user that "time" is a required input to query a certain dataset. Another example, as illustrated in Figure 7.6 where the user specifies only a time frame ("timeFrom" and "timeTo"), the system notifies the user that in order to query the FAF and CFS datasets, the "placeFrom" variable is required, and in order to query the HPMS and NCAST datasets, the roadway highway number (or "link") is required.

Input:

Query Search for: **timeFrom**:2004-01-01 **timeTo**:2012-08-01

Output

FAF : **Needed at least timeFrom, placeFrom**

CFS : **Needed at least timeFrom, placeFrom**

HPMS : **Needed at least link**

NCAST : **Needed at least timeFrom, link**

Figure 7.6: Compulsory Fields Query Input and Output Illustration

The compulsory field rules can be stored in a data table was illustrated in Figure 7.7, where *dataset_id* is a reference to the datasets being used in the system.

id [PK] serial	dataset_name character varying(256)	is_active boolean
1	dataset_faf	TRUE
2	dataset_ncast	TRUE
3	dataset_cfs	TRUE
4	dataset_hpms	TRUE
5	dataset_pvtdata1	TRUE
6	dataset_pvtdata2	TRUE
7	dataset_pvtdata3	TRUE

id [PK] serial	dataset_id integer	compulsory_field character varying(256)	is_active boolean
1	1	timeFrom	TRUE
2	3	timeFrom	TRUE
3	2	link	TRUE
4	2	timeFrom	TRUE
5	4	placeFrom	FALSE
7	1	placeFrom	TRUE
8	3	placeFrom	TRUE
10	4	link	TRUE

Figure 7.7: Compulsory Field Table

7.2.5 Field Mapping Adapter

The field mapping adapter relates the user-specified query fields to the original fields for each dataset. The structure of the field mapping adapter is shown in Figure 7.8 and as follows:

i) *dataset_id* (Foreign key): A reference to the existing dataset as shown in Figure 7.7

ii) *original_field* (Text): The original field name as specified in the dataset

iii) *mapping_field* (Text): The user specified field variable, e.g., timeFrom, timeTo, commodity, mode, placeFrom, placeTo

iv) *is_active* (Boolean): Determines if the field is still available in the dataset

v) *conversion_factor* (Numeric): Used if fields are known to report similar values but in different units, e.g., dollars and million-dollars.

vi) *human_readable_name* (Text): This field displays the human-readable text of a field name, e.g., "dms_orig" in FAF will have a human-readable name of "Domestic Origin FAF Area".

vii) *unit_name* (Text): Field unit of measurement to be used in reporting and conversions

viii) *is_summable, is_averageable, is_minmax* (Boolean fields): These fields are used in the units assignment and aggregation stage

ix) *field_type* (Text): This variable is used in specifying whether a field is an input field or an output field, i.e., if an input field, then the field can be used in querying the dataset and if any output field then the field is only useful for showing querying results.

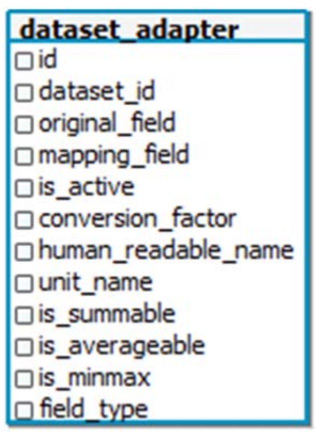


Figure 7.8: Field Mapping Adapter

Figure7.9 shows sample records of the field mapping adapter table.

id [PK] s	datas intege	original_field character varying(256)	mapping_field character varying	is_active boolean	conversion double pre	human_readable_ character varying(	unit_name character vary	is_summable boolean	is_average boolean	is_minmax boolean	field_type character va
10	3	mode_of_transportation__p	mode	TRUE	1	''	''	FALSE	FALSE	FALSE	IN
11	3	sctg_code	commodity	TRUE	1	Commodity	''	FALSE	FALSE	FALSE	IN
12	3	commodity_description	commodity	TRUE	1	Commodity	''	FALSE	FALSE	FALSE	IN
13	3	tons	commodity	TRUE	1	''	tons	TRUE	FALSE	FALSE	OUT
14	3	value	commodity	TRUE	1	''	milliondolla	TRUE	FALSE	FALSE	OUT
15	3	ton_miles	commodity	TRUE	1	''	''	FALSE	TRUE	TRUE	OUT
16	3	value_cv	commodity	TRUE	1	''	''	FALSE	FALSE	FALSE	''
17	3	tons_cv	commodity	TRUE	1	''	''	FALSE	FALSE	FALSE	''
18	3	ton_miles_cv	commodity	TRUE	1	''	''	FALSE	FALSE	FALSE	''
19	1	fr_origin__area_name	place	TRUE	1	''	''	FALSE	FALSE	FALSE	IN
20	1	fr_origin__area_name	placeFrom	TRUE	1	''	''	FALSE	FALSE	FALSE	IN
21	1	dms_orig__area_name	place	TRUE	1	Domestic Origin	''	FALSE	FALSE	FALSE	IN
22	1	dms_orig__area_name	placeFrom	TRUE	1	Domestic Origin	''	FALSE	FALSE	FALSE	IN
23	1	dms_dest__area_name	placeTo	TRUE	1	''	''	FALSE	FALSE	FALSE	IN
24	1	fr_dest__area_name	placeTo	FALSE	1	''	''	FALSE	FALSE	FALSE	IN
25	1	fr_inmode__mode__parent	mode	TRUE	1	''	''	FALSE	FALSE	FALSE	IN
26	1	dms_mode__mode__parent	mode	TRUE	1	''	''	FALSE	FALSE	FALSE	IN
27	1	fr_outmode__mode__parent	mode	TRUE	1	''	''	FALSE	FALSE	FALSE	IN
28	1	year__gte	timeFrom	TRUE	1	''	''	FALSE	FALSE	FALSE	IN
29	1	year__lte	timeTo	TRUE	1	''	''	FALSE	FALSE	FALSE	IN
30	1	sctg2	commodity	TRUE	1	''	''	FALSE	FALSE	FALSE	IN
31	1	tons	commodity	TRUE	1	''	tons	TRUE	FALSE	FALSE	OUT
32	1	value	commodity	TRUE	1	''	milliondolla	TRUE	FALSE	FALSE	OUT
33	1	tmiles	commodity	TRUE	1	''	''	FALSE	TRUE	TRUE	OUT
34	1	trade_type	commodity	TRUE	1	''	''	FALSE	FALSE	FALSE	INOUT
35	2	object_id	unclassifiables	TRUE	1	''	''	FALSE	FALSE	FALSE	''
36	2	cal_date__gte	timeFrom	TRUE	1	''	''	FALSE	FALSE	FALSE	IN
37	2	cal_date__lte	timeTo	TRUE	1	''	''	FALSE	FALSE	FALSE	IN
38	2	timezone	time	TRUE	1	''	''	FALSE	FALSE	FALSE	''
39	2	state__name	place	TRUE	1	''	''	FALSE	FALSE	FALSE	IN
40	2	hwy_link	link	TRUE	1	''	''	FALSE	FALSE	FALSE	INOUT
41	2	hwy_num	link	TRUE	1	''	''	FALSE	FALSE	FALSE	''
42	2	mile_seg	link	TRUE	1	''	''	FALSE	FALSE	FALSE	''
43	2	hwy_dir	link	TRUE	1	''	mph	FALSE	FALSE	FALSE	OUT
44	2	seg_id	link	TRUE	1	''	''	FALSE	FALSE	FALSE	''
45	2	length	link	TRUE	1	''	''	FALSE	FALSE	FALSE	''
46	2	am_peak	link	TRUE	1	''	mph	FALSE	FALSE	TRUE	OUT
47	2	midday_peak	link	TRUE	1	''	mph	FALSE	FALSE	TRUE	OUT

Figure 7.9: Field Mapping Adapter Table Sample

7.2.6 SQL Query

The SQL query layer uses the mapping scheme developed in the field mapping adapter above. Dataset fields that map to the user-specified variables (see *mapping_field* column) are used in querying each dataset and fields with a field type equal to "OUT" form the query results.

Under the current setup, if two dataset fields map to the same *mapping_field* and both are of the field type "IN", both of the fields will be searched for as both fields can contain different values. For example, should the user search for "Texas" using the "placeFrom" variable, only the "origin_state__name" field under CFS will return results even though both the "origin_state__name" and "origin_state_cfs_area" both map to the "placeFrom" field.

7.2.7 Units Assignment and Aggregation

The units assignment and aggregation layer is used in performing simple arithmetic operations to be used in providing summary data. Arithmetic operations include summation and minimum and maximum value determinations.

7.2.8 Output

As discussed earlier, final output from the database can be in the form of

i) summary data as discussed in the units assignment and aggregation section,
ii) ability to download the complete queried data, and
iii) ability to display data in the form of graphs, tables, or online maps for visualization purposes.

7.3 Cost Estimation

Following the processes earlier outlined in the conceptual architecture for an integrated freight database system, the study team proposed a list of items to be considered in estimating the cost for developing and maintaining the system. The task breakdowns and associated unit cost measurements are shown in Table 7.1.

Table 7.1: Integrated Freight Database System Architecture Unit Cost Estimates

Tasks	Unit Costs
1. Project Initialization and Setup	
1.1. Finalize agreements on design specifications and needs with TxDOT	200 man-hours
2. Integrate and use publicly available data	
2.1. Inventory data dictionary elements from publicly available databases	300 man-hours
2.2. Design and test theoretical field mapping adapter that connects to each data dictionary element	200 man-hours
3. Engage private sector stakeholders	
3.1. Correspondence with existing and new freight data partners	40 hours/week
4. Electronic submission of data by freight data sharing partners	
4.1. Design initial system administrator interface	300 man-hours
4.2. Develop and test electronic form submission including secure web connections	200 man-hours
5. Data quality and validation	
5.1. Setup automated data quality checks	400 man-hours
5.2. Conduct manual data quality checks	40 hours/week
6. Automated data scrubbing and aggregation	
6.1. Develop universal data scrubbing adapter	400 man-hours
6.2. Develop universal data aggregation adapter	800 man-hours
6.3. Develop partner specific data scrubber, data aggregator, and field mapping adapters	40 man-hours per data partner
7. Secure data storage and restricted access	
7.1. Acquire secure data warehouse and web servers	Lump sum
7.2. Develop restricted user management system	200 man-hours
7.3. Develop interface to store publicly available data	400 man-hours
7.4. Develop interface to store private sector data	400 man-hours
7.5. Integrate existing TxDOT data centers	300 man-hours per data center
8. Data output and analysis tools	
8.1. Develop and test client-side user interfaces including GIS map integration and graphs	700 man-hours
9. Training	
9.1. Develop training manuals for administrators and general users	300 man-hours
9.2. Conduct training sessions for administrators and general users	40 man-hours per session
10. System Maintenance	
10.1. Dedicate administrators to maintain and update system	80 hours/week

Chapter 8. Recommendations

Effective partnerships are clearly needed between the public and private sectors to ensure adequate freight planning and funding of transportation infrastructure at the state and local levels. Despite this recognition, roadblocks continue to exist in moving these partnerships forward. Traditional primary data collection methods—such as roadside intercept surveys, mail-out/mail-back questionnaires, combined telephone-mail-out/mail back questionnaires, driver trip diaries, personal interviews, etc.—when done correctly, are regarded as very reliable for obtaining freight data. However, these methods can be costly and generally involve a smaller, more select, or targeted sample that may not be appropriate for the population of statewide freight movement databases.

Progress is being made in some states and regions to cost-effectively procure more accurate truck travel data using technological applications instead of the traditional survey methods. In the case of truck data, a number of ITS technologies such as GPS are able to collect various truck travel attributes such as vehicle classification, weight, speed, delay data, incident data, traffic volume data, commodity/cargo type, payload (cargo) weight, truck O-D patterns, trip O-D patterns, average tour length, number of stops per tour, number of truck stops for LTL shipments, travel time, transit time, and travel time reliability. These current technological innovations provide an opportunity for effective data-sharing partnerships between transportation planning agencies and the private sector like never before.

Similar to guidelines outlined in *Freight Data Sharing Guidebook*, the study team recommends a rigorous outreach and follow-up effort to sustain the success of any freight data sharing partnership. The study team recommends that TxDOT use a trusted third party such as an academic institution or a consultant to spearhead the implementation and development of the integrated freight data system. Validation of this sort of partnership was evidenced in the study where stakeholders were more inclined to communicate further with the study team based on their affiliation and trust of the institution. A similar observation was made in the American Transportation Research Institute (ATRI) Freight Performance Measures project involving the Department of Transportation and the FHWA to develop a strategic freight performance tool for identifying and assessing truck mobility issues on our nation's highways (ATRI 2013). As described in *Freight Data Sharing Guidebook,* ATRI is considered a trusted third party by the trucking institution and its relationship with the trucking industry allowed sensitive data to be collected and cleansed before being used in the project. Other observations include Washington State's *GPS Truck Data Performance Measures Program* (McCormack 2011), the *Alabama Statewide Freight Study and Action Plan* involving the University of Alabama at Huntsville and the consulting firm of J. R. Wilburn and Associates (Anderson et al. 2011), and the Texas Transportation Institute's work on using RFID readers to measure wait times at U.S.-Mexico border ports of entry (TTI 2013).

In addition to TxDOT partnering with a trusted third party, the agency should be involved in promoting and advocating the initiative through its relationships with trade associations and industry experts. For example, through TxDOT's Freight Advisory Committee (FAC), private-sector partner organizations can be invited to participate in the data sharing partnership. The FAC provides a convenient avenue for both the public and private sector agencies to define and articulate the purpose, benefits, challenges, and concerns of data sharing. Issues relating to privacy concerns and how the data can be better used in improving upon TxDOT's planning efforts can be adequately addressed.

Also recommended by *Freight Data Sharing Guidebook* is the frequent communication to public and private-sector stakeholders of the potential benefits resulting from the data partnership efforts. TxDOT should follow up on planning programs or efforts being undertaken to address transportation infrastructure issues identified as a result of the data being provided. Partners should also be assured that the data being collected is used for the purposes for which it is intended and will not be used in any way that harms the operations of the partners, such as through regulatory processes. A clear non-disclosure contract is required to reinforce a sense of trust with freight partners and the sample provided in the *Freight Data Sharing Guidebook* is a good option.

Adequate funding should be allocated to sustain the program and cover cost of operations. Any data sharing partnership will require a long-term commitment from TxDOT and the partners and this can be demonstrated through allocation of adequate funding for the program. Adequate funding also ensures the collection of updated data thus making the program more relevant in addressing freight-related issues being faced by TxDOT. A lack of commitment may result in partners not renewing agreements with the agency as they may sometimes not experience any direct benefits from participating in the program.

Through the use of advanced data integration methods, it is possible to overlay publicly available multiple data sources to assist in filling some of the data gaps that currently exist. As noted in a recent National Research Council report, "when separate datasets are collected and analyzed in such a manner that they may be used together, the value of the resulting information and the efficiency of obtaining it can be greatly enhanced" (NRC, 2001). However, industry participation is still required to provide sufficient data to meet needs such as variables relating to air quality (e.g., vehicle fleet age, engine type, vehicle type, roadway speeds), service types (e.g., truckload, less-than-truckload, and just-in-time delivery), trip purpose, and actual production and attraction rates. This study recognizes that effective partnerships are needed between government and the freight community to ensure adequate planning and funding of transportation infrastructure at the state and local levels.

References

3M Company, http://multimedia.3m.com/mws/mediawebserver?6666660Zjcf6lVs6EVs6660fdCOrrrrQ-, 2011

American Transportation Research Institute, National Corridors Analysis and Speed Tool (N-CAST) Available at http://atri-online.org/n-cast/

Anderson, Michael, and Gregory Harris. "Development of Statewide Freight Plan for Alabama Using Integrated Freight Planning Framework." Transportation Research E-Circular E-C158 (2011).

ASTM E1318-94, *Standard Specification for Highway Weigh-in-Motion (WIM) with User Requirements and Test Method*, Annual Book of ASTM Standards, vol. 04.03, West Conshohocken, PA, 1994.

Barnett, Vic, and Toby Lewis. "Outliers in statistical data." *Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics, Chichester: Wiley, 1984, 2nd ed.* 1 (1984).

Beagan, Daniel, Michael Fischer, and Arun Kuppam (2007), *Quick Response Freight Manual II,* Cambridge Systematics, Inc. Sponsored by the Department of Transportation Federal Highway Administration. http://www.ops.fhwa.dot.gov/freight/publications/qrfm2/index.htm Accessed October 8, 2012.

Bureau of Transportation Statistics, 2012, http://www.bts.gov/publications/commodity_flow_survey/2007/state_summaries/html/appendix_c.html

Bureau of Transportation Statistics. *TransBorder Freight Data Program*. http://www.bts.gov/programs/international/transborder/PDF/TransBorderFreightDataProgram.pdf. Accessed October 8, 2012.

Cambridge Systematics 1997. A Guidebook for Forecasting Freight Transportation Demand. Transportation Research Board. No. Project NCFRP-388.

Cambridge Systematics, Inc. and GeoStats, LLP. (2010). "Freight-demand Modeling to Support Public-sector Decision Making." Transportation Research Board. Project NCFRP 8.

Cambridge Systematics, North River Consulting Group, and University of Washington *"NCFRP Report 25—Freight Data Sharing Guidebook"*, Project NCFRP-31 (2013). Available at http://www.trb.org/Publications/Blurbs/169010.aspx

Cambridge Systematics, North River Consulting Group, University of Washington, National Cooperative Freight Research Program, United States. Department of Transportation. Research, and Innovative Technology Administration. Freight Data Sharing Guidebook. Vol. 25. Transportation Research Board, 2013.

Center for Transportation Research, Framework for Collecting Truck Travel Data for the Dallas/Fort Worth Region, The University of Texas at Austin, January 2012

Chase, Keith M., Patrick Anater, and Thomas Phelan. 2013. Freight Demand Modeling and Data Improvement Strategic Plan. SHRP 2 Report S2-C20-RW-2.

Dean, R. B., and W. J. Dixon. "Simplified statistics for small numbers of observations." *Analytical Chemistry* 23, no. 4 (1951): 636–638.

Evans, Timothy, Laura Zayatz, and John Slanta. "Using Noise for Disclosure Limitation of Establishment Tabular Data." Proceedings of Bureau of the Census 1996 Annual Research Conference. US Bureau of the Census, Washington, DC. 1996.

Federal Highway Administration (2010), *The Freight Analysis Framework, Version 3: Overview of the FAF3 National Freight Flow Tables*. http://faf.ornl.gov/fafweb/Data/FAF3ODCMOverview.pdf. Accessed October 8, 2012.

Federal Highway Administration (2011a). *FHWA Vehicle Types,* http://www.fhwa.dot.gov/policy/ohpi/vehclass.htm. Accessed October 8, 2012.

Federal Highway Administration (2011b). *Freight Management and Operation, Chapter 4: Freight Truck Assignment and Calibration*. http://faf.ornl.gov/fafweb/Data/Freight_Traffic_Analysis/chap4.htm#43. Accessed October 8, 2012.

FHWA, A Summary of Vehicle Detection and Surveillance Technologies used in Intelligent Transportation Systems, 2000, http://www.fhwa.dot.gov/ohim/tvtw/vdstits.pdf

FHWA, Evaluation of the Commercial Vehicle Information Systems and Networks (CVISN) Model Deployment Initiative, March 2002, http://ntl.bts.gov/lib/jpodocs/repts_te/13677.html

FHWA, Public Roads, Sept/Oct 1997, http://www.fhwa.dot.gov/publications/publicroads/97septoct/p97sept30.cfm

FHWA, Travel Time Data Collection Handbook, March 1998, http://www.fhwa.dot.gov/ohim/handbook/chap4.pdf

Florida DOT, Traffic Monitoring Handbook, http://www.dot.state.fl.us/planning/statistics/tmh/ch6.pdf

Greaves, S. and M. Figliozzi, Urban Commercial Vehicle Tour Data Collection Using Passive GPS Technology: Issues and Potential Applications, 2008, Transportation Research Record 2049, p.158-166

Grubbs, Frank E. "Procedures for detecting outlying observations in samples." *Technometrics* 11, no. 1 (1969): 1–21.

HowStuffWorks, How Red-light Cameras Work, http://auto.howstuffworks.com/car-driving-safety/safety-regulatory-devices/red-light-camera1.htm

IBM. "Types of constraints." IBM DB2 9.5 Information Center for Linux, UNIX, and Windows. http://publib.boulder.ibm.com/infocenter/db2luw/v9r5/index.jsp?topic=/com.ibm.db2.luw.admin.dbobj.doc/doc/c0020149.html accessed December 4, 2013).

IHS Global Insight, *Transearch,* http://www.ihs.com/products/global-insight/industry-analysis/commerce-transport/database.aspx. Accessed October 10, 2012

Inbound Logistics, 2012, http://www.inboundlogistics.com/cms/logistics-glossary/#C

Intelligent Transportation Systems Joint Program Office: Knowledge Resources—Cost, http://www.itscosts.its.dot.gov/

Liao, Chen-Fu. *Fusing Public and Private Truck Data to Support Regional Freight Planning and Modeling*. University of Minnesota, Minnesota Traffic Observatory. 2010.

Liao, Chen-Fu. *Using Archived Truck GPS Data for Freight Performance Analysis on I-94/I-90 from the Twin Cities to Chicago*. Minnesota Traffic Observatory, University of Minnesota. 2009. http://conservancy.umn.edu/bitstream/97668/1/CTS%2009-27.pdf

Mani, Akshay, and Prozzi, Jolanda (2004). "State-Of-The-Practice in Freight Data: A Review of Available Freight Data in the U.S.". Report No. 0-4713-P2. Center for Transportation Research. The University of Texas at Austin

Martin, Peter T., et al. *Detector Technology Evaluation*. University of Utah Traffic Lab. 2003.

McCormack, Edward and Mark E. Hallenback. *ITS Devices Used to Collect Truck Data for Performance Benchmarks.* Transportation Research Record: Washington, D.C. 2006.

McCormack, Edward and Mark E. Hallenbeck. *Options for Benchmarking Performance Improvements Achieved from Construction of Freight Mobility Projects*. Washington State Transportation Center, University of Washington. 2005.

McCormack, Edward, et al. *Developing a GPS-based Truck Freight Performance Measure System.* WSDOT, Washington State Transportation Center. 2010.

McCormack, Edward, Wenjuan Zhao, and Dale Tabat. "GPS Truck Data Performance Measures Program in Washington State." (2011).

Measuring Personal Travel and Goods Movement: A Review of the Bureau of Transportation Statistics' Surveys—Special Report 277 (2003), Committee on National Statistics (CNSTAT) Transportation Research Board (TRB), Page 111

METRANS (2011). *Using Commercial GPS Data to Quantify Truck Performance*. http://www.metrans.org/nuf/2011/documents/Presentations/McCormack-gps.pdf. Accessed October 8, 2012.

National Research Council. *Principles and Practices for a Federal Statistical Agency*, 2nd ed. National Academy Press, Washington, D.C., 2001.

Nu-Metrics - http://www.fhwa.dot.gov/policyinformation/pubs/vdstits2007/04pt2.cfm

Oracle. "Maintaining Data Integrity in Database Applications." Oracle Database Advanced Application Developer's Guide. http://docs.oracle.com/cd/E11882_01/appdev.112/e25518/adfns_constraints.htm (accessed December 4, 2013).

PrePass, 2012, http://www.prepass.com/services/prepass/pages/whatisprepass.aspx

Prozzi, Jolanda, Dan Seedah, Migdalia Carrion, Ken Perrine, Nathan Hutson, Dr. Chandra Bhat, Dr. C. Michael Walton (2011) *Freight Planning for Texas—Expanding the Dialogue*. Center for Transportation Research: Austin, TX. FHWA/TX-11/0-6297-1

Prozzi, Jolanda, Seedah, Dan, Carrion, Migdalia, Perrine, Ken, Hutson, Nathan, Bhat, Chandra, Walton, Michael. (2011). "Freight Planning for Texas—Expanding the Dialogue." Report No. 0-6297-1. Center for Transportation Research. The University of Texas at Austin.

Prozzi, Jolanda, Wong, Carleton, Harrison, Rob. 2004. Texas Truck Data Collection Guidebook. Center for Transportation Research. The University of Texas at Austin.

Quiroga, Cesar, Koncz, Nicholas, Kraus, Edgar, Villa, Juan. Warner, Jeffrey, Li, Yingfeng, Winterich, David, Trego, Todd, Short, Jeffrey, and Ogard, Elizabeth. 2011. Guidance for Developing a Freight Transportation Data Architecture. Transportation Research Board. Report NCFRP 9

Rhodes, Suzann S., Berndt, Mark, Bingham, Paul, Bryan, Joe, Cherrett, Thomas J., Plumeau, Peter, and Weisbrod, Roberta. 2012. Guidebook for Understanding Urban Goods Movement. Transportation Research Board. No. Project NCFRP 15A.

Ross, Stephen M. "Peirce's criterion for the elimination of suspect experimental data." *Journal of Engineering Technology* 20, no. 2 (2003): 38–41.

SmarTek Systems Inc., 2012 http://www.smarteksystems.com/SASpad.html

Surface Transportation Board. *Economic Data: Waybill*. http://www.stb.dot.gov/stb/industry/econ_waybill.html. Accessed October 8, 2012.

Taylor, John Robert. *An Introduction Error Analysis: The Study of Uncertainties in Physical Measurements*. University science books, 1997.

Terence Fok and William Yan, A Fleet Management and Benchmarking Ontario Power Generation, University of Toronto, 2007 http://www.mie.utoronto.ca/undergrad/thesis-catalog/files/341.pdf

Texas Department of Motor Vehicles (DMV) (2013), TxPROS Information & Tutorials, http://txdmv.gov/motor-carriers/oversize-overweight-permits/txpros-info-and-tutorials

Texas Transportation Institute, Using RFID Readers to Measure Wait Times at the U.S.-Mexico Border. Available at http://tti.tamu.edu/2013/03/01/using-rfid-readers-to-measure-wait-times-at-the-u-s-mexico-border/

Texas Department of Transportation (2013a), Statewide Planning Map, http://www.txdot.gov/apps/statewide_mapping/StatewidePlanningMap.html

Texas Department of Transportation (2013b), DriveTexas, http://drivetexas.org

Texas Department of Transportation (2013c), Texas Statewide Analysis Model—Third Version (SAM-V3)—Overview, Published by the Texas Department of Transportation

Tok, Andre, Miyuan Zhao, Joseph YJ Chow, Stephen Ritchie, and Dmitri Arkhipov. 2011. Online Data Repository for Statewide Freight Planning and Analysis. Transportation Research Record: Journal of the Transportation Research Board 2246 (1), 121-129.

Transportation Research Board (2003). "A Concept for a National Freight Data Program". Committee on Freight Transportation Data: A Framework for Development. Special Report 276.

U.S. Census Bureau, County Business Patterns—How the Data are Collected (Coverage and Methodology) http://www.census.gov/econ/cbp/methodology.htm

Wikipedia contributors, "Data cleansing," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=Data_cleansing&oldid=584731700 (accessed December 4, 2013).

Appendix A: List of Available Freight Databases

DB Name	Sponsor	Performing Organization	Sources	Commodities	O-D Flows	Routing	Shipment	Mode
Airport Activity Statistics of Certificated Route Air Carrier	USDOT	USDOT	Large Certificated Air Carrier	N/A	Enplanement airport only	N/A	Enplanement tons: Freight and Mail	Air
Carload Waybill Sample	Surface Transportation Board	Association of American Railroads	Carload waybills	5-digit Standard Transportation Commodity Codes (STCC) level	Business Economic Area (BEA) codes of O-D of rail shipments	• Interchange states • Railroad and station itinerary	• Tonnage • No. of carloads • Waybill date • Freight revenue • Trailers and containers	Rail
Commodity Flow Survey (CFS)	• Bureau of Transportation Statistics (BTS), USDOT • Bureau of the Census, US Dept. of Commerce (USDC)	• Census Bureau, USDC • Oak Ridge National La (ORNL)	Sample of establishment selected from the Standard Statistical Establishment List (SSEL)	• 5-digit Standard Classification of Transported Goods (SCTG) • Hazardous materials (Hazmat UN/NA codes)	• State • National Transportation Analysis Regions (NTARs) • Foreign country for exports	Port of exit for exports	• Weight • Value • Containerized shipments	• Air • Truck (private & for-hire) • Rail • Inland waterway • Vessels • Pipeline • Surface/air parcel
Exports from Manufacturing Establishments	Bureau of the Census, USDC	Bureau of the Census, USDC	• Annual Survey of Manufacturers (ASM) • US Exports of Merchandise DB • Bureau of Economic Analysis (BEA) I/O accounts of U.S. economy	3-digit North American Industry Classification System (NAICS) codes	State of production	N/A	Free-on-board shipment value for Direct and Indirect Export	All
Freight Analysis Framework (FAF)	• Federal Highway Administration (FHWA) • BTS, USDOT	• Cambridge Systematics • Oak Ridge National Lab • MacroSys Research and Technology • Global Insight • Battelle	• 2002 CFS • Foreign Waterborne Cargo data • Carload Waybill Sample data	2-digit SCTG level	• O, O state/ region (S/R), D, D S/R • 138 O-D regions	TransCAD4.8 was used to assign truck movements on the highway network	• Value • Shipment weight (tons)	• Truck • Rail • Water • Air • Truck-rail intermodal • Parcel, truck-water, & water-rail, pipeline

DB Name	Sponsor	Performing Organization	Sources	Commodities	O-D Flows	Routing	Shipment	Mode
Freight Commodity Statistics (FCS)	AAR	AAR	Commodity Statistics Reports filed with the STB by Class 1 railroads	• 2,3,4,5-digit STCC • Summary 'T' codes for each STCC level of detail	N/A	N/A	No. of carloads and tonnage for originating, terminating, and total freight: U.S. Eastern and Western districts	• Rail • Intermodal (STCC 46)
Fresh Fruit and Vegetable Shipment by Commodities, State, & Months	US Dept. of Agriculture (USDA)	USDA	• Federal Marketing Order Administrative Committees • Federal State Inspection Services (FSIS) • Shippers • Transportation Agencies	• Individual fruits and vegetables • Grouping of minor commodities and mixed-load shipments • Domestic and export shipments	• Origin of shipments (state of origin, foreign country of origin for imports) • Domestic or export destination group	N/A	Weight (Mon, Yr)	• Domestic (refrigerated railcars, truck, trailer-on-flatcar, container-on-flatcar), air, water • Total of all modes
Grain Transportation Report	USDA	Agricultural Marketing Service, USDA	• Shippers • AAR for Class 1 rail shipments • US Army Corps of Engineer (USACE) • Federal Grain Inspection Service • St. Lawrence Seaway Authority	• Total grain • Individual statistics for wheat, corn, and soybeans for exports and barge movements	N/A	• U.S. coastal region for exports (Pacific, Mississippi, Texas, total for all coasts) • "Inland river lock" for barge shipments	• Volume of export shipments inspected • Carloads for rail deliveries to ports by coast • Shipments tons (barge movements, export sales)	• Truck • Rail • Waterway
Less Than Truckload (LTL) Commodity and Market Flow Database	American Trucking Association (ATA)	Martin Labbe Association (MLA)	LTL carriers	• Standard delivery time • Nonstandard delivery time • Special equipment or handling	• 3-digit to 3-digit zip code O-D detail for domestic • Foreign area (Canada, Mexico, Asia, Europe, etc.)	Mileage	• Weight • Volume • No. of shipments • No. of pieces	• LTL • Intermodal • Interline

DB Name	Sponsor	Performing Organization	Sources	Commodities	O-D Flows	Routing	Shipment	Mode
National Transportation Statistics (NTS)	BTS, USDOT	• Volpe National Transportation System Center, USDOT • BTS, USDOT	Standard data sources for each mode from federal gov. agencies, private industries, and associations are used by BTS	N/A	N/A	N/A	Total shipment traffic by each mode of transportation	• All • Vehicle miles of travel • Passenger-and ton-miles by mode
Port Import/Export Reporting Service (PIERS)	United Business Media, Inc.	PIERS reporters at U.S. Custom Houses	Freight vessel manifests	• 6-digit Harmonized System Commodity classification (HS) • Actual bill of landing/manifest commodity description • PIERS 7-digit COMCODE classification	• U.S. shipper/consignee name for export shipments • Foreign shipper name for U.S. imports • U.S. city of O-D • Foreign country of import • Foreign country and city for export shipments	• U.S. port of loading/ unloading • Foreign port of final O-D • Foreign port of transshipment	• Weight • Value	Waterborne
Quarterly Coal Report	US Dept. Of Energy	Energy Information Administration (EIA), USDOE	• Manufacturers consuming coal, coke plants, coal producers, and distributors • US Bureau of the Census, Foreign Trade Statistics	Coal	• Origin (state, foreign country) • Destination (foreign country, U.S. and Canadian sectors)	Customs District for imports and exports	Weight	• Rail • Truck • Inland waterway • Ocean port • Great Lakes • Slurry
Ship Movements Database	Lloyd's Maritime Information Services (LMIS), Inc.	Lloyd's agents stationed at over 5,000 ports worldwide	Merchant vessels engaged in international trade movements	N/A	Port-to-Port	Vessel itineraries based on port-to-port movements	Total shipment tonnage (inferred from vessel capacity and load factor)	Vessel (arrival and departure dates)
State Estimates of Truck Traffic	FHWA, USDOT	State and local highway agencies	Vehicle counts collected by state and local highway agencies	N/A	N/A	N/A	N/A	Truck

DB Name	Sponsor	Performing Organization	Sources	Commodities	O-D Flows	Routing	Shipment	Mode
Transborder Surface Freight Database	• Bureau of the Census, USDC • Federal Railroad Administration (FRA)	• US Customs Service and US Bureau of the Census • BTS	US Bureau of the Census Foreign Trade Statistics program	2-digit Harmonized Schedule commodity codes	• U.S. states by 2-digit postal codes • Canadian province codes • Mexican state codes	US Customs District and port of entry and exit for U.S. imports and exports	• Value • Weight (Imports only) • Containerized designation (U.S. imports only)	• Surface (mail, truck, rail, pipeline) • Others
Transearch	IHS Global Insight	IHS Global Insight	• CFS • ICC Carload Waybill Sample • Corps of Engineers Waterborne Commerce Statistics • FAA Airport Activity Statistics • Census of Transportation—Commodity Transportation Survey (1977) • Bureau of Census Foreign Trade Statistics • AAR Freight Commodity Statistics • Inter-industry trade patterns • Motor Carrier Data Exchange (MCDE)	4-digit STCC (5-digit available for rail and water)	• State • 183 U.S. BEAs • Canadian province data • County or zip-code level	Highway routings imputed from O-D data	• Total weight (tonnage) • Value • No. of loads	• Truck • Rail, domestic air, and domestic waterborne • Intermodal

DB Name	Sponsor	Performing Organization	Sources	Commodities	O-D Flows	Routing	Shipment	Mode
Transportation Annual Survey (Motor Freight Transportation/ Warehousing Survey)	US Bureau of the Census, USDC	US Bureau of the Census, USDC	Commercial Motor Freight Transportation and Public Warehousing Service companies	Primary commodities handled (percentage of annual operating revenue)	• U.S.–U.S. • U.S.–Canada • Canada–U.S. • Other	N/A	LTL shipments	Truck
US Air Carrier Traffic and Capacity Data by Nonstop Segment and On-Flight Market	Federal Aviation Administration (FAA)	• Office of Airline Information (OAI), Bureau of Transportation Statistics • FAA	From 41 Large Certificated Air Passenger Carriers	N/A	N/A	• Nonstop segments • On-flight market	• Nonstop segment: Revenue freight tonnage by carrier and equipment type • On-flight market: enplaned freight and mail tonnage by carrier	• Nonstop segment (departures, aircraft hours) • Carrier • Equipment type
US Air Freight Origin Traffic Statistics	Colography Group, Inc.	Colography Group, Inc.	• Government data sources: County Business Pattern Surveys by the USC for county industry location data • Colography surveys: establishments of firms generating air traffic shipments and air cargo shippers	4-digit Standard Industrial Classification (SIC) codes	• U.S. state, county, and "market area" of origin of air cargo shipments • Destination area reported as domestic of foreign shipment	N/A	• Standard air freight industry classifications • Annual data (weight, value, no. of shipments)	Air
US Exports and Imports Transshipped via Canadian Ports (Annual Report)	Maritime Administration (MARAD), USDOT	• US Customs Service and US Census Bureau • MARAD, USDOT	US Bureau of the Census EA-622 Annual Export tape and IA245 Annual Import tape	4-digit HS commodity codes	• Foreign country of O-D • US Customs Districts of entry/exportation along or in close proximity to the Canadian border	N/A	• Value • Estimated weight	Surface modes (combined)

DB Name	Sponsor	Performing Organization	Sources	Commodities	O-D Flows	Routing	Shipment	Mode
US Exports by State of Origin of Movement ("MISER" State of Export)	Massachusetts Institute for Social and Economic Research (MISER), Univ. of Massachusetts, Amherst	MISER	Magnetic Tapes EQ912 and EA917 (U.S. exports of domestic and foreign merchandise by state of origin) developed by US Bureau of the Census	2-digit SIC codes	• State of origin of movement of export shipment • Foreign country of destination	N/A	• Total value (all modes combined) • Total value and weight (vessel and air modes) • Containerized weight and value (vessel and air)	• Vessel • Air • All other
US Exports of Domestic and Foreign Merchandise by State/Region/Port (State of Export Tapes)	US Bureau of the Census—Data User Services division	US Customs Service	Merchandise export shipment data collected by US Customs at ports of exportation from export shippers	• EQ912 and EA917 (2-digit SIC) • EQ932 and EA937 (4-digit SITC-R3)	• State of origin of export shipment • Region of origin of export shipment • Foreign country of destination	US Customs Districts and Ports of Exportation	• Total value (all modes) • Total value and weight (vessel and air) • Containerized weight and value (vessel and air)	• Vessel • Air • All other
US Imports/Exports of Merchandise on CD-ROM	US Bureau of the Census—Data User Services division	US Customs Service (Automated data collection systems, Entry summary and Shipper's Export Declaration forms)	Import and export shipment data collected by US Customs at ports of entry and export from Customs Brokers and Shippers	• 10-digit Harmonized System (HS) code classification • SIC • Bureau of Economic Analysis (BEA) end-use commodity category • USDA product codes • Advanced Technology Codes (ATCs)	• Export country of destination • Export country of origin	• US Customs District of exit for exports • US Customs Districts of unlading and entry for imports	• Value and quantity (All modes, vessel and air separately) • Shipping weight (Vessel and air)	• Vessel • Air • All modes

DB Name	Sponsor	Performing Organization	Sources	Commodities	O-D Flows	Routing	Shipment	Mode
US Waterborne General Imports (Exports) and Inbound (Outbound) Intransit Shipment	Bureau of the Census, USC	• US Bureau of the Census • US Customs Service	• US Customs Entry Summary (Form 7501) for import shipments • Shipper's Export Declarations (SEDs) for export shipments	SITC and 6-digit HS commodity codes	Foreign country of O-D	US Customs Ports (USBOC Schedule D classification) to Foreign Ports (USBOC Schedule K classification)	• Weight • Value	Vessel
Vehicle Inventory and Use Survey (VIUS)	Bureau of the Census, USC	Bureau of the Census, USC	Vehicle Inventory and Use Survey questionnaire forms (TC 9501/9502) sent to a random sample of truck owners	Annual mileage reported as percentage values for the following classifications	N/A	N/A	N/A	Truck
Waterborne Commerce and Vessel Statistics	USACE	USACE	• Domestic freight carriers who report their vessel operations and cargo activity directly to the COE in the form of Vessel Operation Reports • US Bureau of the Census: US Waterborne Exports and General Imports	4-digit Commodity Classification (domestic waterborne trade flows)	• Master tape: dock-to-dock flows • Port summary report: port throughput, harbor throughput, or channel segment throughput • Public domain database: states of origin and destination of domestic shipments	Only available for a few specific route elements	Weight (tons)	Waterborne
Worldwide (North American) Airport Traffic Report	• Airports Council International (ACI) • ACI for worldwide Airport Traffic Data	ACI	Airports responding to the annual survey conducted by ACI	• Freight • Mail	Domestic/ International flights	Airport	Total weight by direction (enplanement, deplanement)	Air

Appendix B: Sample Workshop Agenda

Agenda

Integration of Data to Optimize
Freight Transportation in Texas
Thursday, April 12, 2012
Dallas, Texas

10:00	**Welcome**
10:10	**Introduction/Study Background**
10:40	**Workshop Objectives**
10:50	**Available Freight Data Sources**
12:00	**Lunch**
1:30	**Freight Data Needs and Usage**
2:30	**Workshop Feedback Survey**

Workshop Feedback

A total of 39 participants attended the five workshops hosted as part of the Integration of Data Sources Project. Workshop participants were asked to complete a Workshop Feedback survey that was used to improve subsequent workshops. Of the 39 workshop participants, 32 completed and submitted a Feedback/Comments. Table B.1 summarizes the feedback obtained by the research team.

Table B.1 Participant Feedback on Workshops

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The workshop was well organized	-	-	-	21	11
The workshop was the right length	-	-	5	16	11
The presentation(s) was clear and understandable	-	-	2	17	13
The speaker(s) were knowledgeable	-	-	-	15	17
The speaker(s) encouraged questions/comments/opinions	-	-	-	11	21
The handout/manual was useful	-	-	5	17	10
I feel this workshop was worth my time	1	-	6	17	8

Note: number of completed surveys: 32

Table B.1 indicates that the workshops were well received by the participants. Specifically:

- 100% of the workshop participants agreed or strongly agreed that the workshop was well organized;
- 84% agreed or strongly agreed that the workshop was the right length;
- 94% agreed or strongly agreed that the presentations were clear and understandable;
- 100% agreed or strongly agreed that the speakers were knowledgeable;
- 100% agreed or strongly agreed that the speakers encouraged questions/comments/opinions;
- 84% agreed or strongly agreed that the handout/manual was useful, and
- 78% of the workshop participants agreed or strongly agreed that attending the workshop was worth their time.

Participants were also asked how the workshop could be improved and were provided the opportunity to add suggestions and comments. Recurrent requests or suggestions included:

- Workshops to follow progress of the project would be desirable.

- Requests of participation from industry leaders and stakeholders.
- Requests to cover other related items. These additional items included cost or price information on the databases, some examples of freight data application using the CD, more interaction with participants, and demonstrations or examples of use or case-scenarios applications and how the database may be locally useful.
- The duration of the presentation comments ranged from the workshop only needing to be a few hour morning session to the appreciation of the pre- and post-lunch session structure. and
- It was noted by a couple of participants that data modeling and forecasting was handled by TxDOT's Transportation Planning and Programming Division and the Districts would potentially not be users of the freight database unless current policy was changed.

Appendix C: Review of Data Collection Methods Used in Currently Available Freight Data Sources

Airport Activity Statistics of Certificated Route Air Carriers

This database was developed by the U.S. Department of Transportation (USDOT), giving detailed enplanement statistics of large certificated air carriers[6] (LCAC). Airline enplanement statistics are reported only for the LCACs; accordingly, small certificated and commuter airlines and foreign-flag carriers are not included in the database. The LCACs are required by the USDOT's Office of Airline Statistics (OAS) to report their monthly enplanement activity statistics as specified in detail in the Code of Federal Regulations (CFR)—Title 14: Part 241. These air carriers supply their monthly activity data in the form of Schedule T-3 reports in Bureau of Transportation Statistics (BTS) Form 41, which has to be filed within the prescribed time period to the OAS.

Quality Control Procedures

Timely filing of Schedule T-3 reports within the prescribed time by LCACs is ensured by the USDOT's OAS and the Office of Aviation Enforcement and Proceedings (Enforcement Office). In the case of persistent delays in monthly filings by a particular air carrier, the OAS and the Office of Aviation Enforcement and Proceedings warn the carrier and may even issue orders for the payment of penalties as high as $200,000.

Limitations of Data Source

The database does not report deplanement statistics for LCACs for any airport. Also, the database does not report in-transit shipment data because only cargo enplanement (loading) data are captured in the Schedule T-3 reports. For each individual airport, cargo enplanement and aircraft activity (departures) data are reported only for the LCACs. Because a significant portion of an airport's traffic is composed of small certificated and commuter airlines as well, the airport activity statistics in the database are not representative of the overall airborne enplanement traffic at the airport.

Carload Waybill Sample

The annual Carload Waybill Sample is developed by the Association of American Railroads (AAR) under contract with the Surface Transportation Board (STB) (previously the Interstate Commerce Commission). The annual database "is a stratified sample of carload waybills for all U.S. rail traffic submitted by those rail carriers terminating 4,500 or more revenue carloads annually," (STB, 2012) maintained in a single ASCII coded data file. This database captures detailed information on total rail traffic, commodities, revenues, O-D flows, and routing information for U.S. railroad shipments.

[6] LCACs are defined as airline carriers having at least 60 seats or a payload capacity of more than 18,000 lbs.

Methodology

The data is collected by the AAR under contract with the STB. As required by the government, freight railway companies of at least 4,500 carload shipments in the past 3 years or more than 5 percent of the state's traffic have to submit carload waybill information. The carload waybills are submitted to the AAR in two formats:

- Hard copy version (*manual system*)
- Machine readable input (*computerized system*)

From these waybill records, a stratified sample is selected by the AAR to compile the Carload Waybill Sample database. The traffic and revenue values collected for the sample are then converted to annual values by using the following expansion factor:

Exact Expansion Factor = Population Count / Sample Count.

Quality Control Procedures

The machine readable input (MRI) format for filing carload waybills reduces the non-sampling errors in data reporting and facilitates easier handling and management of large fields of data for each record. The sampling rate for MRI waybills is higher than for hard copy documents, which also ensures that data processing for the development of the Carload Waybill Sample is faster and more efficient.

Limitations of Data Source

Due to the minimum threshold level (minimum number of carloads) considered in the reporting of carload waybills, some Class II and Class III railroads are often not covered in the Carload Waybill Sample. This omission may lead to sampling errors in the rail shipment data reported. Annual statistics for railroad shipments in terms of number of carloads, shipment tonnage, and revenue of railroads for each sampling category are computed using an expansion factor defined by the ratio of the population to the sample size. The expanded values obtained might not be accurate for the following reasons:

- Railroads are sampled based on the number of carloads and other statistics are not considered.
- Under each sampled category (for example, MRI waybills with more than 100 annual carloads), variations will arise—for example, revenues and tonnage across railroads that cannot be estimated accurately using the expansion factor method.

The Carload Waybill Sample does not report BEA regions of O-Ds for commodity shipments in the following cases:

- BEA regions having two or fewer freight stations
- BEA regions with less than one more freight station than the number of railroads in the BEA region

The Carload Waybill Sample often overestimates the revenues of railroads undertaking contract movements due to the expansion factor method of computation of annual revenues.

Commodity Flow Survey (CFS)

The CFS is conducted by the Census Bureau as a part of its quinquennial Economic Census to capture data on the flow of goods and materials by mode of transportation. Freight transportation characteristics have been estimated in the past in the Commodity Transportation Surveys conducted from 1963 through 1977, but the CFS—with major improvements in methodology, sample size, and scope—is designed to provide more comprehensive information on the movement of freight across the U.S. transportation network.

Methodology

Data sources include a sample of establishments selected from the Standard Statistical Establishment List (SSEL) to provide shipment data. *An establishment is defined as "a single physical location where business transactions take place*." (BTS, 2012) In particular, establishments classified as mining, manufacturing, wholesale, and selected retail industries are chosen for the survey.

Sample selection for the CFS is based on a three-stage design process.

- *First stage*: Sampling of establishments from the SSEL. For example, in the 2007 CFS, the sample consisted of around 100,000 establishments chosen from a population of close to 754,000 establishments with paid employees located in the U.S. The 2002 North American Industry Classification System was used to classify establishments.
- *Second Stage:* Selection of reporting periods in each survey year. There is a one- or two-week reporting period for each quarterly survey. For example, for the 2007 CFS, one-week reporting periods during each of the calendar quarters were employed.
- *Third Stage:* Sampling of shipments for each reporting period. If there are 40 or fewer shipments per week, then no sampling is required. If there are more than 40 shipments per week, a systematic sampling approach is used. If there are between 20 and 40 shipment samples, then sampling is based on the number of shipments per week.

Each establishment selected in the CFS is then mailed a questionnaire for each of its reporting weeks. Two versions of the questionnaire are employed to collect data from the sampled establishments:

- CFS-1000
- CFS-2000

The CFS was first conducted in 1993 as a continuation of the freight data collection surveys conducted between 1963 and 1977. Starting with the year 1997, the survey is being conducted quinquennially as a regular part of the 5-year Economic Census.

Quality Control Procedures

All the establishments in the population cannot be surveyed due to time and cost considerations. Because the CFS is carried out on a sample of establishments, the estimates of freight characteristics for the population are subject to variance depending on the sample size. To reduce the variance of estimation, establishments are sampled on the basis of a certainty/non-certainty design approach. Certainty establishments have a selection probability of 1.0 for the survey while non-certainty establishments are further stratified based on their measure of size and are sampled to meet the constraint on the coefficient of variation.

Similar to the sampling procedure for establishments, shipments are divided into *certainty* and *non-certainty* shipments based on the relative value or weight of each shipment in comparison to the rest of the reported shipments. Certainty shipments are classified as those that have a larger relative value or weight compared to other shipments and for which the survey is able to gather all the information for the entire reporting week. To arrive at an estimate of an establishment's total shipments in a given year, the following weights are applied:

- Shipment weight, shipment non-response weight, quarter weight, and quarter non-response weight
- Establishment level adjustment weight
- Establishment weight = 1 / P[Ei ≡ Sample], where P[Ei ≡ Sample] is the probability of selecting establishment *i* (E_i) in the sample based on the measure of size of the establishment. For certainty establishments, P[Ei ≡ Sample] is equal to 1.0 because they are surely included in the survey sample. For non-certainty establishments, P[Ei ≡ Sample] depends on the size of the establishments and has a higher value for larger size establishments.
- SIC-level adjustment weight

Costs of Data Collection

The FHWA initiated the funding for the CFS in fiscal year 1992. For the initial work on the 1993 CFS, the FHWA transferred $6 million to Census and $0.6 million to Oak Ridge National Laboratory (ORNL) for the freight mileage calculations. The balance was paid by the BTS from its Intermodal Surface Transportation Efficiency Act (ISTEA) budget authorizations. The total costs incurred in the 1993 CFS equaled $12.6 million from the USDOT and $3 million from the Bureau of the Census, U.S. Department of Commerce (Table C.1).

Table C.1: Survey Costs and Sample Sizes: 1993–2002 (Measuring 2003)

	1993	**1997**	**2002**
Survey Costs	$15 million	$19 million	$13 million
Sample Size	200,000	100,000	50,000

Freight Analysis Framework (FAF)

The FHWA developed the FAF to assist in the planning, operation, and management of the nation's freight transportation system. The FAF consists of an O-D database of commodity flows among regions and a network database in which freight flows are converted to truck payloads and assigned to specific routes. It also forecasts future freight flows among regions and assigns those flows to the transportation network. It integrates data from a variety of sources to create a comprehensive picture of freight movement among states and major metropolitan areas by all modes of transportation. With data from the 2007 CFS and additional sources, FAF version 3 (FAF^3) provides estimates for tonnage, value, and domestic ton-miles by region of O-D, commodity type, and mode for 2007, the most recent year, and forecasts through 2040. Also included are state-to-state flows for these years plus 1997 and 2002, summary statistics, and flows by truck assigned to the highway network for 2007 and 2040 (FHWA, 2011).

Methodology

FAF^3 involved a survey sample of 100,000 establishments from the 2007 U.S. CFS (from some 50,000 establishments in 2002). Using PIERS data, improvements were made to support estimates of the internal to the U.S. allocations of imports and exports to FAF domestic zones of freight origination (for U.S. exports) and destinations (for U.S. imports). Additional federal datasets were also incorporated using an improved FAF^3 log-linear modeling/iterative proportional fitting algorithm, as well as the development of out-of-scope estimates. Greater use of U.S. inter-industry input-output ('use' and 'make') coefficients in the development of the FAF^3 out-of-scope (to the 2007 CFS) commodity flow estimates was also employed and FAF^3 provides an O-D-specific treatment of natural gas products, which were evaluated only at the level of national or broad regional activity totals in FAF^2.

Quality Control Procedures

According to the FAF^3 documentation, the purpose of the quality assurance subtask was to manually check the accuracy of merged data using a visual inspection of a scale-based theme developed with TransCAD GIS. The approach is documented as being effective at identifying the major discrepancies between adjacent links (e.g., a significant drop of average annual daily traffic [AADT] between two adjacent highway links) or among various functional classes. For example, if the difference in traffic volume from one link to the next was greater than 20 percent, the original state traffic data collected in the FAF^2 project were consulted to verify if the accurate value had been merged to the network. If the values compared well with the state data, then the more common value for that link was used to ensure continuity in the traffic volume. These abrupt changes could also result from the merging process where aggregation was used. This smoothing process served as a reasonableness check of the traffic data merged with the network (FHWA, 2011).

Fresh Fruit and Vegetable Shipments by Commodities, States, and Months

The Fresh Fruit and Vegetable Shipments by Commodities, States, and Months database is developed by the U.S. Department of Agriculture (USDA). The database captures detailed information on the seasonal variations in domestic (intra- and inter-state), export, and import tonnages of fresh fruits and vegetables in the U.S. The shipment tonnage data are reported in terms of 100,000 lbs. units.

Methodology

Sources contributing to the database include federal marketing order administrative committees, the Federal State Inspection Service (FSIS), shippers, and transportation agencies.

The database sample reports domestic shipment data for all rail-refrigerated and piggyback shipments. The number of records of rail-refrigerated and piggyback domestic shipments represented in the database for each time period (monthly and annual) is not available. Available truck, air, and boat shipment data are reported in the database. No information on the available shipment data for truck, air, and boat modes is included. The truck shipment data for the states of Florida and Arizona are limited to only interstate shipments. Details on the sample size of shippers surveyed at shipping locations are not available. Import data from foreign countries, except Mexico, are collected from the U.S. Department of Commerce. This data pertains to the sample of import shipments (value greater than $1,501) that are reported in the U.S. Customs Entry Summary Form 7501.

Shipment data for rail and piggyback shipments of fresh fruits and vegetables are collected by the USDA from the shipment data reported by rail carriers to the STB in the line-haul revenue waybills. Domestic (inter- and intra-state movements) truck, air, and boat shipments data are collected by various agencies in each state. Statistics on the shipments of fresh fruits and vegetables by these modes are collected from the following sources: federal marketing order administrative committees, the FSIS, and state transportation agencies. Export shipment statistics for fresh fruits and vegetables are collected from the Department of Commerce. Import shipment statistics for fresh fruits and vegetables from all countries, except Mexico, are similarly collected from the import statistics reported by the U.S. Bureau of the Census. Import shipments from Mexico in terms of cross-border truck movements are collected by the Animal and Plant Health Inspection Service of the USDA.

Quality Control Procedures

The shipment statistics for fresh fruits and vegetables included in the database are gathered primarily from agencies other than the USDA, such as the U.S. Bureau of the Census, federal marketing order administrative committees, the FSIS, and the state transportation agencies. Consequently, the quality of the data is largely a function of the accuracy of the data reported by the above sources. Separate quality control procedures adopted by the USDA in their surveys of shippers of domestic shipments or in the data processing stages are, however, not available.

Grain Transportation Report

The *Grain Transportation Report*, developed by the USDA, reports on current and several prior weekly grain transportation activities in the U.S. Tabular summaries are available for total U.S. rail carloads by Class I railroads (U.S. line-haul railroads with operating income in excess of $272.0 million [Inbound, 2012]), volumes of grain shipments inspected for exports, carloads of rail shipments delivered to ports by coast, barge movements of grain commodities based on river lock data, vessel movements at ports by coast, average price of commodities, and the freight rates for ship charters.

Methodology

Sources for data in the report include shippers, the AAR for Class I rail shipments, the U.S. Army Corps of Engineers (barge movements), the Federal Grain Inspection Service (export inspections), and the St. Lawrence Seaway Authority.

The grain shipment statistics are predominantly gathered by the USDA from available freight information. Consequently, the data incorporated into the *Grain Transportation Report* from external sources is limited to the sample sizes of the respective data collection procedures adopted in the development of each source. Information on the sample sizes represented in the external data sources used to develop the *Grain Transportation Report* is not available. Shipment data collected by the USDA at port locations on vessel movements and rail carload delivery statistics are updated weekly and are not subject to any sampling. Because the majority of grain exports from the U.S. are composed of wheat, soybean, and corn, export statistics for these three grain commodities are reported separately in the database. The database does not include import statistics for grain shipments.

The USDA compiles the *Grain Transportation Report* on a weekly basis from data collected by its Agricultural Marketing Service (AMS) and from external sources. Statistics on U.S. rail carloads of grain shipments by Class I railroads are collected from the AAR on a weekly basis to update the database. Data on the export inspection volumes of grain shipments are collected from the Federal Grain Inspection Service. Barge shipment statistics of grains in terms of shipment tonnage are collected from the U.S. Army Corps of Engineers. Statistics on rail carload deliveries and freight rates are collected by the USDA from shippers at ports to update the database weekly.

National Transportation Statistics (NTS)

The NTS database is compiled and developed annually by the BTS (part of the USDOT). The database provides detailed information on the condition and performance of existing transportation facilities, current traffic operations, and fleet sizes by each mode of transport, modal accidents, and fatality statistics; transport-economy relationships; energy utilization in transportation; and the environmental impact of transportation systems.

Methodology

The BTS does not conduct any surveys to collect transportation statistics for the NTS database. The NTS database is prepared from standard data sources developed by federal government agencies, private industries, and associations for each mode of transport. Sample sizes represented in each mode-specific data source for the NTS database are not available.

The BTS publishes the NTS database using the data collected from standard data sources specific to each mode of transport. All the transportation statistics reported in standard published data sources for each mode are collected and compiled by the Volpe National Transportation Systems Center for the BTS and incorporated into the NTS database. Data is collected and updated annually.

Quality Control Procedures

The BTS publishes transportation statistics for each mode in a similar format (multi-modal information) in the NTS database to determine annual variations in statistics for each mode and to ensure a higher degree of quality control.

Port Import/Export Reporting Service (PIERS)

The PIERS database, developed by United Business Media, Inc., is one of the most comprehensive databases on U.S. foreign waterborne imports and exports. The database also reports trade shipment statistics for cargo movements between ports in Mexico and South America to major trade partners around the world.

The PIERS database was originally developed by The Journal of Commerce (JOC) Group before the group was purchased by United Business Media, Inc. in November 2001. With the purchase of the JOC Group, United Business Media not only obtained ownership of the PIERS database, but also the JOC magazine and JOC Web site www.joc.com, thereby becoming one of the leading information service providers in the areas of global trade and transportation sectors.

Methodology

Electronically filed or hard-copy freight vessel manifests submitted to the U.S. Customs Service are the source of information in the database. Vessel manifest data are collected from all foreign trade carriers at major port locations in the U.S., Mexico, Latin America, and Asia. Manual vessel manifests filed at smaller port locations are not reported in the database. No sampling is done and all the statistics in terms of carrier type and shipments are collected for the following:

- Containerized
- Break bulk
- Dry bulk
- Tankers

Thus, data reported in the PIERS database accurately represents the actual shipment statistics because data is collected from vessel manifests of all foreign trade carriers without any sampling procedure. The number of carriers reporting their statistics in the vessel manifests is highly variable and is not available.

Details on import and export shipments are gathered by the U.S. Customs Service at ports of entry and exit from vessel manifests. Electronic vessel manifests are filed by international carriers with U.S. Customs using the Customs' *Automated Manifest System*. PIERS reporters at all major U.S. seaports collect the import and export information from the Customs' vessel manifests and incorporate them into the PIERS database. In total, PIERS reporters gather import and export shipment information from approximately 25,000 bills of lading and vessel manifests at major ports of entry and export. Data is collected and updated monthly.

Quality Control Procedures

To minimize non-sampling errors concerning the data reported in the vessel manifests, PIERS quality assurance staff performs regular audits and cross-checks of the shipment documentation reported by carriers to U.S. Customs before incorporating the data into the PIERS database. In addition, the majority of importers, exporters, and carriers that subscribe to the PIERS database help ensure the accuracy of the data reported in the database by verifying their own shipment information and reporting any discrepancies immediately to Commonwealth Business Media, Inc.

Quarterly Coal Report (QCR)

The QCR is developed and published by the U.S. Department of Energy in fulfillment of its responsibilities in terms of the Federal Energy Administration Act of 1974 as amended, to collect and disseminate statistics on energy-related activities. The QCR provides detailed quarterly summary statistics of U.S. coal production, distribution, foreign trade, receipts, consumption, and stocks to a wide audience including the U.S. Congress, federal and state agencies, the coal industry, and the general public.

Methodology

The QCR derives its data from manufacturers consuming coal, coke plants, coal producers, and distributors, and the U.S. Bureau of the Census Foreign Trade Statistics. The sample is limited to large-scale production, consumption, and distribution. Only manufacturers consuming 1,000 or more tons of coal in the previous year are required to complete the EIA-3 form used by the Energy Information Administration (EIA). Approximately 700 manufacturers are surveyed using the EIA-3 form to obtain information on their quarterly coal consumption. To obtain information on coke and breeze production, distribution, and stocks, the coke plants are surveyed using the EIA-5 form. Around 25 to 30 U.S. coke plants are surveyed quarterly using the EIA-5 form. The respondent list is updated frequently from the information gathered from relevant industry literature. Only coal producers and distributors producing or distributing in excess of 2,000 equivalent lbs. of coal are surveyed. All U.S. industries producing 30,000 or more short tons of coal annually are selected to report their statistics in the Schedule Q, EIA-6 form. All U.S. coal distribution companies with an average coal stock of 10,000 or more short tons per quarter are also selected to report in the Schedule Q, EIA-6 form. Around 630 coal producers and distributors are surveyed quarterly using the Schedule Q of the EIA-6 survey form.

Except for imports and exports, all other statistics pertaining to coal production, distribution, consumption, and stocks are collected by the EIA through mail surveys. Coal consumption and stock statistics are collected by surveying manufacturers that consume coal for all uses other than for coke production. The EIA uses the EIA-3 form to survey these manufacturers. The EIA-3 form captures detailed data on coal consumption, stocks, prices, and receipts from the manufacturers surveyed. Detailed statistics on coke and breeze production, distribution, and stocks are collected by the EIA by surveying all U.S. coke plants. The EIA uses the EIA-5 form to survey all U.S. coke plants. Approximately 30 U.S. coke plants are surveyed using the EIA-5 form to obtain data on their coke production, distribution, and stocks. The EIA also collects data on coal productions and distributions for each coal-producing state by surveying all coal-producing and distributing companies that produce or distribute in excess of 2,000 lbs. of short tons. The EIA surveys all the sampled coal producing and distributing companies on a quarterly basis using the Schedule Q of the EIA-6 survey form. Schedule-Q of Form EIA-6 collects data on coal productions, producer stocks, and distributor stocks for each coal-producing state. The EIA also collects, on an annual basis, coal distribution statistics from all U.S. companies that own or purchase and distribute more than 50,000 short tons of coal annually. Exceptions on the threshold limit are, however, allowed for the states of Arkansas, Maryland, Oklahoma, and Pennsylvania that have a 10,000 short ton annual threshold limit. Import and export coal-trade statistics are collected by the U.S. Department of Energy from the foreign trade statistics developed by the U.S. Bureau of the Census. Data is collected quarterly and updated annually.

Quality Control Procedures

To minimize the non-sampling errors in the data collection and ensure comprehensive coal statistics, the EIA follows up with the non-respondents using written and telephone requests. The EIA maintains all the coal data in a computerized system and conducts frequent edits to ensure consistency and accuracy with minimum reporting errors. The data collected with the EIA-3 form are compared with the coal distribution statistics reported in the annual survey of companies using the EIA-6 form to identify under-coverage problems in the in EIA-3 form. The list of coal consuming manufacturers is updated frequently, using data collected from State Air Quality and Energy Offices, to ensure that the sample surveyed with the EIA-3 form is selected from the entire population of coal consuming manufacturers. The list of coal producers to be surveyed with Schedule Q of Form EIA-6 is also updated frequently using data reported by the Mine Safety and Health Administration (MSHA), U.S. Department of Labor, and the lists maintained by various state agencies.

Ship Movements Database

The Ship Movements database is a comprehensive data source containing worldwide merchant ship movement information. The database is developed by Lloyd’s Maritime Information Services, Inc., a joint venture of Lloyd’s of London Press and Lloyd’s Register of Shipping. Lloyd’s Maritime Information Services, Inc. is the only single source that provides detailed maritime information on vessel movements among major ports around the world.

The Ship Movements database reports current movements of a large number of merchant vessels (around 30,000) that transport international freight. One significant feature of this maritime database is the daily updating of ship movement information, made possible by up-to-date information gathered by Lloyd’s agents stationed at principal ports worldwide. The database is used extensively in the planning and development of vessel service patterns on international waterborne trade routes.

Methodology

Merchant vessels engaged in international trade movements provide information for the database. The Ship Movements survey is administered by Lloyd’s agents at major ports around the world to a sample of the following ship categories:

- Tankers and combination vessels weighing more than 6,000 dead weight tons (DWT)

- Dry Bulk Carriers (DBC) weighing more than 10,000 DWT
- All other vessel types weighing more than 5,000 DWT.

In total, the database covers approximately 2 million ship movements in a calendar year.

The Ship Movements database is developed and updated by LMIS, Inc. from information gathered on current worldwide movements of merchant vessels by Lloyd's agents at principal ports around the world. The information is collected by conducting surveys of vessels at each major port location worldwide. The vessel surveys are designed to gather detailed information on vessel movement itineraries between ports (including arrival and departure dates of shipments), and vessel characteristics (including name, type, and vessel capacity). Copies of survey forms used by Lloyd's agents are not available. Data is collected and updated in real time.

Quality Control Procedures

High quality and efficiency in data collection is ensured by the Lloyd's Agency Department that provides central office support to the Lloyd's agents stationed at ports worldwide. The Lloyd's Agency Department conducts periodic inspections and hosts conferences for Lloyd's agents around the world to improve the professional surveying skills of the agents. The daily updates of the Ship Movements database are accomplished by electronic transmission of survey reports and standard forms by Lloyd's agents to the LMIS, Inc.

State Estimates of Truck Traffic

Estimated AADT volumes for single unit and combination trucks and vehicle miles traveled estimates for nine classes of trucks by highway functional systems are available from individual state highway agencies. The agencies provide the FHWA with the latter data. The data are included in the Highway Performance Monitoring System (HPMS), which is available to the public. Apart from the AADT and VMT (vehicle miles traveled) data in the HPMS, additional AADT estimates for trucks can be obtained directly from the individual state highway agencies. These data may be available online (database format) or the hard-copy versions can be acquired by contacting the respective state highway agency. The state highway agencies also report total AADT values for trucks, percent values for single-unit trucks, and percent values for combination trucks for a sample of non-local highway sections to the FHWA in ASCII format on magnetic tape or diskettes.

Methodology

Data is provided through vehicle counts collected by state and local highway agencies. Statewide truck traffic data collection involves selecting a sample of highway site locations for collecting vehicle count data and the selecting of time periods for vehicle classification data collection (permanent count or temporary count sites). Vehicle classification data are collected at a small number of permanent count sites in each state using AVC devices, which are set up at a small number of permanent count locations and at a larger number of temporary count sites. AVC devices use an axle spacing interpretation algorithm to interpret vehicle axle spacing and classify vehicles into the appropriate FHWA specified vehicle categories. Because vehicle axle spacing varies from state to state, each state develops its own algorithm, specific to its needs, for the AVC system. The sample size for permanent count sites range from 50 to 100 for each state.

Vehicle classification counts from permanent count sites are collected daily and do not require any time-of-day sampling. Vehicle classification data are also collected at a larger number of temporary classification sites (approximately 200). Traffic count data, without vehicle classification, is collected at a large number of temporary count sites in each state. Sampling by time of day at temporary count sites aims to collect vehicle counts during periods of maximum traffic. Counts at temporary sites are usually sampled over a 48-hour weekday period once every 3 years.

VMT data are calculated for nine to ten classes of trucks from the vehicle classification data by highway functional system. The VMT data for different truck classes by highway functional systems are reported annually by state highway agencies to the FHWA in Lotus 123 file format. The AADT estimates are computed from the vehicle counts at permanent and temporary count locations for single-unit and combination trucks on selected non-local road sections.

Truck classification count data from permanent count sites are not factored because the traffic counts are collected by AVCs daily. Consequently, traffic data from AVCs are directly used to compute AADT values. Traffic counts collected from temporary count sites over a 48-hour weekday period need to be factored to estimate truck traffic over the entire week.

The factors used to calculate the AADT from temporary count data are calculated by computing the numerator and denominator of the factor separately. The numerator of the factor is taken as the AADT value calculated from permanent count sites using AVCs. The denominator of the factor depends on whether monthly or weekly factors are used in the factoring procedure. A detailed discussion of the factoring procedure for estimating AADT values from temporary count data is available in the Traffic Monitoring Guide (TMG) at the following Web address: http://www.fhwa.dot.gov/ohim/tmguide/pdf/tmg2.pdf. Because truck travel patterns vary significantly compared to other vehicles, separate factors are calculated for estimating total AADT values and AADT values of trucks. Data is collected and updated annually.

Quality Control Procedures

To ensure accurate AADT estimates, separate factors are computed for trucks and total AADT (dominated by car traffic). Because truck travel patterns vary significantly across months compared to passenger travel, using separate factors for trucks and cars ensures more reliable AADT estimates.

In computing AADT values from classification data gathered using AVCs at permanent count sites, AASHTO recommended using the average of averages method instead of the more common simple averages method.

The simple averages method for computing AADT does not provide accurate values in the case of missing data during periods of equipment down time. The average of averages method recommended by AASHTO provides more reliable and accurate estimates of AADT.

For temporary traffic counts, at least 48-hour weekday periods are recommended to capture traffic variations over a 24-hour period in the calculation of AADT.

Transborder Surface Freight Database

The Transborder Surface Freight database is developed on a monthly basis by the BTS at the USDOT under a contract with the U.S. Bureau of the Census. The Census Bureau provides the BTS with detailed reports of U.S. international trade statistics collected as part of its Foreign Trade Statistics program. Using the Census reports, the BTS develops tables of U.S. import and export trade flows with Canada and Mexico, including shipment characteristics by commodity type and surface modes of transportation.

Development of the Transborder Surface Freight database was initiated in 1993. The objective was to study the impacts on U.S. surface trade flows with Canada and Mexico as a result of the North American Free Trade Agreement (NAFTA) signed by the U.S., Canada, and Mexico in December 1993, and enacted on January 1, 1994.

Methodology

The U.S. Bureau of the Census *Foreign Trade Statistics* program provides information for the database.

Statistics of U.S. import and export shipments with Canada and Mexico reported in the Transborder Surface Freight Database are sample estimates of the total shipments. Transborder Surface Freight database is derived from the U.S. import and export trade data developed by the U.S. Census Bureau as part of its FTS program, which does not include import and export shipments of values less than $1,251 and $2,501 respectively. Sample size of shipments for which data is reported in Transborder is highly variable and is thus, not available.

U.S. international trade statistics are collected by the U.S. Census Bureau for its Foreign Trade Statistics program using the following means:

- Electronic data collection procedures
- U.S.—Canada Data Exchange, Automated Broker Interface (ABI) and Automated Export Reporting Program (AERP)
- U.S. Customs paper documents at ports of entry and exit

The BTS acquires U.S. international trade data from the Census's FTS program and extracts U.S. import and export data with Canada and Mexico by commodity type and surface mode of transportation. Data is collected in monthly files and reported annually.

Quality Control Procedures

Information on transshipped shipment trade statistics has been removed from the Transborder Surface Freight database to make it more comparable with other U.S. international freight databases. To obtain consistent data on the state of physical origin of U.S. export shipments, the reporting of the shipment origin by state of exporter has been terminated. To ensure accuracy and reliability in the estimates, frequent analytical reviews are conducted by the BTS with the help of the U.S. Census Bureau to check for inconsistencies and to make timely improvements to the database.

Costs Incurred in Data Collection

The Transborder Surface Freight database is developed from the data collected by the U.S. Bureau of the Census for the U.S. foreign trade statistics program. No separate surveys are conducted by the BTS to collect U.S. trade statistics with Canada and Mexico.

TRANSEARCH

The TRANSEARCH database is one of the most widely used commercial sources of freight movement data in the U.S. It contains U.S. county-level freight-movement data by commodity group and mode of transportation. The historical database combines primary shipment data obtained from 22 of the nation's largest freight carriers with information from public sources, and is accompanied with 30-year forecasts consistent with IHS Global Insight's macro forecasts. TRANSEARCH is compiled and produced annually.

Methodology

TRANSEARCH database is constructed from a sample of commercial, public, and proprietary freight data sources including the following:

- CFS
- ICC Carload Waybill Sample
- Corps of Engineers Waterborne Commerce Statistics
- FAA Airport Activity Statistics
- Census of Transportation — Commodity Transportation Survey (1977)
- Bureau of Census Foreign Trade Statistics
- AAR Freight Commodity Statistics
- Inter-industry trade patterns
- Motor Carrier Data Exchange (MCDE)
- Commodity, spatial (O-D), and shipment size (tonnage, value) data collected from truckload carriers and distributors

Data source selection is developed from over 100 freight traffic data sources sampled for four major modes: truck, rail, air, and water. Sampling for the *Motor Carrier Data Exchange* is based on geographical detail (at the state level, and 3-digit and 5-digit zip codes) and truckload carriers of manufactured goods participating in the program. Statistics for a total of around 75 million shipments per year are reported.

The available freight data sources are inconsistent in terms of commodity classifications, geographical detail, and base years. Accordingly, the procedure involves combining each of the data sources together, checking their validity and applicability, assigning commodity geography and mode descriptions, and arriving at a common framework for freight traffic representation.

Due to the complexity of truck freight movements at the state and national levels, the *Motor Carrier Data Exchange* program is an integral part of development of the database to arrive at the O-D flows of truck freight traffic. Carriers participating in the *Motor Carrier Data Exchange* program provide information on their annual traffic flows in terms of commodity, statewide O-Ds, and shipment sizes. Data is compiled and produced by IHS Global Insight on an annual basis, and is available about 15 months after the end of the collection period.

Quality Control Procedures

The database uses the disaggregation of state-to-state O-D flows into BEA and county area O-D flows within each state to get more accurate spatial distribution of freight. This occurs in a two-stage process. First, origin-state shipment production volumes are distributed to BEA areas based on employment statistics. Additional volumes due to imports are added to BEA areas having ports. Secondly, destination-state shipment consumption volumes are distributed to BEAs based on employment and population shares. Disaggregation is conducted on the basis of minimum volume criteria to avoid unrealistic fragmentation of traffic flows.

Waterborne Commerce and Vessel Statistics

The Waterborne Commerce and Vessel Statistics database, developed annually by the U.S. Army Corps of Engineers: Navigation Data Center (NDC) provides comprehensive shipment statistics data for domestic and foreign waterborne trade flows across U.S. ports and waterways. The database is the only comprehensive source of data for both domestic and foreign waterborne trade shipments in and out of the United States.

Domestic shipment data are collected specifically for the database by the Corps of Engineers from Vessel Operating Reports obtained from domestic carriers. Foreign trade statistics are directly obtained from the U.S. Census Bureau's U.S. waterborne import and export trade statistics. Further enhancements are, however, made to the database in terms of vessel movements.

Methodology

Domestic freight carriers who report their vessel operations and cargo activity directly to the Corps of Engineers in the form of Vessel Operations Reports, as well as the U.S. Bureau of the Census: U.S. Waterborne Exports and General Imports are the sources of data.

Domestic waterborne shipment statistics are collected by the U.S. Army Corps of Engineers from the Vessel Operations Reports submitted by all domestic carriers at U.S. port locations. Accordingly, there is no sampling involved in the data collection for domestic waterborne trade flow statistics. A total of more than 1,500 domestic vessel operating companies provide their shipment statistics in the vessel operations reports submitted to the U.S. Army Corps of Engineers. Foreign waterborne trade flow statistics are developed from Customs' manifest data beginning with calendar year 2000. Waterborne import and export shipments of value less than $1,251 and $2,501 respectively are not included in the survey by the U.S. Customs Service. Information on the number of records of waterborne import and export shipments reporting their statistics to the U.S. Customs is not available.

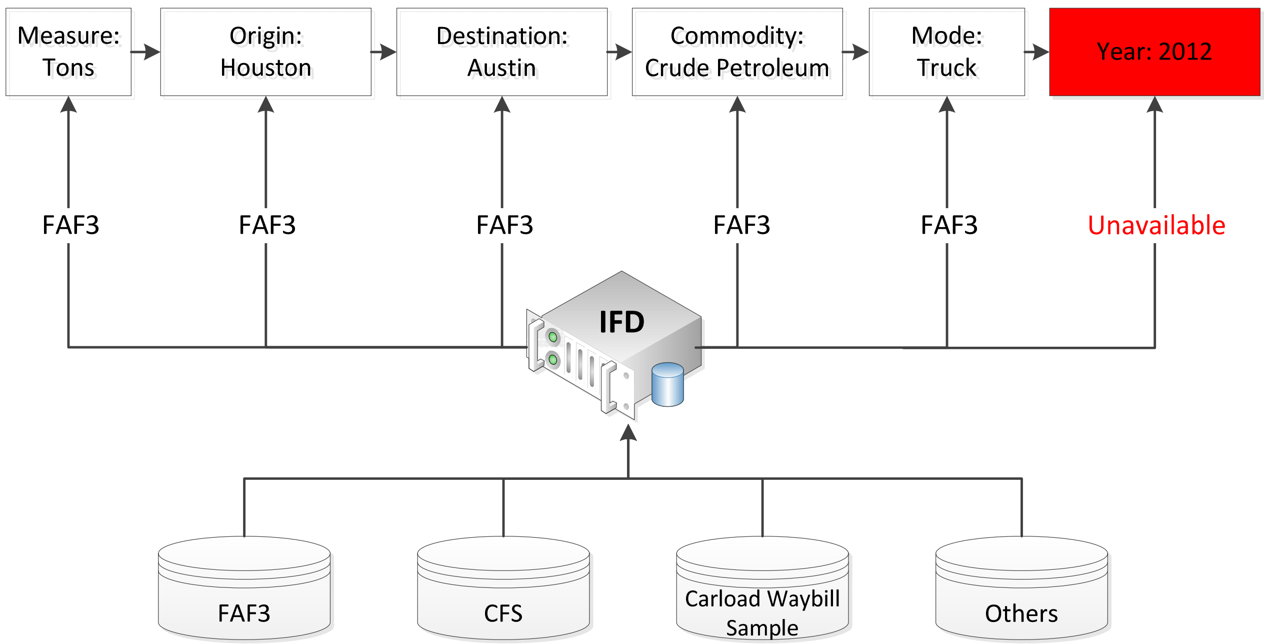
Quality Control Procedures

The U.S. Army Corps of Engineers has increased its efficiency in processing the domestic shipment filings (Vessel Operation Reports) to avoid delays in reporting waterborne commerce statistics to the U.S. Congress and other key federal agencies. Vessel Operation Reports from all the domestic carriers ensures that the domestic waterborne shipment statistics reported in the database are accurate estimates without any sampling errors (bias) in the annual estimates.

Appendix D: Data Gap Scenarios

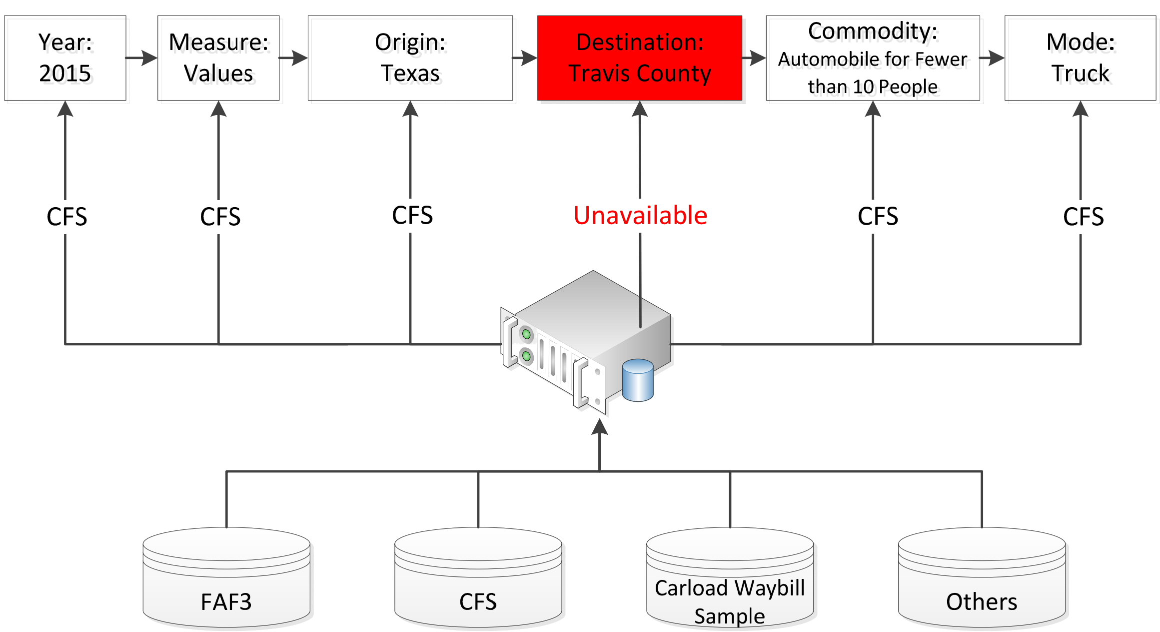
Scenario 2

How many tons of crude petroleum was moved from Houston to Austin by trucks so far in 2012?

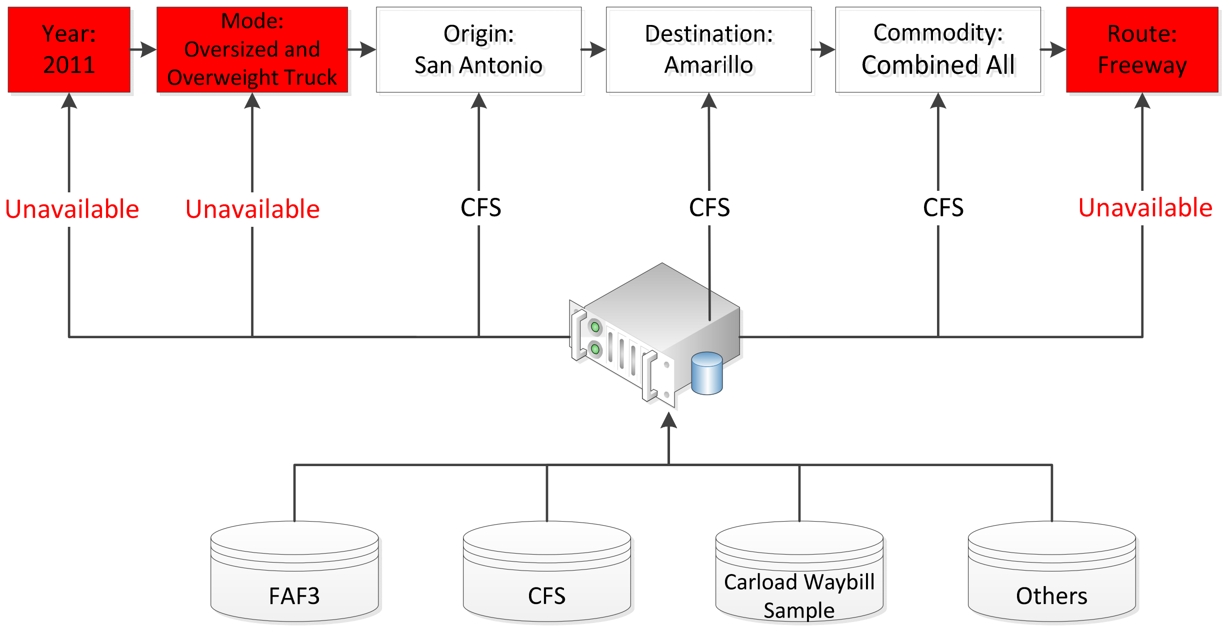


Scenario 3

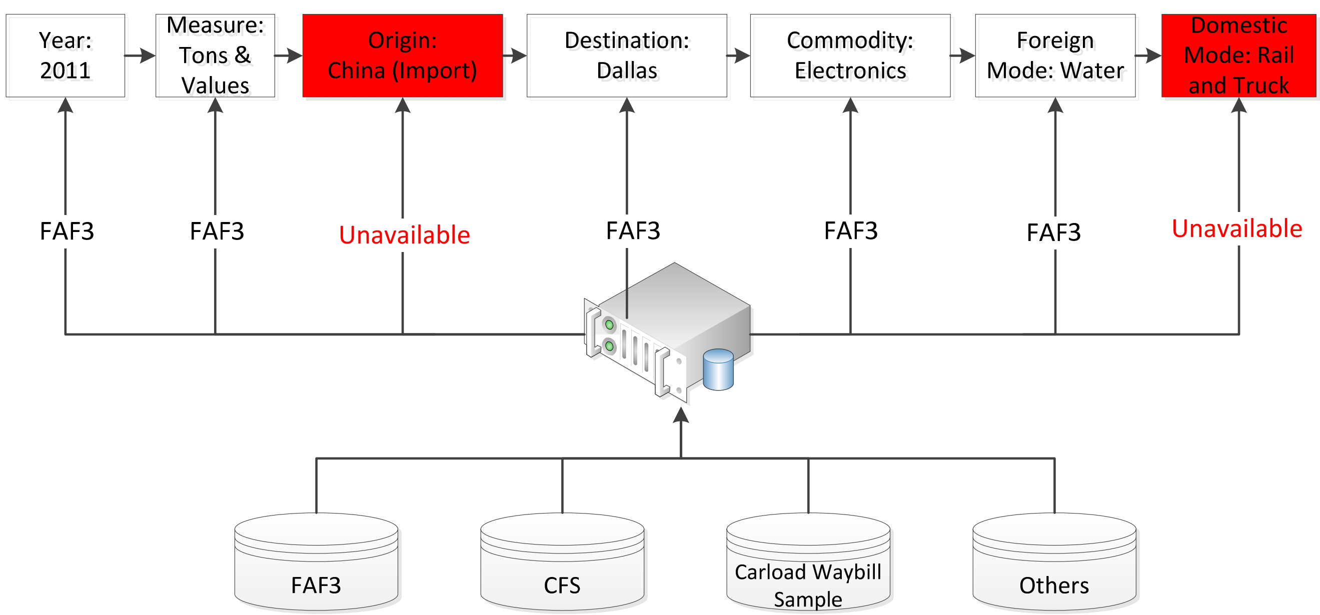
How much value of Sedans will be moved to Travis County within Texas by trucks in 2015?



Scenario 4
How many oversized and overweight vehicles were moved from San Antonio to Amarillo using freeways in 2011?

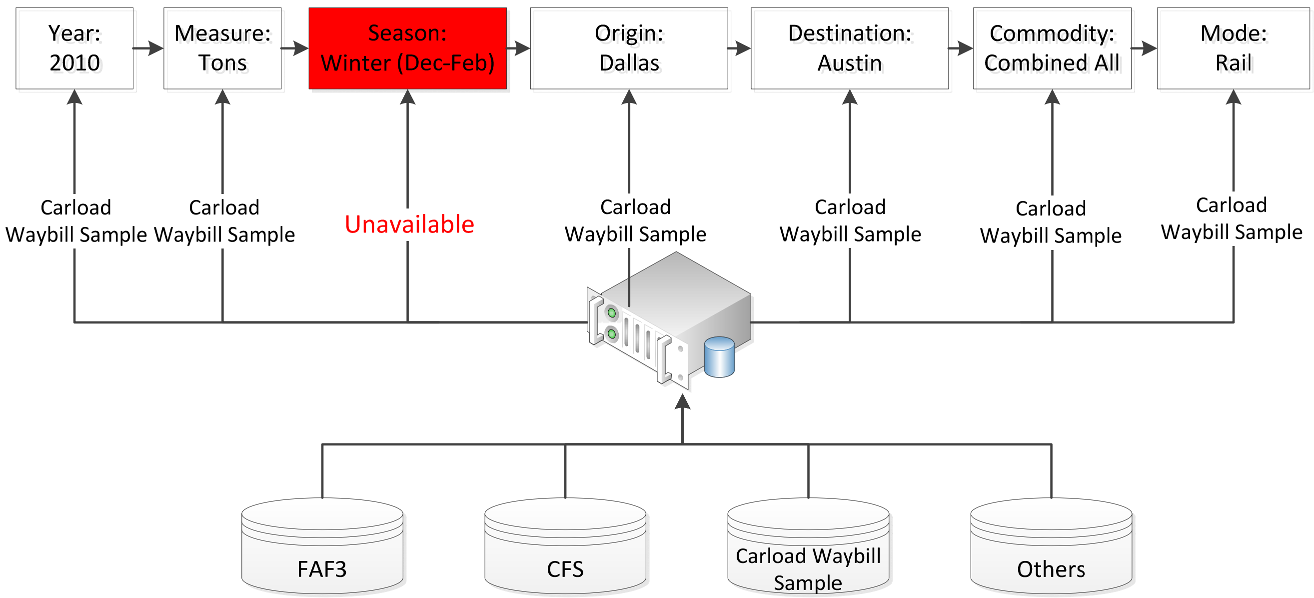


Scenario 5
How many tons and value of electronics were imported from China to Austin by rail and truck in domestic mode and by water in foreign mode in 2011?



Scenario 6

How many tons will be moved by rail from Dallas to Austin during winter season (Dec – Feb) in 2015?



Appendix E: Bilingual Brochures

Project No. 0-6697
Integration of Data Sources to Optimize Freight Transportation in Texas

Freight transportation is critical to the economic prosperity of any region. The challenge lies in disaggregating freight transportation demand to flows that can be assigned onto a state and region's transportation network. Disaggregated freight flows are necessary to:

- provide a clear picture of freight movements on the transportation system;
- determine the impact of freight on a region's infrastructure and the implications in terms of funding;
- evaluate strategies for improving freight mobility;
- forecast system performance;
- mitigate impacts of truck traffic on general mobility, and
- improve the safety performance of the transportation system.

The passage of the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991 initiated an increasing interest in freight modeling within statewide planning efforts, particularly the evaluation of current and future freight transportation capacity necessary to ensure freight mobility. Although freight models have started to emerge as tools to inform transportation policies, a critical challenge in the development of these models remains insufficient and inferior quality data.

The Texas Department of Transportation (TxDOT) has contracted with the Center for Transportation Research at The University of Texas at Austin (CTR) to:

- develop a strategy for collecting and integrating available freight data;
- explore the feasibility of entering into a data sharing partnership with the freight community for the collection of detailed and robust freight data that will satisfy the needs of transportation planning agencies;
- develop and pilot a prototype Freight Data Architecture and Handbook, and
- advise TxDOT on the cost-effectiveness of acquiring and maintaining a freight data sharing partnership to populate the Freight Data Architecture.

The outcome of this research will be a report that provides a strategy for integrating different freight data sources and for collecting and compiling primary freight data from Texas's freight stakeholders through a data sharing partnership. A prototype Freight Data Architecture with accompanying Handbook will also be provided. The Handbook will provide guidance for populating and maintaining the Freight Data Architecture.

Proyecto No. 0-6697
Integración de Fuentes de Datos para la Optimización del Transporte de Carga en Texas

El transporte de carga es fundamental para la prosperidad económica de cualquier región. El desafío consiste en desagregar la demanda de transporte de carga a flujos que se puedan asignar a la red de transporte estatal o a nivel regional. Obtener información desagregada en relación a los flujos de transporte de carga es esencial para:

- proporcionar una imagen clara de los movimientos de carga en la red;
- determinar el impacto del transporte de carga en la infraestructura de la región y sus implicaciones en términos de costos;
- evaluar estrategias para mejorar la movilidad de las mercancías;
- predecir el rendimiento de sistema;
- mitigar los impactos que genera el tráfico de carga en términos generales de movilidad, y
- mejorar el desempeño de la seguridad del sistema de transporte.

La aprobación de la Ley de Eficiencia en el Transporte Intermodal Terrestre de 1991 (ISTEA, por sus siglas en inglés) provocó un creciente interés de las dependencias a cargo de la planeación estatal en relación a los modelos de carga; especialmente en relación a la evaluación de la capacidad para transporte de carga actual y futura necesarias para garantizar la movilidad de mercancías. Aunque los modelos de carga han comenzado a surgir como herramientas para formular políticas de transporte, el reto fundamental radica en el hecho de que en el desarrollo de estos modelos sigue siendo insuficiente y los datos escasos o de poca calidad.

El Departamento de Transporte de Texas (TxDOT) contrató al Center for Transportation Research de la Universidad de Texas en Austin para llevar a cabo este estudio, cuyos objetivos son:

- desarrollar una estrategia para recopilar e integrar datos de carga que ya se encuentren disponibles;
- explorar la viabilidad de acordar el intercambio de información con distintos actores involucrados en el transporte de carga para recolectar información detallada necesaria para las dependencias a cargo de la planificación de la red de transporte;
- desarrollar un prototipo piloto de Manual para la Arquitectura de Datos de Transporte de Carga, y
- asesorar a TxDOT acerca del costo-efectividad de adquirir y mantener acuerdos de intercambio de información para rellenar la Arquitectura de Datos de Transporte de Carga.

Un informe final documentará los resultados de este estudio y recomendará una estrategia para la integración de las fuentes de datos de carga y los métodos más efectivos para la recopilación de datos a través de acuerdos de intercambio de información celebrados con actores involucrados en el transporte de carga en Texas. El reporte también incluirá un prototipo de Arquitectura de Datos de Transporte de Carga, así como su el Manual correspondiente que brindará información de cómo poblar y mantener al día la Arquitectura de Datos.

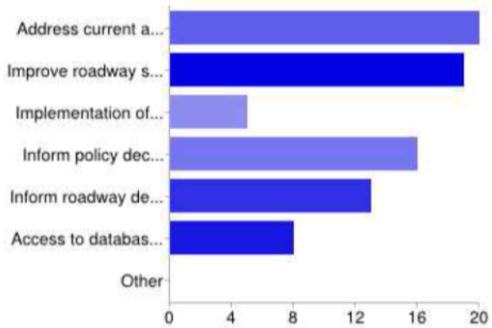
Appendix F: Online Survey Responses

32 responses

Summary See complete responses

Data Sharing Partnership Benefits and Concerns

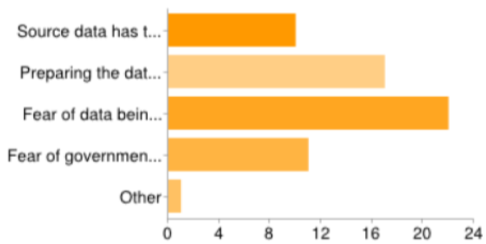
1. What benefits do you seek should you enter into a data sharing partnership with TXDOT?



Address current and anticipated transportation issues such as congestion and capacity constraints
Improve roadway safety and public education on freight
Implementation of ITS technologies to improve freight mobility efficiency
Inform policy decisions such as designation of truck-only lanes, local ordinances, access restrictions
Inform roadway design decisions such as wider turning lanes, traffic signal design and timings, route
Access to database (except proprietary data)
Other

People may select more than one checkbox, so percentages may add up to more than 100%.

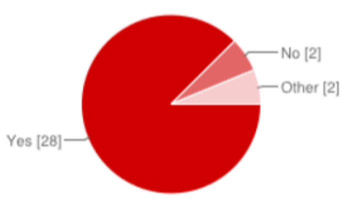
2. What are you main concerns for entering into a data sharing partnership?



Source data has to be modified to meet needs
Preparing the data will burden on already overworked staff
Fear of data being mishandled or being improperly used
Fear of government involvement and law enforcement
Other

People may select more than one checkbox, so percentages may add up to more than 100%.

3. If guaranteed that the information will never become public or shared with others, will you be more willing to enter into a data sharing partnership with TXDC

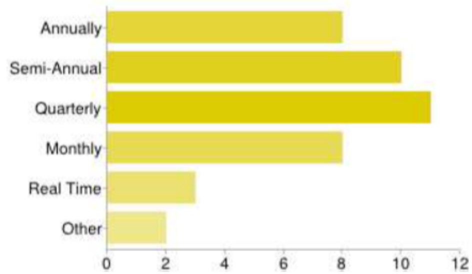


Yes	**28**
No	**2**
Other	**2**

4. Have you previously entered in to a data sharing partnership? If Yes, how were your concerns addressed?

No no. NO NO no Yes, shared data with federal officials regarding hours of service regulatory economic impact. N/A no No. I confidentially constraints. No no No No

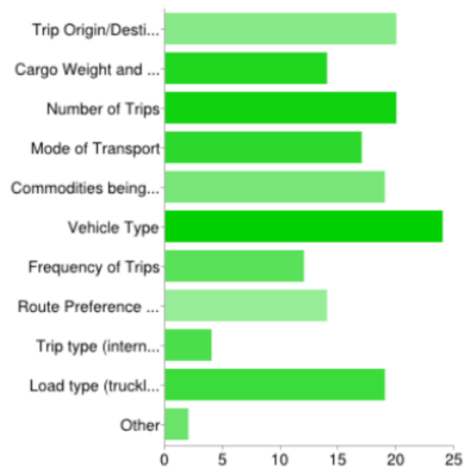
5. If you do agree to a data sharing partnership, how often will you be willing to share data with TXDOT without additional effort from your end (please check one)



Annually	8
Semi-Annual	10
Quarterly	11
Monthly	8
Real Time	3
Other	2

People may select more than one checkbox, so percentages may add up to more than 100%.

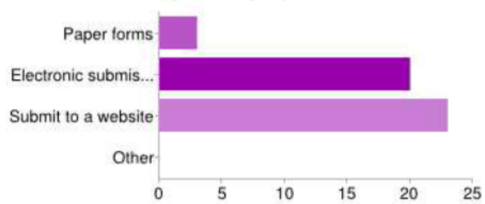
6. What kind of data are you willing to share?



Trip Origin/Destination	20
Cargo Weight and Value	14
Number of Trips	20
Mode of Transport	17
Commodities being transported	19
Vehicle Type	24
Frequency of Trips	12
Route Preference (e.g. interstate, toll, state highways)	14
Trip type (internal, external)	4
Load type (truckload, less-than-truckload, service truck)	19
Other	2

People may select more than one checkbox, so percentages may add up to more than 100%.

7. In what format are you willing to provide the data



Paper forms	3
Electronic submission e.g. e-mail	20
Submit to a website	23
Other	0

People may select more than one checkbox, so percentages may add up to more than 100%.