EL PASO REGIONAL TRANSPORTATION DATA WAREHOUSE AND TRAFFIC MONITORING

by

Swapnil Samant and David Salgado

Project performed by

Center for International Intelligent Transportation Research

185917-00014

185917-00014 – El Paso Regional Transportation Data Warehouse and Traffic Monitoring

September 2017

Report prepared by

Center for International Intelligent Transportation Research

Texas A&M Transportation Institute

4050 Rio Bravo, Suite 212

El Paso, Texas 79902

TEXAS A&M TRANSPORTATION INSTITUTE

The Texas A&M University System

College Station, Texas 77843-3135

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

DISCLAIMER AND ACKNOWLEDGMENTS

This research was performed by the Center for International Intelligent Transportation Research, a part of the Texas A&M Transportation Institute. 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.

CHAPTER 1: INTRODUCTION

The Paso del Norte (PDN) region that comprises of the City of El Paso, City of Las Cruces, and Ciudad Juarez is a multijurisdictional and binational environment (see Figure 1). Traffic conditions in this region are influenced by various agencies (e.g., Texas Department of Transportation [TxDOT], City of El Paso, Instituto Municipal de Investigacion y Planeacion, and U.S. Customs and Border Protection [CBP]). However, only the City of El Paso and TxDOT have traffic monitoring capabilities on their own right of way (ROW).

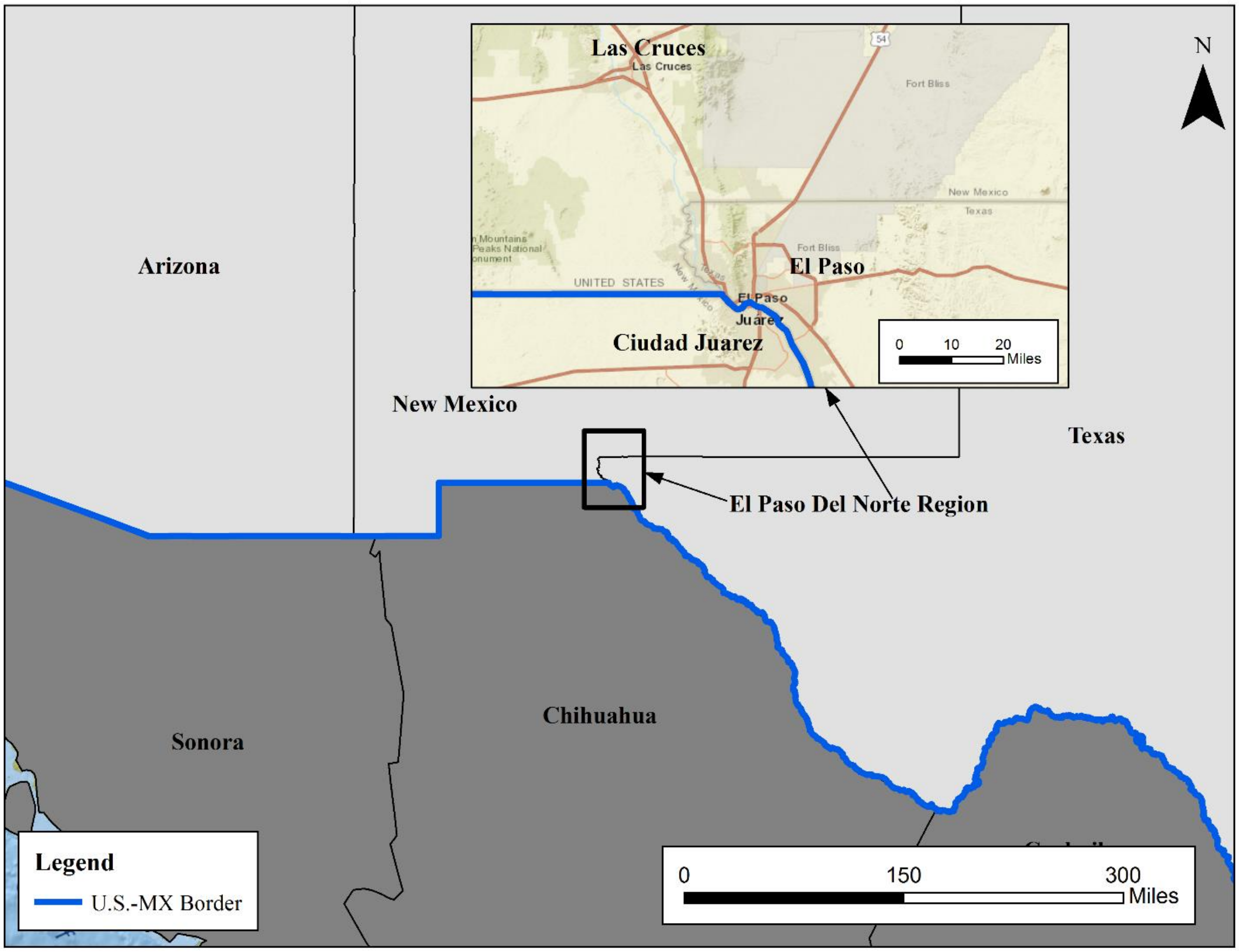


Figure 1. El Paso del Norte Region

The traffic information for this region is distributed by various data sources. TxDOT reports travel times, traffic incidents, and road closures along I-10 and other major state roads. The City of El Paso also reports traffic incidents and road closures on its own ROW. Regarding border crossing operations at land ports of entry (LPOEs), TTI manages the Border Crossing Information System (BCIS). BCIS monitors and reports border wait times and crossing times Ysleta LPOE for passenger and commercial vehicles. BCIS also monitors wait time and crossing time for commercial vehicles at Bridge of the Americas (BOTA). CBP reports wait times and number of lanes opened for all LPOEs in the PDN region: Tornillo LPOE, Ysleta LPOE, BOTA LPOE, PDN Bridge LPOE, and Santa Teresa LPOE. Sun Metro is the public transportation provider that serves El Paso. Sun Metro provides transit information for the City of El Paso. Local media and websites provide information about special events in the region. Finally, the National Weather Service (NWS) monitors and reports weather alerts and conditions in the PDN region.

The goal of this project is to develop a data warehouse for transportation data in the PDN region. The data warehouse system development was influenced by the evolution of intelligent transportation systems (ITS) in the region. The system will have the capabilities of collecting, processing, fusing, and storing transportation data to generate traffic information. The system will serve as centralized repository of transportation data and traffic information for the PDN region. Additionally, this system will provide the ability to monitor regional mobility (including cross-border mobility).

This project has two objectives:

1. Document the state of the art in the use of the data warehouse concept in the transportation field.
2. Develop a framework for the PDN Regional Data Warehouse System.

Traffic information can be used to help commuters and travelers to plan their trip. Transportation data and traffic information obtained from the PDN Regional Data Warehouse System can also help researchers calibrate traffic forecasting tools. The PDN Regional Data Warehouse system is critical for expanding the Center for International Intelligent Transportation Research's (CIITR's) research capabilities.

This report consists of five chapters. The present chapter (Chapter 1) serves as the introduction. Chapter 2 defines the data warehouse concept and describes its role in transportation as a critical part of traffic management systems (TMCs). Chapter 3 documents the state of the practice on virtual TMCs in the United States. Chapter 4 presents the proposed framework for the PDN Regional Data Warehouse system. Finally, Chapter 5 documents the conclusions and findings of the project.

CHAPTER 2: DATA WAREHOUSE IN THE TRANSPORTATION FIELD

As mentioned in the introduction, this chapter defines the data warehouse concept and describes its role in transportation as a critical part of TMCs. The data warehouse concept has been around for more than two decades. A data warehouse system is a repository of integrated information available for querying and analysis. Figure 2 presents the typical warehouse architecture. The various sources of data are processed, fused, and stored in the data warehouse system. The information stored in the data warehouse can be queried by different user applications (Gupta 1997).

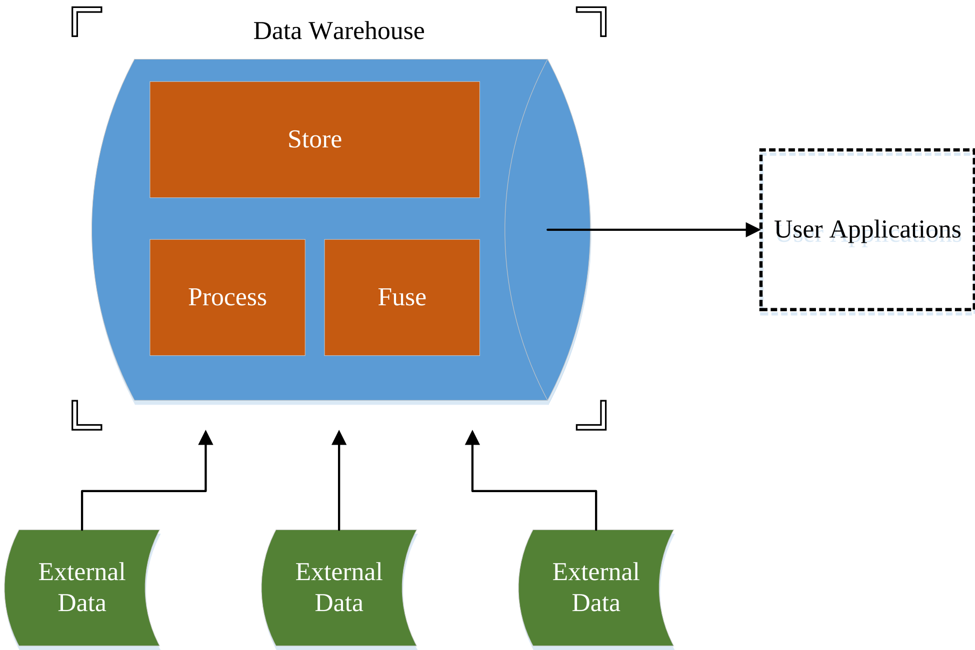


Figure 2. Typical Data Warehouse Architecture

Data warehouses are powerful data-driven systems that can be used as decision support systems (DSS). In other words, information obtained from a data warehouse system can be incorporated in the decision-making process of any industry or sector to control risks and uncertainty. ITS data repositories can be considered as data warehouses that support operations of traffic monitoring, management, and traffic information dissemination.

The literature revealed the existence of a data warehouse system at the Unites States-Canada border. The Cascade Gateway Data Warehouse provides traffic information at four LPOEs between the Lower Mainland, British Columbia, and Whatcom County, Washington State. The information elements include:

- Volume.
- Delay.
- Service rate.
- Vehicles in queue.
- Queue length.

The information can be queried by LPOE, direction, month, and day. Figure 3 shows a sample of the query screen.

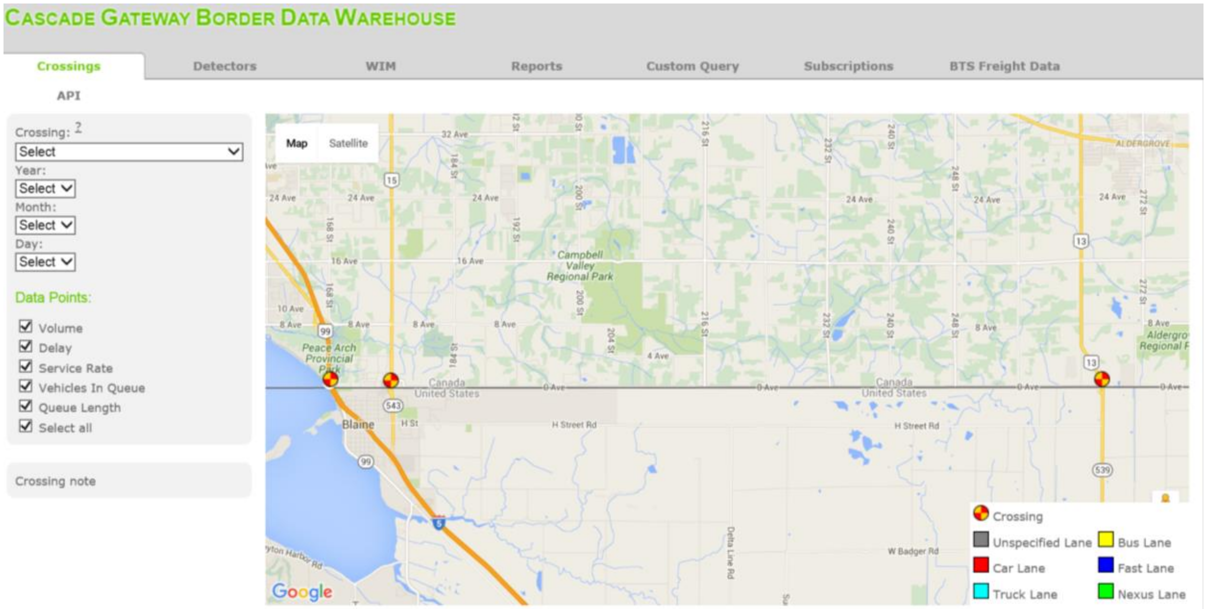


Figure 3. Cascade Gateway Data Warehouse Crossings (Cascade Gateway Border Data Warehouse 2017)

It also includes additional data collected from readers on I-5 and Washington State Route 539 as shown in Figure 4. The information elements include:

- Volume.
- Occupancy.
- Speed.
- Vehicle length.

This information can be queried by detector or weight in motion station and vehicle type. Figure 4 shows a sample of the query screen.

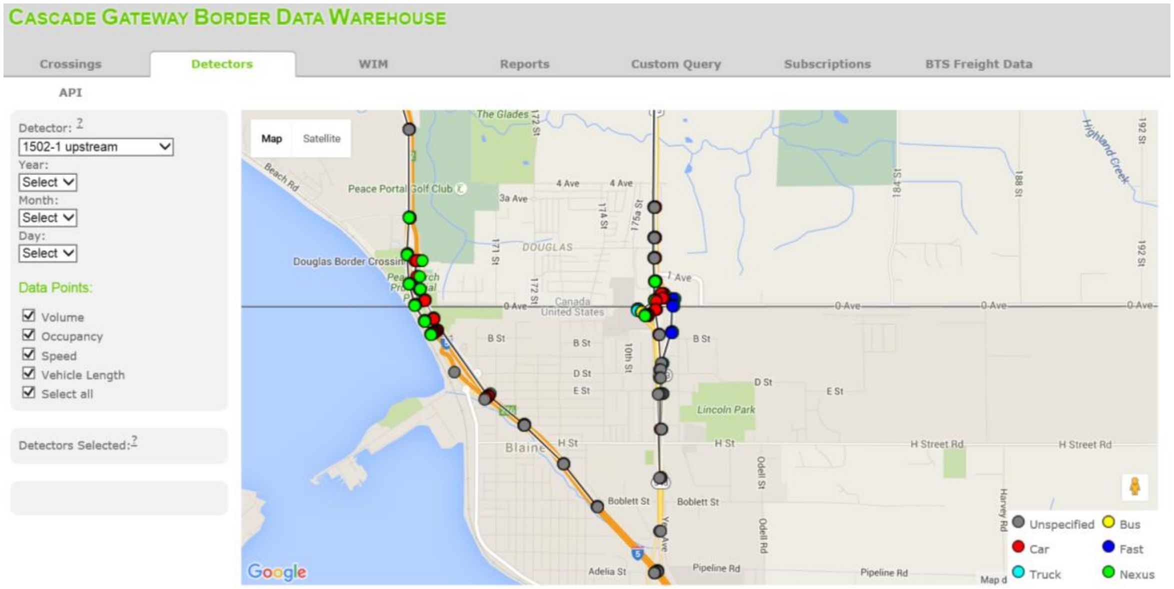


Figure 4. Cascade Gateway Data Warehouse Detectors (Cascade Gateway Border Data Warehouse 2017)

Data warehouse systems are a critical part of the TMCs. According to the Federal Highway Administration (FHWA):

> The TMC is the hub or nerve center of most freeway management systems. It is where the data about the freeway system is collected and processed, fused with other operational and control data, synthesized to produce "information," and distributed to stakeholders such as the media, other agencies, and the traveling public. TMC staff uses the information to monitor the operation of the freeway and to initiate control strategies that affect changes in the operation of the freeway network. It is also where agencies can coordinate their responses to traffic situations and incidents.

Data warehouse systems are the heart of TMCs and are specifically used to process and fuse transportation data to generate traffic information. There are four TMC models currently used in the United States (Lukasik, et al. 2014):

- Centralized—In this model, an agency manages the TMC with a clearly defined mission and objective. Cooperative efforts with other agencies are regulated by means of inter-agency agreements. In this TMC model, all systems are situated in a specific location, which normally is the TMC facility. Personnel can work at the TMC facility or remotely, and some work can be outsourced under the supervision of the agency.
- Distributed—In this model, multiple agencies establish a joint program to manage the TMC. The agencies reach agreements to define the role of each agency, policies, practices, funding, staffing, etc. This TMC model is commonly used in large metropolitan and multijurisdictional areas. Systems and staff are situated in different TMC facilities. Every TMC facility can have exclusive functions or functions can be shared between multiple TMC facilities.
- Virtual—In this model, no physical TMC is necessary. A virtual TMC model allows users to perform all TMC operations using computers and networks without being

present at a physical nerve center or without the existence of such a physical nerve center. Staff can work from any agency facility. However, there is still the need for physical communications to ITS field devices and between centers.

- Hybrid—This model is a combination of the virtual TMC model and the centralized or distributed models. This happens when some of the TMC functions in these two models are virtual.

Figure 5 shows where the different TMC models presented above are deployed.

Figure 5. TMC Model Deployment Locations (Lukasik, et al. 2014)

Table 1 shows the main TMC functions and provides a description of each one (Lukasik, et al. 2014) (Smith and Venkatanarayana 2005). All these functions can be performed by a TMC with the required systems regardless the TMC model adopted.

Table 1. TMC Functions

Function	Description
DSS	• Data Warehouse System that process, fuse, and store information • Distribution System for delivering the information to stakeholders
Transportation Asset Management	• Reporting scheduled construction and other related events • Monitoring the condition of Critical Social Overhead Capital • Detecting Asset Failure
Traffic Flow Monitoring	• Monitoring traffic volume, speed, truck weight, and delays on roads and high-occupancy vehicle Lanes • Traffic Signal Management
Public Transit Management	• Monitoring and evaluating transit vehicle operations to ensure schedule adherence and to identify and minimize delays
Incident Counteraction	• Detecting or receiving incidents and notify it to officials and drivers
Other	• Weather Management, Railroad passing control, etc.

A formal definition of virtual TMC is provided by FHWA:

> A virtual TMC is the function of monitoring, controlling, and managing the functional elements of a transportation management system through the use of computers and computer networks without being present at a physical nerve center or without the existence of such a physical nerve center. This includes the functions of monitoring, collecting, processing and fusing transportation system data; disseminating transportation information to outside entities; implementing control strategies that effect changes in the transportation system; and coordinating responses to traffic situations and incidents.

The migration from centralized or virtual TMC model requires significant resources. Therefore, agencies need to know in advance how virtual TMC models could help their needs. Table 2 presents the advantages and disadvantages of virtual TMCs.

Table 2. Advantages and Disadvantages of Virtual TMC Model (Lukasik, et al. 2014)

Virtual TMC Advantages	Virtual TMC Disadvantages
Capital cost savings – A virtual TMC does not require the construction of a TMC main building, which is translated into a significant savings in capital investments.	**Requires Broader Staff Capabilities** – The implementation of a virtual TMC model requires training the personnel to be able to work independently, and to be proficient in the virtual TMC software. Additionally, virtual TMC personnel must know all TMC functions, not some of them, which is the case in centralized and distributed TMCs.
Recurring Cost Savings – Centralized and distributed TMC facilities incur in high maintenance costs associated to the hardware located at TMC buildings, which are not required in a virtual TMC. One example of these hardware pieces are videowalls that need to be replaced after a certain number of years or special lighting and protection systems.	**Expanded/Revised Training Programs –** Virtual TMC require significantly more trainings for personnel than centralized and distributed TMCs.
Operate from any Location – Virtual TMC operators only need a computer with a certain software to work from any location. Hence, TMC personnel can immediately operate when an emergency occurs after hours.	**IT Security** – Since virtual TMC are permanently connected to the internet, the investment in network security systems (software and hardware) and protocols require a higher capital, operations, and maintenance resources.
Staff Flexibility – TMC personnel can work anytime anywhere with a computer and the specific TMC software. For example, this allows TMC personnel to work during weekends if needed.	**More Difficult Multi-Staff Coordination** – Since virtual TMC personnel can work independently from any location, the interaction and communication between personnel may be reduced. Consequently, coordination between different agencies and TMC personnel may be more difficult.
Staff Security – In cases where the TMC building is located next to an at-risk government facility, a virtual TMC could increase staff security because staff are not required to go to a certain building to work.	**Existing Agency Agreements** – Existing agreements to share systems and facilities would need to be revised and adapted to the virtual TMC necessities.
Multiagency Operations – Virtual TMCs facilitates the cooperation between different agencies.	**Higher Computer Software and Systems Integration Costs** – The virtual TMC model requires more complex and highly integrated software and systems.
	Need for Physical Command Control Center – Many agencies have the desire for command and control centers where many videos and applications can be displayed and viewed simultaneously on one common display. This will be lost in the virtual TMC model.

CHAPTER 3: VIRTUAL TMC MODELS IN UNITED STATES

Researchers envision the PDN Regional Data Warehouse system as the heart of a future virtual TMC that would monitor, control, and manage transportation operations in the PDN region. Hence, researchers focused on identifying the role of data warehouse systems on virtual TMCs in the United States. This chapter documents the state of the practice on virtual TMCs in United States.

Two of the most important virtual TMCs are associated with integrated corridor management system (ICMS) efforts carried out in San Diego (California) and Dallas (Texas). The following paragraphs provide a detailed description of the Virtual TMC models used in the San Diego and Dallas ICMS.

SAN DIEGO ICMS

The ICMS is a fully automated virtual TMC that runs 24 hours a day 7 days a week. The San Diego ICMS consists of three key components (Dion and Skabardonis 2015):

- Data hub—This component is the ICMS data warehouse system. Data captured by the stakeholders and transportation agencies are sent to the ICMS data warehouse system via a standardized regional communication network. The data are contextualized (i.e., processed and fused) and stored. As a result of this contextualization, data become traffic information. Then, traffic information is sent to the 511-traveler information system and disseminated to travelers via telephone, a website, or a mobile device application. The traffic information is also sent to the DSS.
- DSS—This system helps TMC operators to select response plans for diverse types of incidents. Response plans are pre-approved by all stakeholder agencies, and the DSS will recommend these plans based on real-time traffic information. The recommended plan is then subjected to real-time simulation and predictive analysis to confirm its improvement in traffic conditions.
- System services—This key component assists with data management, system management, system maintenance, and training activities by means of the ICMS Data Storage System and the Corridor Performance Measurement System.

Figure 6 is extracted from the report named "San Diego I-15 Demonstration Integrated Corridor Management System PATH Report on Stage 3: Site Demonstration and Evaluation," and presents how the San Diego ICMS components are related.

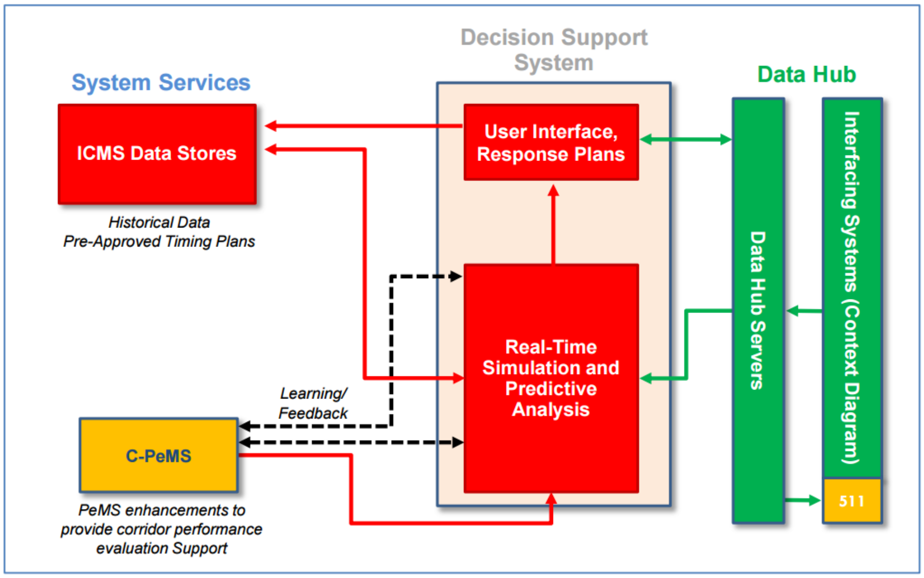


Figure 6. ICMS Key Components (Dion and Skabardonis 2015)

The following bullets lists and describes the systems associated with the San Diego ICMS, which are depicted in Figure 7:

- Ramp Meter Information System (RMIS)—This system is used by the California Department of Transportation (Caltrans) to manage metering signals that control the access to freeways. RMIS sends real-time information about the current metering plans, and the ICMS may request changes on metering plans based on traffic conditions.
- Lane Closure System—This system sends information to the ICMS identifying the number of lanes closed, and the status of the construction or maintenance activities in those lanes.
- Advanced Traffic Management System 4.1—This system is used by Caltrans to manage traffic detectors, changeable message signs, and closed-circuit television (CCTV) Cameras. Caltrans use this system to monitor traffic congestion and incidents, measure travel times, and to monitor traffic conditions at planned events. The ICMS can request express lanes control and the display of certain messages on the CMSs.
- Congestion Pricing System (CPS)—This system is used by Caltrans to change toll prices based on traffic conditions. The CPS sends current toll process and the status of variable traffic signs to the ICMS, and the ICMS sends express lanes congestion levels to the CPS.
- Regional Arterial Management System (RAMS)—This system was deployed by San Diego's Regional Planning Agency to coordinate traffic signals along major arterial corridors operated by different agencies. Thanks to this system, subscribing agencies can view signal timing plans of all signals along arterial corridors regardless the agency who operates them. Subscribing agencies can also develop, propose, and implement timing plans spanning multiple jurisdictions. This system also monitors traffic congestion and intersection status. RAMS exchanges current timing plan information with the ICMS, and the ICMS may perform timing plan requests.

- Regional Event Management System—This system is a XML-based web interface to the California Highway Patrol's Computer-Aided Dispatch Media Server. The system sends incident or event information along with the response status to the ICMS.
- Regional Transit Management System (RTMS)—This system is used by the Metropolitan Transit System and the North County Transit District to support fixed-route transit operations. This system sends transit operational data entailing vehicle location, schedule adherence, selected stop/route data and panic alarms to the ICMS. The ICMS may request service changes to the RTMS.
- Smart Parking System—This planned system will collect real-time parking data, set dynamic parking rates and provide real-time parking information to travelers. This system will send parking availability information to the ICMS.
- Real-Time Simulation System (RTSS)—This system executes traffic simulations in the corridor of the strategies proposed by the DSS and sees the possible outcomes of each. RTSS receives real-time data from the ICMS and sends various performance metrics resulted from the traffic simulation.
- Network Prediction System (NPS)—This system predicts origin-destination flows along the corridor to support the DSS. NPS receives real-time data and sends traffic predictions to the ICMS.
- Weather NWS—This system sends weather information to the ICMS.
- Traveler Information System (511)—This system allows motorists to access traffic information via web, phone, or public access television. The ICMS sends alerts to the 511 system, and the 511 system sends usage statistics to the ICMS.
- Arterial Travel Time System (ATTS)—This system travel time measurement in the arterial network. ATTS sends travel times to the ICMS.

The ICMS also sends data to third-party information service providers in XML format. Additionally, the integrated virtual corridor TMC is the system that allows TMC personnel to access the ICMS functionalities in this virtual TMC model.

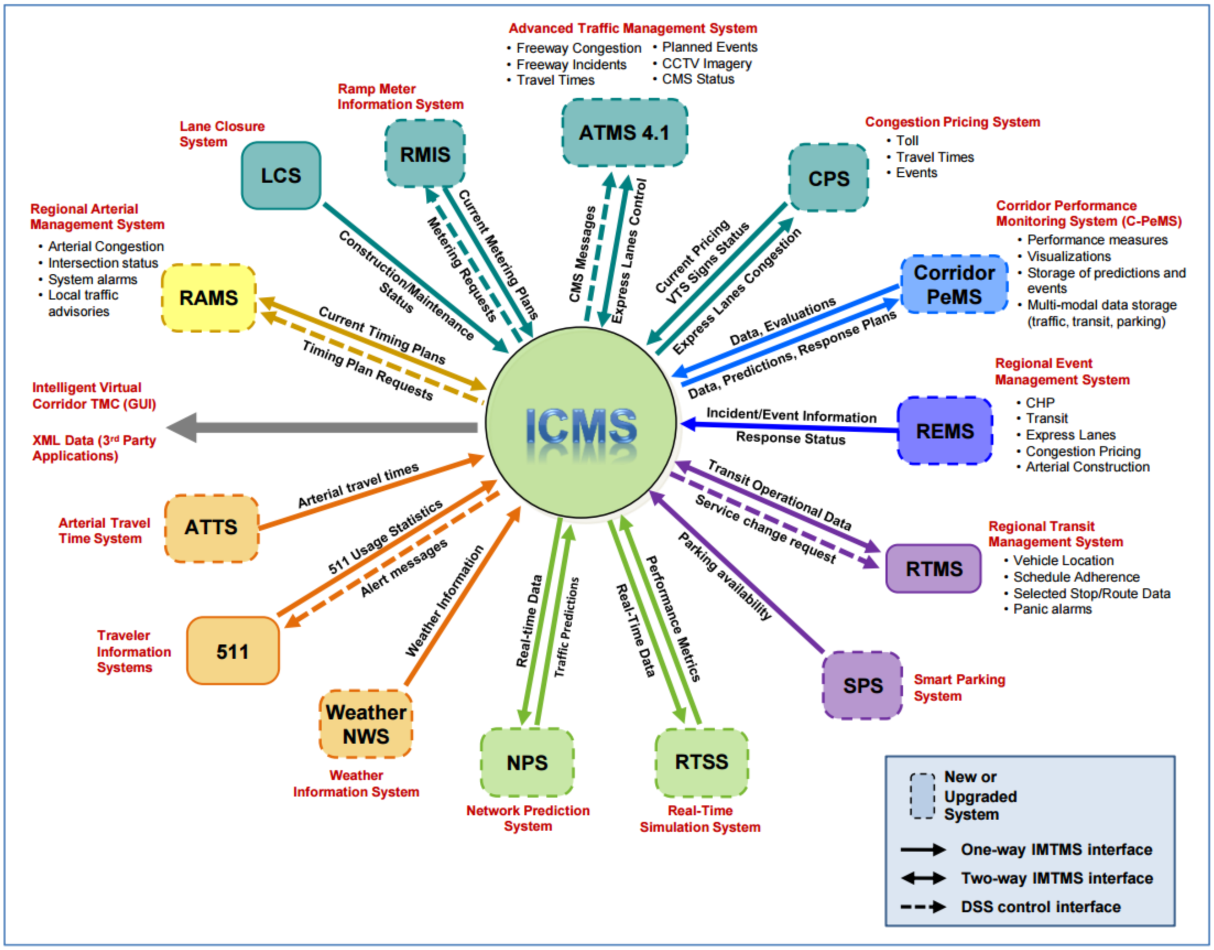


Figure 7. ICMS Context Diagram (Dion and Skabardonis 2010)

DALLAS ICMS

The Dallas ICMS is comprised by three main systems: SmartFusion, DSS, and SmartNET. The SmartFusion system is continuously collecting transportation data to monitor the network. The SmartFusion system process and fuse transportation data generating traffic information as any other data warehouse system. Then, traffic information is stored and disseminated to the 511-traveler information system (i.e., website, phone and mobile application), social media, and the other two systems in which the ICMS consists of (i.e., DSS and SmartNET).

The Dallas ICMS collects the following transportation data:

- Transit and event information from the Dallas Area Rapid Transit. Data consist of:
    - Transit schedule.
    - Stops and Times.
    - Automated Vehicle Location data.
    - Automatic Passenger Counter data.
- Weather information.
- Parking information.

- Speed information:
    - Real-time speed data from HERE®.
    - Travel times from Bluetooth® sensors.
- Advance Traffic Management Systems information from systems owned and operated by TxDOT in the City of Dallas and the City of Fort Worth. Data consist of:
    - Incident and lane closure information.
    - CCTVs.
    - Dynamic message signs.

The DSS obtains the real-time traffic information from SmartFusion and SmartNET for their evaluation. A set of strategies (plans) is proposed after evaluation. The plans are previously developed by the stakeholders and are associated with certain traffic conditions and forecasted benefits resulted from the traffic simulation subsystem. Once the plan is proposed, it will be evaluated by the DSS prediction modeling system to corroborate that it will improve traffic conditions. Finally, the DSS provides the recommended plan to the ICMS coordinator and partner agencies via SmartNET. Then, the agency can accept or reject the proposed plan. SmartNET is a graphical user interface that allows the ICMS coordinator and the agencies to exchange information among them. Additionally, SmartNET allows agencies to provide data to SmartFusion using center to center connections. Figure 8 presents the structure of the ICMS.

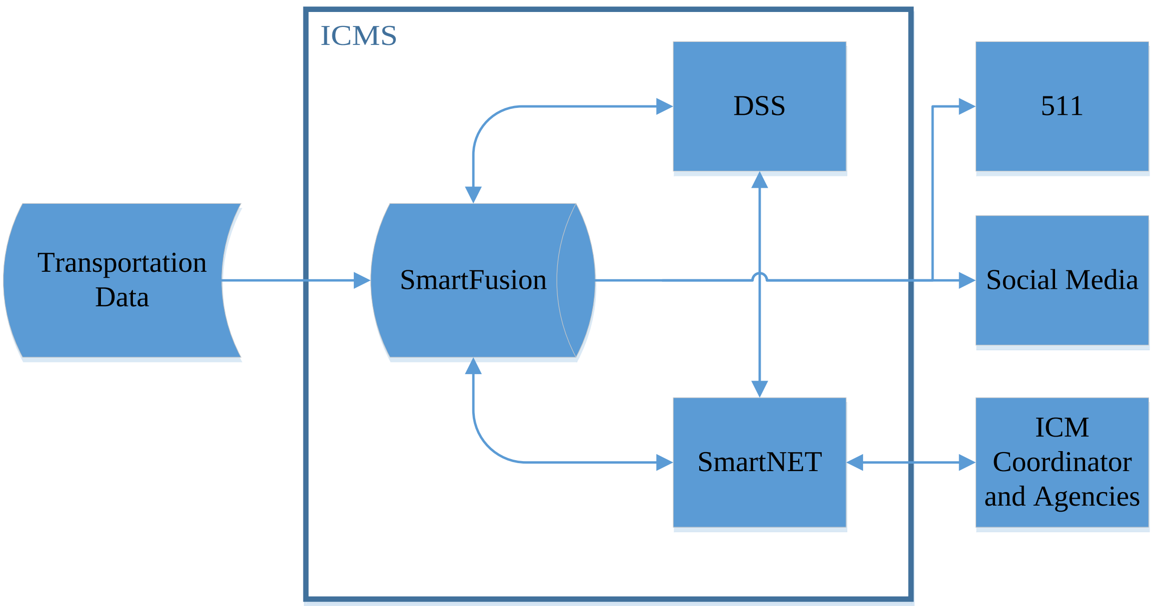


Figure 8. Dallas US 75 ICMS

Since December 2016, the DSS was discontinued due to lack of operation and maintenance funds. However, SmartFusion and SmartNET are still operating to provide support to the 511-traveler information system. Operations and maintenance costs associated to the 511-traveler information system are covered by the North Central Texas Council of Governments.

CHAPTER 4: PDN REGIONAL DATA WAREHOUSE FRAMEWORK

As mentioned in Chapter 1, the PDN region is a multijurisdictional and binational area where traffic information is distributed by various data sources. Currently, there is no physical or virtual infrastructure that offers a complete picture of traffic conditions in the region. The implementation of a virtual TMC could fill this infrastructure gap. Virtual TMCs help commuters make informed decisions about their route. Hence, a virtual TMC could improve mobility in the PDN region and could serve as a DSS for stakeholders and decision makers.

Researchers contacted TTI researchers involved in the Dallas ICMS project. These researchers stated that stakeholders and decision makers find the DDS system very useful. They also stated that travelers value 511-traveler information system. Based on these inferences, researchers recommend starting discussions with stakeholders and decision makers in the PDN region to implement a virtual TMC in the region. The implementation of a virtual TMC in the PDN region starts with the implementation of a data warehouse system capable of processing, fusing, and storing transportation data from various sources. This chapter describes the proposed framework for the PDN Regional Data Warehouse system.

The PDN Regional Data Warehouse system will have the capabilities of collecting and storing traffic data, non-traffic data, and network data. Traffic data consist of a set of variables that describes traffic conditions (e.g., traffic volumes, speed, travel time). On the other hand, non-traffic data consist of a set of variables that describes exceptional circumstances that may affect traffic conditions (e.g., traffic incidents, work zones). Network data consist of a set of variables that characterize roadway sections. A roadway section can be defined as the smallest piece of infrastructure with transportation data and traffic information associated to it. Additionally, the PDN Regional Data Warehouse system will have capabilities of processing and fusing transportation data. In other words, data will be contextualized to generate traffic information that can be used by diverse user applications. These user applications aim to meet stakeholders and decision makers' needs.

Figure 9 depicts the PDN Regional Data Warehouse system. The system will collect traffic and non-traffic data in their original format (i.e., raw data) from various sources in 15-minute intervals. Network data will be updated as needed based on changes on roadway sections. Next, the system will store raw data and later process the data. After processing, the data will be fused and converted in traffic information ready to be queried by the user applications developed to meet stakeholder's and decision makers' needs.

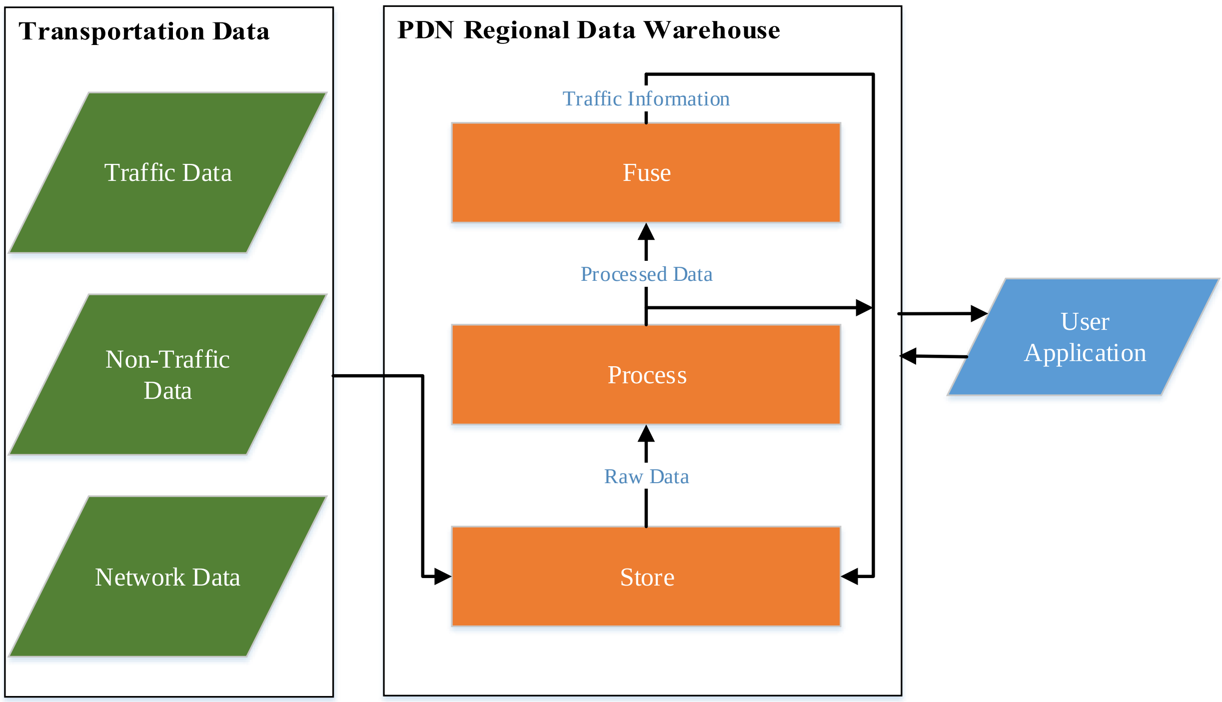


Figure 9. PDN Regional Data Warehouse System

A detailed description of each transportation data element and the computational processes performed by the PDN Regional Data Warehouse system follows.

TRANSPORTATION DATA

Transportation data collected by the PDN Regional Data Warehouse system will consist of traffic data, non-traffic data, and network data. The system will collect traffic data and non-traffic data for each roadway section every 15 minutes. Network data will be updated as needed based on roadway section changes. The following paragraphs define each variable collected and identify their sources.

Traffic Data

The system will collect the following for each roadway section: date, time, travel time, average speed, number of lanes operating, wait times at LPOEs, and traffic volumes. A brief description of the data to be collected and their sources is as follows:

- Date and time: The system will collect time and dates (i.e., time stamps) associated to each traffic data collection from the official U.S. time website (https://www.time.gov/).
- Travel time and speed: The system will collect travel times from TxDOT El Paso District TMC (TransVista), BCIS, INRIX, and Google. TransVista captures and monitors travel times along major freeways in El Paso region by means of Bluetooth sensors. Similarly, BCIS captures and monitors border wait times and crossing times at Ysleta LPOE for passenger vehicles traveling northbound and southbound directions.

BCIS also captures and monitors border wait times and crossing times at BOTA and Ysleta LPOE for commercial vehicles (trucks) traveling northbound. The system will also collect (if feasible) INRIX travel time data for TxDOT roads within the PDN region. Moreover, the system will collect travel times for major roads in the PDN region from Google (including Ciudad Juárez and Las Cruces). Finally, the system will compute average speed by using travel times collected.

- Number of lanes operating and wait times at LPOEs: The system will collect the number of lanes open and wait times for pedestrians, passenger vehicles, and commercial vehicles traveling northbound from the CBP Border Wait Times website (https://bwt.cbp.gov/index.html). Number of lanes opened and wait times will be collected for the six LPOEs in the PDN region: Tornillo LPOE, Ysleta LPOE, BOTA LPOE, Stanton LPOE (only for SENTRI), El Paso del Norte LPOE, and Santa Teresa LPOE. The system will also complement these data with wait times and crossing times obtained from BCIS for Ysleta LPOE and the BOTA.
- Number of lanes operating at roadway sections: The system will estimate the number of lanes operating based on the maximum number of lanes of each roadway section and lane closures due to traffic incidents, construction zones, and special events.
- Traffic volumes: The system will collect annual average daily traffic (AADT) reported by TxDOT via Statewide Planning Map website (http://www.txdot.gov/apps/statewide_mapping/StatewidePlanningMap.html) for major roads in the PDN region within Texas. Regarding LPOEs, the system will collect northbound traffic volumes in real time for commercial vehicles at the Ysleta LPOE from CIITR sensors. Moreover, the system will collect monthly traffic volumes for commercial vehicles, passenger vehicles, and pedestrians traveling northbound from CBP for the six LPOEs in the PDN region. The system will also collect traffic volumes for commercial vehicles, passenger vehicles and pedestrians traveling southbound from the City of El Paso via Ysleta LPOE and Stanton LPOE. Finally, the system will collect total northbound traffic volumes per month for commercial vehicles, passenger vehicles and pedestrians aggregated for the entire PDN region from the Bureau of Transportation Statistics website (https://transborder.bts.gov/programs/international/transborder/TBDR_BC/TBDR_BC_Index.html).

Non-Traffic Data

The system will collect the following non-traffic data for each roadway section: date, time, special events, weather alerts and conditions, traffic incidents, construction zones, and commodities transported from Mexico to the United States. A brief description of the data to be collected and their sources is as follows:

- Date and time: The system will collect time and dates (i.e., time stamps) associated to each non-traffic data collection from the official U.S. time website (https://www.time.gov/).
- Special events: On the U.S. side, the system will collect special events information from the "Visit El Paso" website (http://visitelpaso.com/) and the "Visit Las Cruces" website

(https://www.lascrucescvb.org/). On the Mexican side, the system will collect special events information from the "Visita Juárez" website (http://www.visitajuarez.com/).

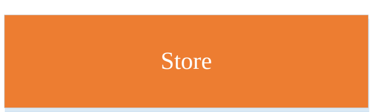
- Weather alerts and conditions: The system will collect weather alerts and conditions for the United States from NWS website (http://www.weather.gov/). On the Mexican side, the system will collect weather conditions from the Mexican national weather agency (Servicio Metereologico Nacional) website (http://smn.cna.gob.mx/es/).
- Traffic incidents: The system will collect traffic incident location in from the City of El Paso and El Paso County Public Alert and Incident Notification System website (http://legacy.elpasotexas.gov/traffic/). The system will also collect number of lanes closed due to traffic incidents from TransVista website (http://its.txdot.gov/elp/elp.htm).
- Construction zones: The system will collect lane closure information location due to programmed construction works from TransVista website (http://its.txdot.gov/elp/elp.htm).
- Commodities transported from Mexico to the United States: The system will collect total northbound monthly commercial value (in U.S. dollars) of commodities traveling via any commercial LPOE (northbound and southbound) in the PDN region from the U.S. Census Bureau Foreign Trade website (https://www.census.gov/foreign-trade/index.html).

Network Data

The system will generate a roadway section identification number (ID) and a section name per roadway section. Additionally, for each roadway section the system will collect: starting point (from), end point (to), and number of lanes. Network data will be collected once and the system will periodically check if need to be updated. Network data can be obtained from local transportation agencies or in the field.

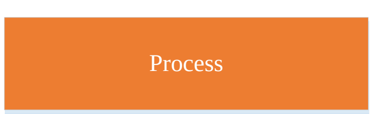
PDN REGIONAL DATA WAREHOUSE

Data Storage



The system will store raw or unprocessed traffic data, non-traffic data, and network data. The system will also store processed data and traffic information. These raw data, processed data, and traffic information will be available for being queried by user applications.

Data Processing



The system will process raw data previously stored. Raw data have different sources and structure. Thus, they should be homogenized prior to be fused. The way of processing each data set will differ based on the structure in which it is collected. The system will generate a table for each type of data: network data, traffic data, and non-traffic data. Once generated, these tables will be ready to be fused.

Network Data Table

Table 3 presents the structure of network data tables.

Table 3. Example of Network Data Table

Network Data						
Section ID	**Road name**	**Direction**	**From**	**To**	**Length**	**Total lanes**
00001	I-10	EB	31.83948, -106.570496	31.824785, -106.559286	2.63	4
00002	Mesa	EB	31.815274, -106.549858	31.809322, -106.538736	1.65	3
00003	Dyer	NB	31.83948, -106.570496	31.809322, -106.538736	2.01	2
00004	El_Parque	WB	31.824785, -106.559286	31.815274, -106.549858	2.35	2

Network data tables consist of seven fields:

- **Section ID**: It is a five-digit unique ID number that will be assigned to each roadway section. Section ID never changes, and it will be used as common field to fuse network data, traffic data, and non-traffic data.
- **Road name**: It is the name of the road in which the road section is located. This field is coded as text with a maximum of 10 characters using underscores instead of spaces.
- **Direction**: This field identifies the directionality of the roadway section: EB for eastbound, WB for westbound, SB for southbound, and NB for northbound.
- **From**: This field defines the exact location of the point where the roadway section starts. This location is defined by means of numeric values of latitude and longitude expressed in decimal degrees.
- **To**: This field defines the exact location of the point where the roadway section ends. This location is defined by means of numeric values of latitude and longitude expressed in decimal degrees.
- **Length**: This field contains the length of the roadway section expressed in miles with two decimals.
- **Total lanes**: This field contains the number of lanes of the roadway section expressed by mean of a two-digit numeric value.

Traffic Data Table

Table 4 presents the structure of the traffic data tables.

Table 4. Example of Non-Traffic Data Table

Traffic Data									
Date	**Time**	**Travel time**	**Average speed**	**Lanes operating**	**Hourly Traffic volumes**	**Veh. type hour**	**AADT**	**Veh. type AADT**	**Section ID**
02/03/2017	13:22:11	120	45	02	NA	NA	NA	NA	00001
10/15/2017	22:25:42	50	60	03	NA	NA	7000	Mix	00002
02/07/2017	01:23:00	70	55	02	200	Ped	NA	NA	00003
10/31/2017	03:05:01	90	53	02	260	Pass	NA	NA	00004

Traffic data tables consist of 10 fields:

- **Date**: This field contains the date in which traffic data/non-traffic data were collected. The format is month/day/year.
- **Time:** This field contains the time in which traffic data/non-traffic data were collected. The format is hour:minute:second expressed in military time.
- **Travel time**: This field contains the travel time for the roadway section expressed in seconds.
- **Average speed**: This field contains the average speed for the roadway section expressed in miles per hour.
- **Lanes operating**: This field contains information about the number of lanes operating along the roadway section.
- **Hourly traffic volumes**: This field contains traffic volumes expressed in number of vehicles per hour (if available).
- **Veh. type hour**: This field identifies the type of vehicle for which hourly traffic volumes are presented. The codes used are “car” for passenger vehicles, “truck” for commercial vehicles, “ped” for pedestrians, and “bus” for buses.
- **AADT**: This field provides the annual average daily traffic for the roadway section.
- **Veh. type AADT**: This field identifies the type of vehicle for which AADT are presented. The codes used are “Mix” for mixed traffic and “Truck” for trucks.
- **Section ID**: It is a five-digit unique ID number that will be assigned to each roadway section. Section ID never changes, and it will be used as common field to fuse network data, traffic data, and non-traffic data.

Non-Traffic Data Table

Table 5 presents the structure of the non-traffic data table.

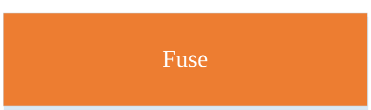
Table 5. Example of Non-Traffic Data Table

Non-Traffic Data							
Date	**Time**	**Special event**	**Weather condition**	**Weather alert**	**Traffic incident**	**Construction zone**	**Section ID**
02/03/2017	13:22:11	0	1	0	0	1	00001
10/15/2017	22:25:42	0	1	0	0	0	00002
02/07/2017	01:23:00	1	3	1	0	0	00003
10/31/2017	03:05:01	0	3	3	0	0	00004

Non-traffic data table consists of eight fields:

- **Date**: This field contains the date in which traffic data/non-traffic data were collected. The format is month/day/year.
- **Time**: This field contains the time in which traffic data/non-traffic data were collected. The format is hour:minute:second expressed in military time.
- **Special event**: This field identifies the existence of a special event within a 5-mile radius from the roadway section. The field will contain the numeric value of “1” if there is an event, or “0” if there is not.
- **Weather condition**: This field identifies weather conditions. The field will contain a numeric value that defines weather conditions as follows: (1) sunny, (2) cloudy, (3) windy, (4) rainy, and (5) snowy.
- **Weather alert**: This field identifies the existence of a severe weather alert activated by the NWS. The field will contain a numeric value that defines the type of weather alert as follows: (0) No alert activated, (1) winter weather/cold weather (i.e., winter storm) (2) fire weather (i.e., weather conditions that propitiates wildfire), (3) fog/wind/severe weather (i.e., dense fog, extreme wind, or tornado), (4) marine (i.e., special marine weather conditions), (5) flooding, (6) excessive heat, and (7) tropical (i.e., tropical storms or hurricanes).
- **Traffic incident**: This field identifies the existence of a traffic incident by means of a binary code as follows: (0) no traffic incident, and (1) there is a traffic incident.
- **Construction zone**: This field identifies the existence of a construction zone by means of a binary code as follows: (0) no construction zone, and (1) there is a construction zone.
- **Section ID**: It is a five-digit unique ID number that will be assigned to each roadway section. Section ID never changes, and it will be used as common field to fuse network data, traffic data, and non-traffic data.

Data Fusion



Finally, the system will generate traffic information by fusing processed data. After processed, the data are fused and converted in traffic information ready to be queried by the user applications developed to meet stakeholder’s and decision makers’ needs. Table 6 presents an example of the data fusion process.

Table 6. Example Data Fusion

Segment Data						
Segment ID	**Road name**	**Direction**	**From**	**To**	**Length**	**Total lanes**
00001	I-10	EB	31.83948, -106.570496	31.824785, -106.559286	2.63	4
00002	Mesa	EB	31.815274, -106.549858	31.809322, -106.538736	1.65	3
00003	Dyer	NB	31.83948, -106.570496	31.809322, -106.538736	2.01	2
00004	El_Parque	WB	31.824785, -106.559286	31.815274, -106.549858	2.35	2

Traffic Data								
Date	**Time**	**Travel time**	**Average speed**	**Lanes operating**	**Hourly Traffic volumes**	**AADT**	**Veh. type hour**	**Segment ID**
02/03/2017	13:22:11	120	45	2	NA	NA	NA	00001
10/15/2017	22:25:42	50	60	3	NA	7000	NA	00002
02/07/2017	01:23:00	70	55	2	200	NA	Ped	00003
10/31/2017	03:05:01	90	53	2	260	NA	Pass	00004

Non-Traffic Data							
Date	**Time**	**Special event**	**Weather condition**	**Weather alert**	**Traffic incident**	**Construction zone**	**Segment ID**
02/03/2017	13:22:11	0	1	0	0	1	00001
10/15/2017	22:25:42	0	1	0	0	0	00002
02/07/2017	01:23:00	1	3	1	0	0	00003
10/31/2017	03:05:01	0	3	3	0	0	00004

CHAPTER 5: CONCLUSIONS

A data warehouse system is a repository of integrated information available for querying and analysis. Data warehouse systems are the heart of TMCs and are specifically used to process and fuse transportation data to generate traffic information ready to be queried by the user applications developed to meet stakeholder's and decision makers' needs.

The PDN Regional Data Warehouse system is critical for expanding CIITR's research capabilities. Traffic information can be used to help commuters and travelers plan their trip. Transportation data and traffic information obtained from the PDN Regional Data Warehouse System can also help researchers calibrate traffic forecasting tools. PDN Regional Data Warehouse System will enhance CIITR's research contributions to the region.

REFERENCES

Cascade Gateway Border Data Warehouse. 2017. *Crossings.* August. http://www.cascadegatewaydata.com/Detector.

—. 2017. *Detectors.* August. http://www.cascadegatewaydata.com/Crossing.

Dion, Francois, and Alexander Skabardonis. 2015. "San Diego I-15 Demonstration Integrated Corridor Management System: PATH Report on Stage 3: Site Demonstration and Evaluation."

Gupta, Himanshu. 1997. "Selection of views to materialize in a data warehouse." *Database Theory—ICDT'97* 98-112.

Lukasik, Dan, Monica Castellanos, Adam Chandler, Elliot Hubbard, Ramanujan Jagannathan, and Teresa Malone. 2014. *Guidelines for Virtual Transportation Management Center Development.* Washington, DC: U.S. Department of Transportation Federal Highway Administration.

Smith, Brian L., and Ramkumar Venkatanarayana. 2005. "Realizing the Promise of Intelligent Transportation Systems (ITS) Data Archives." *Journal of Intelligent Transportation Systems*, Oct: 175-185.