COMPANION POWERPOINT PRESENTATION TO
UNITY DATABASE

TxDOT Project 0-6697-CTR: Integration of Data Sources to Optimize
Freight Transportation in Texas

DECEMBER 2013; PUBLISHED SEPTEMBER 2014

<table>
<thead>
<tr>
<th>Performing Organization:</th>
<th>Sponsoring Organization:</th>
</tr>
</thead>
</table>
| Center for Transportation Research  
The University of Texas at Austin  
1616 Guadalupe, Suite 4.202  
Austin, Texas 78701 | Texas Department of Transportation  
Research and Technology Implementation Office  
P.O. Box 5080  
Austin, Texas 78763-5080 |

Performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.
INTEGRATION OF DATA SOURCES TO OPTIMIZE FREIGHT TRANSPORTATION IN TEXAS

TXDOT Research Project 0-6697
December 2013

DRAFT
COMPANION PRESENTATION
Study Team

Dr. C. Michael Walton (RS)  Dr. William J. O'Brien (Co-PI)
Dan Seedah (Co-PI)          Bharathwaj Sankaran
Alejandra Cruz-Ross(Co-PI)  Prateek Agarwal
Jolanda Prozzi             Meredith Cebelak
Peter La Fountain           Sarah Overmyer
Background

“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.”

– Jolanda Prozzi
Background

Disaggregated freight flows are necessary to:

• provide a clear picture of freight movements
• determine the impact of freight on infrastructure and funding implications
• evaluate strategies for improving freight mobility
• forecast system performance
• mitigate impacts of truck traffic on general mobility
• improve transportation system performance and safety
Background

Numerous freight data sources exist but are found to be incompatible due to:

- different protocols in assigning origins and destinations
- different commodity classifications
- different assumptions to estimating or dealing with missing data
- different expansion factors and control totals
- different procedures used for data aggregation
Study Objectives

1. Develop a strategy for collecting and integrating available freight data

2. Explore the feasibility of entering into a data sharing partnership with the freight community

3. Develop a prototype Freight Data Architecture

4. Advise TXDOT on the cost-effectiveness of acquiring and maintaining a freight data sharing partnership
Related TXDOT and Federal Studies

- NCFRP 9 – Guidance for Developing a Freight Transportation Data Architecture (2011)
- NCFRP 25: Freight Data Sharing Guidebook (2013)
- NCFRP 47 - Freight Transportation Data Architecture: Data Element Dictionary (ongoing)
Study Objectives

1. Develop a strategy for collecting and integrating available freight data
   I. Identify Texas Freight Data Needs
   II. Examine Existing Databases
   III. Examine Freight Data Collection Methods
   IV. Identify Current Data Gaps
Identify Texas Freight Data Needs

1. Conducted five workshops statewide
2. Participants include TXDOT, MPO and city planners
3. Discussions include:
   - What freight data do you use?
   - For what purpose(s) do you use the freight data?
   - Issues experience in obtaining reliable data?
   - What freight data variables do you need?
   - What level of detail do you require?
   - How would you use the proposed integrated Statewide Freight Database (what queries will you run)?
Identify Texas Freight Data Needs

4. Main sources of freight data include:
   - TXDOT truck volume counts including PMIS
   - Confidential Carload Waybill Sample
   - Freight Analysis Framework
   - Marine port data
   - TRANSEARCH
   - CRIS accident database
Identify Texas Freight Data Needs

5. General concerns include:
   – Assumptions used to disaggregate the data
   – Lack of disaggregated origin-destination data and routes (city, zip code, street level)
   – Lack of transportation planning involvement in industry decision-making
   – High costs of acquiring data (TRANSEARCH)
Identify Texas Freight Data Needs

5. General concerns (continued):
   - Outdated data
   - Reliability of “free public data”
   - Lack of traffic counts on local infrastructure
   - Lack of commodity information
   - Lack of information about seasonal movements
Existing Databases

• Mode of transport information is the most readily available in most of the databases.

• For roadway movements,
  – origin/destination points are insufficient to meet the needs of TXDOT
  – None of the databases contain information at a city or zip code level
  – Only one database (TRANSEARCH) contains information at a county level
Existing Databases

• For roadway movements (continued)
  – Lack of 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,
  – routing data is being collected but may be confidential (e.g. Carload Waybill Sample)
Existing Databases

• Advanced data integration methods may assist in filling some data gaps.

• Need for industry participation to provide data 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
  – actual production and attraction rates
  – model validation
Examine Freight Data Collection Methods

• Survey Data Collection Methods
  – Telephone interviews, mailout/mailback surveys, combination, trip diaries, roadside/intercept interviews, personal interviews

• Technological
  – Loop detectors, sensors, video imaging, GPS, toll tags, etc.
## Identify Current Data Gaps

### Strategy for Collecting & Integrating Available Freight Data – Step IV

<table>
<thead>
<tr>
<th>Database</th>
<th>Country</th>
<th>NTAR/BEA/CSA</th>
<th>Origin-Destination</th>
<th>State</th>
<th>City</th>
<th>Zip Code</th>
<th>Port Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commodity Flow Survey (CFS)</td>
<td>0</td>
<td>0</td>
<td>Import</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Freight Analysis Framework (FAF)</td>
<td>0</td>
<td>0</td>
<td>Export</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Transearch / Reebie (Private)</td>
<td>0</td>
<td>0</td>
<td>Country</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Carload Waybill Sample (Public Use Waybill Sample)</td>
<td>0</td>
<td>0</td>
<td>NTAR/BEA/CSA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Waterborne Commerce Statistics</td>
<td>0</td>
<td>0</td>
<td>Origin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Air Carrier Statistics (Form 41 Traffic)</td>
<td>0</td>
<td>0</td>
<td>Destination</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Border Crossing/Entry Data</td>
<td>0</td>
<td>0</td>
<td>NTAR/BEA/CSA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>North America Transborder Freight Data</td>
<td>0</td>
<td>0</td>
<td>Origin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maritime Administration (MARAD) database</td>
<td>0</td>
<td>0</td>
<td>NTAR/BEA/CSA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>USA Trade</td>
<td>0</td>
<td>0</td>
<td>Origin</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fresh, Fruit and Vegetables</td>
<td>0</td>
<td>0</td>
<td>NTAR/BEA/CSA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>National Transportation Statistics (NTS)</td>
<td>0</td>
<td>0</td>
<td>NTAR/BEA/CSA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Annual Coal Report</td>
<td>0</td>
<td>0</td>
<td>NTAR/BEA/CSA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PIERS (private)</td>
<td>0</td>
<td>0</td>
<td>NTAR/BEA/CSA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Texas Permitting &amp; Routing Optimization System (TxPROS)</td>
<td>0</td>
<td>0</td>
<td>NTAR/BEA/CSA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
# Identify Current Data Gaps

<table>
<thead>
<tr>
<th>Database</th>
<th>Commodity Classification</th>
<th>Industry Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commodity Flow Survey (CFS)</td>
<td>2 Digit</td>
<td>NAICS</td>
</tr>
<tr>
<td>Freight Analysis Framework (FAF)</td>
<td>2 Digit</td>
<td></td>
</tr>
<tr>
<td>Transearch / Reebie (Private)</td>
<td>4 Digit</td>
<td></td>
</tr>
<tr>
<td>Carload Waybill Sample (Public Use Waybill Sample)</td>
<td>2 Digit</td>
<td></td>
</tr>
<tr>
<td>Waterborne Commerce Statistics</td>
<td>4 Digit</td>
<td></td>
</tr>
<tr>
<td>North America Transborder Freight Data</td>
<td>2 Digit</td>
<td></td>
</tr>
<tr>
<td>Motor Carrier Management Information System (MCMIS) - Census File</td>
<td></td>
<td>O</td>
</tr>
<tr>
<td>USA Trade</td>
<td>HS</td>
<td>NAICS</td>
</tr>
<tr>
<td>National Transportation Statistics (NTS)</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>Annual Coal Report</td>
<td></td>
<td>NAICS</td>
</tr>
<tr>
<td>PIERS (private)</td>
<td></td>
<td>HS</td>
</tr>
</tbody>
</table>
Identify Current Data Gaps

<table>
<thead>
<tr>
<th>Database</th>
<th>Trip Frequency</th>
<th>Frequency of Data Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freight Analysis Framework (FAF)</td>
<td></td>
<td>2007</td>
</tr>
<tr>
<td>Carload Waybill Sample (Public Use Waybill Sample)</td>
<td></td>
<td>Yearly</td>
</tr>
<tr>
<td>Waterborne Commerce Statistics</td>
<td></td>
<td>Yearly</td>
</tr>
<tr>
<td>Air Carrier Statistics (Form 41 Traffic)</td>
<td>O</td>
<td>Monthly</td>
</tr>
<tr>
<td>Border Crossing/Entry Data</td>
<td>O</td>
<td>Monthly, since 1995</td>
</tr>
<tr>
<td>North America Transborder Freight Data</td>
<td>O</td>
<td>Monthly, since 1994</td>
</tr>
<tr>
<td>Maritime Administration (MARAD) database</td>
<td>O</td>
<td>Annual</td>
</tr>
<tr>
<td>PMIS</td>
<td></td>
<td>Annual</td>
</tr>
<tr>
<td>Texas Crash Records Information System (CRIS)</td>
<td>O</td>
<td>Daily</td>
</tr>
<tr>
<td>PIERS (private)</td>
<td>O</td>
<td>Weekly</td>
</tr>
<tr>
<td>Texas Permitting &amp; Routing Optimization System (TxPROS)</td>
<td></td>
<td>Daily</td>
</tr>
</tbody>
</table>
Identify Current Data Gaps

Matching O-D and Commodity Code Data

FAF3
- Default Data Source
- Foreign_Origin
- Zone_Origin
- Zone_Destination
- Foreign_Destination
- Foreign_inMode
- Zone_mode
- Foreign_outMode
- SCTG_2
- Trade_Type
- Value
- Tons

CFS
- Hazardous Materials Data

IFD

Carload Waybill Sample
- Yearly Updated Rail Data
  - BEA_Origin
  - BEA_Destination
  - STCC (converted to SCTG_2)
  - Value
  - Tons
  - No. of Loads

Matching

FAF3 - Zone(TX)
- 481 Austin
- 482 Beaumont
- 483 Corpus Christi
- 484 DFW
- 485 El Paso
- 486 Houston
- 488 San Antonio

FAF3 - SCTG_2
- 26 Wood Prod.
- 39 Furniture
- ...

CWS - STCC
- 24 Lumber or Wood Prod.
- 25 Furniture
- ...

Matching

CWS - BEA(TX)
- 087 Beaumont-Port Arthur
- 127 DFW
- 130 Austin-San Marcos
- 131 Houston-Galveston-Brazoria
- 132 Corpus Christi
- 134 San Antonio
- 157 El Paso

- 128 Abilene
- 129 San Angelo
- 133 McAllen-Edinburg-Mission
- 135 Odessa-Midland
- 136 Hobbs
- 137 Lubbock
- 138 Amarillo

Data Gaps
Identify Current Data Gaps

System Output
- XML based Report

Mediators
- Key Factor for Selecting Default Data Source:
  Commodity - Hazardous Material
- Default Data Source: CFS

Data Sources
- FAF3
  - Updated Every 5 Years
  - Foreign_Origin, Zone_Origin, Zone_Destination, Foreign_Destination, Foreign_inMode, Zone_mode, Foreign_outMode, SCTG_2, Trade_Type, Value, Tons
- CFS
  - Updated Every 5 Years
  - Foreign_Origin, Zone_Origin, Zone_Destination, Foreign_Destination, Foreign_inMode, Zone_mode, Foreign_outMode, SCTG_2, Hazardous_Material, Trade_Type, Value, Tons
- Carload Waybill Sample
- Others

Unavailable (Data Gaps)

Integrated Freight Transportation Database

User Input
Freight Data Sharing Partnership

• Most stakeholders interviewed considered that a partnership would be beneficial

• Majority of stakeholders concerned with
  – the mishandling or improper use of data
  – time commitment required in scrubbing and preparing data in-house
  – new government regulations and law enforcement measures
Freight Data Sharing Partnership

• Lightening the information technology (IT) requirements for stakeholders is highly recommended.

• If guaranteed that the information would never become public, 88% of survey respondents were willing to participate in a data-sharing partnership.

• None of the respondents interviewed or surveyed are currently participating in a data-sharing partnership.
Freight Data Sharing Partnership

• Data variables that stakeholders were willing to share (by rank)
  – trip origin/destination
  – number of trips
  – vehicle type,
  – load type (truckload, less-than-truckload)
  – route preference,
  – commodity being transferred
  – cargo weight
  – mode of transport.

• A clear non-disclosure contract is required
• Support from trade associations such as TXTA was found to be invaluable in outreach efforts
Prototype Freight Data Architecture

- Examine existing architectures
- Develop conceptual architecture
Existing Architectures

Freight Analysis Framework 3
Existing Architectures

Oregon Freight Data-Mart System Architecture

- Incidents
- Bottlenecks
- WIM Stations
- Freeway Sensor Data
- Truck Volume Locations
- Truck Generator Locations

PORTAL Data Archive
Stores textual and numeric data and geolocation information

Query Results

Web Server
queries database and generates HTML contents

Request for Google Maps Application Programming Interfaces
Maps

Web Browser
Displays map

(File modified from Figliozi and Tufte, 2010)
Existing Architectures

• Oregon Freight Data-Mart
Proposed Texas Freight Data Conceptual Architecture

1. Integrate 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 Texas traffic data centers
7. Data output and analysis tools
Mediator Architecture

1. Client Query
   [time], [place], [mode], [commodity]

2. Example Databases
   - FAF
   - CFS
   - HPMS
   - PRIVATE DATA
   - NCAST

3. Compulsory Fields
   [e.g. time, place, mode]

4. Field Mapping Adapter
   [FAF: dms_origin => placeFrom]
   [CFS: origin_state_cfs_area => placeFrom]

5. SQL Query
   [select “value, tons” from FAF where dms_origin = “Austin”]

6. Units Assignment & Aggregation
   [tons, aadt, $USD] [sum, min, max, average]

7. Output
Unity DB
an integrated multimodal freight database

conceptual demo system is currently accessible at http://www.unitydatabase.com
Unity DB

• Currently includes the following databases:
  – Freight Analysis Framework
  – Commodity Flow Survey
  – TXDOT Highway Performance Monitoring System Traffic Data
  – ATRI National Corridors Analysis and Speed Tool (N-CAST)
  – Three private sector database samples
Unity DB Screenshot
Study Recommendations

1. Effective partnerships with private sector is needed to ensure adequate freight planning

2. Rigorous outreach and follow-up efforts will be required

3. Data sharing partnership will require a long-term commitment from TXDOT

4. Data from existing ITS technologies should be tapped
Study Recommendations

• Through advanced data integration methods, it is possible to overlay publicly available data sources to assist in filling some existing data gaps