User Guide
The Eastern Transportation Coalition
Freight Data Matrix

August 2020

The Eastern Transportation Coalition is an alliance of 100 public agencies including state departments of transportation, regional and local transportation agencies, toll authorities, law enforcement, port, transit and rail organizations, from 17 states and the District of Columbia. The Coalition provides a neutral forum for key decision and policy makers to address multi-modal transportation management and operations issues of common interest and concern. For more information, please visit our website www.tetcoalition.org

Executive Summary

Because all modes share a region's infrastructure, the movement of all modes of transportation must be evaluated. For example, state departments of transportation are required to monitor transportation use for various modes such as passenger cars, transit, or commercial movements. Freight is an important mode to consider and similar to other modes, practitioners have a variety of data sources at their disposal, but there exists a need to understand and evaluate those sources based on particular needs.

Most reports written to help freight data users tend to fall into two broad areas concerning 1) how to manage a freight data architecture (knowledge content) or 2) what a freight study should include, without discussing specific data needs. Both approaches have merit and should continue to be used, but neither addresses where to find data and how to evaluate it.

This document is intended to assist practitioners in identifying data needs and databases that can provide that needed data. A Freight Data Matrix and associated White Paper: "Demystifying Freight Data" were developed by The Eastern Transportation Coalition (the "Coalition", formerly the I-95 Corridor Coalition), which categorizes freight data into distinct groups, such as time and location, or other discussions on costs and usage. This allows practitioners to more readily identify relevant sources of freight data for planning and other activities

The need for a data matrix is discussed first, followed by detail on how the matrix is structured by an attribute type. Each attribute type is introduced, detailed, and followed by some considerations for practitioners. Finally, the user guide concludes with practical steps on identifying data needs and using the data matrix to evaluate datasets.

Contents

List of Figures

The Need for a Data Matrix

Why do people want to know freight data? Freight data itself has little merit, but the information that using freight data provides is meaningful. For example, all these elements have implications for freight research needs:

- Where is the state (region) growing/declining?
- Is manufacturing, agricultural, and/or mineral extraction expanding?
- Is trade increasing or decreasing?
- Any significant expansion of intermodal rail facilities, ports, industrial parks, or airports in the state?

Each of these study areas requires some different domain knowledge. Over the past 50 years, the planning process has become more information-intensive, requiring non-engineering data on environmental or economic concerns. But the challenge is that we discuss "freight" confronted with a universe of data without a context to evaluate what is needed. The following list shows some potential needs to evaluate various freight movements to provide insights for state and local planners.

- Time (date) associated with the shipment movement itself;
- Mode (truck, rail, water) and submodal;
- Product origin and destination, including international shipments;
- Facility or equipment interchanges, including intermodalism;
- Type of equipment used to move the product;
- Product weight, density (measured in pounds per cubic foot) and value;
- Shipment size;
- Route used for domestic shipments. For international trade, the inland movement to/from a port, airport or gateway and the movement to/from foreign markets;
- Shipper and receiver relationship (contractual);
- Transportation rates, fees, and costs;
- Time sensitivity (just in time) or perishability of the product;
- Equipment movements, including repositioning empties and backhauls;
- Other products moving on the same piece of equipment (multiple products from either the same or different shippers);
- The economic multipliers associated with the shipment (tied to other modeling efforts);
- Cargo ownership, including the names and addresses of the shipper, receiver, and carriage provider;
- Tax and fuel payments tied to shipment;
- The relationship between goods movement to the economy and jobs;
- Timely data collection and reporting of the shipment event to others (the information is reported quickly after the shipment occurred) identifying the actual product that was shipped;
- Determining where did the truck stop, when and for how long?

There are many reasons why this level of detail is not required for different studies, and even if this wealth of information was available, any research project may only require a subset of that information. The gap between having freight data and its application to satisfy federal requirements or meet other agency goals

remains an important but challenging, role for public sector transportation planners. The gap was highlighted by a previous report published by the Coalition on state freight planning.[1]

The report recommended leveraging data from other agencies, where appropriate, to improve analysis and acceptance by others. But it was data, ranging from the costs of acquiring, usage restrictions, interpretation, and application, that presented unique challenges.[2] States had "low freight knowledge", which required agencies to "get up to speed" on freight issues, despite several states having completed freight planning activities previously. This was complicated by the inability of staff to find the time to pursue freight domain knowledge and skill development.

For example, most Coalition members used the following databases: the Freight Analysis Framework (FAF), Transearch, National Performance Management Research Dataset (NPMRDS) or the American Transportation Research Institute (ATRI) (**Error! Reference source not found.**). States augmented their w ork with other datasets or industry inputs. Databases were collected, purchased, etc., as part of the freight planning consultant procurement process. Only one coalition member possessed a small annual freight data budget.

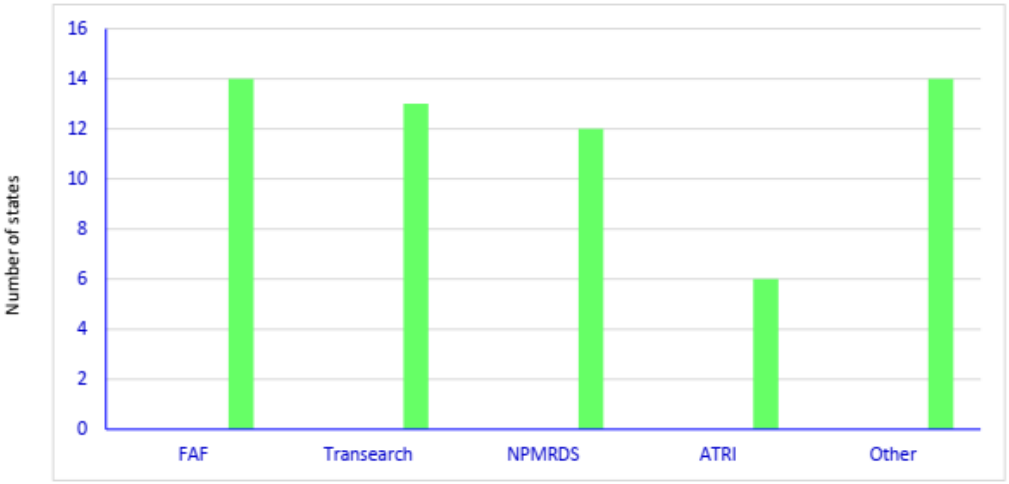

Figure 1 - Data Sources Used By selected States in the I-95 Corridor for State Freight Plans (source: I-95 Corridor Coaltion. (2018). *Member State FAST Act-Complain Freight Planning Activities, Issues and Recommendations)*

The inability to secure funding for freight data-limited staff access. This also prevented the states from adding value to the state/region's general transportation planning efforts through possible collaboration, such as through data-sharing agreements. As a result, most state freight plans are conducted by consultants, directed by in-house staff, who reported feeling unsure considering their "self-perceived" lack of freight knowledge.

Also, the lack of a single, national, user-friendly data source for freight was reported as a challenge for planners. While the coalition members want more trade flow information, there was also a need for economic

[1] I-95 Corridor Coaltion. (2018). *Member State FAST Act-Complain Freight Planning Activities, Issues and Recommendations.* https://i95coalition.org/wp-content/uploads/2018/06/I95CC-State-Freight-Plan-Report-Final.pdf?x70560

[2] Many of these same elements were listed in National Freight Cooperative Planning Program, No. 49, https://www.ncfrp49-newfreightdata.com/resources/

data, especially when considering network elements and ways to improve coordination between state and local planners on integrating freight data on movements and system needs.

The report suggested that states examine ways to pool data purchases, or other economic information, to do other required US DOT reports. Other recommendations suggested disaggregated commodity flows to the county level and a freight network. Most relevant to this report and the associated Coalition Freight Data Matrix is the request for support in accessing, current, and user-friendly freight data. To respond to this need, the Eastern Transportation Coalition has assembled a Data Matrix to provide a consolidated source for freight data. This tool is detailed in the next section, The Data Matrix.

There remains a complexity regarding freight ranging from a focus on the individual supply chains and the aggregated movements based on some criteria, such as mode, industry or region. This complexity requires organizing freight attributes to understand the research question and to identify potential information/guidance. Recently, AASHTO published such a report under the SHRP2 Program entitled "Freight Data Guide for Improved Transportation Planning."[3] The report identified seven broad categories of freight-related data research needs:

1. Freight Movements, movements by region, by mode and by origin and destination,
2. System Condition and Performance, information on system underperforming (bottlenecks), condition of system, freight system assets.,
3. Trends for the Future, scenario planning, and forecasts,
4. Freight Generators and Land Use, mainly location-based information,
5. Freight Corridors and Last Mile, How does freight use the system, routes, and truck type,
6. Safety, where are safety hotspots and rail grade crossing areas,
7. Truck Parking, where do trucks stop during their movement.

Context matters, especially when determining what approaches are applicable to specific research needs. Given there is no single "freight study", or "freight database", complexity is common. Focusing on attributes that can be considered for both operations and planning, there may only be a need for six broad categories related to freight shipments. These six categories are:

- The Truck, Including Weight, Registration, Configuration,
- The Commodity, Including Weight, Value, And Origin-Destination
- The Safety Of The Movement, Such As Permitting, Registration, Hazmat,
- The Location Of The Movement, Such As Speed, Roadway, Time Of Day,
- The Costs Of The Movement, Such As The Cost To The Shipper, Tolls, Rates,
- Time Refers To Information Related to the Shipment, such as the Day, Month, Year That Something Occurred, But Also Including Future Predictions.

These categories form the basis of the Data Matrix which is detailed in the next section.

[3] https://www.fhwa.dot.gov/goshrp2/Solutions/Design/C20/Freight_Demand_Modeling_and_Data_Improvement

The Data Matrix

Purpose

Given the diversity of freight shipments, the development of a data matrix was seen as a way to distill some attributes that could be understood and used by researchers. This is not an exhaustive list, but the data matrix and the corresponding spreadsheet are designed to assist planners in framing their research needs and understanding how to compare information within and among freight databases.

Each section below will have a short introduction, followed by the data matrix table attributes and an associated description as well as a Considerations For Practitioners section.

First Level Structure

In creating the data matrix, categories were selected to provide some organizational structure when comparing specific attributes. The categories are:

Information about the database (Metadata)

1. *Summary:* Overview of the database, including its primary uses, strengths, and weaknesses.
2. *Ownership:* Who owns the database, including access to information on the data source, licensing agreements, release information, etc.
3. *Data Collection and Usage:* Thoughts concerning how the data is developed.
4. *Training and Summarized Data:* Additional insights into the database and the availability of summarized data concerning its use.

Information within the database (Attributes)

5. *Time:* Information related to when the truck is operating, including travel times, and speed.
6. *Truck:* Information on the truck itself, such as its age, weight and/or size.
7. *Commodity:* Often used as a proxy for who benefits from the movement of the cargo (shipper/carrier), but there are special considerations, such as hazardous cargo, weight, value, that are also important to know.
8. *Safety:* Information related to crashes/incidents, but can also include information on the trucking firm or driver, hard braking, inspections, and other attributes.
9. *Location:* Information related to the geospatial attributes of the truck, such as route choice, facilities.
10. *Cost/Revenues/Finances:* Information directly to operations, such as fuel, taxes, tolls. There are many indirect costs, such as congestion or lost productivity due to system unreliability, that are not captured in these datasets and require some modeling.

Information about the application of the database

11. *Usage and Application:* Provides the researcher with a starting point to consider the usefulness of the data.

Items not discussed are things outside of the freight shipment information itself, but that exist in a public space, such as emissions and other environmental factors. Many of the items referred here may relate to other modal databases, however, reviewing those resources will be useful.[4] Private and trade association data is also available.

Second Level Structure

Summary

The summary metadata provides an overview of the data set, focusing on key applications, attributes, and strengths and weaknesses. The goal is not to delineate how a database should or should not be used, but rather to provide a starting point for determining the usefulness of a database for specific needs.

Field	***Attribute description (Summary)***
Database Name	The data name, including any abbreviations
Summarized	Is the information available in a summarized manner for quick access?
What Are the Units of Measure?	What are the most important variables being measured?
Lowest Geography	What is the lowest level of geography collected or modeled by a database?
Strength	The pros of a database, especially in the topic areas were the database excels.
Weakness	The cons of a database
Limitations on Data Use	A catchall for items related to use, release, training, etc.,
Website	Where to get more information on the data source

Considerations for Practitioners:

- The analyst can examine the key elements of the database to determine its application or usefulness regarding a particular research need.
- One should develop a "beginner's mindset," focusing on this action of continual learning. For example, the data matrix is designed to be a tool, and adjusted to fit a particular need.

[4] Rail information https://www.stb.gov/stb/industry/econ_waybill.html, Water information https://www.iwr.usace.army.mil/About/Technical-Centers/NDC-Navigation-and-Civil-Works-Decision-Support, https://www.maritime.dot.gov/data-reports, Aviation (https://www.bts.gov/topics/airlines-and-airports-0)

Ownership and Use

Every database has an "owner," someone responsible for the release of that information. That vendor/agency may create or model the information based on other data sources. Every database requires some data wrangling/cleaning management, but the data owner interfaces with the data user. In doing so, the researcher could understand how the data owner sees their relationship to the data and its support, but also how they manage their intellectual property.

Field	***Attribute description (Ownership and Use)***
Data Owner	There is a mix of public and private databases using in freight analysis. Generally, Federal Databases are free to use. Some venders collect data from various sources, but have limitations on what data may be released.
Free or Fee to Access?	The cost to a user (not the agent who produces/processes/releases the data).
Is the Data Proprietary?	Proprietary information may require some signatures concerning the use, application, and data ownership. Some databases are based on publicly available sources, but the researcher pays for the aggregation, cleansing, etc., which may not have been done with the original data release.
Licensing Agreements	How is the data licensed for use? Can multiple users access the database, can you share the data?
Primary Audience for Which Data Was Developed	Was the database developed for private sector or public sector clients, or based on capturing some new data point, not necessarily focusing on a market segment, or for a certain industry or sector? Knowing this can provide some understanding concerning how the database evolves in the future.

Considerations for Practitioners:

- New datasets are emerging all the time. The questions are "can I use these databases", "what are the costs", and "how to evaluate them in comparison to existing databases"?
- Do I need to mention the data source when it is presented?
- The internet, and the corresponding desire for "pushing/capturing data", has created a new paradigm concerning the valuation of information. More data elements are being recorded that were unavailable in the past, but the associated costs with storing, managing and monitoring data quality have decreased, generating more bias towards a perception of "cheap" or "free" data. In many ways, these data development/maintenance costs are borne by the firm/agency responsible for the dissemination of that data. For the federal government, these data programs are funded by the respective agencies' budget elements. The multiyear process of designing, collecting and processing

the CFS runs $20 million per cycle, while the FAF costs two million dollars to pay for services to merge and assign existing sources.[5]

Data Collection and Coverage

A database's usefulness depends on its ability to describe what is being measured. In some ways, most vendors will promote/focus on these metadata elements to differentiate themselves from other data programs, explaining how they have more inputs, or better data management systems, or some process that makes their information more valuable to the research community.

Field	Attribute description (Data Collection and Coverage)
Sample/Universe	Universe means that all elements are reported, while sampled means that a portion of the information is collected. Most transportation data will reflect "the universe" (i.e., all data points are collected), or be partial samples to gather information.
What Is the Sample Size (%), If Known?	What percentage of the universe is sampled? Sampling requires more scrutiny to address small sample sizes or statistical errors.
Primary, Secondary, Modeled	Primary data refers to data produced, processed, and distributed by the agency or firm. Secondary data refers to information that is collected from diverse sources, which provides more information, but is one step removed from the data collection process. Modeled data are produced by supplementing a source with additional information to fill in statistical gaps or to add attributes that were not explicitly stated in the original data sources. Primary data are collected directly from the transaction/movement, while secondary data is processed by someone else, external to the original data set, and finally, modeled data can be blended from multiple primary and/or secondary datasets.
Collection Frequency	Data can be collected on an ongoing basis (submitted paperwork, paid tolls) or is provided through any reporting format (monthly, quarterly, annually). How often the data collection occurs may influence the accuracy of the data reported.
Timeliness of Data	The further data moves in time from when it was collected, the more perishable the data becomes. This may require some additional modeling to "refresh the data".
Publication Frequency	Data may be continually collected, but it may not be released immediately after collection. Knowing when data will be released can assist in planning schedules, purchasing decisions, etc.

[5] These figures represent program costs and does not include staff resources. For the FAF, this does not include the costs associated with the various data programs, such as the CFS, that are integrated into the FAF. See White Paper "Comparison of the CFS to the FAF – A Lesson in Data Architecture" to better understand how the FAF depends upon multiple databases, showing the value of the work required to prepare this one database. The same could be said for most other databases that collect and integrate data from many sources.

Considerations for Practitioners:

- These elements represent the attributes that may result in your data being rejected as "unworthy" for a specific study (and lead to uncertainty concerning the overall research effort). Arguing that historical database attributes remain relevant or that the data does not have the granularity one requires or that it is an incomplete representation of the universe, etc., often results in an analysis being discredited not by the analytical approach, but by the quality of the data. A common saying that applies here is "Garbage In-Garbage Out".

- Federal databases have specific time guidelines for collection and dissemination. As such, they may be fair "late" when released, but often are the only database available on a topic, requiring additional information/processing to "update" or provide near term forecasts.

- If possible, understand the originating paper trail of the database. (For example, the FAF is based on the CFS, but the CFS survey document includes some attributes that are not included in the FAF.)

Training and Summarized Metadata

There are many ways to comprehend a database's usefulness. By seeing what materials are posted online to assist users, or to discuss with others who have worked with the same topic/data, one could learn the nuances given some patience and curiosity. Other researchers can benefit from seeing how peers have used a database or similar databases for various reports.[6] For example, some databases will provide summarized information, either as static maps/charts or summary reports. Other databases, due to the complexity of the data/information, are unable to provide summarized information for general users. Finally, some databases may provide tools to manipulate the database without having to download and process raw data files.

Field	**Attribute description (Training and Summarized Metadata)**
Training Materials	Are training materials, videos, etc., available?
Data Development Reports	Does the database have Frequently Asked Questions, user forums, social media, etc., concerning, methodology, structure, analytical approach, etc.?
Online Processing Tools	Can the user process the database through a portal, allowing the user to access the data and produce simple data calculations, tables, etc.?

Considerations for Practitioners:

- Two challenges practitioners face are: "do they have the time to engage in reviewing the training materials" or "are they willing to understand the pros/cons of a database?" This leads to questions about data management/managing user expectations. This may also prevent researchers from falling into the common fallacy that a single database can answer every question.

[6] The development of a Coalition repository, reflecting on common research topics and what analytical tools were used could potentially elevate the Coalition's memberships understanding of freight transportation data.

- Data visualization tools – some databases are largely used through a GIS software application, while other software applications are data visualization platforms. These software applications are not databases but depend on developing and formatting the data to generate outputs. For example, in Figure 2, we see data as blocks, and while arranged and sorted, without context, we do not know what to do with the information regarding the relative values between the color units.

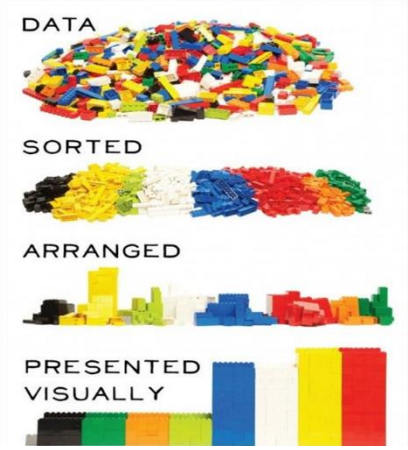

Figure 2 - Hierarchy of Data into Visualization

- Non-researchers may see a chart/table/figure developed by a database, which they may expect duplicated for their region. The same question should be raised when consultants deliver reports, namely, "can the staff replicate/update that information?"

  Most data providers are proud of their work and are delighted to talk to their user community, but many questions could be answered if a user consulted a Frequently Asked Questions, tried to manipulate the data, or relied on user/training materials.

Time Attributes

Every database contains some element of time (such as weekly, quarterly, annually). The time element contains not only information about the past or present condition, but in some cases, datasets can also include forecasted variables as well. These attributes are separate "time attributes" related to the actual processing and other metadata elements that are not specifically used in a freight research effort.

Field	Attribute description (Time Attributes)
Time of Day Information	The database reports time of day for the truck movement
Duration of Trip	Operations at a gate, through a facility, last-mile delivery, driving tours, etc.
Parking Duration	How long was the engine stopped? For example, is the truck stopped for staging purposes or for overnight rest?
Time Element	How is time reported, such as annually, monthly, quarterly?
Forecasted Time Periods	Are there forecasts associated with the database? If so, what is forecasted, and by whom? Planners require forecasts to validate their long-term models. Understanding when the forecast was made (and its economic assumptions) is critical when reviewing a forecast's applicability.

Considerations for Practitioners:

- Some operational data, such as dwell and gate times, may not be available from a terminal operator, but represent an import consideration for estimating truck performance. The development of Electronic Logging Devices (ELDs) may provide staging and other locational information related to truck movements, including hard braking and stoppages.
- Not all databases report time as a column but instead will list the time attribute in the file name. As such, the researcher should get in the habit of putting time attributes in the data, especially if doing multiyear studies.
- Forecasts can become "stale." For example, the economy or a key sector within a state can experience change, but the forecast, not being updated, may not capture those key "turning points." As such, using forecasts may be problematic for some studies, especially when discussed with the public sector, who may operate under a different time horizon. However, a 30-year forecast or some similarly long forecast will assume these variations are reflected by focusing on average changes, not yearly changes, over the forecasted period.
- For some products, such as agricultural goods, seasonality can influence freight activity within a region.

Truck Attributes

By definition, freight moves on something. For the most part, freight moves in special vehicles. While the category says "Truck." many of these same attributes can be applied to other modes.

Field	Attribute description (Truck Attributes)
% Of Truck in Traffic Volumes	Does the database compare truck volumes to the total traffic, such as Average Annual Trips or units?
Truck Classification	Does the database report truck classification data by USDOT Guidance[7]
By Axle	Does the database report truck information by the number of axles?
Truck Weight	Information on the weight of the truck and its cargo.
Truck Registration Number	Do you want to link to the International Fuel Tax Agreement (IFTA) or some other resources?
What Modes Are Represented	Several databases will have other modes in addition to trucking.
Relationship to Other Modes	Databases that do not have specific truck data can be used to evaluate truck movements related to drayage operations, traffic generation, etc. Trucks interact with all modes and where data is not available specifically for trucks, it can be teased out by understanding other modal traffic.
Name of Carrier	What is the name of the trucking company?

Considerations for Practitioners:

- To discuss truck data, one must define a truck and its operations. Using Trucking 101, there are several broad categories for trucking, based on different criteria related to ownership of the truck or the cargo being carried.[8] For example, the truck may operate on a "For Hire" basis, meaning the cargo will carry freight for anyone willing to purchase their services. Other carriers may operate as "Private Carriers", which means their trucks only carry their cargo and do not offer services to other companies. Other considerations include whether the company operates as a specialized carrier, such as a car carrier, or general carrier. Furthermore, how far does the truck travel, such as local deliveries, within the region or state, or for interstate trade? Private fleets operate largely within their own network, while Truckload carriers tend to sell point to point movements. A Less than Truckload (LTL) will consolidate loads, moving on a hub and spoke network. There are parcel and express delivery carriers, often called integrated services.

[7] https://www.fhwa.dot.gov/policyinformation/tmguide/tmg_2013/vehicle-types.cfm

[8] TRB Electronic Circular E-C146: Trucking 101: An Industry Primer, http://www.trb.org/Main/Blurbs/164560.aspx

- Trucking information can be enhanced by working with other state agencies, such as the state police, revenue agencies, but there may be some limits regarding confidentiality or access.
- For multimodal studies, most modes may have better O-D flows than trucks, resulting in trucks often treated as the difference between the top-line estimate and the known traffic on other modes.
- Many databases will have "trucking" information but will often require converting tonnage into trips. Most practitioners use the information from the VIUS, last published in 2002. The U.S. Department of Transportation hopes to recreate the series in the near future.[9] Some states have developed their own VIUS updates[10].

[9] Information on the VIUS is posted here https://www.census.gov/econ/overview/se0501.html

[10] California presented on their VIUS Program at the TRB Innovations in Freight Data Conference, 2017. http://www.trb.org/Main/Blurbs/176391.aspx

Commodity Attributes

Commodity reflects not only the physical commodity shipped but the relationship to the beneficial cargo owner, shipper or an intermediary. The commodity also includes things describing the commodity itself, such as its weight, size, and value. Other details, such as if the commodity is hazardous or possess some special regulatory control (military shipments) could also be discussed here. However, not every study requires knowing who owns the cargo, or even great detail concerning the commodity beyond some broad description. For example, shipping often describes items such as containerized (where the cargo moves in international containers), bulk cargos (such as grain (dry bulk) or oil (liquid bulk), or breakbulk (such as cars or logs, where the cargo is individually packed and shipped. Different databases will describe cargo based on operational attributes (such as the ocean shipping example) or some other clustering, such as manufactured or agricultural goods.

Field	***Attribute description (Commodity Attributes)***
Commodity Detail/Codes	What commodities are reported here, such as broad categories of cargo or tied to a specific commodity structure? There are multiple freight codes, so the researcher would benefit from understanding these classification codes.
Shipper	Is the name of a shipper or some other agent reported in the database? For private companies, the truck itself represents the shipper (such as grocery trucks are mostly privately owned.[11]
Value of Cargo	The value of the cargo, for international shipments this is required, although some domestic databases will estimate the value or may have access to domestic bills of lading.
Weight of Cargo	Weight (in English or Metric units) of cargo.
Ton Miles	The measure of system usage - ton-mile is defined as one ton moving one mile. It is useful information concerning system use of transportation that captures some elements of not only activity but movement.
Special Cargo Concerns	Some cargos have special criteria concerning their shipment, such as hazardous cargo or military shipments, or grandfathered cargos by state (such as agricultural, mining or timber).
International Data	Reports information on international shipments, such as origin or destination, or movement through a facility.
O-D Pattern	The cargo shipment's origin and destination. The origin and destination are influenced by where the paperwork begins (in the case of international shipments) not necessarily the origin of the product.

[11] See Trucking 101 http://www.trb.org/Main/Blurbs/164560.aspx for a discussion of privately and publicly owned truck operations.

Considerations for Practitioners:

- Not all commodity codes capture the same information. Some tend to be based on shipment characteristics, while others may be based on industry groupings. Furthermore, different modes use alternative classification structures. The different commodity classifications often require the use of crosswalks, which may lead to some misalignment when reporting out summarized numbers.[12]

- Most researchers criticize that agricultural shipments or local delivery information are not necessarily captured at a project level. Part of the problem is the information of that movement is normally not recorded in most modal flow databases (not geographically focused) and must be constructed/modeled. For example, the Shipper Export Declaration records where the product began its international movement, not its original point of production.[13]

- Customs district and port information may not align with actual port authority/terminal boundaries. Customs and Border Protection see ports and terminals through an administrative lens. For example, the Port of South Louisiana lies within three different Customs port areas, complicating any research seeking to break out tonnage along the lower Mississippi River by the port authority, not the Customs Port area.[14]

- Some databases will include intraregional shipments. Without addressing this, the researcher could easily double count intraregional shipments, as the Outbound Intrastate movement and the Inbound Intrastate movement become added as separate data points.

- Common Classification Codes:

    - North American Industrial Classification System (NAICS) - https://www.census.gov/eos/www/naics/
    - Standard Transportation Commodity Code (STCC) - https://www.railinc.com/rportal/standard-transportation-commodity-code
    - Standard Classification of Transported Goods (SCTG)- https://bhs.econ.census.gov/bhs/cfs/Commodity%20Code%20Manual%20(CFS-1200).pdf
    - Harmonized Commodity Codes-https://www.usitc.gov/tata/hts/index.htm

[12] There is a good writeup comparing usage of Commodity Codes on the FAQ section of the Transborder Freight Database https://www.bts.gov/statistical-products/transborder-freight-data/north-american-transborder-freight-data-faqs Question 17.

[13] Census – how to fill out the SED (copy and paste this into your browser) https://www.census.gov/foreign-trade/schedules/b/2011/correctwayforb.pdf

[14] Customs and Border Protection Port Codes https://www.cbp.gov/contact/ports

Safety Attribute

This refers to data related to crashes/incidents but can also include information on the trucking firm or driver, or operations, such as hard braking, and/or inspections and other regulatory activities. While freight studies can focus largely on commodities and economic activity, understanding the safety implications remain of public interest.

Field	***Attribute description (Safety Attribute)***
Inspections	Does information about inspections exist, tied to a location, a company, or a driver? Does that information relate to other secondary information, such as cargo weight, origin/destination?
Regulatory	How and what regulatory actions have been reported concerning truck movements?
Hours of Service	Any hours of service violations, and if so, where?
Speed	Truck speed. The focus is not on trucks speeding, but more on where truck speeds are declining, indicating bottlenecks or congestion.
WIM	Weight in Motion Data.

Considerations for Practitioners:

- In addition to internally generated safety information, freight data researchers may consider rail grade crossing and other safety data resources.[15]
- Federal Motor Carrier Safety Administration posted additional safety/operational information that can be useful in discussing truck/passenger safety concerns.[16]
- Reach out to other public agencies concerning freight safety. For example, Virginia DOT included State and local police when conducting truck parking studies.[17]

[15] Federal Railroad Administration, https://safetydata.fra.dot.gov/OfficeofSafety/Default.aspx, and FHWA https://highways.dot.gov/research-programs/safety/safety-data-and-analysis

[16] https://www.fmcsa.dot.gov/safety/data-statistics

[17] *Virginia Truck Parking Study.* http://www.virginiadot.org/projects/resources/VirginiaTruckParkingStudy_FinalReport_July2015.pdf

Location Attribute

Transportation depends upon infrastructure. These assets require maintenance, planning and program considerations. To effectively do this, freight movements must be linked to specific roadways and transportation assets. In some ways, location tied to physical roadways becomes the most essential element for some freight studies, but the question concerning what locational data is available can be problematic. Location reflects not only the infrastructure asset used by a carrier, but the location of a facility, the economic geography of the region, or other geospatial attributes.

Many databases, by their collection, processing, release, etc., will have some geospatial information such as state or national data, or at a more granular level, such as a border crossing or toll facility. The problem is that not all levels of geography correspond with every researcher's needs, resulting in the use of other databases to disaggregate freight data or to assign freight traffic to different segments.[18]

Field	***Attribute description (Location Attribute)***
Route Used	Information about specific highway or route, or flows through a particular facility. For some datasets, this may be the number of cars at a certain facility, which is tied to a specific route.
Permits	Type or information about permitted cargo (exceptions, over-dimensional, etc.) Permits can be done at a state level, or for particular segments (industry). Having the ability to review these activities with specific facilities can assist in developing pavement conditioning monitoring systems, or other corridor improvements that may be necessary to support trucking.
Work Zone	Did the truck transit a work zone? While work zones are temporary, knowing the geography and duration of work zones could influence modeling and user costs.
Parking	Information about operations, location, or search time concerning truck parking.
Geography Detail	What is the lowest level of geography reported? For some databases, this may be national or state level. Other databases may possess information on a segment or facility level or even at a specific lat/long position.
Facility Information	Specific information tied to a facility, port, or some other infrastructure asset.

[18] NCFRP 20: Developing Subnational Commodity Flow Databases, (copy and paste this into your browser) http://onlinepubs.trb.org/onlinepubs/ncfrp/ncfrp_rpt_026Dev.pdf

Considerations for Practitioners:

- The focus on performance metrics can be tricky, as data collection and aggregation may not adequately capture all the relationships of a system. For example, the ATRI Bottleneck report highlights many problematic intersections, arguing that lower speeds indicate a structural problem in the roadway[19]. However, there may be land use or other considerations outside of the control of the DOT to address the roadways' operational speed.

- Tour information is becoming available through various vendors, although that information may still require the researcher to understand how to manipulate GIS data. This may involve data cleaning efforts, in addition to linking the GIS databases to specific highway networks.[20]

- All of these become important to know what the level of geography is, even if the database does not provide the same level of granularity as is required. Furthermore, there is an assumption that data aggregated for a total is more robust than taking broad data and "disaggregating" the information to a local level.

- Electronic Logging Device information is becoming available, providing additional information about the spatial nature of the transportation movement. There are other questions concerning accessing ELD information, namely, can providers begin selling the information to State DOTs without compromising the relationship between the technology company and the trucking industry.[21]

Cost/Revenues/Finance Attributes

Transportation decisions are based on the costs between shippers and carriers. Transportation costs become hard to quantify for public sector planners, as costs and rates are not the same for all involved in transportation activity.

When discussing rates and costs, there is a general set of criteria one can use when considering freight shipments and modal choice. Heavier, homogeneous products, which are not time-sensitive, will move by water, rail, or pipeline. As the product's shipment sizes get smaller and more valuable on a price per ton basis, shippers are willing to pay for this added service and speed. The comparison outlines mode choice, but not why shippers select different modes or carriers for certain services.

Shipment characteristics do not discuss the "why" concerning how transportation rates and costs are developed. The question of costs, while germane to the discussion on transportation usage, is not necessarily available for all shipments consistent with the other databases listed here. Most databases reported in the data matrix do not contain any information on costs, but the cost attributes are included here to represent a potential avenue for discussion, especially related to working with regional tolling authorities.

---

[19] https://truckingresearch.org/2019/02/06/atri-2019-truck-bottlenecks/

[20] Daniel Hulker discussed how he linked the HERE datasets to the Kentucky Transportation Highway network. https://youtu.be/gQTGvbbGiGc

[21] https://www.fmcsa.dot.gov/hours-service/elds/electronic-logging-devices

Field	*Attribute description (Cost/Revenues/Finance Attributes)*
Costs-Rates	Information on costs or rates. Costs and rates provide good barometers comparing the costs of using a particular route, node, etc., but also costs/rates serve as market indicators concerning traffic demand, willingness to pay, modal choice options, and other transportation considerations. In some ways, costs may also indicate other market signals, such as congestion valuation.22
Tolls	Estimated tolls paid. The question of tolls also indicates the willingness to pay for the adoption of technologies to manage tolls, consider peak versus non-peak costing.
Permitting	Estimated Permit fees. Each state has provisions for permitting over-dimensional/overweight loads, or routes, which may provide additional information concerning truck operations.
Taxes	Estimated taxes for a truck.

Considerations for Practitioners:

- Information on permits, tolls, and other direct costs are captured by various agencies but represent a potential resource for planners.
- The International Fuel Tax Agreement (IFTA) could provide some additional resources, but due to confidentiality data governance rules, may not be accessible to various agencies.
- Tolling data also represent a resource that could be considered in examining peak/non-peak staging and other operational considerations, especially if the truck is able to be tracked between multiple facilities (pending the ability to manage data confidentiality).
- Transportation costs and rates are not necessarily equal, and are fairly fluid, depending upon market conditions and business operations.[23]
- Many of the items here could be utilized as information to assist in preparing studies, such as standardized costs for benefit-cost analysis, economic impact studies, and other economic multipliers, that while not freight planning, are important for seeking/allocating program funding.[24]

Usage Metadata

Of the various attributes, this is the most dependent on individual practitioner skills and abilities. For example, if one has a strong excel skillset, but limited GIS knowledge, then GIS data could be cumbersome. Others may have domain knowledge that is not necessarily transferable between categories, such as a local

[22] The ATRI report on truck parking represents not a database, but an information resource. "An Analysis of the Operational Costs of Trucking: 2016 Update. American Transportation Research Institute."

[23] ITTS Working Paper WP 20, "The Potential for Domestic Waterways to Handle Non-Traditional Cargos" included appendix article on using rates and costs in freight studies.

[24] https://www.transportation.gov/office-policy/transportation-policy/benefit-cost-analysis-guidance-2017

planner having to address economic forecasts. The table is presented as a starting point for individual practitioners to assess their understanding of a data resource, which may provide a structure to evaluating one's understanding/range of skills in a certain research effort.

Field	Attribute description (Usage Metadata)
Potential Uses of Data	What were some applications where you used the data? Were these good or bad, and what other databases were used, if necessary, to supplement the report.
How Easy Is the Data To Use?	How much manipulation was required to prepare and utilize the data? Do I have the database structured for additional queries, updates, etc., if required?
How Easy Is the Data To Share?	Did I have problems sharing the data, files, etc., with others, and if so, was it a formatting problem, or a training question?
How Easy Is The Data To Analyze?	Did my organizational structure, queries, etc., make the data analysis easy to conduct? Do I need to change anything concerning my knowledge management software/skills to better utilize this information?
How to Integrate W/ My Other Tools?	Do I need to integrate this database with other resources, either at a detailed or summary level? Can I integrate this information with agency resources, such as performance metrics or real-time data access, to provide additional value to my senior leadership?
For What Research Have I Used This Data?	Have I, or my agency, used this data in the past, and if so, how and for what reports/outputs/etc.?

Considerations for Practitioners:

- This may require talking to others to evaluate not only your data programs but also their use of the data. Building institutional awareness of data programs is an ongoing task, but necessary to develop the information and skills required to support the ever-evolving policy and planning needs. Furthermore, the adopting and integration of new data sources and analytical tools require that agencies understand their current practices to provide a starting point for making a strategic investment in new analytical capacities.

How To Use The Data Matrix

When discussing the use of the data matrix, the question becomes one of data/knowledge management. There are many resources for freight data that can be explored to consider developing a robust freight data analytical program, but the data matrix can be used for quick checking when considering what data is required for a planning need.

1. Use a scorecard, as discussed below, to identify data needed to complete the study.
2. Assess in-house data.
3. Identify what internal data sources are missing, and how important missing pieces are for need.
4. Use the data matrix to identify potential sources of data to fill in missing needs identified above.
5. Evaluate and compare that information in the data matrix to evaluate which databases contain information that supports the research efforts.

Using a Scorecard: Example

The table below shows a scorecard template that can be used to select the critical attributes for the particular planning need and write out ideas concerning analytical requirements for each attribute.

Scorecard Template

Critical Attribute	*Attribute Need (*write objective and attributes required/desired)
✓ ✗	Truck:
✓ ✗	The commodity:
✓ ✗	Safety:
✓ ✗	Location:
✓ ✗	Costs of Movement:
✓ ✗	Time:

Truck Parking Scorecard

✓	Truck: I want to know the number of trucks.
✗	The commodity: not important, unless I want information on shippers, cargos, etc.
✗	Safety: While I may want to know any operational issues, regulations, etc., they may not be mission-critical.
✓	Location: Critical, I want to know the movement starts/stops.
✓	Costs: May want to look at truck times through tolling facilities to see if there is a behavior change (variable rates, time of day, etc.)
✓	Time: I want to know the time of day, duration of staging, and overnight parking.

Using the Matrix

The Data Matrix is set up in an Excel Spreadsheet with two sheets. The Vertical database is set up to analyze the databases by attribute, while the Horizontal Workbook is better for data entry. The left two columns are frozen panes to assist in the navigation of the database.

The Vertical sheet, shown here, allows the user to enter new database attributes. There is a yellow band highlighting the different metadata elements. One could hide or resize the columns when entering new databases. The files were manually copied and transposed into the Vertical database.

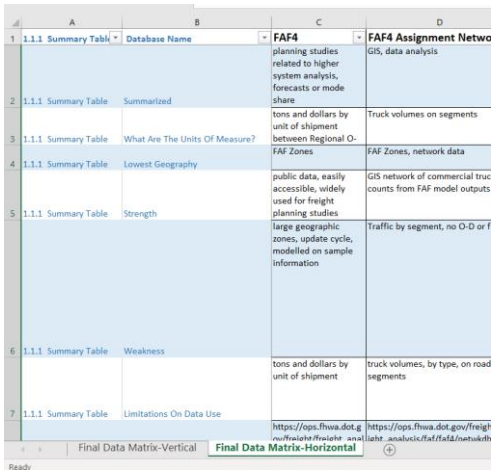

Figure 3 – Data Matrix Overview

The Horizontal sheet allows the user to filter each attribute. Filters can only be used on columns, not rows. Once data needs have been identified using the scorecard, the Data Matrix can provide a quick way to review databases when considering research needs.

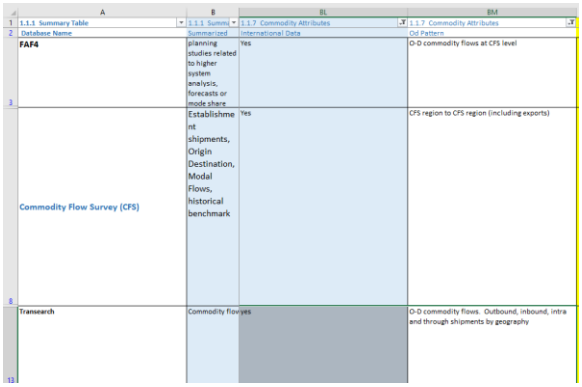

Figure 4 – Filtered Databases for International and O-D Flows

There is no single database that will address all data needs concerning understanding freight. Practitioners must continue to improve their understanding of freight logistics and economic conditions to plan freight data/analytical research needs. This will require both internal support and funding, and the ability of the practitioners to continue to expand their knowledge regarding the nuances of freight-related research and analytical skill development. Composition of various data sources will certainly evolve, including those used as examples in the Coalition's Freight Data Metadata Matrix, which was developed on behalf of the

Coalition's Freight Committee as a resource but also as an area for discussion and collaboration. To that end, the Freight Data Matrix should be viewed by practitioners as a resource but also as a "work in progress" and with anticipation for feedback on and adjustment of and addition to its content. A longer-term goal of the Coalition is to create an on-line repository for the Freight Data Matrix where the Coalition and its members can continue to populate the matrix with sources and examples of data effectively used by practitioners.

Appendices

A: Current Data Sources in the Freight Data Matrix

- Airsage
- American Transportation Research Institute
- Border Crossing/Entry Data
- BTS TransBorder freight data
- Commodity Flow Survey (CFS)
- FAF4
- FAF4 Assignment Network
- FHWA Truck Parking-Jason's Law
- INRIX (5)
- Intermodal Freight Facility Database
- Statistics of U.S. Businesses (SUSB)
- Streetlight
- Streetlytics (Citilabs)
- Summarized Air Cargo Statistics
- Transearch
- Vehicle Inventory and Use Survey (VIUS)

B: Freight Data Fields Map

Part A. Metadata Elements	1.1.1 Summary Table	Database name
		Summarized
		What Are the Units Of Measure?
		Lowest Geography
		Strength
		Weakness
		Limitations on Data Use
		Website
	1.1.2 Attributes Related to Ownership and Use	Data Owner
		Free or Charge Is There A Fee To Access This Data?
		Is the Data Proprietary?
		Licensing Agreements
		Primary Audience for Which Data Was Developed
	1.1.3 Attributes Related to Data Collection and Coverage	Data Source: What Is the Source Of The Data?
		Sample/Universe
		What Is the Sample Size (%), If Known?
		Primary, Secondary, Modelled
		Collection Frequency
		Timeliness of Data
		Publication Frequency
	1.1.4 Attributes Concerning Summarized and Training Materials	Summarized Data Available
		Training Materials
		Data Development Reports
		Online Processing Tools
	1.1.11 Attributes Related to Usage	Key Tips For Analyzing
		Potential Uses of Data
		How Easy Is the Data to Use?
		How Easy Is the Data to Share?
		How Easy Is the Data to Analyze?
		How To Integrate W/ My Other Tools?
		For What Research Have I Used This Data?
Part B. Freight Categories	1.1.5 Attributes Concerning Time	Does the Database Contain Any Temporal Information Yes/No
		Time of Day Information
		Duration of Trip
		Parking Duration
		Time Element
		Forecasted Time Periods
	1.1.6 Attributes Concerning Truck Trips	Truck Volumes – Yes, No
		% Of Truck in Traffic Volumes
		Truck Classification
		By Axle
		Truck Weight
		Truck Registration Number
		What Modes Are Represented
		Relationship to Other Modes
		Name Of Carrier

	1.1.7 Attributes Concerning Commodity	does It Include Commodities (Yes/No)?
		Commodity Detail/Codes
		Shipper
		Costs-Rates
		Value of Cargo
		Weight of Cargo
		Ton Miles
		Special Cargo Concerns
		International Data
		OD Pattern
	1.1.8 Attributes Related to Safety	Safety Data (Yes/No)
		Inspections
		Regulatory
		Hours of Service
		Speed
		WIM
	1.1.9 Attributes Related to Location	Location Data (Yes/No)
		Route Used
		Permits
		Work Zone
		Parking
		Geography Detail
		Facility Information
	1.1.10 Attributes Related to Cost	Cost-Rates (Y/N)
		Costs-Rates
		Tolls
		Permitting
		Taxes