Demystifying Freight Data: Theory, Terms, and Practical Recommendations

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

Data is a topic often on the minds of agency staff as the world becomes ever more measured and data-rich. Evaluating and effectively using these new sources of data is a challenge for public sector planners. The growing emphasis on understanding freight activity, coupled with evolving software and data resources, can be overwhelming for busy staff. This White Paper, "Demystifying Freight Data" has been prepared by the Eastern Transportation Coalition, (formerly the I-95 Corridor Coalition), hereafter referenced as "the Coalition", on behalf of its Freight Committee. The White Paper seeks to provide some background on the theoretical context around data, in particular, to freight data, and to begin to provide practical recommendations for public agency transportation planners/practitioners. It also includes a discussion around data terms which are often confused.

The Coalition recognizes that this White Paper and related Coalition documents to this paper are intended for informational and instructional guidance and do not purport to answer all questions and guidance on understanding and applying freight data. The Coalition views this white paper as part of on-going Freight Committee activities including sharing of information, knowledge, and best practices as transportation planners navigate the complex world of freight data for planning and related purposes.

While a standalone document, this White Paper supplements the Coalition's additional document "The User Guide to The Eastern Transportation Coalition Freight Data Matrix" and the associated "Freight Data Matrix Metadata" spreadsheet.

Contents

List of Figures and Tables

Introduction

Freight is the movement of any cargo, i.e., non-passenger traffic (excluding public service vehicles) carried over a physical network. Cargo can move on publicly or privately funded infrastructure for the purposes of satisfying internal logistics needs or as a service to others. From the public sector perspective, freight movements interact or utilize publicly owned and maintained systems.[1] This requires public sector leaders to monitor and assess how their actions can influence freight activity (policies), improve the movement of freight through the system (programs), understand future needs (planning), and finally reach out to the community to ensure these actions are consistent with broader societal goals (promotion).[2] These 4P's of policy, programs, planning, and promotion can be used to organize research and understanding of freight activities.

There is a growing need among the public sector to understand freight operations. This need stems in part from concerns over system usage/congestion and resiliency, but largely from the recognition that goods movement depends upon well-maintained infrastructure and supports the nation's economy.

Planners require data and the necessary tools to analyze it to support effective decisions about freight movement. Learning to review and adopt data and analytical tools can provide a valuable contribution to discussions on infrastructure development, use, and maintenance.[3] With the emerging challenge on the nation's infrastructure, a suite of data, models, and guidance may be required to provide the necessary intelligence needed for sound public (and private sector where appropriate) leadership.

This white paper first provides background on important terms and concepts related to freight research before discussing challenges associated with that research. It concludes by providing information about evaluating and adopting freight data and internalizing a data-driven approach.

Background

The need for data and analysis arises from a question that must be answered, perhaps related to predesign for a roadway or a grant application. These needs come from diverse groups, some internally generated from other programs or leadership, while others can originate from outside the agency. Some requests may follow a scripted schedule outlined by statute (such as annual performance metrics or planning goals), while others may be generated in an ad hoc manner. Each may also require analytical support, resulting in possible purchases of data or consulting services. Finally, not all reports become "actionable," such as the inability of the agency to commit to implementing a project or changing priorities. One could argue that from a state DOT perspective, some elements will be different when compared to Federal oversights and requirements, or to working with MPOs or local governments.

Given the mix of diverse needs, geography, users, modes, etc., there remains no one way to frame the question concerning freight research data/analytical needs. Not only do these differences exist from an analytical need, but other differences also exist regarding the institutional requirements to examine freight

[1] NCFRP Report 8 – Freight Demand Modeling to Support Public Sector Decision Making
[2] Private sector needs will look similar, but will require more in person per marketing, industry meetings, etc., with different parameters regarding domain knowledge, price points, market knowledge, etc.
[3] The paper focuses on data and knowledge management, but the same framework could be applied in reviewing analytical tools or processes.

data. There are differences ranging from who requested the information, and why, as well as the degree of analytical support that an answer may require. Sometimes someone simply wants a bullet point, while another may want a full study. Finally, not every report has the same "shelf-life." Some studies are tied to program (funding) activities, while others serve as inputs to decision-makers. One way to frame freight research is by using the 4Ps discussed earlier; Table 1 provides more detail.

TABLE 1. THE 4 P'S OF FREIGHT RESEARCH NEEDS FOR PUBLIC SECTOR INFRASTRUCTURE PLANNING AGENCIES

Category of Research Program	*Subcategories*	*Requesting Agent*	*Frequency*	*Perception of Analytical Support*	*Agencies' Level of commitment at the end of a project*
Policy	BCA, white papers, economic research, etc.	Internal, State, Federal	Highway Authorization bill (major) State Legislative (minor), market studies	Data requirements vary, may need outside consulting support	N/A Varies (internal medium, legislative, high)
Planning	STIP, TIP, Corridor, LRP, Special Programs	Federal, State	FAST (5 years), other modes have different schedules, State Rail Plans	Maybe most data-intensive. Mostly requires consulting the support	Yes High
Programs	Operations, permitting, environmental, modal	State, Federal, Local	As needed (or, "ongoing")	Will require some new information, may require some consulting support	Yes High
Promotion	FAC, Roadshows, etc.	Internal (satisfy other programs), the responsible agency	In response to Policy, Planning Document, and Programs	Little new information created, use existing, often requires some consulting support	N/A Varies

One critique of the 4P's is that the categories do not address the geopolitical reality of freight planning, which often constrains financial/programming options because of jurisdictional authorizations (not least of which is the ongoing equity question between urban and rural areas). Because the public sector must plan for freight, the question of geography matters.[4] For example, the public sector planner for a state will have a different

[4] These groupings are not fixed. There exists some degree of overlap among these elements.

need to examine the higher functional class than an MPO planner or even a city planner. Each has guidance, programs, etc., that bias them towards understanding some segments more than other roadways. For example, the Georgia DOT needs to understand movement into, from and within the State, but does not necessarily need to know the traffic moving on neighborhood streets in Athens, Georgia. Municipalities are unable to improve a roadway in the neighboring county, even if that roadway is the county's primary access. Conversely, the same focus occurs in the private sector, as the shift from Mega tends to be more planning, to the local level, which becomes more operational.

TABLE 2. THE RELATIONSHIP BETWEEN INFRASTRUCTURE, ANALYTICAL NEEDS, AND FREIGHT MOVEMENT RESEARCH FOCUS

	Infrastructure	*Political Level of Analysis*	*Public Sector Analysis*	*Freight Movement Focus*
Mega	Interstate	National	Global Supply Chains, Ports, Intermodal	National, International, trade lanes,
Macro	Principal Arterial	State	Movement between Higher-level facilities (intermodal, unit trains, etc.)	Within State/Neighboring States, corridors
Meso	Minor Arterial	Region (MSA)	The movement to regional warehouses, transfer/collector facilities	Regional networks, routes, drayage
Mini	Collector (Corridor)	City	Business deliveries, collection, etc.	Routes, last mile
Micro	Local Road	District	ECommerce, consumer deliveries	Local, last mile,

While the 4P's can be addressed for any infrastructure provider, freight operates in many ways that ignore the 4P's that decision-makers must address. Freight is geographically neutral, as carriers tend to offer services that are constrained by networks, not political boundaries. Freight will use any mode that satisfies a shipper's service and cost concerns. Finally, freight can range from a push tug with 20 tows to a post office delivery vehicle. The size and scope of these shipments vary tremendously. And while there exist many public sector agencies engaged in the "freight industry", such as the military, this is often a private sector activity. To engage the private sector, one needs to understand some elements of the following dimensions: modal systems, commodities and shipper needs, and location determinations. For example, waterways operate in a much different regulatory and funding environment than railroads, but both compete for similar cargos. An inland terminal may promote a rail spur and may only be concerned with regional development if it adds additional competition to their existing or planned facilities.

From the private sector perspective, there are three M's that may be addressed (Table 3). The first is markets, which refers to what type of cargo is moving to and from what locations (expressed as demand for transportation services). The second is on which mode that cargo moved (infrastructure usage and capacity). And the final M is mobility (can the service be provided, what are the costs, and what bottlenecks exist).

Using that matrix, the key elements of the economic/analytical needs also change between the variables. The market question is often, what is moving, and will that movement continue, i.e., is my region gaining or losing traffic. If so, the next step is to identify the market share, shippers, etc., to determine what actions can be taken. These actions can be through changing policies (incentives) or planning for new capacity. Once the markets are identified, the next question tends to discuss, what segments is this traffic using? For example,

are trucks moving along Highway 72, and if so, how many trucks, what are their classifications, and any other questions to assist in planning for projects in the corridor. These studies, as planning documents, are necessary to support projects that are implemented by a public sector infrastructure provider. These tend to have longer-term project implementation schedules, so understanding future needs become important also. Finally, there are mobility questions. In this context, the cargo is moving on the network, but the focus is on how to make that movement more efficient, given the existing physical constraints of the network. In all three "Ms", the risk to innovative technologies exists, disrupting several aspects of the freight system, and the resulting planning and programming for system activity. Finally, there is often an unspoken reality, namely freight industry locations exist because of other decisions (land use, incentives, etc.) and over time, these facilities get "locked-in" to a location, unable to bear the switching costs associated with moving to a different location, keeping freight moving through regions that may be deemed to be undesirable when compared to their neighbors.

TABLE 3. THREE M'S OF FREIGHT ATTRIBUTES

Attribute	Key Element	Economic Consideration	Analytical Challenge	Subcategories	Public Sector Response	Innovative Technologies/Emerging Themes
Markets	What is moving?	What is the economic contribution?	Market information, especially at a shipper level, fragmented	Mode (Water, Rail, Airports, Trucking), Industry (agricultural, automotive, etc.)	Determine traffic, estimate economic activity	3-D Printing, Trade Policy, Sourcing
Modes	What network did the shipment use, valuation of roadway usage	Long Term, structural costs	Modal capacity available, but traffic on modes, segments not well defined (key exceptions were governed data is collected, i.e., tariffs)	Economic Development, Site Certification, land use	Investment, grant applications,	Tolling, Bridge Performance Monitoring, Usage fees
Mobility	What influences its movement? And what are the outcomes from mobility? Time of movement?	Short term valuations	Growth of performance metrics may ignore structural land use/economic decisions, but the growing role of resiliency, mobility treats modes and markets as fixed	Supply Chains, Drayage, Truck Parking Resiliency, Regulation	Operations, mode shift, permitting, ITS deployment, Performance Metrics	Autonomous/Connected Vehicles, ITS

Defining a single "freight" research area for public sector agencies becomes problematic without identifying some cross-sectional structure between one of the 4P's and one of the 3M's. Looking at the six attributes in Table 1, the connection to the 3M's becomes problematic. For example, costs or other safety regulatory actions can determine Mobility, i.e., people select services based on availability. These services have the authorization to operate based on some regulatory decision. The Modes are the actual networks and units engaged in the operational movement of that cargo, so knowing information on time and location provides insights into system usage, and by extension, planning and operations research. The last category, Markets, describes the "why" that the movement occurred. But even among the 3M's, the six attributes must be considered in any study.

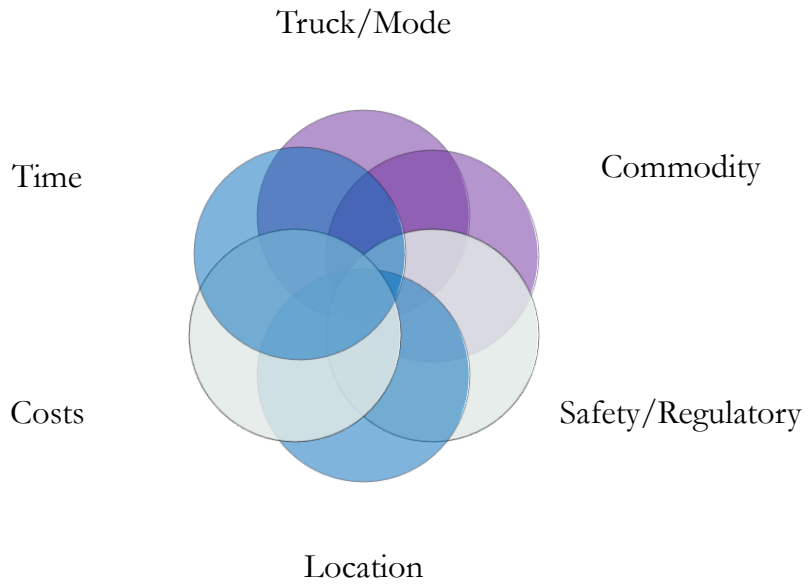


FIGURE 1. SIX ATTRIBUTES OF DATA MATRIX COMPARED TO 3M'S (MARKETS, MOBILITY, MODE)

With this background information in mind, the important question remains how public sector entities evaluate what data options are available to them for examining truck movements. Part of that effort would be to prioritize the need and the benefit of that information for other users. One must evaluate freight data to satisfy program and "as needed" research, but the quality of the data management program depends upon the effort associated with understanding data, how to institutionalize an internal knowledge-based management process, and to evaluate new databases. The following sections will discuss these topics.

Data

To satisfy transportation planning requirements, practitioners must have data and the associated analytical tools to make sense of that data. This section seeks to lay out some terms concerning data and knowledge work.

The term "data" complicates the discussion. At one level, data can be defined as the smallest unit that can be counted, reported, experienced, or observed, though the term is often inflated to include not only preprocessed raw data but also modeled or processed information. As data consists of discrete individual

units, without any organization/ categorization, data alone remains useless for decision-making. The process of categorizing the data creates information, summarized in a manner to understand it. Metadata, Category, and Attribute all fall into the gray space between data and information, forming the bridges between the two elements. Once the data are organized, the information is combined with other insights, information, etc., to create the necessary reports required to assist a decision-maker. The decision-maker could use that intelligence in making planning decisions or other actions.

For the practitioner, each step becomes revealed by specific actions. For data, this may mean purchasing, arranging storage, data wrangling, reviewing, and cleaning datasets, in addition to simply storing the data where it can be accessed in a timely manner. Once the data is ready to be converted into a database structure, one can do so through various software queries or some other process to prepare the data for providing intelligence.[5] Intelligence requires the ability to access and review the data for its intended purpose to provide insights for making decisions. Figure 2 shows the progression from data to wisdom.

Key Terms: Metadata, Category, Attribute

These terms refer to the data elements themselves and provide a format for discussing the data elements in an objective manner.

Metadata: information about the data source, describing its attributes, categories, its origin, use and application.

Category: A structural approach to group data, describing a common attribute between two elements. For example, the blue car follows into the category of color.

Attribute: The actual characteristic of an element, such as a blue car, blue is the attribute.

The following picture shows the difference

Metadata-Vehicle

Category- Truck, Color, Location

Attribute- Class 5 Truck, White, Parked in St. Tammany Parish

The movement from left to right also results in a shifting skillset. The person who functions in the data category will focus on data cleaning, data integrity, and data maintenance. That person may not have extensive knowledge of the elements in the data as an analysis, preferring instead to examine the quality of the data, not its knowledge content. The person who accesses the data to create information may need to understand what is or is not in a database (See Appendix "Freight Data Architecture: Comparing CFS to the FAF ").

In some ways, there is a tradeoff, as the data domain expert may not possess planning knowledge. This tends to lead to challenges concerning managing research programs, as the specialized domain knowledge worker's manager may be unable to evaluate either the work or the required efforts to sustain a data program.

[5] A database is a series of related files that contain data elements but has been indexed, categorized, cleaned, to provide a source of processing the data in an efficient manner.

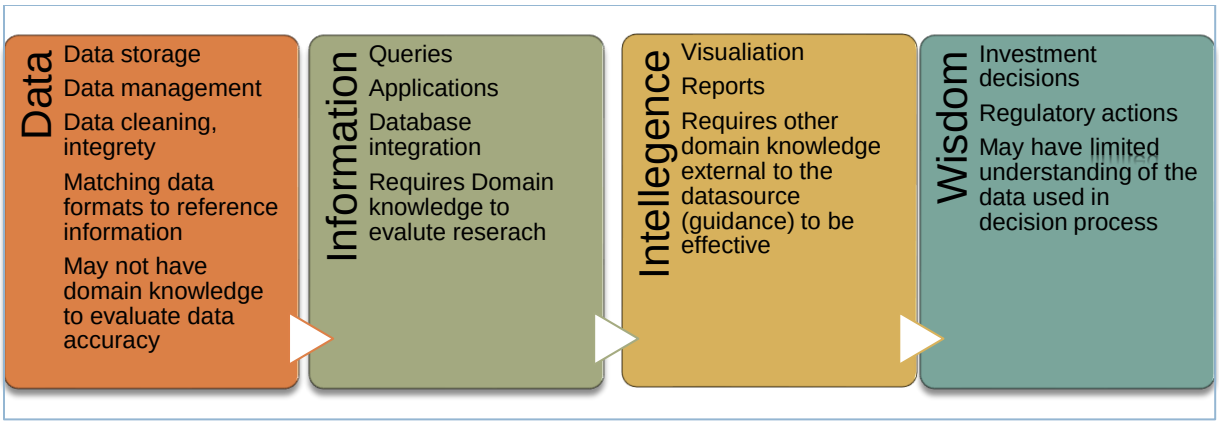


FIGURE 2. THE EVOLUTION FROM DISCRETE DATA TO WISDOM.

Using the progression framework from Figure 2 for the example of truck counts on I-95 (Figure 3), counting the trucks provides data, but this data has limited meaning. The organization may need to know how many trucks are operating on I-95 in Virginia daily. With this information request, the data must be organized into some structure to process the query. That organization will include information specific to the request, namely data categorized by time and geography. In processing that data into information, the format must also be considered. Is a chart, table, or simply a single number required?[6]

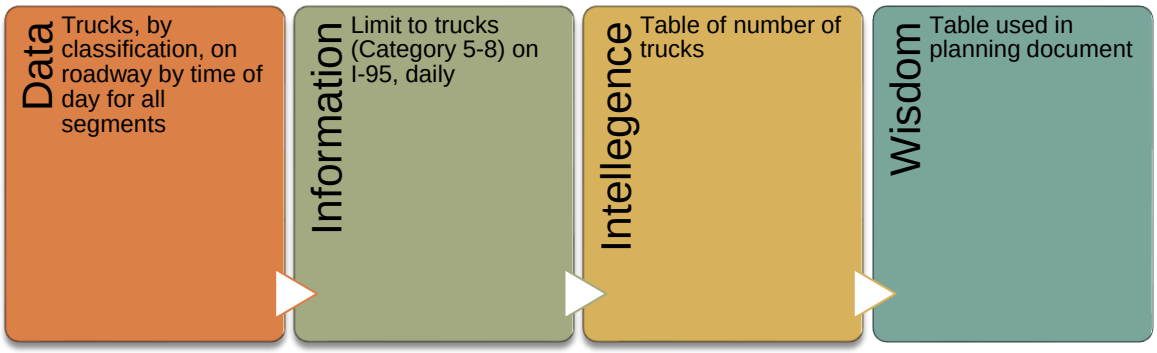


FIGURE 3. EXAMPLE OF TRUCK DATA BECOMING WISDOM

Truck count data in the example could be obtained from a variety of sources, including:

- Virginia Department of Transportation Traffic Website[7],
- FHWA Map of Highways[8], or
- FHWA Virginia Highway System Performance Monitoring System files[9].

[6] The data matrix scorecard could be helpful for organizing the research framework
[7] http://www.virginiadot.org/info/2018_traffic_data.asp
[8] https://ops.fhwa.dot.gov/freight/freight_analysis/nat_freight_stats/nhsmajortrkrts2040.htm
[9] https://www.fhwa.dot.gov/policyinformation/hpms/shapefiles.cfm

Each will produce data that requires additional processing activity and time. Knowing that similar intelligence can derive from many sources, complicates the research while allowing more flexibility in answering questions.

Transformation of data from discrete units into actionable intelligence can be enhanced through structured collaborative activities. As mentioned in the I-95 Corridor Coalition report "Member States' FAST Act-Compliant Freight Planning Activities, Issues and Recommendations, June 2018", every state must complete many of the same research programs concerning freight shipments. Finding a forum to engage—such as through webinars, peer exchanges, or some other mechanism—that focuses on topical research areas could provide additional insights. The group could outline how others have approached the research question, including discussing the analytical processes and data resources used.

Evaluating New Databases

The world is becoming more instrumented, measured, and "data-rich." This has resulted in the development of new databases that were not available ten years ago. The push for performance measures has led to a desire to capture more information to assist decision-makers. When examining a new database, the researcher should also question how the new data supplement or replace existing data and should first assess existing data to determine whether new data is indeed necessary.

The development of a robust framework to advance and maintain a knowledge database program leads to the question of adding new data. The question of new data will come from many diverse sources, such as expanding domain knowledge, outside vendor calls/consulting studies, or from other internal users. Each will expect the researcher to be able to approve, utilize, or purchase their database! (Having them fill out the Data Matrix maybe a good first step – *see user guide*.)

The following process provides a good summary of the elements one must consider.[10] The researcher must determine what is needed, why, and how much to commit to the acquisition of that resource. Once that is identified, the process of determining needs through some structured approach begins.

[10] C20 Product Implementation for Technical Freight SME Support Services—Phase II Freight Data Guide for Improved Transportation Planning

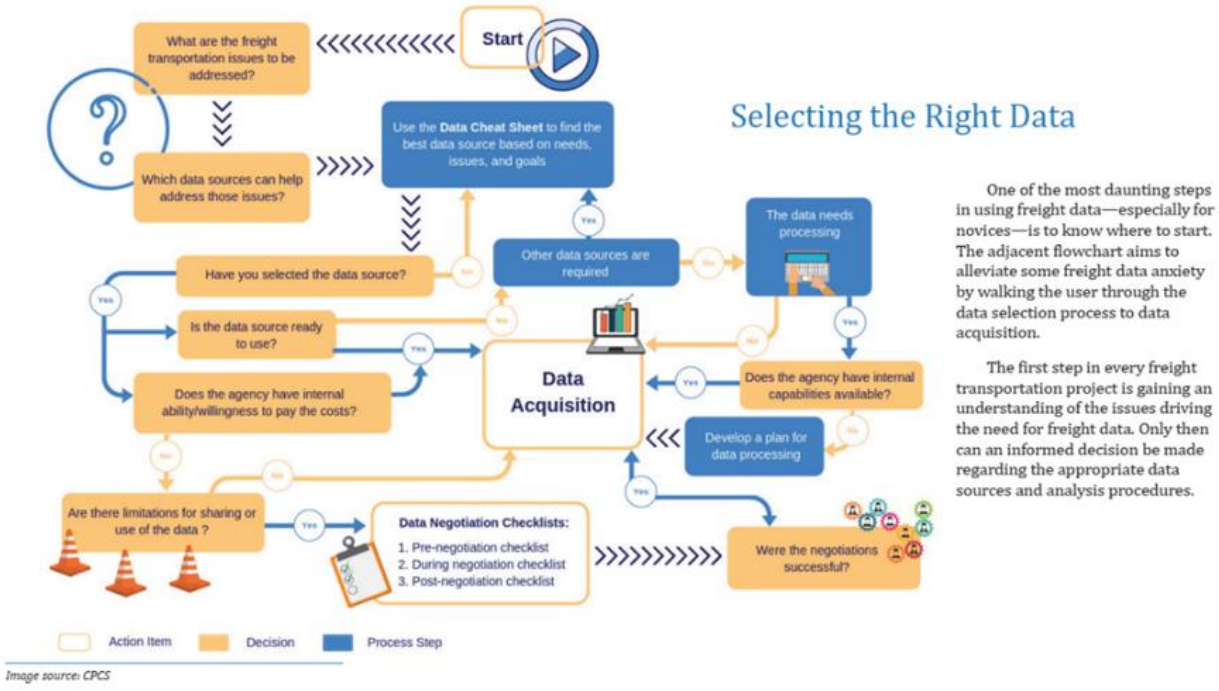


FIGURE 4. SELECTING THE RIGHT DATA

However, once the data purchase acquisition decision is made, there remain other decisions concerning the data resource. This is an area most people fail to evaluate, often substituting short-term research goals that are not coordinated properly with other requests. An alternative evaluation method is to use the checklist in Table 4.

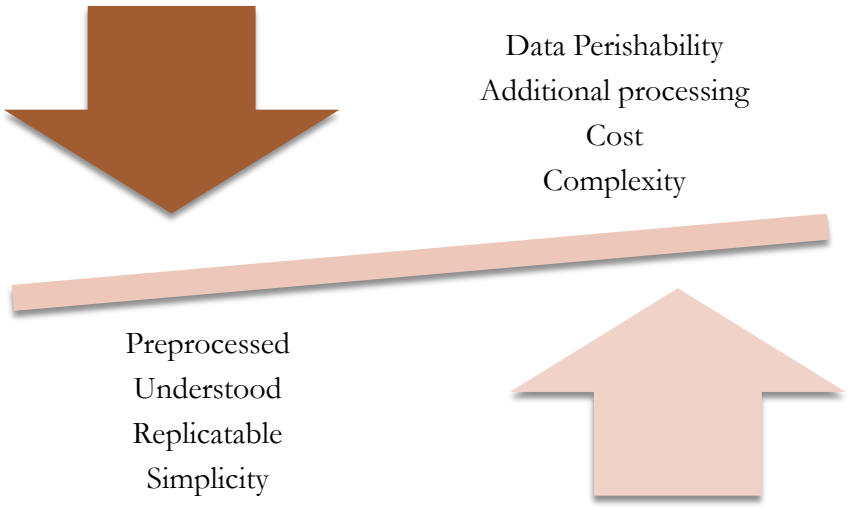


FIGURE 5. TRADEOFFS WHEN ADOPTING NEW DATA RESOURCES

TABLE 4. CHECKLIST OF EVALUATING NEW DATA RESOURCE ACQUISITIONS

Attributes	*Checklist*	*Yes*	*No*
Institutional Direction	Did our mission change regarding this topic?		
	Does this new data provide insights that are aligned with our new mission?		
	Does this new data align with changing guidance?		
Meta Data Comparison	Is data captured closer to the event (timelier)?		
	Does the data supplement existing data?		
	Does this data replace existing data?		
	Does this data fulfill an existing analytical gap?		
	Do data processing and acquisition costs change?		
	Does the collection process for this data provide better data than current programs?		
Staff/ Resource Costs	Are there any training and software acquisition costs?		
	Is this a one-time purchase?		
	Is there an internal, external, learning curve		
	Does this adoption change my data or contract budget		
	Are there any costs for maintenance and renewals?		

Developing Data Capacity

A disclaimer: The following assumes these are internally generated questions. While the same approach could be used for evaluating service consulting requests, there exist other program elements one would add beyond these questions.

Organizations must also encourage a commitment to developing staff capacity concerning freight domain knowledge and skill acquisition. This section discusses some ideas concerning how staff can lead that internal effort.

The response to a need for insight from data may fall into one or more of four broad categories: Institutional, Skills, Costs, and Review. The **Institutional** category links the need to the organization's goals and values. The **Skills** category relates to the ability to provide the answer, while **Costs** outline what (if any) additional resources may be needed. Finally, the last category is **Review**, i.e., what can be done better or differently in the work process. Rearranging the four categories results in RISC, an appropriate reminder of the possible consequences of misinformed research.

Institutional: The objective is to provide timely intelligence to support the organization's mission. In many ways, knowing the right answer, but the wrong question does not help anyone, and practitioners must guard against biases. To accomplish this, practitioners can ask the following questions:

- Who needs the information?
- Who asked the question?
- When do they expect an answer?
- What are their expected outcomes?
- Will this require an internal review, and if so, who would do that work?
- Will this intelligence be used internally or externally?
- Who will review this work?
- How important is this request compared to others?
- In what format is the report needed (chart, text, etc.)?

- Is the protocol prescribed by law or guidance?
- Will this work require a presentation/training when completed?
- What level of confidence is needed?

Skills: Without an honest assessment of skills and abilities, the practitioner may struggle to produce the required answer. Some questions here may include:

- Do I have the time to spend on this analysis?
- Do I have access to the data to complete the task?
- Do I have the software/skills to complete the task?
- Is this like previous questions I (or others) have answered in the past?
- Can someone else answer this question better than me?
- Do I have the domain knowledge to understand the topic?
- Do I need a collaborator?
- Do I need some training to answer this question?

Costs: Sometimes there are costs associated with analysis. Not all data is accessible in the format one needs and, in some cases, data must be purchased. The practitioner must understand the resource costs, but these may matter little to the person who is requesting the information.

- Do I need to purchase data/information services?
- Do I need to get a license or right to access the data?
- Do I need to purchase software or hardware?
- Do I need to hire a consultant because I do not have the skills, the time or bandwidth to complete this project?
- Can I legally share this data, or does it have to be summarized, etc.?
- Do I need to pay for training to respond to this request?

Review: After the work is delivered, sometimes it is helpful to review with the requester to understand how well the information met their needs. These questions may include:

- Will I be asked similar questions in the future?
- Do you want this information for yourself/or are others to access this information directly without asking me?
- Do they need the training to access the data themselves?
- Do you or I need more domain knowledge?
- Did the information satisfy our organization's needs?

Conclusion

The challenge concerning new data is not any one element, but the fact that a holistic approach is not taken. With an ever-increasing amount of valuable data available, regional collaboration efforts, such as The Eastern Transportation Coalition (formerly I-95 Corridor Coalition), could play a role in being the domain expert in certain areas to assist member states. The following slide (Figure 6) from "NCHRP: Active Implementation: Moving Research into Practice" remains highlights the increasing demands for the adoption of data and the need for a robust approach in the transportation community.[11]

[11] http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP_ActiveImplementation.pdf

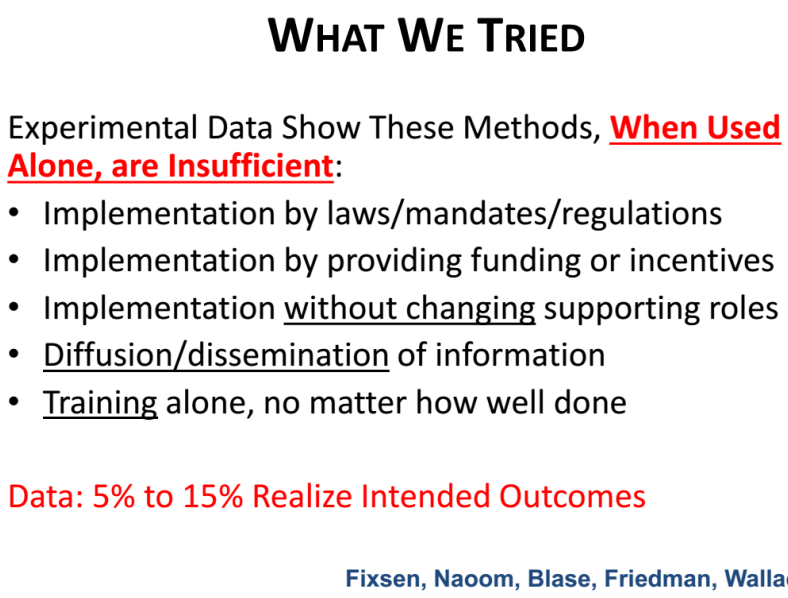


FIGURE 6. ACTIVE IMPLEMENTATION OF RESEARCH

No database, despite its cost, is beneficial for planning if it is not used effectively. Often users do not develop the skills to utilize a database due to many factors, such as training, lack of clear direction as to how to utilize the database, and the linkage to program goals. Furthermore, in today's discussion on "Big Data," people often assume that all intelligence is two clicks away and free. This simplistic view has led to both a deviation in the importance of solid data analytical development as well as an overstatement in the desirability of applications, which often requires its own training for the tools to be utilized effectively. This also tends to undervalue the need to develop staff and domain expertise. This document sought to provide both theoretical context and practical recommendations for addressing these issues. The Data Matrix and User Guide should be consulted for more information.

Appendix

Relevant Bibliography of Freight Data Studies

Freight data has been a topic for many years, as highlighted by the "Data Needs in the Changing World of Logistics and Freight Transportation" conference held in 2001.[12] During that discussion, there were two sessions of interest, namely, "Critical Issues Facing Freight Data Collection and Analysis," and "The Future for Freight Transportation Data Collection and Analysis." In many ways, these two topics have remained part of the discussion on freight. There have been many studies over the past twenty years calling for ways to examine freight data. For example, in 2003, TRB, Special Report 276 was released, focusing on a national freight data program.[13] The second finding, on page 6, listed that "The current disjointed patchwork of freight data sources is costly to generate and maintain, but does not provide decision-makers with the data they require." NCFRP Report 9, Guidance for Developing a Freight Transportation Data Architecture only further highlighted the inability to describe freight data in regard to supporting the needs of state and local planners. Even as planners recognize that this growing emphasis on freight requires decision-makers to understand the changing relationship of freight, we remain unable to fully adopt these evolving technologies and reshape freight data programs.

However, many groups are seeking to address these concerns, including various freight studies being undertaken by FHWA and The Eastern Transportation Coalition regarding freight fluidity. Other resources include the Commodity Flow Survey User Conferences organized by TRB and the TRB Innovations in Freight Data Conferences. There are many freight-related reports published under the NCFRP program that have a wealth of useful resources.[14]

- FHWA Freight and Land Use Handbook, https://ops.fhwa.dot.gov/publications/fhwahop12006/index.htm
- FHWA Statewide Freight Plan Template https://permanent.access.gpo.gov/gpo21292/sfp_template.pdf
- Federal Register Notification on State Freight Planning, Vol. 81, No. 199 / Friday, October 14, 2016 https://www.gpo.gov/fdsys/pkg/FR-2016-10-14/pdf/2016-24862.pdf
- U.S. Maritime Administration, https://www.maritime.dot.gov/sites/marad.dot.gov/files/docs/outreach/data-statistics/6841/guidemarketresearchmarinetransportationsvcs.pdf
- U.S. Army Corps of Engineers, Planning Guidance Notebook, http://www.publications.usace.army.mil/Portals/76/Publications/EngineerRegulations/ER_1105-2-100.pdf
- U.S. Army Corps of Engineers, Planning Toolbox, https://planning.erdc.dren.mil/toolbox/index.cfm
- National Cooperative Freight Research Program reports:
  22: Freight Data Cost Elements
  25: Freight Data Sharing Guidebook
  26: Guidebook for Developing Subnational Commodity Flow Data
  29: Making Trucks Count: Innovative Strategies for Obtaining Comprehensive Truck Activity Data
  35: Implementing the Freight Transportation Data Architecture: Data Element Dictionary
  37: Using Commodity Flow Survey Microdata and Other Establishment Data to Estimate the Generation of Freight, Freight Trips, and Service Trips: Guidebook

[12] https://www.dot.ny.gov/divisions/policy-and-strategy/darb/dai-unit/ttss/data-changing-world
[13] National Research Council (U.S.) Committee on Freight Transportation Data: A Framework for Development, Transportation Research Board Special Report 276, 2003.
[14] http://www.trb.org/NCFRP/NCFRP.aspx

"Freight Data Cheat Sheet" and Developing A Repository

Given the need to understand freight planning, here is a cheat sheet one can use to evaluate freight research or could be used to develop additional research ideas. The structure could also be used to review existing work tasks to identify how these items are being addressed, and by what database/information sources the research utilized.[15]

TABLE 5. TEMPLATE FOR ATTRIBUTES AND TOPIC AREAS

Research Topic	Attributes						Other Domains	Reference Document, source
	Time	Truck	Commodity	Safety	Location	Costs		
Resilience								
Mode Choice								
Supply chains								
Last-mile								
Congestion pricing – tolls								
Barges-traffic								
Bottlenecks								
Economic Impacts/BCA								
Safety								
Emissions								
Autonomous Trucks								

[15] This concept is discussed in the User Guide to the Data Matrix.

Freight Data Architecture: Comparing CFS to the FAF

When one can understand the details of a database, it becomes easier to understand both what it contains or where it may be useful for a certain application. For example, the Freight Analysis Framework depends upon the Commodity Flow Survey as its primary dataset, but no state reported using the Commodity Flow Survey (CFS) in their state freight plan when surveyed by the Coalition. Here is an example of how not understanding a database's sources can lead to not leveraging additional information if required.[16]

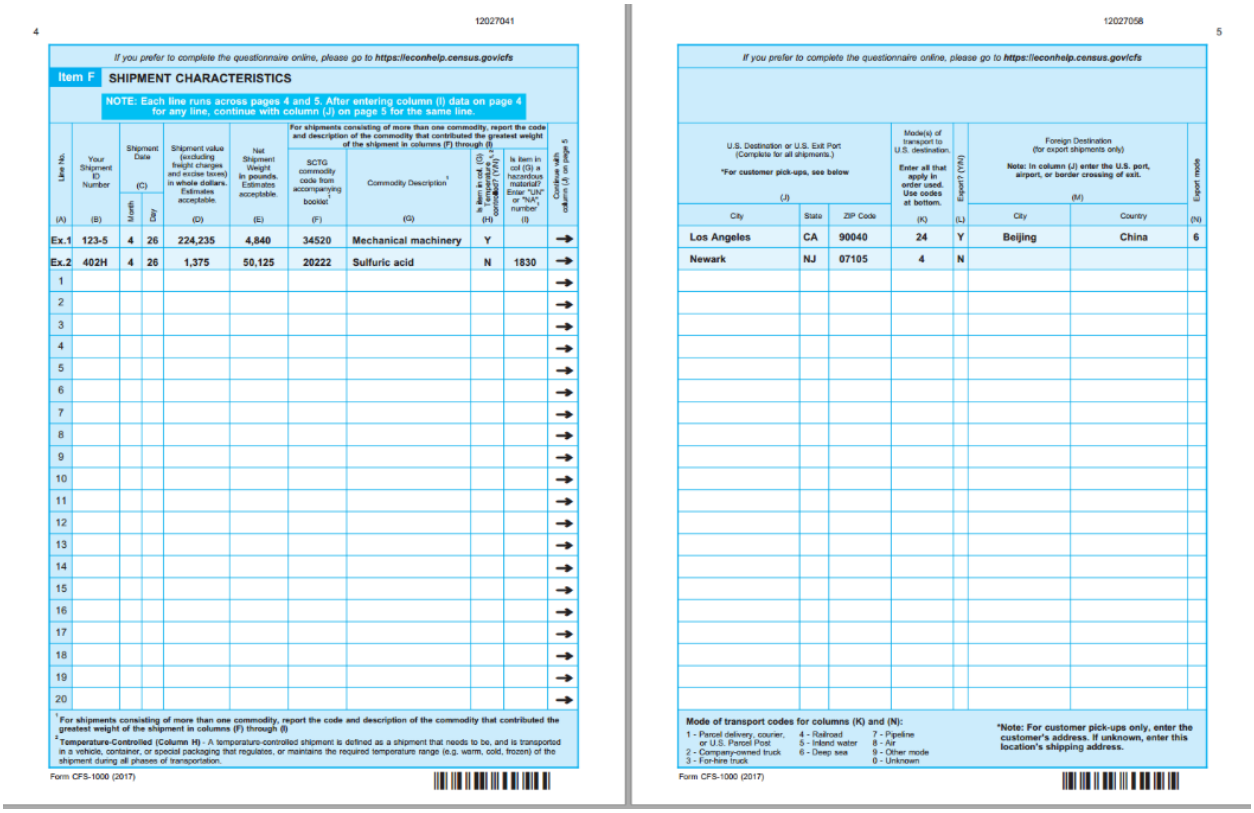
The Bureau of Transportation Statistics: The Commodity Flow Survey is conducted throughout the survey year with establishments selected receiving four questionnaires, one during each calendar quarter of the survey year. The establishments are asked to provide shipment information about a sample of their individual outbound shipments during a prespecified one-week period of the calendar quarter. Generally, the four questionnaires are the same.

The following information is collected:

- Shipment ID number
- Shipment date (mm/dd)
- Shipment value
- Shipment weight in pounds
- Commodity code from Standard Classification of Transported Goods (SCTG) list
- Commodity description
- Hazmat number [United Nations (UN) or North American (NA) number]
- U.S. destination (city, state, zip code) - a gateway for export shipment
- Modes of transportation
- Foreign destination (exports only - city and country)
- Export mode
- Temperature control

The CFS focuses on outbound shipments using the following forms, which are posted online.[17] The forms, of which a few sample pages are copied here, show the details regarding the information that is released for the public files. Understanding this structure, especially from the perspective of the person processing the form, may help the analyst understand the nuances of the submission and the database. The report has information on the shipper, their location, as well as the sample rate for their shipment size. Finally, the CFS has many useful data elements that are not captured in other databases, such as information on temperature control or Hazmat codes. When we examine the forms, we see the detailed structure available to survey respondents.

[16] This same approach could be used to compare all the databases that are integrated into the FAF.
[17] The form is posted at https://www.census.gov/programs-surveys/cfs/technical-documentation/questionnaires.html



If you prefer to complete the questionnaire online, please go to https://econhelp.census.gov/cfs

Item F SHIPMENT CHARACTERISTICS

NOTE: Each line runs across pages 4 and 5. After entering column (I) data on page 4 for any line, continue with column (J) on page 5 for the same line.

Line No. (A)	Your Shipment ID Number (B)	Shipment Date (C) Month	Shipment Date (C) Day	Shipment value (excluding freight charges and excise taxes) in whole dollars. Estimates acceptable. (D)	Net Shipment Weight in pounds. Estimates acceptable. (E)	For shipments consisting of more than one commodity, report the code and description of the commodity that contributed the greatest weight of the shipment in columns (F) through (I). SCTG commodity code from accompanying booklet[1] (F)	Commodity Description[1] (G)	Is item in col. (G) temperature controlled (TCNS)?[2] (H)	Is item in col (G) a hazardous material? Enter "UN" or "NA" number (I)	Continue with column (J) on page 5
Ex.1	123-5	4	26	224,235	4,840	34520	Mechanical machinery	Y		→
Ex.2	402H	4	26	1,375	50,125	20222	Sulfuric acid	N	1830	→
1										→
2										→
3										→
4										→
5										→
6										→
7										→
8										→
9										→
10										→
11										→
12										→
13										→
14										→
15										→
16										→
17										→
18										→
19										→
20										→

[1] For shipments consisting of more than one commodity, report the code and description of the commodity that contributed the greatest weight of the shipment in columns (F) through (I)

[2] Temperature-Controlled (Column H) - A temperature-controlled shipment is defined as a shipment that needs to be, and is transported in a vehicle, container, or special packaging that regulates, or maintains the required temperature range (e.g. warm, cold, frozen) of the shipment during all phases of transportation.

Form CFS-1000 (2017)

12027041 4

12027058 5

If you prefer to complete the questionnaire online, please go to https://econhelp.census.gov/cfs

U.S. Destination or U.S. Exit Port (Complete for all shipments.) *For customer pick-ups, see below (J) City	State	ZIP Code	Mode(s) of transport to U.S. destination. Enter all that apply in order used. Use codes at bottom. (K)	Export (Y/N) (L)	Foreign Destination (for export shipments only) Note: In column (J) enter the U.S. port, airport, or border crossing of exit. (M) City	Country	Export mode (N)
Los Angeles	CA	90040	24	Y	Beijing	China	6
Newark	NJ	07105	4	N			

Mode of transport codes for columns (K) and (N):
1 - Parcel delivery, courier, or U.S. Parcel Post
2 - Company-owned truck
3 - For-hire truck
4 - Railroad
5 - Inland water
6 - Deep sea
7 - Pipeline
8 - Air
9 - Other mode
0 - Unknown

*Note: For customer pick-ups only, enter the customer's address. If unknown, enter this location's shipping address.

Form CFS-1000 (2017)

FIGURE 7. COMMODITY FLOW SURVEY FORM

The information from the CFS is merged with other data resources to create the Freight Analysis Framework (FAF). As such, elements are lost as they are not easily joined, or appropriate for databases focusing on the national level origin and destination flows. This is evidenced in the FAF file structure:

- Foreign Origin
- Domestic Origin
- Domestic Destination
- Foreign Destination
- Foreign Inbound Mode
- Domestic Inbound Mode
- Freight Outbound Mode
- Commodity Code (SCTG)
- Trade Type (Import, Export, Domestic)
- Value
- Tons
- Ton-Miles
- Year.

The FAF possesses a different file structure, as many variables reported in the CFS were not included. Also, the FAF does not have a direct link to the CFS microdata set. So, in this regard, knowing not only the FAF and the CFS could provide more detail for some applications.[18]

[18] For other major data sources see FAF4 documentation.