TRANSPORTATION RESEARCH
CIRCULAR
Number E-C269
January 2021

2020 Commodity Flow Survey Workshop

September 24, 2020
TRB Virtual Event

The National Academies of
SCIENCES · ENGINEERING · MEDICINE
TRB
TRANSPORTATION RESEARCH BOARD

TRANSPORTATION RESEARCH CIRCULAR E-C269

2020 Commodity Flow Survey Workshop

September 24, 2020
TRB Virtual Event

Sponsored by
Bureau of Transportation Statistics

Organized by
Standing Committee on Freight Transportation Data

Rapporteur
Sarah Hernandez
University of Arkansas

Transportation Research Board
500 Fifth Street, NW
Washington, D.C.
www.trb.org

TRANSPORTATION RESEARCH CIRCULAR E-C269
ISSN 0097-8515

The **Transportation Research Board** is one of seven major programs of the National Academies of Sciences, Engineering, and Medicine. The mission of the Transportation Research Board is to provide leadership in transportation improvements and innovation through trusted, timely, impartial, and evidence-based information exchange, research, and advice regarding all modes of transportation

The **Transportation Research Board** is distributing this E-Circular to make the information contained herein available for use by individual practitioners in state and local transportation agencies, researchers in academic institutions, and other members of the transportation research community. The information in this E-Circular was taken directly from the submission of the authors. This document is not a report of the National Academies of Sciences, Engineering, and Medicine.

Transportation Research Board
500 Fifth Street, NW
Washington, D.C.
www.trb.org

Preface

The Commodity Flow Survey (CFS) Workshop represents the joint efforts of the Transportation Research Board (TRB), the U.S. Department of Transportation (DOT) Bureau of Transportation Statistics (BTS) and the TRB Standing Committee on Freight Transportation Data. The 2020 CFS Workshop was timed to coincide with the release of the 2017 CFS data tables and the planning of the 2022 CFS. This was the first all-virtual CFS Workshop to be hosted by TRB. In contrast to the previous CFS Workshops, this virtual workshop brought together federal, state, and local level planners and researchers, academics, and practitioners from public and private sectors to discuss efficient ways to align the CFS with the needs of data users and producers. The goal of the workshop was to provide a platform for discussion of critical needs and uses to improve future iterations of the CFS. Participants met over a half-day virtual meeting to work together to identify how the CFS can be used to answer freight data questions, share information and knowledge, and highlight emerging issues surrounding freight movement.

Chaired by Katherine Lawson of the University at Albany, State University of New York, the planning committee represented CFS data producers, data users, analysts and modelers. The ad hoc committee developed the conference agenda and pre-conference surveys. The workshop program is available in the appendix. The breakdown of the 139 attendees was as follows:

Affiliation Group	Number of Attendees
U.S. Census Bureau	38
Consultant/Private Sector	31
Universities	19
State governments	16
U.S. Department of Transportation	11
Local and regional governments	3
Other	21

The virtual workshop provided a platform for open discussion and synthesis of the challenges, uses, and needs for CFS data products. Among the many topics discussed, the workshop participants and organizers considered the benefits and trade-offs of freight flow attributes including the types and grouping of commodities, level of geography, and categories of modes that are included in the public release of the CFS and its derivatives, e.g., the Freight Analysis Framework (FAF). FAF is a commodity-based freight flow forecasting model produced through a partnership between BTS and Federal Highway Administration (FHWA). FAF uses data from the CFS, among other sources, to forecast freight tonnage and value by region, commodity, and mode. Workshop attendees shared opportunities and challenges of current uses, applications, and approaches related to CFS and FAF. Data producers shared advances in processing methods and data collection techniques.

This e-circular contains summaries of the three main conference sessions including the three breakout groups hosted to enable discussion among attendees. No language should be construed as consensus findings or recommendations on the part of the conference; the planning committee; the rapporteur; the National Academies of Science, Engineering, and Medicine; or the Bureau of Transportation Statistics.

The planning committee thanks the TRB staff, session moderators and recorders, representatives from the Bureau of Transportation Statistics and Census, and attendees for their contributions to the success of the workshop and its many planning iterations.

DISCLAIMER

This e-circular has been prepared by the workshop rapporteur as a factual summary of what occurred at the workshop. The planning committee's role was limited to planning and convening the workshop. The views contained in the proceedings are those of individual workshop participants and do not necessarily represent the views of all workshop participants, the planning committee, TRB, or the National Academies of Sciences, Engineering, and Medicine. *This document went through an internal TRB peer review process.*

Contents

Introduction

CATHERINE LAWSON
University at Albany, State University of New York

Conducted as a partnership between the Bureau of Transportation Statistics (BTS) and the U.S. Census Bureau (Census), the Commodity Flow Survey (CFS) provides key data for many state and federal freight planning activities. The CFS is a shipper survey used to gather freight flow information including commodity shipped, origin and destination, value and weight, and mode. The survey is collected from more than 100,000 establishments including industries of mining, manufacturing, wholesale trade, warehouse and distribution centers, and select retail and service trade industries. The 2017 CFS is the sixth survey since the program started in 1993 and the next CFS will take place in 2022.

Results from the 2017 CFS are provided publicly as data tables, and new to the 2017 CFS, through an online visualization dashboard. Like preceding CFS releases, the 2017 CFS also includes the Public Use Microdata Sample (PUM), renamed the Public Use File (PUF). The PUM/PUF is an experimental data product of the CFS data release and differs from the published estimates. It contains 20 variables for all usable shipment records collected by the CFS including a shipment tabulation weighting factor that can be used to expand the PUM/PUF to represent the total population of in-scope U.S. shipments. Due to noise-infusion to maintain confidentiality in the PUM/PUF data file, estimates derived from the PUM/PUF may not necessarily match the CFS estimates.

A brief review of the 2017 CFS results shows that 81% of freight shipments (93% by tonnage) move by a single mode over an average of 243 miles, primarily by truck. Looking at finer resolutions of the 2017 CFS shows that for-hire truck shipments traveled 369 miles on average representing 42% of shipments by tonnage and 39% by ton-miles, while company-owned trucks traveled 45 miles on average with 29% by tonnage and 5% by ton-miles. Comparing the trends established from the 2012 CFS, the number of shipments by all modes increased by 5%, tonnage increased by 10%, ton-miles increased by 5%, and average miles per shipment increased by 8%. Again, drilling into finer details, comparing 2012 to 2017 CFS results shows that for-hire trucks average miles per shipment decreased by 27% while the average miles per shipment for company-owned trucks decreased by 22%. Between 2012 and 2017, average miles shipped by rail decreased by 28%, and shipments moved by water decreased by 71%. Multimodal modes, parcel, and courier shipments increased 25% by value, 33% by tonnage, and 3% by average miles per shipment. Looking at temporal trends in distance shipped, the number of shipments traveling less than 50 miles increased by 27%, representing an increase of 34% in ton-miles while the number of shipments traveling 50 to 99 miles increased by 19%, representing an increase of 21% in ton-miles.

The CFS data provides a comprehensive picture of in-scope national multimodal freight flows and is the only publicly available source of this data for highway freight. CFS results are used to analyze freight trends, commodity and vehicle flows, and for forecasting freight demands in order to guide multimodal freight infrastructure planning, operational management, and policy decisions. The CFS serves as input to the national freight forecast policy tool, the Freight Analysis Framework (FAF) which is used by planners at the federal, state, and regional levels. FAF is produced as a partnership between BTS and the Federal Highway Administration

(FHWA) and includes estimates of freight movements not covered in the CFS. In this way, FAF provides a comprehensive depiction of freight movement across the U.S., including all modes of transportation.

The 2020 CFS Workshop represents a continuing history of convening data producers and users to identify benefits and challenges of data collection, uses, and needs. The 2020 CFS Workshop is the fourth such workshop, with the first hosted in 2005 and subsequent workshops continuing in 2010 and 2015. Each workshop aims to capture new data collection technologies, analytical methods, and applications of CFS data and its related products like the FAF. The goal of the 2020 CFS Workshop was to spur discussion to improve upon the 2017 survey so that it meets the needs of data producers and users from industry, the public sector, and academia. The 2020 CFS Workshop welcomed 139 attendees from the following areas:

Affiliation Group	**Number of Attendees**
U.S. Census Bureau	38
Consultant/Private Sector	31
Universities	19
State governments	16
U.S. Department of Transportation	11
Local and regional governments	3
Other	21

The 2020 CFS Workshop was hosted as a virtual event sponsored by the Transportation Research Board (TRB) and BTS. The half-day event was held on September 24, 2020 and brought together participants from local, regional, state, and federal transportation agencies, industry, research organizations, and academia. To address the goals of the workshop, the event was organized into listening sessions with question and answer opportunities as well as breakout sessions for discussion among attendees. The traditional poster sessions and other lectern sessions seen in prior CFS Workshops were not included in the 2020 event due to the virtual format.

First, an opening session welcomed attendees and outlined the charge to fully engage in workshop discussions by providing feedback and use cases, and highlighting new and evolving changes associated with the 2017 CFS. Next, representatives from BTS, Census, and research organizations gave presentations on 2017 CFS data, methods, and history. This was followed by three breakout sessions titled: (1) Commodities Revisited, (2) Geography Matters, and (3) Emerging Modes. Each session tasked participants to discuss the necessary trade-offs between geographical resolution, commodity grouping, and mode detail. For example, in the Commodities Revisited breakout session, attendees were asked to discuss the importance of commodity detail in contrast to added detail on mode and geography. With each breakout session, moderators engaged with workshop attendees to highlight and discuss data uses, needs, and challenges related to the topic area of the session. Attendees were pre-sorted into breakout groups based on their responses to a pre-workshop survey which included questions on attendees' professional demographics, uses of the data, and desired breakout session topic area. Finally, the workshop ended with a Plenary Session in which breakout session moderators reported back to the larger group what had been discussed during the breakouts. The workshop closed with comments from BTS and Census leadership.

The remainder of this document details the activities of the workshop. This e-circular serves as an information source for the broader community of existing and potential CFS data users as well as an invitation for those interested in providing input for future CFS iterations.

Welcome and Workshop Purpose

CATHERINE LAWSON
University at Albany, Conference Chair
Moderator

WORKSHOP VISION AND EXPECTATION

PATRICIA HU
Director, Bureau of Transportation Statistics, U.S. DOT

The purpose of this workshop is to shape the next CFS to be conducted in 2022 and to help ensure it is aligned with the changing freight and economic landscape, the needs and challenges of the data users, and the reporting duties of the Bureau of Transportation Statistics (BTS). The BTS was established by the Intermodal Surface Transportation Efficiency Act (ISTEA) in 1991, and has since served to administer transportation data collection, analysis, dissemination, and reporting for all modes of transportation. BTS is one of 13 principal statistical agencies given legislative authority to collect and disseminate transportation data for statistical purposes. BTS is policy-neutral and objective. Data produced by BTS are not to be used for regulation or enforcement. The goals of BTS are to fill long-term data gaps by the development of surveys such as the CFS and the Vehicle Inventory and Use Survey (VIUS).

Traditionally the CFS Workshop is timed to coincide with the most recent CFS data release as a way to share noted trends and insights garnered from the data. This year, 2020, BTS challenges CFS Workshop attendees to consider a different purpose for this gathering—that is, to look toward the next CFS and ask what directions the CFS should take to maintain relevance, shape future decision-making, and address data user needs today and into the future. As a data producer, BTS is looking to the data users to consider what questions the 2022 CFS should address, what decisions the 2022 CFS should inform, and how the future needs differ from today.

The 2020 CFS Workshop falls at a unique time in the United States, as the country addresses the challenges presented by the global COVID-19 pandemic and the emergence of digital technologies. These current forces work to potentially transform freight by 2025, and the 2022 CFS will need to transform accordingly. Concerning the global pandemic, initial data shows that the freight sector was more resistant than the passenger sector to pandemic responses and mitigation actions. Future questions may be asked of the CFS data such as will the freight system fully recover and what will be the "new normal" for freight. Thinking about updates to the 2022 CFS in this light, the 2020 CFS Workshop attendees are asked to consider how the CFS can be designed to answer these questions. Moreover, emerging digital technologies may transform the manner in which freight moves, the efficiency of the freight system, and overall shipping behaviors and trends. Technology may also transform the way BTS receives CFS responses by changing the reporting format and the quantity of data that is reported. The 2022 CFS may transform to be able to answer the question of how emerging digital technologies impact freight movement.

Two main challenges arise from consideration of transformations to the CFS. The first challenge is how to design CFS to reflect the new normal for freight movement in light of the pandemic and emerging digital technologies. This includes a discussion about what additional information should be collected by the CFS to meet this challenge. The second challenge is how

BTS can address changing demands on the CFS data in light of budget constraints. This includes a discussion of what data generated by the CFS is no longer needed and what details can be removed. Further, the CFS can be designed with considerations of the minimum data required by users in terms of geography, commodity, and mode. Through this 2020 CFS Workshop, the data producers ask the data users to think outside the box and envision what the 2022 CFS should look like while addressing the two challenges presented here. The workshop provides the platform on which to discuss the future of the freight transportation data program in ways that make it more relevant and useful.

REVIEW OF 2017 CFS

JULIE PARKER
CFS Project Manager, Bureau of Transportation Statistics

As a component of the U.S. Economic Census, the CFS is a comprehensive multimodal shipper survey conducted every five years to collect data on commodity, value, weight, mode of transportation, destination, temperature control shipments, hazardous materials (HAZMAT) shipments, and exports. The Freight Analysis Framework (FAF) that estimates and forecasts freight flows is a product of the CFS. In fact, the CFS contributes 70% of the data in FAF. CFS covers 43 commodity classes and aims to provide equitable coverage of commodity movements in the United States. Geographically, the CFS covers all 50 states plus Washington D.C. and must provide equitable coverage of all 50 states as a federal data product. The purpose of this presentation is to review the 2017 CFS data collection and note changes to survey methodology and results. The 2017 CFS data was made available online in July 2020.

The 2017 CFS data collection included 104,000 sampled establishments. Of these, 72% responded and 81% of the responses were usable. The resulting CFS data contains almost 6 million shipment records, an increase of 1.1 million records from the 2012 CFS. In total, the shipments in the CFS data represent 12.5 billion tons of goods amounting to $14.5 trillion.

To disseminate the results of the CFS, the BTS, in conjunction with Census, provides several data products including data tabulations, a print report, and a Public Use File (PUF). Three new mode specific data tables are provided with the 2017 CFS results. These include tables for exclusive truck, rail, and water modes and depict where the dominant mode is part of a multimodal movement. Unique to the 2017 CFS, the domestic mode share of a specific export mode is indicated in the data tables. The CFS print report for 2017 was enhanced by removing duplicate tables and estimates, including a HAZMAT division table (there is a HAZMAT table in 2012), and expanding the modes in inter- and intra-state table and export regions. Specifically, export regions now include the rest of the Americas (South and Central America), Europe and Africa, and Asia and Oceana. The PUF, formerly known as the Public Use Microdata (PUM) file, will ultimately be incorporated into an online data visualization tool using Tableau. This platform will allow users to get specific data elements that are part of the compiled PUM/PUF without requiring advanced programming skills. Overall, the PUF is the same product as the PUM, but with a different name. The online tool continues the trend established in 2012 of providing an API for user access to data.

In addition to changes in result dissemination tools and formats, there were three notable changes to the 2017 survey methodology. The changes include (1) mileage calculations for

modal shifts, (2) increased target sampling, and (3) imputation. These are important to note as changes in survey results may reflect changes in the survey methodology and not necessarily changes to shipment patterns. A full description of the methodology and changes can be found on the BTS website at www.bts.gov/cfs/methodology_2017.

First, changes to mileage calculations were made in the 2017 CFS. To address modal shifts, a commodity requirement was added that used observed data to support routing decisions. Briefly, the dock or station of the transfer was required to service the specific commodity when estimating routing. In the past, the closest dock or station was used regardless of possible mismatch between commodity shipped and the dock or station handling abilities. For the 2017 CFS, the mismatches were restricted. For example, for rail, in the 2012 CFS if there was a grain storage facility five miles away and a steel yard a mile away, the flow of grain would be assigned to the closer facility (e.g., the steel yard) even if it did not match the commodity (e.g., grain would more appropriately be transferred at a grain storage facility). Now that is corrected, as there is a match between commodity and facility. In other words, the facility used in the 2017 CFS was based on three factors—commodity, volume, and distance from origin or destination—while for the 2012 CFS, the only factor considered was distance from origin or destination. If the new assignment or routing crosses a zip code boundary then a transfer is required which would produce a multimodal shipment. A shipment may cross a zip code boundary to be matched to a transfer facility that is appropriate for the commodity. Thus, the values for multi-mode shipments may differ between the 2017 CFS and prior years. This applies to truck-water shipments as well, and the change reflects a better ability to capture the true modal nature of the shipment. Water and rail data for this effort was gathered from the U.S. Army Corp of Engineers and the U.S. Surface Transportation Board, respectively. To address water export shipments errors noted in earlier CFS releases, the domestic water portion of exports is included in the mileage calculations for the 2017 CFS.

Second, the 2017 CFS saw increased target sampling for establishments that ship 6,400 or more shipments in a sampled week. Additionally, some establishments worked with BTS to find ways to share data more efficiently. Reducing respondent burden, this resulted in establishments sharing their data in alternative ways, referred to as consolidated reports, which often contained more data than what was required.

Third, and last, there was an increase in imputation in the 2017 CFS, although the overall imputation rates remain low. Imputation is the process of replacing data that are missing from a dataset due to nonresponse in the survey. Increasing imputation allows BTS to increase the amount of usable information and improve the quality of the data. Reported imputation rates were provided for the following data elements.

- Value (2.4%)
- Tonnage (3.7%)
- Destination zip code (1.6%)
- Commodity (2.3%)
- HAZMAT (4.2%)
- Temperature control (1.7%)

Beyond changes to survey methodology, the 2017 CFS also saw changes in response to user feedback provided in the 2015 CFS Workshop. As noted, electronic reporting increased for the 2017 CFS evidenced by 86% of respondents choosing to electronically report. Motivated by

challenges in accurately capturing commodity transported, a machine learning classification system was introduced. The commodity classification scheme follows the Standard Classification of Transportation Goods (SCTG) codes but should the CFS transition to other commodity schemes, the machine learning platform can assist with that transition. BTS piloted subarea geography using the 2012 data and will produce the same subarea aggregates for the 2017 data.

Like the 2015 CFS Workshop, the 2020 Workshop seeks to facilitate discussion among data producers and users regarding how best to align the CFS with the needs and challenges expressed by all parties. Some items discussed in prior workshops remain relevant today and the goal is to continue to address these issues while ensuring they are still on point with the changing freight landscape. These items relate to changing industry practices, new data collection and processing methods, small area estimation possibilities, the rise in e-commerce, and linkages to foreign trade data. BTS seeks to find out how users interact with CFS data products like the PUF and preliminary estimates, what these products are applied to in practice, and ways in which the products can be made more relevant.

CFS WORKSHOP QUESTIONNAIRE SUMMARY FINDINGS

CHA-CHI FAN
Director of the Office of Data Development and Standards, Bureau of Transportation Statistics

Prior to the workshop, pre-registered attendees were asked to complete a survey. The purpose of the survey was to 1) measure use of CFS data products including FAF by CFS data users of diverse professional backgrounds, and 2) sort attendees into pre-selected workshop breakout sessions. The survey was completed by 43 attendees (an approximate 50% response rate), 35% of which were from state and local government agencies, 30% from the private sector, 16% from academia, 14% from federal government agencies, and 5% from trade organizations. In this presentation, Cha-Chi Fan summarized the survey results and highlighted how the responses can affect the development of future CFS deployments.

In terms of data product usage, more survey respondents reported using the FAF relative to the use of the CFS. Overall, 50% of respondents reporting using FAF and CFS, 80% reported using FAF and among these users 30% use only FAF, and 55% reported using CFS and among these users 7% use only CFS. When asked about alternative data sources, respondents mentioned the Transearch database, a proprietary data product and planning tool that predicts U.S. freight flows by commodity, geography, and mode. Those that use the CFS or FAF reported mostly using CFS and FAF to analyze patterns of commodity and vehicle flows. The second most common use of both FAF and CFS was to develop models and/or tools for policy analysis. For FAF, forecasting related to infrastructure, energy, or the environment was reported as a common use while for CFS respondents reporting use of CFS to conduct economic analysis. Considering the common uses of CFS and FAF data (Figure 1), it is curious why more respondents use FAF over CFS. Responses to this question show that 20% of respondents were unfamiliar with the CFS dataset and 17% of respondents found the FAF data more readily usable.

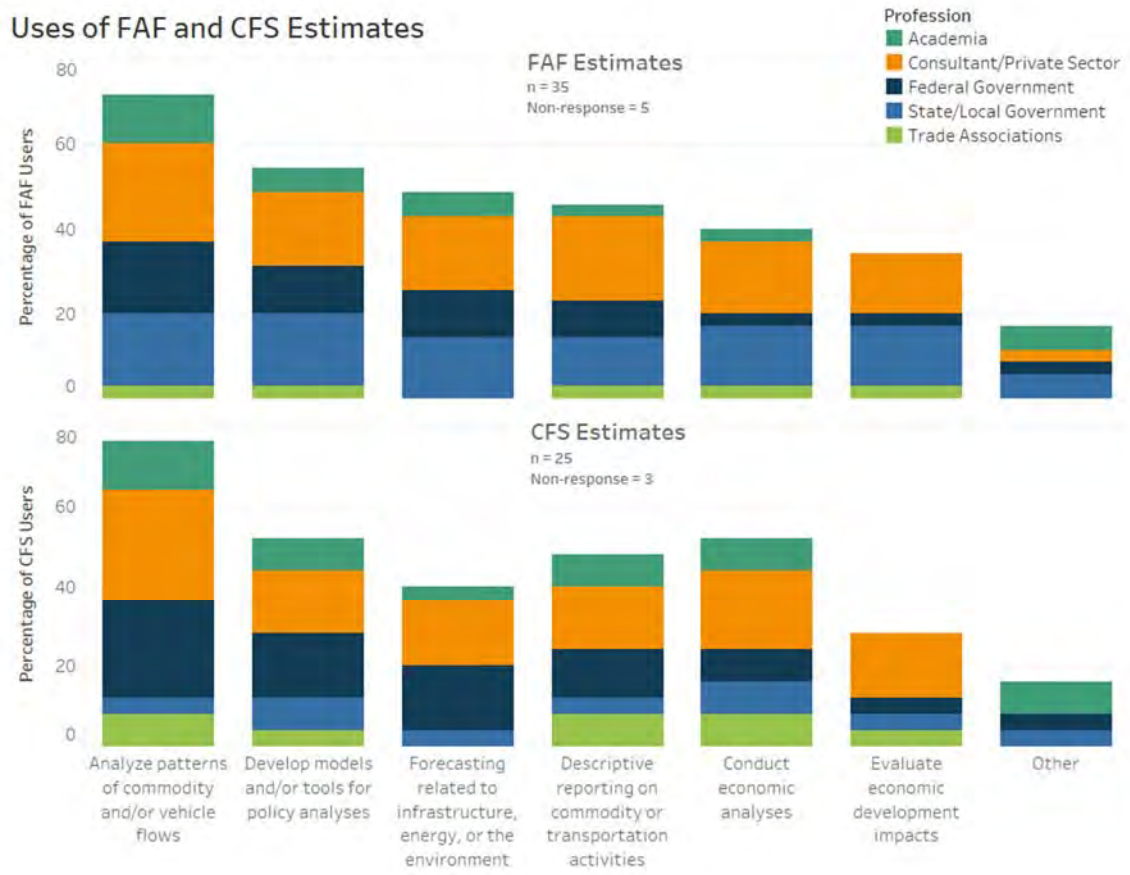

FIGURE 1. Pre-workshop responses for uses of FAF and CFS data.

In terms of product usage, 12% of respondents use the CFS microdata files (e.g., the PUMS file in 2012 and the PUF in 2017), 32% of respondents use only the CFS published data tables, and 44% use both data tables and microdata files. Of the provided CFS data tabulations, distance shipped has the highest usage among respondents (40% of responses), followed by exports (30%), HAZMAT (25%), and temperature control (20%). Of the provided FAF data, all geographic levels are reported to be used including regional OD, state level, and national-level estimates, with state-level estimates reported as the most commonly used. Of the time periods supplied by FAF, respondents reported using current year estimates most often.

Looking specifically at the CFS and FAF data products used, 90% of respondents reported geography as the most commonly used dimension; 68% reported use of the mode dimension; and fewer than 40% reported use of shipment weight, truck volumes, distance shipped, industry, export, HAZMAT, temp control, or other data dimensions (Figure 2). Although commodity ranked second in overall dimension use, it ranked third behind mode for state and local government respondents. Among the data estimates provided by the CFS products, geography (location) was ranked as the most important when compared to commodity, mode, HAZMAT, truck volumes, export, and distance shipped. Thirty-two percent of respondents reported geography as the most important estimate. For those reporting the geography dimension as the most important, sub-state aggregation was ranked as most important. When asked about trade-offs for finer geographic detail, most respondents (46%) would trade commodity and modal detail for finer geography detail.

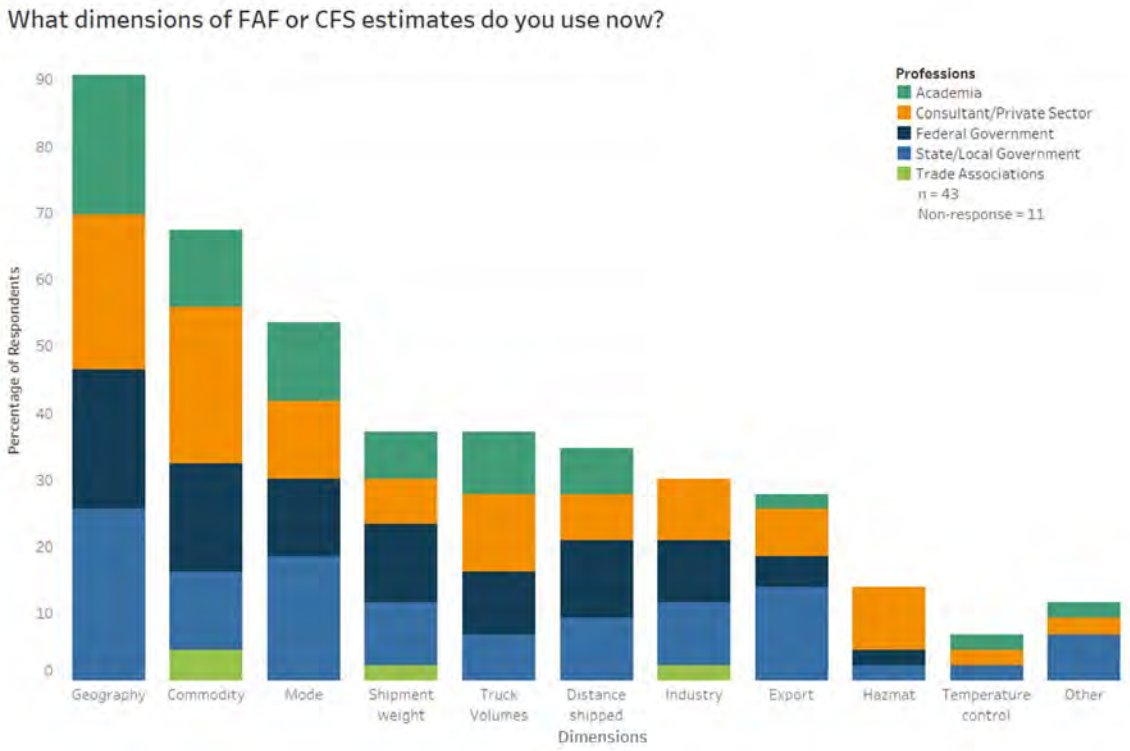

FIGURE 2. Pre-workshop survey dimensions of FAF and CFS used.

Lastly, the pre-workshop survey asked respondents about anticipated use of the Vehicle Inventory and Use Survey (VIUS). The VIUS is the principal data source for the physical and operational characteristics of the U.S. vehicle population. The VIUS originated in 1963 as the Truck Inventory and Use Survey and was last conducted in 2002 with a new survey expected to be conducted in 2022 to gather vehicle activity data for 2021. Data from the 2021 VIUS will be released in 2023. The 2021 VIUS will include new questions about modernization of vehicle features including physical and operational characteristics. FHWA vehicle classes 3 to 8 will be included in the sample frame. Some vehicles in classes 1 to 2 (pickups, mini vans, other light vans, SUVs) will be included in the sample frame as well. Over 50% of the respondents noted that VIUS would be useful. Respondents reported that two- or four-digit SCTG commodity detail would be most useful and that sub-state geographical details are the most important for the commodity dimension.

QUESTION AND ANSWER SESSION WITH AUDIENCE

PANELISTS

Cha-Chi Fan, Patricia Hu, Julie Parker, Chester Ford, Berin Linfors

At the conclusion of the presentations, the panelists provided additional detail on CFS, FAF, and VIUS data collection efforts. This section summarizes the audience questions and highlights the responses of the panelists where appropriate. The questions surrounded three main topic areas: survey methodology, uses of survey data, and new data products (e.g., the VIUS).

Survey Methodology

- Did the workshop presurvey response match with the expectations of the BTS staff?

Julie Parker and Cha-Chi Fan responded that there were a couple notable surprises in the survey responses. For instance, BTS anticipated that everyone would use the new VIUS but only 50% of respondents reported planned use of VIUS. BTS also noted from the pre-workshop survey results that geography was a primary interest, but it was unclear the degree to which sub-state data would be used if available.

- Is the 2017 data projected from 2012 or is it from Census?

Chester Ford responded by clarifying that the 2017 CFS data is all based on 2017 responses, not projected from 2012 CFS. FAF has estimates for 2017 and 2018, which are based on the 2012 CFS. Now that the 2017 CFS is available, FAF will have a base year of 2017 and be projected to 2018, 2019, and 2020.

- Has the shift to electronic reporting had a positive impact on the survey and in what ways?

Julie Parker and Berin Linfors responded affirmatively in that with electronic reporting, more data is available, and that data is cleaner. In addition, with electronic reporting, commodity descriptions can be used to obtain finer and more accurate commodity classifications through a newly developed machine learning approach. Overall, the small percent of paper responses has been burdensome and the data quality can be lacking, so moving forward BTS will increase electronic reporting.

- How does the FAF convert between value and volume? Is there a price database?

Chester Ford responded by noting that while there is not a price database, relationships have been established between commodity value and volume. For domestic data from CFS, value and tonnage are directly reported. For exports, sometimes value to weight ratios across multiple years are used. With water, the value and tonnage data is used directly.

- What were the most noticeable changes in commodity flows or shipment characteristics for the new CFS?

Julie Parker and Chester Ford responded by stating that shifts in commodity flows were notable but stable. This can be attributed in part to new categorizations resulting from the machine learning method. Specifically, the commodity category labeled “miscellaneous” could be appropriately assigned to a unique commodity group. Julie and Chester noted that users should be careful about comparing routing across CFS years, since the 2017 has more precision.

Uses of Survey Data

- What percent of flows can be classified as middle or last? Or is it all long-distance flow?

Julie Parker and Chester Ford responded to this question by stating that the CFS does not include flow classification by segment of trip**.**

- How are you planning to release a freight emissions inventory? Are there other sources for truck or rail volumes that can be used to estimate emissions?

Julie Parker, Cha-Chi Fan, Patricia Hu, and Chester Ford responded to this question by stating that the CFS, FAF, and VIUS do not collect data on freight emissions. While VIUS does ask about VMT, fuel type, and other physical and operational characteristics of the vehicle, it does not ask about measured emissions. The FAF considers distance traveled ranges and equipment in the estimation of highway flows. Documentation for the FAF methodology can be found online and may be suitable for high-level emissions analysis.

- How can a user request more data for commodity classification and/or for geographical estimates at the zip code level?

Julie Parker and Chester Ford suggested that to obtain finer geographical detail (currently the data is provided at the state level) or commodity detail (currently the data is provided at three- and four-digit SCTG codes) than what is provided in the public files and data tables, a user needs to make a request for internal data through the Research Data Center (RDC). The CFS data is protected for privacy so not all requests for finer detail can be met and in no instances is the respondent level data shared.

Vehicle Inventory and Use Survey (VIUS)

- Will VIUS include e-commerce delivery vehicles? Will the VIUS distinguish between for-hire and owner-operated vehicles? Will VIUS capture electrification and sleeper berth design?

Cha-Chi Fan responded to this question by stating that the VIUS will include vehicles in classes 1 through 8, which may not capture smaller e-commerce delivery vehicles. The VIUS will include designations for for-hire versus owner-operated vehicles and for various engine fuel types and cab characteristics (sleeper berth, etc.).

2022 and Beyond
Prioritization Strategies for Moving Forward at Top Speed

ALISON CONWAY,
City College of New York
Moderator

MINIMUM REQUIREMENTS OF FUTURE CFS AND FAF

ROLF SCHMITT
Deputy Director, Bureau of Transportation Statistics

The minimum data requirements of the CFS and FAF have evolved over time to meet the needs of data users and to comply with privacy protections of data producers. Changes to survey collection purpose, methods, and resulting data products reflect the changing freight landscape and the needs of data users to apply the data to freight planning analyses. This is evident in the history of the CFS and its predecessor: the Commodity Transportation Survey (CTS). This presentation describes the history and evolution of the CFS, highlighting its early origins as the CTS, and reflects on how the requirements for data collection have changed and lead to the current CFS and FAF data products.

Both the CTS and the CFS measured shipments from business establishments, sampled by sector of the economy, to estimate region-to-region movements of commodities. When initiated in 1963, the CTS focused on shipments by the manufacturing sector. Recognizing three decades later that substantial flows of freight were shipped from other sectors of the economy, planners of the CFS expanded coverage to mining, wholesale, and the portion of retail in the mail order business. At the time, these sectors represented most of the physical freight flows in the United States (Figure 3). Farm-based shipment were not included since farms are covered by the Census of Agriculture that is conducted by the Department of Agriculture rather than the U.S. Census Bureau. The focus on outbound shipments rather than both inbound and outbound avoids double counting. This focus captures exports, not imports.

The CFS is a shipper survey rather than a carrier survey to capture the true origin and destination of multimodal shipments and to capture shipments by shipper-based trucks and other equipment. Shippers were assumed to have more knowledge than the manufacturer or carriers about the use of third-party shipment practices and models (Figure 4). The parcel delivery category was also introduced as a means to better depict multiple mode shipments since shippers were not likely to know mode choices of parcel carriers.

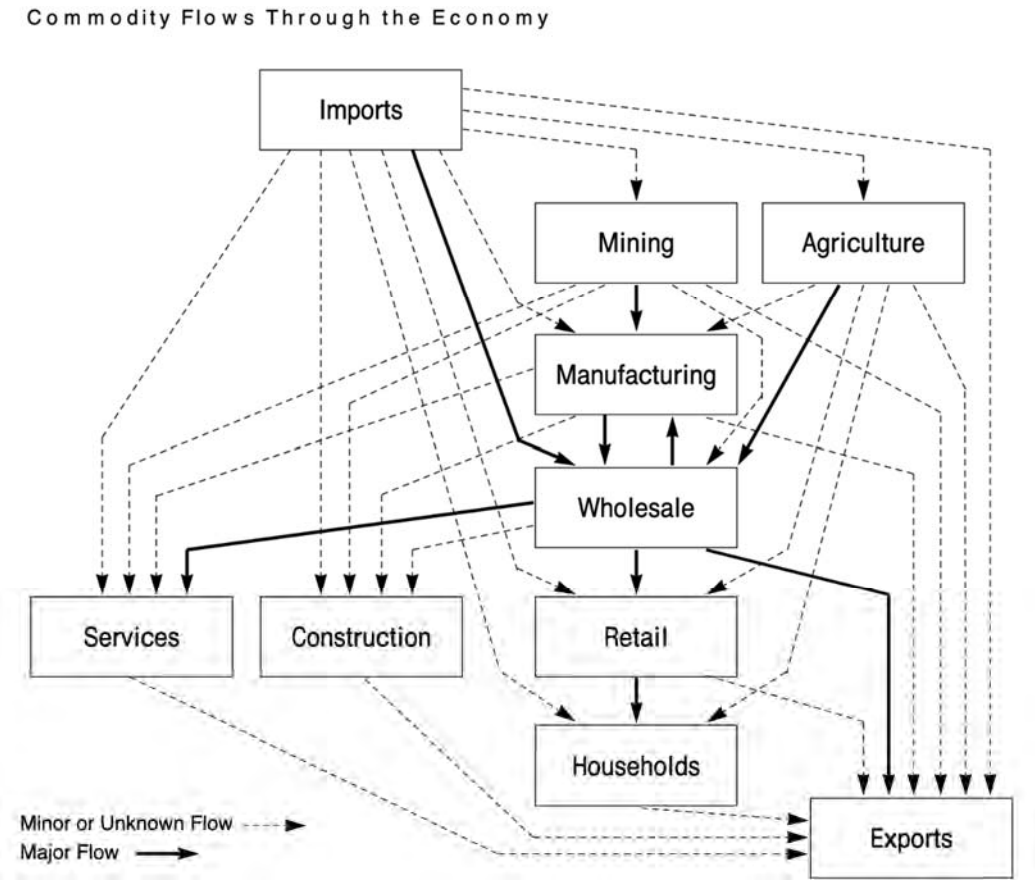

FIGURE 3. Commodity flows through the economy (image from Rolf Schmitt).

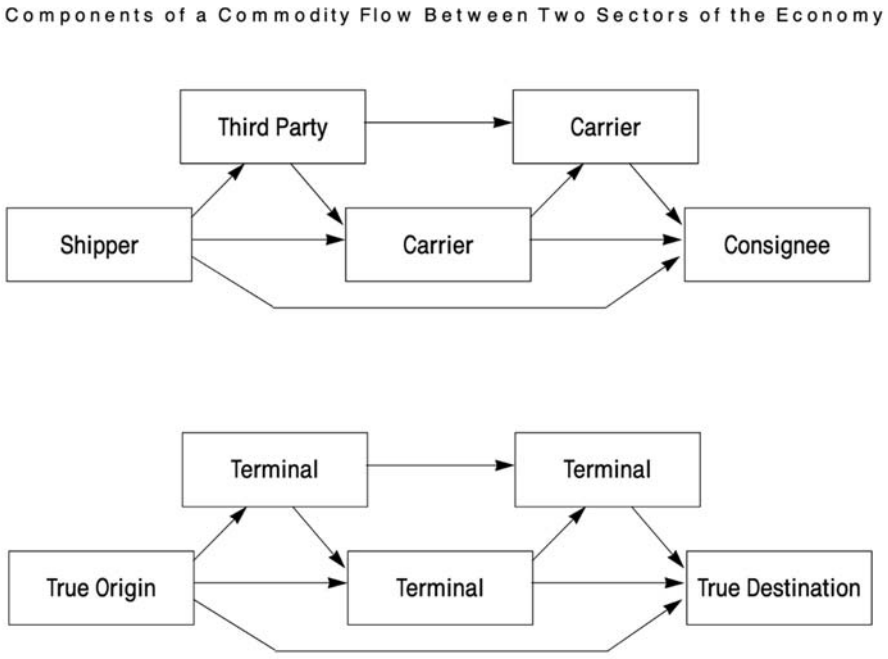

FIGURE 4. Components of a commodity flow between two sectors of the economy (image from Rolf Schmitt).

Although the CFS made several advances over the original CTS, the CFS still has some data gaps. The FAF was created to fill these coverage gaps, including imports, farm-based shipments, petroleum in pipelines, and shipments by governments and households. The FAF was also created to provide annual estimates for years between each edition of the CFS, to assign region-to-region flows to the transportation network, and to make multi-decade projections of future flows. FAF projections assume that future change is driven primarily by general economic growth and not industry restructuring. While the FAF informs industry-to-industry freight flows in addition to its primary focus on region-to-region flows, the FAF does not measure company-specific supply chains.

Ongoing changes in the economy will affect the 2022 CFS, raising at least three questions. First, will the 2022 CFS be able to capture the spread of e-commerce throughout retail and the replacement of consumer-provided transport with commercial transport? Movements from brick-and-mortar establishments in personal vehicles are typically measured as shopping trips in household surveys while shipments from retail outlets and distribution centers in for-hire or shipper-owned vehicles are measured as freight in business establishment data such as the CFS. Second, should the CFS drop the mail and parcel mode for specific modal combinations including truck-and-mode-unknown, and should the FAF break out the category of multiple modes and mail into specific combinations? The multiple mode option was introduced to handle intermodal shipments for services like UPS and FedEx which decide on the mode after picking up the shipment in a truck. Third, will FAF users who want county-to-county flows be willing to accept estimates that exceed the geographic detail published in the CFS? What methods for disaggregation or small area estimation still meet data quality requirements within the constraints of confidentiality?

TRADE-OFF: COMMODITY, MODE, AND GEOGRAPHY

CHESTER FORD
Analyst, Bureau of Transportation Statistics

Just as data users express particular needs and requirements for the CFS, data producers have practical requirements to consider when creating CFS data products. As a federal data product, the CFS should equitably cover all 50 states (plus Washington, D.C.) and commodities. However, the raw data is naturally concentrated by commodity and geography and thus does not naturally provide equitable coverage. This presents a challenge for data producers. The purpose of this presentation is to describe the trade-offs between commodity, mode, and geography in the context of constraints for creating CFS data products.

The CFS can be distilled into three key dimensions: commodity, geography, and mode. Considering the fixed resources of the program, the trade-offs among these dimensions can be represented by a constraint triangle (Figure 5). Only one (or two) of the dimensions can be expanded (disaggregated), but at the expense of consolidating (aggregating) the other two dimensions. Some of the existing hazardous materials tabulations already function in this manner, with expanded detail for UN commodity codes, but decreased detail on mode and geography. As expressed by data users in prior CFS Workshops, further breakdown by mode and destination (geography) helps to meet the application needs of data users and stakeholders.

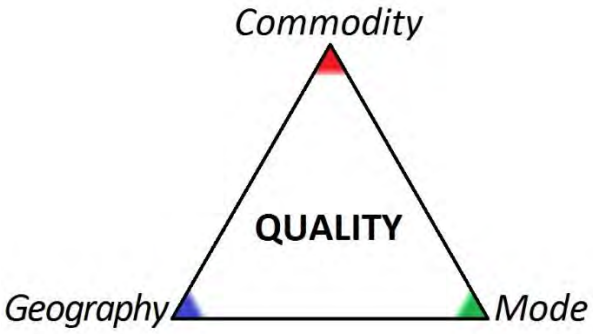

FIGURE 5. Constraint triangle paradigm.

The 2017 CFS sample included 104,000 establishments of which 60,100 establishments provided usable data corresponding to 5.98 million shipment records. On average, there are 99 shipment records per establishment and 455 reporting establishments per area. At a finer geographical resolution, consider that there are 3,200 counties in the United States. By employment, the top 32 of these 3,200 counties contain a quarter of the employees in industries covered by the CFS. This concentration of activity shows up in the responses for the CFS and points to the challenges data produces consider when providing equitable coverage by commodity, geography, and mode.

Examples from the 2012 CFS responses and results help illustrate and explain the trade-offs inherent to the data. As it pertains to the geography dimension, a look at the Texas metro areas shows that Harris County dominates the Houston geographical area in the CFS, providing approximately three quarters of the information for that zone. Likewise, Dallas and Tarrant counties dominate the Dallas geographical zone in CFS. As it pertains to shipment characteristics concerning geography, St. Louis presents an example in which more information is available as the destination areas are rolled together (Figure 6). Specifically, local shipments currently dominate the St. Louis geographical zone in the CFS. If the size of the zone were to increase, this would have an accelerating impact on the availability of information such that larger destination areas would be more likely to have more estimates (shipments). Especially at longer distances, the geographical granularity may become spotty as a result since smaller destination areas would have less estimates (shipments). This is the effect of data quality and its relation to the number of shipments. When thousands of shipments are provided for a zone, commodity, or mode, estimates are almost always possible, but as the number reduces to tens of shipments, the likelihood of publishable estimates decreases. BTS supports a cutoff for minimum quality based on relative standard error. Estimates are not provided when the relative standard error is less than 50%. However, all shipments get used in higher-level estimates.

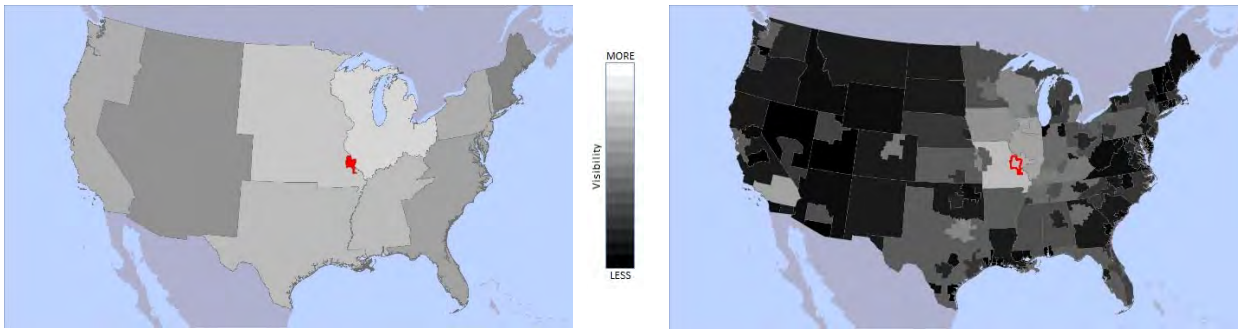

FIGURE 6. Example of data quality for St. Louis from the 2012 CFS.

Inequity across modes is also pronounced. In the 2017 CFS, there were 240,000 shipments by rail, air, water, and other modes. This represents only 4% of all records. Trucking accounts for 70% of records and parcel shipments accounts for 36%. While modal shifts may evolve in the future, there will most likely be imbalanced mode distributions and this can affect the level of mode disaggregation provided in CFS data products. Parcel shipments, for example, tend to be multimodal and are expected to increase in the future, so future CFS data may disaggregate or reconsider how parcel shipments are reported.

Thinking about the constraint triangle and balancing data user needs with the challenges faced by data producers, it is key to consider how CFS data products may be presented differently in future years. The following examples suggest how the balance among survey dimensions may be reimagined. The current CFS data products follow the mode structure which parses parcel shipments from other modes (Figure 7). Parcel shipments (36% of shipments in 2017) can be multimodal. Data users and producers may consider dropping the parcel shipment mode and shifting parcel shipments to singular modes (e.g., move parcel by truck-to-truck mode). Input from users is needed to guide this change as drayage (first/last mile) trips may be more important than the long-haul, dominant mode portion. The geography dimension likewise can be altered. There is user demand for county-level data, and data producers recognize that counties are the building blocks for metropolitan statistical areas (MSA) and “rest of state” regions. User input is needed to guide the decision on balancing geography aggregation. For instance, is there a preference for splitting MSAs into smaller areas based on county or sub-county geography? The commodity dimension has already seen improvement with the introduction of the machine learning code to classify commodity into SCTG codes. Currently there are 42 two-digit SCTG classes, with some information available at the three-digit SCTG level (Figure 8). While this has been the norm, input from data users is needed to determine if other commodity classification schemes would be more useful.

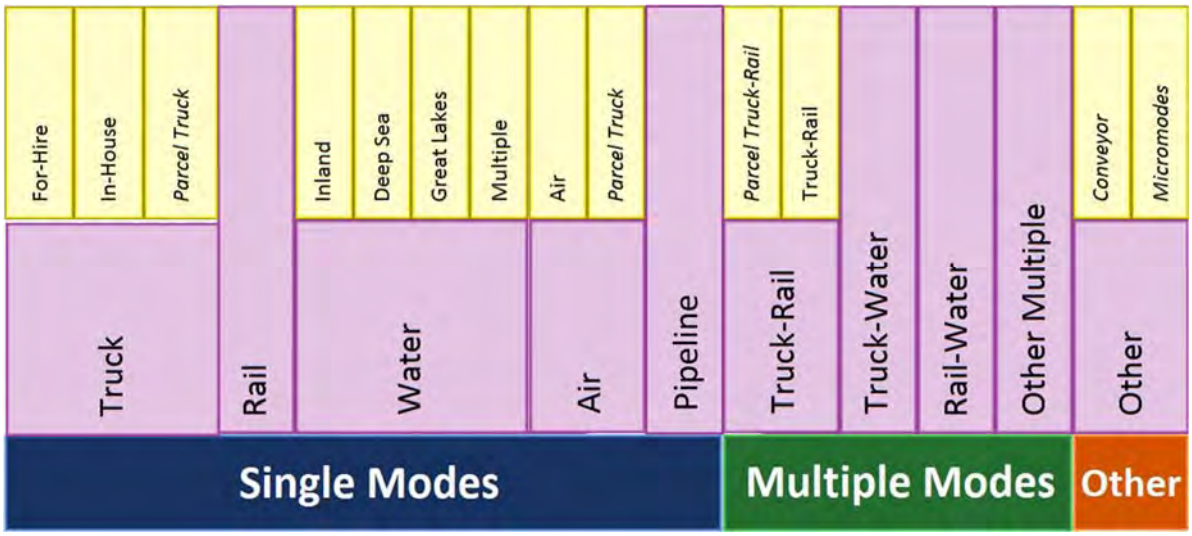

FIGURE 7. Current CFS data definitions structure for mode.

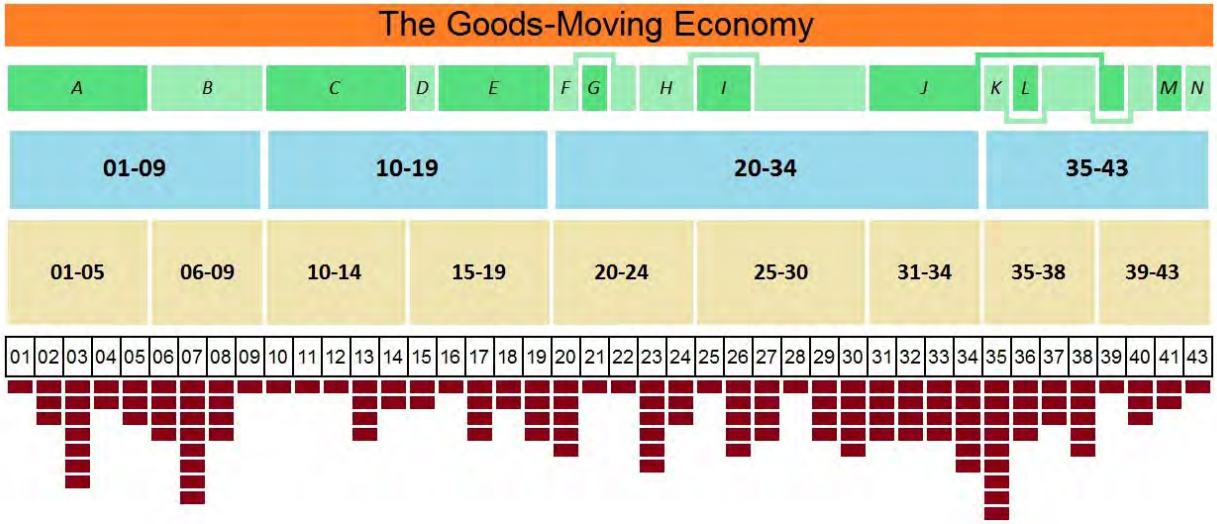

FIGURE 8. Current data definition structure for commodity using the 42 two-digit level SCTG classes in the CFS.

In addition to questions for data users related to the constraint triangle, data producers ask for input on more general questions.

- What should the CFS and FAF data programs look like to maximize the usefulness to the users?
- What other data tabulations, tools, and visualizations would users find helpful if they were not limited to geography, commodity, and mode?
- What is the level of comfort users have with synthesized data and how is synthesized data used?

WHAT USERS NEED NEXT AND HOW TO MEET THE NEEDS

BILL EISELE
Senior Research Engineer and Mobility Division Head, Texas A&M Transportation Institute

Up to this point in the workshop, perspectives from data producers commonly reflect on constraints that limit or challenge data user needs. This is best illustrated by the constraint triangle in which the three key dimensions of the CFS data must be balanced to maintain data protection and accuracy standards. This presentation reflected on data user perspectives by highlighting current and forecasted industry trends that impact CFS data usage and suggesting ways in which current data limitations can be countered by new data collection and use opportunities (Table 1).

TABLE 1. Industry and User Need Categories

Trend and/or Public-Agency Topic of Interest	Data Needs and/or Related Questions
Global Economy (trade)	Import/export flows
Pandemic impacts	Supply chain impacts
Big data are everywhere	How to integrate?
Spectacular increases in e-commerce	Trip info, travel behavior changes, land use impacts
Technological advances (C.A.S.E.)	Impacts on travel behavior? Maturation rates?
Multimodal performance needs	Multimodal trip info—speeds, volumes, by commodity
Policy, program and project prioritization	Multimodal trip info—speeds, volumes, by commodity
Micro-mobility (people and goods)	Impacts on travel behavior?

Overall, CFS is a valuable tool for providing high-level freight movement situational awareness, and it is important to maintain this data source and its related data products like FAF. However, data users from local, state, and regional transportation agencies and the consultants and researchers who work with these agencies, suggest that multimodal trip information reflecting supply chain level movements at the local level is needed. But, as mentioned by Rolf Schmidt, the CFS is not designed to gather supply chain movements, as it is a shipper survey.

Nevertheless, public agency data use cases indicate that local-level supply chain data would benefit from many planning and operational applications. At the local level of analysis, agencies are increasingly interested in issues such as the micro-mobility of people and goods; policy, program, and project prioritization; and multimodal performance needs. At the state and national/regional levels, agencies are increasingly focusing attention on the uses of advanced technologies like connected, autonomous, and electric vehicles, increases in e-commerce, and uses of big data to better understand and measure the impacts of freight on transportation infrastructure. At the international scale, agencies continue to seek a better understanding of the global economy and how domestic transportation systems and infrastructure affect imports and

exports at the international level. Moreover, the 2020 global pandemic has created a need to better understand national and international freight.

In common to these use cases is the need for data depicting the multimodal supply chain, travel behavior at the trip, route, and segment level, and system performance (speed, travel time, volume, and the variance in these measures) at the route level. CFS data meets many of the data needs at the national/regional or state levels but new ways of gathering and sharing the data are needed to better address local and international needs. Two suggestions were put forth to address CFS data challenges.

First, to gather multimodal trip information, the CFS can integrate import/export data from alternate sources. Specifically, to address many of the data needs expressed above, it would be helpful if the CFS produced multimodal trip details including the total supply chain movement from origin to destination as intact trips. Currently, as reported by Chester Ford in the prior presentation, half of the trips reported through the CFS are less than 230 miles in length suggesting that many long-distance trips may be missing. To add in the long-haul portion, the international portion of a shipment is needed but is not available from CFS as it is a U.S. shipper survey. To supplement, it may be necessary to gather cross-border information from statistical reports of other countries.

Second, since the CFS is conducted in five-year cycles, intervening years between surveys are not included. This presents issues with temporal coverage. The length of the data cycle may also result in gathering and sharing timely data. Users often desire to apply commodity flow data to periods by month (season), week, or daily intervals, which is a challenge given that the CFS contains annual estimates. There are likely opportunities to automate the shipper survey to increase data timeliness. For example, the use of automated exchanges of data from sampled entities' enterprise data systems or other sources of electronic business data could be leveraged to address temporal continuity. As Julie Parker mentioned in her earlier presentation, increases in electronic reporting brought increases in the amount of data shared by shippers. Data users are encouraged by the possibility of increasing data sampling to the point that local-level estimates can be shared without compromising confidentiality.

User Breakout Sessions

Following the formal presentation sessions, workshop participants were invited to join one of three parallel (concurrent) breakout groups. Each breakout session had a moderator and a recorder. Workshop participants had indicated their preference a breakout session in the pre-workshop survey. Of the 43 pre-workshop survey respondents, 17 selected to join the breakout session titled "Commodities Revisited", 12 selected to join the session titled "Geography Matters", and 14 selected to join the session titled "Emerging Modes". Subdivisions of participants in each breakout session by profession are shown in Table 2 below.

TABLE 2: Breakout Session Participants by Profession

Session	U.S. Census	Private Sector	State Gov.	University	Other	Total
1: Commodity Revisited	1	4	2	6	4	17
2: Geography Matters	4	4	3	0	1	12
3: Emerging Modes	2	3	2	1	6	14
Total	**7**	**11**	**7**	**7**	**11**	**43**

Each group was asked to discuss the trade-offs between commodity, geography, and mode detail provided by the CFS data products including the FAF. The purpose of the breakout sessions was to gather data user insights and preferences for trade-offs so that BTS can get a greater understanding of user needs. Moderators provided an overview to open the sessions and then asked for participant comments and discussion of the following topics.

Breakout 1: Commodities Revisited

In this breakout session, participants discussed the multiple levels of SCTG commodity codes that are commonly used in transportation planning and forecasting. As commodity detail increases, compromises may have to be made in geographic and modal specificity. This session examined how commodity information is used reliability at finer levels of commodity, geographic, and modal detail.

Breakout 2: Geography Matters

In this breakout session, participants focused on the various levels of geographic detail that could be provided by the CFS from states to zip codes. The discussion also prompted participants to discuss how geography be aggregated for different purposes, and the trade-offs between geographic, commodity, and modal specificity.

Breakout 3: Emerging Modes

The emergence of new modes and the evolution of existing modes pose challenges for collecting, processing, and using CFS data. In this session, participants were asked to discuss issues

surrounding modal definitions, particularly distinctions between trucking sectors and the classification of multimodal and parcel shipments.

Summaries of each breakout session are presented in the following sections.

COMMODITIES REVISITED

SCOTT DRUMM
CPCS, Moderator

JA LEE
Bureau of Transportation Statistics, Recorder

Introduction

The commodity classification scheme is a hierarchical system that can be further disaggregated to up to seven-digit SCTG codes. The pre-workshop survey showed that at least 60% of participants use CFS commodity data at the two-digit SCTG level and 50% of participants use the three-digit level. With larger swaths of geography, it is possible to further disaggregate the commodity classification. Geography can be split at such a level that commodity groupings and destination information are available while maintaining variability at a reasonable level. Conversely, to obtain very detailed geographic data for shipment value, it is necessary to lose commodity detail as shown in the example for shipment value by detailed origin and destination zones in the Portland, OR area (Figure 9). A challenge in commodity disaggregation is that commodities may not share transportation characteristics so there is inherent value in keeping some version of a disaggregated commodity scheme.

Value of Freight Movements within Portland, OR-WA Metro Area

Origin		Destination: W411A	W411B	W411C	W532Z
Multnomah and Clackamas Counties OR	W411A	11,617	3,562	2,260	2,258
Greater Portland Northwest OR	W411B	5,618	1,646	755	410
Greater Portland South OR	W411C	1,305	726	2,190	326
Portland WA	W532Z	1,316	428	261	2,277

FIGURE 9. Pilot dataset for shipment value for geography disaggregation for Portland, OR.

Commodity detail is reported with a description and then respondents provide one of the 43 two-digit SCTG classification codes. Because commodity data is reported with details by the establishments and it is noted that the respondent can select an inappropriate SCTG code, the future CFS will no longer ask respondents to look up their SCTG code. Rather, the code will be derived by a Machine Learning Commodity Classification tool built by BTS and Census. This means that it is possible to move to another commodity classification system. This breakout session asked data users to consider what commodity classification schemes are desired and what trade-off would be viable to get finer commodity detail.

Discussion

The key themes emerging from this breakout session included the following and are discussed below in more detail.

1. There is a desire to allow flexibility in the commodity classification scheme to ensure comparisons to prior years (that only had the SCTG scheme) but realizing that commodity groupings may need to change over time. This is suggested to be accomplished using online tools like Tableau and through furthering the machine learning analysis for commodity group labeling.
2. The commodity grouping structure is important for distinguishing the varied transportation characteristics among commodity types. It would be beneficial for the commodity groups to be aggregated/disaggregated according to transportation characteristics.
3. There is value in providing geographic detail at both ends (origin and destination) of the commodity trip as well as commodity value as it would contribute needed data for corridor management (e.g., it would provide a lens on the value of a corridor or segment).

Commodity Classification Considerations

The group discussed possible high-level commodity grouping schemes that would be of value across applications. In general, some individuals in the group suggested that the commodity classification scheme recognize and be distinguishable by vehicle type. Knowing the commodity type, even at an aggregate level, would help planners determine the type of equipment involved in shipping and would aid in the ability to tie commodity shipped to investments needed. As an example, one participant suggested that knowledge of equipment (vehicle type) determined from the commodity classification could be used to assess the needs of transload facilities for agricultural products. Another participant mentioned, as an anecdotal example, that soybeans are not in the same classification with corn and wheat even though they are transported in a similar manner. Several participants suggested the following high-level commodity categorizations.

- Liquid bulk (tankers) versus dry bulk (vans or hoppers)
- Refrigerated goods (refrigerated containers or vans) versus non-refrigerated (vans or containers)

Participants also highlighted the increasing prevalence of e-commerce and its impacts on planning activities. One participant noted that more data specifically tied to e-commerce shipments is needed to answer questions about local transportation system performance, capacity of regional distribution facilities, and interactions between land use and transportation. This

participant thought that other datasets such as the Institute of Transportation Engineers (ITE) Trip Generation Manual do not provide accurate data for distribution facilities and thus had the opinion that there is a data gap that may be closed by commodity re-classification in the CFS.

SCTG Code System

The current commodity classification in CFS is the SCTG system. Participants discussed other possible established schemes such as the North American Product Classification System (NAPCS), Harmonized System, and Standard Transportation Commodity Code (STCC) and considered whether these schemes would be more relevant than the SCTG system. The STCC system is used by the railroad industry and there are existing crosswalk tables between STCC and SCTG that could be used for mapping between schemes. One participant noted that NAICS is used in conjunction with SCTG for producing more disaggregate commodity detail. Participants agreed that the SCTG system is useful but that crosswalk tables between SCTG and other common schemes like the STCC would be valuable. Although some participants from the railroad industry suggest that SCTG to and from STCC crosswalks are available, other participants noted the challenge of providing cross walk tables can be a complex issue attributed to different planning applications. A repeated suggestion was to allow the user to select the classification scheme through the online data portal.

A caveat to a new classification scheme was expressed by several participants. One of those challenges was understanding and learning a new system. A challenge with a new scheme is that it will be an issue to compare to prior year datasets if clear crosswalk tables are not provided. Thus, rather than introducing a new scheme, participants suggested the creating of different levels of "roll ups" (aggregations) that could be customizable. Participants also suggested perhaps any new system could be built based on an existing system. Further, with the introduction of machine learning for commodity classification in the 2017 CFS, participants were curious how the same or similar techniques could be applied to translate results into different classification schemes that may be more in line with their application areas. Lastly, participants suggested that when the data in the national level CFS does not provide direct observation of local-level commodity flows, it can be used instead to help guide local-level sampling. Thus, the national commodity scheme could help shed light on more locally relevant follow up studies.

Origin-Destination Details and Commodity Classification

A challenge of providing origin-destination along with commodity classification is that it may violate the nondisclosure clause for some specific commodities. A participant suggested that if data users were to extract the origin and destination along with commodity code, then this would provide enough detail to identify the shipping establishment. This would be of more concern for unique commodities, for example, aircrafts, natural gas, and asphalt. There was concern expressed that the level of readily available public data could allow a level of reverse engineering that would deanonymize the CFS results.

While recognizing these challenges in providing origin-destination data along with detailed commodity data, participants repeatedly suggested that such data would be helpful in conducting corridor or segment level analysis. Knowledge of commodity moved along a study corridor from a specific origin to a specific destination would help with project identification and

prioritization. One participant stated that private sectors (consulting firms) will want commodity reported at a level of detail to be able to understand the parts of the transportation network that are impacted. In further support of this statement, another participant remarked that commodity detail can be used to infer value, and if linked to specific origins and destinations, can then be used to understand the flow of value (economy). This data, as suggested by the participant, would be a good resource for explaining the value of the transportation system for legislative level decision-making.

GEOGRAPHY MATTERS

DANIEL SMITH,
The Tioga Group, Moderator

KELSEY ELLENWOOD
Bureau of Transportation Statistics, Recorder

Introduction

The BTS follows the "rule of three" when publishing CFS results. There must be at a minimum three reporting firms in any breakdown. This allows anonymity of the reported results and complies with the nondisclosure conditions. When there are fewer than five shipments, estimates might not be possible. Once there are 10 to 15 shipments, conditions for publishing data are sufficient but data quality may be low. When there are more than 30 shipments, there is a much better foundation for confidence in the data.

As it pertains to geography, there is a reported origin zip code and a destination zip code. Issues arise in anonymizing the data for origin zip code. This is because it can be difficult to conceal the shipping establishment's identity at small geographical scales. Tabulations below the county level can pass nondisclosure thresholds, but only for the largest counties. For the smallest counties, which are a majority of the 3,300 counties in the United States, publishing results below the state level becomes challenging.

This breakout session explored methods to create sub-metro area geographies to enable further analysis of CFS data. An example was provided for the Portland, Oregon area that can be split into sub-geographies while still enabling commodity grouping and destination information to be available and variability low (Figure 10).

Discussion

The key themes emerging from this breakout session included the following and are discussed below in more detail.

1. There is a perceived imbalance in state-level estimates for states that do not have MSAs within their boundaries.
2. There is a desire to obtain county-level estimates and to do so participants are willing to give up details on commodity and mode.

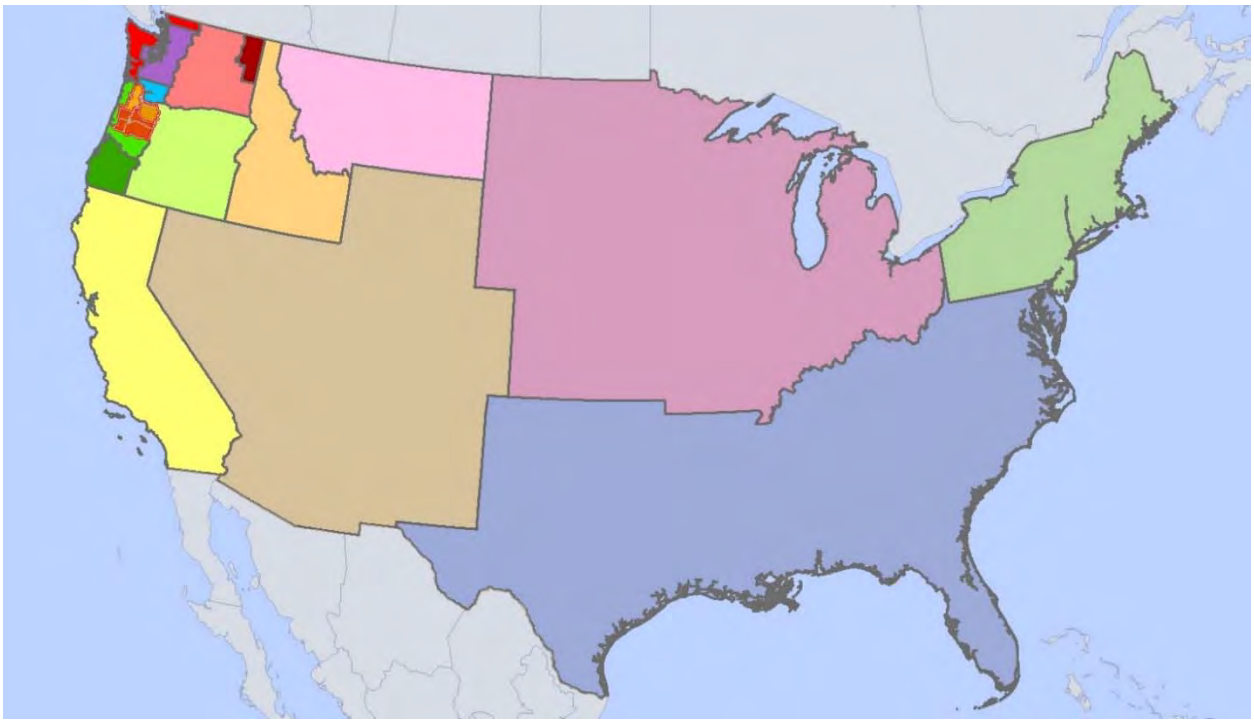

FIGURE 10. Illustrative example of geographical detail for the Portland area.

State-Level Data Challenges

From all participants, there was a desire to obtain county and zip code level estimates. Participants from state, regional, and local governments commented that when only state-level data is available, they must look (and purchase) elsewhere to get the finer level of geography needed for planning studies. For example, in Vermont, and many other states, only state-level data is available from CFS since there are not any within state MSA regions. The participants reflected that it is a continual challenge to address.

County-Level Estimates and Trade-Offs

Many participants believed it would be best to trade off detail in mode and commodity to get finer level geography estimates. Several participants commented that their agency or company has its own means of disaggregated geographical detail to the county level using mathematical models for imputation. In some cases, participants suggested that they use proprietary data from third-party data providers such as Transearch or truck GPS data to obtain usable data at the county or finer level. It was noted that Transearch, in particular, pulls data from the CFS.

Representatives from Census and BTS highlighted work underway to accomplish this request. For example, researchers are looking at producing truck and parcel estimates of areas with high shipping volumes using “rolled-up” (aggregated) commodity classification levels. Moreover, it was suggested that detailed geography may be obtainable from business-level microdata, although this is not widely distributed.

Lastly, a few participants suggested that inherent in the trade-offs among commodity, mode, and geography are the natural boundaries and groupings of the data that prevent further disaggregation due to nondisclosure protections. In some cases, a county boundary may not match the MSA, or the MSA-level of reporting may provide more detail that the county level. In these situations, participants suggested the possibility of collecting more data (more establishments or shipments) for specific areas.

EMERGING MODES

PAUL BINGHAM
IHS Markit, Moderator

DAN BIGGIO
Bureau of Transportation Statistics, Recorder

Introduction

Modes reported in the current CFS are organized according to a hierarchy for a categorical set of freight modes including the main modes of truck, rail, water, air, and pipeline, and a category to distinguish multiple mode shipments. Modal data starts with noting the mode reported by the establishment. If the shipment mode is not logical, then the mode is imputed. To aid this process, there are checks that ensure that shipments fall into typical use. Shipments outside these established bounds are scrutinized.

This breakout session explored methods to further distinguish existing mode categories and to establish new and emerging modes that may be included in future CFS. With changes in the economy and supply chains, such as the continued growth of e-commerce, new mode categories may better describe the changing freight mode landscape. For example, the multiple mode category does not necessarily provide the detail that is needed to fully understand e-commerce shifts in freight mode use. Data users are asked, would different modal categories be more helpful?

Discussion

The key themes coming out of this breakout session were:

1. There is a need to allow flexibility to collect and report on emerging modes since it is a challenge to define new mode penetration into the freight market at this time.
2. The CFS reporting structure impacts how emerging modes can be considered and tracked over time.
3. The highlighted emerging modes included drayage, drones, and e-commerce and micro-freight deliveries.

Flexibility in Selecting Emerging Modes

Participants commented that it is a challenge to define new modes for future CFS given the time horizon (5 years) between surveys and the dynamic changes to freight movements resulting from unanticipated events (like pandemics). Participants commented that because the future of freight is dynamic and reactionary, instead of defining new or emerging modes at the onset of the CFS, it would be beneficial to define processes that make it easier to integrate new modes into future CFS.

One participant suggested that the challenge in recognizing freight modes lies in the need to define what constitutes a freight shipment given emerging behavioral trends in online shopping. The example was provided of new grocery and restaurant delivery services that grew and emerged during the pandemic. In the National Household Travel Survey (NHTS), transporting a pizza or groceries from the store to home was considered household trip. However, if the pizza or groceries are now transported by a third-party provider like Uber or Postmates, they may be considered a form of micro-freight. Participants reflected on this definitional issue and how it may influence what modes should be considered and distinguished in the CFS. A participant from BTS commented that the survey framework can consider any number of new modes given there is enough activity to report on and maintain nondisclosure requirements.

CFS Reporting Structure

Participants noted that the issue of naming and recognizing new and emerging modes in the CFS is tied to the challenges associated with the CFS reporting structure. For example, Amazon, Walmart, Target and other large companies may offer for-hire services in the future. This could be a challenge to capture with the current CFS sampling methodology since these large companies are not currently considered shippers and are thus not necessarily sampled in the current CFS. However, participants noted that future sample frames may need to address non-traditional shippers to capture a comprehensive representation of shipments. One participant suggested that the VIUS may be able to provide data to address this challenge.

The CFS reporting structure related to organization of modes may also be influenced by emerging modes. For instance, the moderator posed the question of whether multiple (rail-truck, rail-water) or individual (truck, rail, water) modes should be changed to accommodate emerging modes. One participant stated that changes to the multiple mode aggregation may not provide the level of detail needed to track shipments at the level required for state, regional, and local-level planning studies. Further, participants commented that tracking changes in mode shifts over time resulting from emerging business practices and behavioral changes is difficult since the level of modal resolution in the CFS is low. The example of parcel carrier mode distinction was provided. Parcel is considered a service, and not necessarily a mode, however it is included in the CFS reporting structure. Participants commented that there is likely a need to redefine parcel shipments in light of emerging and growing prevalence of e-commerce. For example, Amazon has redefined e-commerce and blurred the lines distinguishing parcel shipments.

Finally, when asked about willingness to make trade-offs among commodity, mode, and geography, a few participants suggested that the CFS should continue to focus on commodity detail over other dimensions (geography and mode). Participants recognized that CFS provides national level estimates well across all dimensions and that users should disaggregate the data as needed to expand any one dimension.

Possible Emerging Modes to Consider

Emerging modes mentioned by participants included:

- Drayage
- Land-drone deliveries, e-bikes, and golf carts ("micro-modes")
- E-commerce delivery systems

The question was posed as to the threshold that users would apply to distinguish truck drayage within a multiple mode shipment. A participant from BTS suggested that although truck drayage is known to be over shorter distances, there is no specific mileage threshold to qualify truck drayage, thus presenting a challenge in reporting this mode. The current CFS defines drayage to be between the zip code of the origin or destination to the dock or railhead. One participant noted that drayage is important to the Environmental Protection Agency (EPA).

Land-drone deliveries, e-bikes and golf carts were mentioned in the discussion and generally referred to as "micro-modes" systems; although a proper definition of "micro-modes" was suggested by participants as an important need. Micro-modes were noted as emerging in places like Oregon. One participant highlighted the issue for Oregon and suggested that CFS create a framework to define and integrate new modes into CFS.

E-commerce delivery systems were a consistent focus of conversation in the session. Some participants noted the pervasiveness of Amazon shipments and their potential impact on transportation system activity. The prevalence of e-commerce will affect the types of modes identified in the CFS (individual modes versus multiple modes versus parcel shipments), and the scope of establishments to include in the survey. As an example of the potential change in transportation system activity, and as a suggestion on how important it may be to capture e-commerce modes, a participant highlighted the change in package return services. It was stated that e-returns are three times more frequent than in-store returns. However, this data is not captured in the CFS.

Plenary Panel
Breakout Debriefing Discussion and Next Steps

KATHLEEN HANCOCK
Virginia Tech
Moderator

BRIAN BONNER
Bureau of Transportation Statistics
Recorder

In the Plenary Panel session, moderators from each breakout session reported on the highlights of participant discussions. This was followed by a question and answer session with participants and representatives from BTS and Census. Consistent across all sessions was the mention of e-commerce and its impacts on the desired structure of the CFS mode, commodity, and geography dimensions. Participants noted that the prevalence of e-commerce would affect mode detail specifically in how parcel modes are reported, and in commodity classification specifically in distinguishing retail. Further, there was a general desire to better understand network, corridor, or sub-state region flows and vehicle movements. To do so with CFS data, it was recognized that commodity and mode detail may need to be forgone. The following summarizes the discussion during the Plenary Session.

BREAKOUT DEBRIEFING DISCUSSION

Panelists

- Scott Drumm*, CPCS*
- Dan Smith*, The Tioga Group*
- Paul Bingham, *IHS Markit*

Commodity

The desire to obtain commodity flow stems from the need to understand vehicle flows and their impacts on transportation infrastructure. Participants supported the idea of structuring commodity classification to distinguish vehicle or vessel operational characteristics. The current commodity structure reported as two- or three-digit SCTG codes is somewhat adaptable given the introduction of a machine learning model to (re)classify to alternate schemes and groupings. While SCTG code classification has drawbacks (namely, not necessarily matching with vehicle operational patterns), it is translatable to other common schemes like STCC or NAICS. Participants noted that crosswalk tables between existing schemes (e.g., SCTG) and any alternative scheme of future CFS, would be valuable. There was a desire to leverage the future VIUS to better align commodity groupings with vehicle operational patterns.

Trade-offs among the three CFS dimensions (commodity, mode, and geography) were expressed in regard to a desire to obtain origin-destination data for shipments. Detail on both ends of a trip was requested. It was suggested that “rolled-up” (aggregated) commodity groups

would be acceptable if detailed origin-destination was available. Lastly, participants expressed a need for value of a shipment in addition to the quantity (tonnage) shipped.

Geography

The current CFS produces state-level data with some reporting for larger MSA regions. Participants noted that county-level data is desired. Most uses of CFS products tended to be for sub-state analysis, with many participants suggesting their own methods for state to sub-state disaggregation. County-level data had been used for project prioritization based on truck usage, for instance.

Trade-offs for more aggregate commodity and mode distinction were acceptable to obtain higher granularity in geography. However, it was noted that there is a lower bound on commodity aggregation such that enough commodity detail would still be needed to link it to vehicle operations and movements. Participants suggested the use of alternative data sets like truck GPS data to supplement gaps in our understanding of vehicle movements. It was noted that in the future it would be beneficial to integrate GPS and CFS data to understand physical patterns of commodity flows.

Mode

Reorganizing the mode structure in the CFS to capture e-commerce flows and parcel mode shipments was discussed. Some participants expressed concern that the current single and multiple mode structure in CFS may not adequately depict increases in e-commerce flows making it difficult to track changes over time. Likewise, drayage operations were noted as a potential mode to consider; however, questions remain as to what constitutes a drayage trip from a multiple mode shipment. When discussing new and emerging modes, a number of participants were generally supportive of the idea of creating clearer definitions for what represents a freight mode and less emphasis was placed on listing specific future modes. Specific emerging modes mentioned included micro-modes like e-bike and drone deliveries. In this context, participants suggested that larger trends in micro-modes, large retailers serving as their own shippers, and e-commerce have the potential to shift the ways in which the CFS classifies shipment modes and creates a sampling frame. For instance, the scope of the sampling frame may need to capture large online retailers like Amazon which may be considered shippers.

NEXT STEPS

CHA-CHI FAN
Bureau of Transportation Statistics

BERIN LINFORS
Census

This workshop brought together users from industry, consulting, public agencies, and research institutes for a lively discussion on the uses of current and future needs for commodity flow data. Now is an interesting time for freight data users and producers. In this workshop, data users

shared their views on expected changes in freight operations considering the global pandemic, increases in e-commerce, and shifts to electronic reporting. Meanwhile, data producers are looking for ways to improve the collection and dissemination of the CFS. This discussion also extended to FAF and the upcoming VIUS, which will serve to fill key data gaps noted by users.

Through question and answer sessions and breakout groups, the BTS data producers noted preferences expressed by data users for adaptations to commodity classification schemes, finer geographical resolution, and the inclusion of new modes by restructuring mode classification schemes. BTS data producers were able to share methodological improvements made in the 2017 CFS through machine learning applications for commodity classification and new methods to produce small area estimates.

Looking forward to the 2022 CFS, participation and direction from data users is key to defining effective release schedules for data products such as the preliminary data release. BTS needs insight from users on how the preliminary data release is used, how it can be more useful, and how it should be changed. These needs are taken into account and balanced against fiscal constraints for future CFS products.

CONCLUDING REMARKS

CATHERINE LAWSON
University at Albany, Conference Chair

The Workshop Planning Committee thanks participants, panelists, and organizers for contributing to a synergistic discussion on CFS data products and its uses. This workshop format as a virtual meeting had a different look and feel as compared to prior CFS Workshops but had the same high level of participation. As data users, it is important to continue to communicate with BTS and Census to provide expert feedback on how the data is used and how each dimension of the data can be improved.

There are many avenues for following up with the conversations started today including the TRB Annual Meeting and the 2021 Innovations in Freight Data Workshop. Through TRB and the Standing Committee on Freight Transportation Data, users are encouraged to develop and submit Research Needs Statements and work with the committee as members and friends to shape the Annual Meeting sessions and workshops on these important topics. Based on conversations today, participants are eager to get access to the VIUS which will be a topic of discussion at the upcoming TRB Annual Meeting and the Freight Data Workshop.

It is anticipated that the impacts of the Covid-19 pandemic will last a while with varied degrees of effect on freight vehicle and commodity flows. Data like that generated by the CFS and VIUS will help generate recovery solutions. With global pandemics, the impacts on freight are far reaching. As data users, this can be seen as an opportunity to find new partners for freight data applications. The Federal Emergency Management Agency (FEMA) has the potential to be such a partner. Applications such as Hazard Mitigation Plans can benefit from the expertise of freight data users and from freight data related to designating critical infrastructure by its relevance to freight movements.

Thank you to all those that joined in the discussion and will continue to join in critical discussions that help to shape freight data and its many applications.

APPENDIX

Conference Program

Welcome and Workshop Purpose

Date: 9/24/2020
Time: 11:00 AM to 12:00 PM ET
Location: Online

Moderated by Catherine T. Lawson, University at Albany, Conference Chair
Session Description:
The initial, brief session will introduce the virtual workshop, its purpose, and deal with any "housekeeping" issues relating to the virtual format. (11:00 – 12:00)

Workshop Vision and Expectation
Presenter: Patricia Hu, Bureau of Transportation Statistics (10 minutes)

Review of 2017 CFS
Presenter: Julie Parker, Bureau of Transportation Statistics (20 minutes)

CFS Workshop Questionnaire Summary Findings
Presenter: Cha-Chi Fan, Bureau of Transportation Statistics (15 minutes)

Live Q&A Audience (15 minutes)

2022 and Beyond: Prioritization Strategies for Moving Forward at Top Speed

Date: 9/24/2020
Time: 12:00 PM to 12:30 PM ET
Location: Online

Moderated by Alison Conway, City College of New York
Session Description:
This session focuses on the decisions that will need to be made for the next CFS deployment, including minimum requirements, data collection and product trade-offs, user needs, and potential solutions from newly developing data science. (12:00 – 12:30)

Minimum Requirements of Future CFS And FAF
Presenter: Rolf Schmitt, Bureau of Transportation Statistics (10 minutes)

Trade-off: Commodity, Mode, Geography
Presenter: Chester Ford, Bureau of Transportation Statistics (10 minutes)

What Users Need Next And How To Meet The Needs
Presenter: Bill Eisele, Texas A&M Transportation Institute (10 minutes)

Minimum Requirements- Commodities (Breakout Session 1)

Date: 9/24/2020
Time: 12:45 PM to 1:45 PM ET
Location: Online

Moderated by Scott Drumm, CPCS
Recorder: Ja Lee/Chat monitor
Session Description:
The breakout session has multiple levels of SCTG commodity codes that are commonly used in transportation planning and forecasting. As commodity detail increases, compromises may have to be made in geographic and modal specificity. Breakout 1 will examine how commodity information is used and reliability at finer levels of commodity, geographic, and modal detail.

Minimum Requirements- Geography (Breakout Session 2)

Date: 9/24/2020
Time: 12:45 PM to 1:45 PM ET
Location: Online

Moderated by Daniel Smith, The Tioga Group
Recorder: Kelsey Ellenwood
Session Description:
This breakout session will focus on the various levels of geographic detail from states down to zip codes, how they may be aggregated for different purposes, and the trade-offs between geographic, commodity, and modal specificity.

Minimum Requirements: Emerging Modes (Breakout Session 3)

Date: 9/24/2020
Time: 12:45 PM to 1:45 PM ET
Location: Online

Moderated by Paul Bingham, IHS Markit
Recorder: Dan Biggio
Session Description:
The emergence of new modes and the evolution of existing modes pose challenges for collecting processing and using CFS data. Breakout 3 will concentrate on issues surrounding modal definitions, particularly distinctions between trucking sectors and the classification of multimodal and parcel shipments.

Plenary Panel: Breakout Debriefing Discussion and Next Steps

Date: 9/24/2020
Time: 2:00 PM to 3:00 PM ET
Location: Online

Moderated by Kathleen Hancock, Virginia Tech

Recorder: Brian Bonner
Session Description:
This Debriefing Discussion will recap the three topics from the breakout session. The discussion is anticipated to address trade-offs, cost impacts, and the ability to accomplish future goals. (2:00 – 3:00)

Panelists (5 minutes each): Scott Drumm, CPCS; Dan Smith, Tioga Group, Inc.; Paul Bingham, IHS Markit (15 minutes)

Panel Discussion with Audience (30 minutes)

Next Steps (5 minutes each): Cha-Chi Fan, BTS; Berin Linfors, Census (10 minutes)

Concluding Remarks: Catherine T. Lawson, University at Albany, Conference Chair (5 minutes end at 3:00)

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, non-governmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. John L. Anderson is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at **www.nationalacademies.org**.

The **Transportation Research Board** is one of seven major programs of the National Academies of Sciences, Engineering, and Medicine. The mission of the Transportation Research Board is to provide leadership in transportation improvements and innovation through trusted, timely, impartial, and evidence-based information exchange, research, and advice regarding all modes of transportation. The Board's varied activities annually engage about 8,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation.

Learn more about the Transportation Research Board at **www.TRB.org**.

TRANSPORTATION RESEARCH BOARD
500 Fifth Street, NW
Washington, DC 20001