TRANSPORTATION RESEARCH

CIRCULAR

Number E-C088 January 2006

Commodity Flow Survey Conference

July 8–9, 2005
Boston, Massachusetts

TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

TRANSPORTATION RESEARCH BOARD
2005 EXECUTIVE COMMITTEE OFFICERS

TRANSPORTATION RESEARCH BOARD
2005 TECHNICAL ACTIVITIES COUNCIL

TRANSPORTATION RESEARCH CIRCULAR E-C088

Commodity Flow Survey Conference

July 8–9, 2005
Boston, Massachusetts

Kathleen Hancock, *Editor*
Virginia Polytechnic Institute and State University

Transportation Research Board
Freight Transportation Data Committee and
Travel Survey Methods Committee

January 2006

Transportation Research Board
500 Fifth Street, NW
Washington, DC 20001
www.TRB.org

TRANSPORTATION RESEARCH CIRCULAR E-C088

ISSN 0097-8515

The **Transportation Research Board** is a division of the National Research Council, which serves as an independent advisor to the federal government on scientific and technical questions of national importance. The National Research Council, jointly administered by the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine, brings the resources of the entire scientific and technical communities to bear on national problems through its volunteer advisory committees.

The **Transportation Research Board** is distributing this 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 Circular was taken directly from the submission of the authors. This document is not a report of the National Research Council or of the National Academy of Sciences.

Transportation Research Board
500 Fifth Street, NW
Washington, DC 20001
www.TRB.org

Ann E. Petty, Production Editor; Jennifer Correro, Proofreader and Layout

Preface

The Commodity Flow Survey (CFS) is a key data source for a myriad of freight planning activities. Two Transportation Research Board (TRB) standing committees, the Committee on Freight Data and the Committee on Travel Survey Methods Committee, initiated a conference to understand the survey better, see how other data sources are being used to supplement it, and explore possible improvements to future iterations of the survey. The conference, convened 6 months after the release of detailed CFS data and coinciding with planning for the 2007 CFS conference, offered an interactive format for a diverse set of users to engage in productive dialogue.

An ad hoc committee, chaired by Arnim Meyburg of Cornell University and selected by the sponsoring committees, carried out the detailed planning for the conference over a year-long period. This circular consists of individually attributed papers and summaries. No language should be construed as consensus findings or recommendations on the part of conference, the planning committee, or the sponsoring committees.

The planning committee represented CFS producers, analysts, and modelers. The 106 persons attending reflected organizational diversity as follows:

U.S. Department of Transportation	24%
Census Bureau	9%
State government	16%
Consultant and private sector	12%
University	15%
Other	23%

The Bureau of Transportation Statistics provided funding to support travel and on-site expenses. The conference was cosponsored by Federal Highway Administration, the United States Census Bureau, and the American Association of State Highway and Transportation Officials.

Contents

APPENDIXES

Introduction

KATHLEEN HANCOCK
Virginia Polytechnic Institute and State University

ARNIM MEYBURG
Cornell University

Information resulting from the Commodity Flow Survey (CFS) provides a cornerstone for existing freight planning activities. As a national survey subject to limited resources, the CFS has been pushed to its limits by stakeholders attempting to meet their national, regional, state, and to some extent, local freight planning needs. This conference was designed to engage these stakeholders in a dialogue to understand the survey better, see how other data sources are being used to supplement it, and explore ideas to improve its future iterations. Toward this, the conference was divided into three types of interactions:

1. Presentations and papers from key stakeholders intimately involved with the CFS,
2. An opportunity to see current applications of the CFS and interact with the people involved, and
3. Interactive workshops focused on issues associated with the existing CFS and potential improvements to both the current uses of and future design for the survey.

Rational decisions have to be based on good information and data to be effective and efficient. It is clear that transportation in the U.S. economy represents multifaceted and complex interactions between public and private sector concerns. It is critical for the functioning of the U.S. economy and, in a broader sense, for the support of our lifestyle.

Hence, reliable and relevant data about this sector are crucial for public and private investments and planning in support of freight movements. Ideally, a national freight transportation policy encompassing local, state, and national concerns would be a desirable framework and should be an objective for the freight community. Such a policy would undoubtedly give rise to a comprehensive freight data collection plan in support of such a policy, as well as to a set of standards for supplemental data collection efforts.

In the absence of a comprehensive freight data collection plan, every data collection effort is likely to have serious gaps. Therefore, it is necessary to work with what is available and attempt to improve current efforts. The objectives of this conference follow this approach to freight data collection and use.

The focus of the conference was to engage in dialogue, not to arrive at consensus or recommendations. Therefore, the following summary captures individual ideas from the participants that recurred during the conference. Detailed information about the topics for the conference is provided in the workshop resource papers and more detail about discussions from conference participants is provided in the workshop summary sections.

SCOPE, COMPARABILITY, SHIPMENT CHARACTERISTICS, AND SPECIAL MEASUREMENT ISSUES

A fundamental issue for effective use of the current CFS is understanding the gaps that exist in the data. These gaps result from commodities, services, or both that are out of scope; from ambiguity, uncertainty, or inability of the current CFS to capture desired information, and the inability of Census to release information because of statistical rigor or private sector confidentiality issues.

Another issue is the compatibility of the CFS internally across time and externally across other data sets. Changes to the CFS between surveys have created opportunities for different uses but have also resulted in different commodity and industry classification systems and different geographies. A good, well-documented crosswalk between surveys could provide an important tool to link information over time. Because the CFS is a survey of shipments and many other surveys are carrier surveys, much of the information is incompatible across data sets without major efforts to resolve issues such as differing commodity coding schemes, differing reporting of imports, exports and intermodal shipments, and differing geographies. Coordination between agencies performing the various surveys could reduce these incompatibilities.

The CFS was never conceived as providing all information to all users. Therefore, most stakeholders will need to combine data from the CFS with other data sources. Several factors necessary to make this possible were discussed including identification of data elements that would allow merges; identification of methodologies to account for double counting, overestimating, or both; development of techniques to use origin–destination data to perform links; and development of techniques to capture through-shipments. Another interest was establishing methods for accessing and mining information from private-sector data while maintaining the security of that information.

IMPROVING CFS DATA PRODUCTS

Data products for the most recent 2002 CFS have recently been released, and these, along with data products from the 1993 and 1997 surveys, have been used by stakeholders to address their needs. One of the discussions at this conference was to explore possible improvements or additions to these products that could make them more useful. Ideas included developing origin–destination tables emphasizing more geographic and industry data, reverse geographic flows and additional aggregations; a loaded network product that would identify freight corridors and trade and market areas; and trip generation and trip distribution variables to facilitate modeling of missing data.

ENHANCING SURVEY METHODS

Planning for the 2007 CFS includes the following goals: reduce survey costs through more efficient sampling and data collection; achieve larger sample size at finer levels of geography to improve utility for a wider range of users; and mitigate respondent burden to improve response rates. Discussion from participants focused on sampling methods, instrument design, mode of collection, nonresponse, and future design considerations.

IMPROVING THE USE AND ACCESSIBILITY OF THE 2002 CFS

To ensure the continuation and support for a major freight data survey, increasing the number of CFS data users and the number of data applications is critical. Conference participants discussed alternative data formats, increased awareness of the data, improved access to the data, the expansion of the data through data fusion, and incentives to encourage more creative and extended uses as methods for accomplishing this goal.

KEY OBSERVATIONS FROM THE CONFERENCE

The CFS, by itself, is necessary but not sufficient for a national freight information program. It is one part of the larger national freight data picture. Previous TRB special reports have detailed this larger need and a framework for improvement, but so far, little has occurred. If we are to make informed decisions, a necessary foundation such as the CFS should be acknowledged by the policy and planning community as an important and necessary foundation.

How CFS Fits in the World of Freight Data

ERIC C. PETERSON
U.S. Department of Transportation

I represent a new agency, the Research and Innovative Technology Administration, or RITA. RITA is the U.S. Department of Transportation (USDOT) administration dedicated primarily to research, development, and technology coordination and management and was created last year in a bill proposed by Transportation Secretary Norman Mineta, passed by the House and Senate, and signed by President Bush. They are all on record supporting innovative solutions to transportation challenges, including the statistics program. RITA is the new home of the Bureau of Transportation Statistics.

In addition to its research focus, RITA provides research and analytical support to the secretary. RITA helps the secretary understand major trends, anticipate future conditions, and analyze major issues. As a consequence, I can speak as both a provider and a user of transportation data.

High on the list of major issues today that we're focused on is freight. It is clear from both empirical and anecdotal data that not only are freight volumes increasing, the type of freight and the way it moves also are changing as our economy changes.

Transportation moves the American economy. We have a strong economy—low unemployment at the same time new jobs are being created. But we must continue to look to the future and be prepared to face what President Bush calls "the challenges of a rapidly changing economy."

And the CFS can help us prepare to meet those challenges that the president talks about. The CFS documents the way transportation moves the economy today, and it provides us information that can help guide research and investments that will equip our national transportation system to handle the demands of the economy of tomorrow.

We face congestion challenges; FHWA and the Texas Transportation Institute (TTI) say congestion lasts longer, affects more roads in more places, and creates more extra travel time than in the past. TTI estimates that congestion costs us $63 billion a year in travel delays and wasted fuel, not counting the impacts on America's economy and general well-being.

In our supply chain network, even the most conservative forecasts suggest that overall freight volumes will grow by 70% by the year 2020. Much of that growth is for just-in-time deliveries of lighter and more valuable goods, which require fast and reliable delivery. International trade, which accounted for only about 8% of our gross domestic product in 1970, accounts for almost 30% today and continues to rise, putting new strains on our transportation network.

So Secretary Mineta has set "capacity and congestion" as one of the Department's top priorities for the coming years. Our challenge is to focus our freight data program, including the CFS, on providing the information to make informed decisions on meeting the demands that a growing economy places on the transportation system.

Information on the flow of commodities among regions and along major intercity transportation links is essential to understanding key trends and issues such as growth in freight transportation activity throughout the United States and the pressures created by that growth on the nation's transportation systems; patterns of merchandise trade with domestic and international partners and the economic growth potential associated with that trade; volumes of traffic passing through a location between distant origins and destinations, indicating the effects of external traffic on local transportation facilities and the importance of local facilities to distant places; markets served by different modes of transportation and intermodal combinations; locations exposed to risks of hazardous materials incidents and other safety aspects of freight transportation; energy use and environmental consequences of freight transportation; efficiency and productivity of logistical systems supporting the nation's economy; and likely impacts of transportation policies on efficiency, economic productivity, safety.

As Abraham Lincoln said on the subject of federal funding for internal improvements:

> Statistics will save us from doing what we do, in the wrong places.... The surplus, that which is produced in one place to be consumed in another; the capacity of each locality for producing a greater surplus; the natural means of transportation, and their susceptibility for improvement; the hindrances, delays, and losses of life and property during transportation, and the causes of each, would be among the most valuable statistics in this connection.

The CFS has been a main source of freight data for the USDOT since the program began in the 1990s and will serve as a foundation for our future efforts in freight data. At this conference and in the coming weeks, we must ask how our freight data program is supporting the needs of the secretary and other decision makers in the federal government, in state governments, and in local transportation agencies. Are we giving them the information they need? The CFS is central to our understanding. It is the only source of freight data that attempts to cover all modes of transportation and intermodal combinations. What we in USDOT expect from the CFS is statistically sound data on what commodities are being shipped, by what mode or intermodal combination are they being shipped, and where they are going.

We understand the many challenges of obtaining this information from a huge and diverse economy. We are asking for information from small businesses and industrial giants, from producers of paper clips and power plant generators, and from local wholesalers to players in the global economy. We care about the shipments of small businesses and local establishments because they play a significant role in the economy and place significant demands on the transportation system when added together.

We expect to understand from the CFS how far shipments are sent, whether across town or across the continent. The distribution of commodities and mode choice by distance is valuable for modeling and policy analyses.

We expect to understand from the CFS how much moves by what means between states and major metropolitan areas, but not for small areas. If we include 40 kinds of commodities and six modes and intermodal combinations, we need to fill a matrix of over 3 million cells to measure flows among the 114 regions used in the CFS. If we tried to estimate flows among the 3,141 counties and county-equivalent jurisdictions, they would need to fill a matrix of more than 2 billion cells. The sample requirements and confidentiality issues involving such a large matrix are daunting.

The CFS does not do it all. You will learn later in this conference how the CFS is linked with other data sources to estimate flows not covered by the CFS, make provisional estimates of flows for the current year, and forecast future freight activity. You will learn how USDOT and others deal with geographic detail and other limitations of the CFS.

With this understanding, we can work together over the next 2 days to make the CFS as effective as possible to meet our current and future information needs. Our goal is to make the CFS a solid cornerstone, doing well the things it can do best, and assuring that we can link CFS results with other data sources to obtain the complete picture of commodity flows and freight activity in the country.

We should consider improvements to the CFS in the context of a broader freight data program. The question is, if we do a better job within our current structure, are we providing the secretary and other decision makers with the information they need? Or do we need to do something different? To this end, we are working with our partner, the Census Bureau, to conduct research on the CFS and identify improvements that can be made to the upcoming 2007 CFS.

Some things to think about: Are we identifying major, long-term and emerging trends, especially related to global trades? Are we giving decision makers the right data in useful forms and in a timely fashion? Are we doing the most we can with the numbers we have? Are we doing the right kinds of analyses? How can we best combine national data programs with local understanding so that we can truly "think globally and act locally"?

We have huge task ahead of us. We all must work together to get this task accomplished. Our new agency is here to work with you and support you. We are reaching out to promote collaboration throughout our transportation community. We are working as partners and we want to hear your ideas. Let us know where you believe RITA can make a difference for you.

Secretary Mineta said, "Understanding the role freight plays in our economy is crucial if we are going to sustain today's fast-growing economy in the years ahead." Fortunately, people have tracked freight movements for centuries. If they hadn't, we wouldn't know about the rum, fish, salt, and tobacco that moved through the Port of Boston in the 18th century. But never has tracking been done in such a complex world with different modes and different combinations of modes carrying so many different kinds of freight to so many different places.

All this freight movement takes place in the rapidly changing environment of the 21st century. Our transportation world today is far different from the world that existed when the freight railroads, the nation's ports, and even the Interstate Highway System were built. From 1992, when work began on the first CFS, transportation has changed tremendously. The 2007 CFS should be modified and adjusted to the extent that it is practicable, to capture and reflect these changes.

We have the opportunity to add to the understanding of freight's role that Secretary Mineta said is needed. Let's work together to make sure we don't miss that opportunity.

Commodity Flow Survey Plans for 2007

DEBORAH JOHNSON
Bureau of Transportation Statistics

An old saying: If you don't know where you're going, any road will take you there. Where we're going is partly a function of where we've been. Here is a brief history of U.S. commodity surveys.

A predecessor survey series was far more limited than the current Commodity Flow Survey (CFS). The Commodity Transportation Survey (CTS) consisted of fewer industry sectors (only manufacturers) and limited shipments (about 20,000 establishments). The CTS was last published in 1977. After the last CTS was published the United States entered into an era of deregulation. A majority of the economic regulatory data were scaled back dramatically. This left a significant gap in freight statistics for about 15 years.

The Bureau of Transportation Statistics (BTS) was created December 18, 1991, to fill some of the major gaps in transportation data. One of the major reasons for the establishment of BTS was to address the lack of freight flow data. BTS and the Census Bureau partnership began in 1992. This partnership made sense because Census

- Had past experience in conducting CFSs and other establishment surveys;
- Had an establishment list for constructing a frame and sample;
- Has mandatory reporting authority (as component of economic census); and
- Had requisite resources in house for conducting a major establishment data collection effort.

Objectives of the CFS are to

- Estimate shipping volumes by commodity and mode of transportation at various levels of geography and
- Measure flows of goods between states and regions by mode and commodity.

Coverage includes

- Manufacturing, mining, wholesale, auxiliaries, and selected retail, and
- All modes, including private and for-hire trucking.

In general, the objectives of the CFS are limited by sample size. We would need an incredibly large sample (probably larger than the entire BTS budget would allow) to provide low-level geographic estimates or very detailed flow data.

Coverage includes industry sectors more likely involved in shipping activities. Some establishments or sectors are excluded because of other operational constraints in data collection (e.g., military). Other gaps in coverage include imports, most retail, and agriculture to the point of retail.

For the most recent CFS conducted in 2002, 50,000 establishments were sampled (of approximately 760,000 eligible CFS establishments on the Business Register). More than 3 million establishments are on the Business Register in total (most are not shippers or are otherwise out-of-scope for CFS).

The goal was to sample between 20 to 40 shipments made during the establishment's designated reporting week, 1 week per quarter. If an establishment shipped fewer than 40 per week, we took all of the shipments. Otherwise, the respondent was to select a sample per the form's instructions.

In 1993, we had very limited time for research and development before fielding. In 1997, there was more time so research was conducted and modest changes were made. Because of budget uncertainties in 2002, a very limited amount of time and resources were available to conduct research and make changes.

Examples of improvements BTS and Census made in the CFS were

- Sample design:
  1. Created measure of size estimate, improving stability for sampling establishments, and
  2. Identification of certainty establishments based on specific modes using prior survey reports.
- Sampling instructions: Created better instructions to ease respondent burden and help alleviate errors.
- Processing improvements**:**
  1. Use of Docuprint to print forms and materials and
  2. Use of imaging of returned forms.
- Editing improvements:
  1. Increased review and analysis of results and
  2. Additional edits such as quarter-to-quarter checks and value range checks.
- Reporting period: Shortened from 2 weeks to 1 week to reduce respondent burden.
- Decreased sample size:
  1. Reduced the sample size from 200,000 establishments in 1993 to 100,000 in 1997 to allow for more intensive follow-up and ensure higher quality data and
  2. Reduced to 50,000 in 2002 on the basis of budgetary limitations.
- Questionnaire design: Replaced Y/N check off for hazardous materials with UN/NA number.

From the 2002 CFS data, a variety of reports have been or will be produced such as

- United States Report (December 2004),
- Hazardous Materials Report (December 2004),
- 50 State Reports (December 2004),
- Export Report (March 2005),
- Metro Areas/Rest of States (June 2005),
- CD (all tables and analysis tool) (July 2005),
- State Summaries Report (preliminary release July 2005), and
- Freight Shipments in America Report (January 2006).

Most of these products can be ordered on the BTS website. The BTS website also includes downloadable tables and other information. We look forward to comments in the data products workshop on improvements to these products, and suggestions for additional products.

The CFS methodology has gradually evolved and improved over the past few surveys. However, more could and should be done, especially in light of the significant changes that have taken place in the transportation industry and the world in general.

Now is a critical time for reassessment. Here are some of the reasons why this reassessment is necessary:

- Globalization of the world's economy (China and European Union),
- Downsizing of domestic production establishments,
- North American Free Trade Agreement and now the Central American Free Trade Agreement,
- Consolidation of railroad and trucking industries,
- Contracting out of transportation functions—sharp growth in third-party logistics providers, and
- Technological advancements, including the Internet, Global Positioning System, and radio frequency identification device tags, and change in how cargo is tracked and processed.

BTS is committed to continuing to improve the CFS. BTS previously commissioned the Transportation Research Board (TRB) and the Committee on National Statistics to form an expert panel of transportation specialists, statisticians, and survey experts to review the CFS (and other BTS survey programs). There were four main recommendations:

1. Conduct the 2007 CFS;
2. Initiate a research program to investigate survey methods;
3. Solicit user input into the design process; and
4. Reevaluate roles and responsibilities of BTS and Census to use expertise and experience effectively.

These recommendations are taken seriously. We have worked with TRB and its committees to sponsor this conference and are looking forward to hearing your feedback and getting your input during the next 2 days.

Here is where we are now and what we've got planned over the next 18 months:

- Finalize industry coverage (November 2005);
- Initiate questionnaire pretest (December 2005);
- Begin precanvass operation (April 2006);
- Submit Office of Management and Budget (OMB) clearance (June 2006);
- Construct sample frame (September 2006);
- Start initial mailout (December 2006); and
- Begin data collection (January 2007).

These dates are not set in stone and might vary slightly, but this provides a broad timeline of when major decisions must be made and key operations started. The OMB date is significant since forms, instructions, and procedures need to be "near final" by this date.

BTS and Census have evaluated lessons learned and built on those experiences to develop research areas. These include forming joint investigative teams (JITs) to perform the research. The JITs are researching such areas as scope and industry coverage, questionnaire and data collection, electronic reporting options, mileage calculations, precanvass, and a variety of other major areas that impact the 2007 CFS.

Here are highlights of some research areas that are already under way. What improvements can reasonably be made in coverage? This involves evaluation of coverage changes because of standard industrial classification (SIC) to North American Industrical Classification System (NAICS) conversion. In the SIC to NAICS conversion, the most prominent areas are logging, printed products, and fuel oil (propane), and research may identify others. We also need to be able to answer the question, How can we better identify intermodal shipments in the CFS?

The BTS and Census JITs have developed an internal assessment of questionnaire and instructions based on the 2002 experience and other indicators of issues such as, items with significant edit failures rates. Now the JITs are engaged with cognitive researchers in developing a protocol for testing and improving the form and procedures, as well as evaluating the feasibility and usability of a web reporting option.

In a precanvass survey, we are trying to maximize the benefit of this operation by assessing the most beneficial objectives and identify industry sectors of most need. Precanvass could yield better size information to improve sample selection process and to screen non-shipping auxiliaries; it could also be used to obtain contact information for large certainty establishments (with a large impact on results) to maximize response. Better measures of size information improves efficiency of first-stage sampling (i.e., the better this information, the better the selection process for choosing establishments) and ultimately improves final survey estimates).

Changes have been made to the third-stage sampling process, but it is always difficult when survey respondents are asked to perform the sampling operation. We are still looking at ways to improve this process to lessen burden on respondent and minimize errors caused by incorrect sampling procedures.

JITs are just one method to ensure that we have a balance of transportation expertise, statistical knowledge, and survey experience from both BTS and Census. For all critical operational areas, these teams are actively involved in research, recommendations, and implementation.

This brings us to the conference today. We are interested in hearing your suggestions for improvements for the CFS.

In addition to what happens here at this conference and in the various workshops, we would like to establish a continuing dialogue so that feedback does not stop here. Please continue to provide your comments to Michael Cohen at BTS at michael.cohen@dot.gov and John Fowler at U.S Census Bureau at john.l.fowler@census.gov.

Census Bureau CFS 2007 Planning Issues

THOMAS L. MESENBOURG
U.S. Census Bureau

It's a pleasure to have the opportunity to participate in this groundbreaking user's conference on the Commodity Flow Survey (CFS).

Let me start with just a brief overview of the Economic Census. The Economic Census is the single largest compilation of economic statistics undertaken by the federal government and second only to the Decennial Census in its scope and complexity. The Economic Census is a multifaceted program consisting of the core establishment-based economic censuses that cover all nonagricultural sectors of the economy and a series of related programs focusing on topics of national interest such as the CFS, the Vehicle Inventory and Use Survey, and the Survey of Business Owners.

The CFS, conducted jointly by the Census Bureau and the Bureau of Transportation Statistics, is an important component of our Economic Census program. While the CFS is unique in many ways, it leverages the underlying economic census infrastructure, uses the Business Register as its sampling frame, and takes advantage of the data we collect from individual establishments. Later, I will discuss how we are using some of the new information we collected in the 2002 Economic Census to help us improve the 2007 CFS, but first I want to briefly discuss some of the challenges all statistical programs will be facing over the next several years.

PROGRAM CHALLENGES

Securing adequate resources for statistical programs is going to be increasingly difficult. Over the past 3 years the Census Bureau has fared very well, obtaining new funding for the American Community Survey (ACS), full funding for the 2002 Economic Census data collection and processing, and additional monies for new service and e-business statistics. Unfortunately, it appears the future will not be as rosy. Based on the House and Senate mark for FY 2006, the Economic Census budget will probably take about a $3 million cut or about 5% of the request. This cut will not impact 2006 CFS planning activities. However, the Senate mark, if it prevailed, would terminate the ACS, kill the re-engineered short form Decennial Census, and eliminate seven existing current economic statistics programs. In FY 2007, the Economic Census will request about $84 million, $16 million, or almost 24% more than the likely FY 2006 appropriation. This is a critical year for the CFS since data collection will begin in December 2006, a full year earlier than the core census programs. Things will get even more difficult in FY 2008, the Economic Census collection year, when we will be requesting about $63 million more than the FY 2006 request and Decennial will be requesting about $350 million more than in 2006. Bottom line—we are going to have to develop plans that can respond to various funding scenarios. We also have to be proactive in selling these budget requests and one way we can do this is by being able to point to program improvements in the CFS.

The second challenge is ensuring our programs are relevant and improve even in the face of constrained resources. This means that the Economic Census and the CFS need to reflect our changing economy. For the 2002 Economic Census expanded content related to new measures of e-commerce, leased employees, expanded expenses data, new service product data, and first time collection of information on changing supply chain activities all played an important role in "selling" the census as relevant and forward thinking. Demonstrating that census program components such as the CFS are dynamic and reflect our constantly changing economy is going to be critical in selling our future budget requests. The results from this CFS conference can make an important contribution to achieving this objective.

A third challenge is to make sure our data requests align more closely with accounting conventions and companies recordkeeping practices. We can not afford to take the participation of businesses for granted. Careful review of 2002 CFS reporting patterns and characteristics, a fresh look at our report forms, cognitive interviews with individual businesses, and field testing of new inquiries are important areas of research that will help ensure businesses can comply with our data requests. So over the next 2 days as we consider various improvement opportunities, let us put ourselves in the shoes of a respondent and ask if we understand what is being requested and if this information would be available in establishment records.

IMPROVEMENTS TO THE 2007 CFS

Let me now discuss some of ideas BTS and the Census Bureau are exploring related to improving the scope and methodology of the 2007 CFS.

The shift from the antiquated, albeit long-playing standard industrial classification system to the new, conceptually based North American Industry Classification System (NAICS) was not a trivial or painless process for any statistical program. The 2002 CFS was no exception. A number of industries formerly included in the CFS, such as logging and publishing with printing shifted into sectors, namely agriculture and the information sector, traditionally not covered in the CFS. The NAICS classification of auxiliaries (establishments providing services to other establishments of a company) based on their primary activity rather than according to the activities of the establishments they supported also raised additional coverage issues. Finally, new boundaries for retail and wholesale trade caused additional complications such as fuel oil dealers moving to retail trade from wholesale. The good news is that we have two Economic Censuses under NAICS, and the changes for NAICS 2007 are going to be modest with minimum significant impact on the 2007 CFS.

The one area that we will not be able to address is logging which is now in agriculture and not surveyed in census or current surveys or reflected in the Business Register. Fuel oil dealers, and publishing are areas that we are evaluating carefully right now and I am very hopeful that we will be able to address most of the coverage issues related to these special cases.

For auxiliary establishments, we have a plan to identify those that provide shipping and transportation services to other parts of the company. On our Business Register—the source of the frame for the CFS sample—auxiliaries also include establishments that provide management, data processing, accounting and other services to their company. The Business Register doesn't include an indicator of whether the auxiliary establishment ships, so we are planning an advance mail out to these establishments to better identify their business activity and obtain a measure of their shipping magnitude. This information will permit much more efficient sampling of auxiliaries actually engaged in shipping goods.

A more complicated issue relates to American businesses adaptability and dynamism. Establishments increasingly are taking on additional activities, activities that very often are not captured by the industry code. In the 2002 Economic Census we added a new supply inquiry, Item 28, to report forms sent to manufacturers, wholesalers, retailers, and transportation and warehouse facilities. The check box inquiry included some questions related to local and long-distance delivery, warehousing, and logistics consulting that may be of interest to the CFS. Establishments were asked if they performed the activity, whether it was provided by another company, or if the activity was not performed at all. A word of caution: the numbers I cite are unweighted responses from those establishments that answered the Item 28 inquiries and the information has undergone only cursory review.

Here are some highlights related to the supply chain check boxes:

- In manufacturing of the 105,000 responses:
  - Almost 49,000 delivered locally; another 26,000 contracted it out.
  - About 31,000 did long-distance delivery; 44,000 contracted it out.
  - 32,000 warehoused final products at the establishment; 2,300 contracted it out.
- In wholesale trade of the 235,000 responses:
  - 125,000 did local delivery; 49,000 contracted it out.
  - 58,000 did long-distance delivery; 81,000 contracted it out.
  - 143,000 warehoused at this establishment; 26,000 contracted it out.
  - 29,000 did logistics consulting; 17,000 contracted for it.
- In retail trade of the 430,000 responses:
  - 142,000 did local delivery; 37,000 contracted it out.
  - 33,000 did long-distance delivery; 52,000 contracted it out.
  - 90,000 did warehousing at this facility; 30,000 contracted it out.
  - 23,000 did logistics consulting; 13,000 contracted for it.
- In transportation of the 50,000 establishments responding:
  - 21,000 did local delivery; 4,600 contracted it out.
  - 19,000 did long-distance delivery; 5,700 contracted it out.
  - 7,000 had warehouses, 1,700 contracted warehousing out.
  - 5,100 did logistics consulting; 850 contracted for it.

I find the supply chain data fascinating in that they point out the diverse activities being performed by establishments in different sectors and industries. I would counsel against drawing inferences from very aggregated and unweighted data, but they point out areas requiring additional exploration and research related to the coverage of the CFS. We are in the process of generating more detailed tabulations to help in CFS planning activities.

A second section of Item 28 asked a series of check box questions about who owned and managed the establishment's inventory and where the inventory was held. While 90% of manufacturing and retail establishments that responded indicated they owned and managed the inventory held at their location, only 72% of wholesalers owned and managed inventory at their establishment. New inventory practices where vendors own and manage inventory held at the customer's location is not only complicating inventory measurement but also making it extremely difficult to identify changing business practices. The increasing attention given to third-party logistics providers (3PLs) is a good example of these changing practices.

In fact Tuesday's *Wall Street Journal* featured an article on 3PLs. It is clear that firms are leveraging core competencies into entirely new business opportunities, but no one industry code captures 3PL providers activities. I had staff perform a quick tabulation of the industry codes of some of the companies featured in the article and we found a diverse set of codes, with most establishments of these companies coded in 4841, general freight trucking; 48,851, freight transportation arrangements; 492, couriers and messengers; 493 warehousing; and a handful in 541,614, process, physical distribution, and logistics consulting services where establishments provide operating advice and assistance. Given the diversity of industry codes for 3PLs, we face a huge challenge in identifying efficient ways of sampling this activity. This is an issue we are struggling with right now and we are looking at Economic Census product lines and in some cases kinds of business inquiries that can help us target major 3PLs. This is definitely an area where we are looking for some innovative approaches. Once we figure out methods for identifying possible 3PLs, next we will have to determine whether the 3PL can actually report CFS shipment, mode, origin-and-destination–related information or if this information available only from the client business. We will only be able to answer these questions through additional study of census data combined with a number of visits to 3PL providers and their customers.

IMPROVEMENTS IN METHODOLOGY

Let me give a quick overview of some of the methodological improvements we are researching. We have developed nearly a dozen Joint Investigative Teams.

As Deborah mentioned, one area we think we can substantially improve is our CFS survey instruments. While we have made incremental improvements along the way, our paper questionnaire, instruction guide, and commodity coding manual have never undergone formal cognitive testing. We have work already underway to accomplish this testing beginning later this year. A critical component of the testing involves face-to-face interviews with companies to discover answers to questions such as what steps do they take to complete our form? What kinds of problems did they encounter? And did they interpret our instructions the way we want them to? We will conduct multiple rounds of company interviews, with each round involving anywhere from 10 to 30 companies. In both the 2002 Economic Census and in our current surveys, cognitive testing of this kind has produced significant improvements in the quality of data responses, and reductions in respondent burden. We are looking forward to the testing results for our CFS materials with great anticipation.

In addition to improving our paper questionnaire, we also are investigating the possibility of providing an electronic reporting option for CFS respondents. Developing an electronic reporting option will require significant resources, so we need to ensure that electronic reporting will significantly lower respondent burden, increase response rates, and improve data quality. We plan to coordinate testing of an electronic reporting pilot instrument with the cognitive testing of the paper form. The results of the testing will help us decide if electronic reporting is appropriate for CFS respondents. Our experience from other programs indicate that electronic reporting will be embraced only if it provides significant value added to the businesses filling out the form.

KEYS TO SUCCESS

Today and tomorrow are all about exploring a wide-range of opportunities to expand and improve the CFS, but we also want to keep in mind the conditions that will help ensure a successful implementation of a survey the size and complexity of the CFS. As you can see from both Deborah's and my remarks, we have an ambitious and very full platter of things to accomplish in a relatively short period of time. After all, the first 2007 CFS forms will be mailed in December 2006, a mere 17 months from now.

To implement the improved 2007 CFS, I suggest that we keep in mind the following keys to success.

- Ensure we clearly understand each agencies roles and responsibilities. A document delineating roles and responsibilities for a whole host of activities is being developed.
- Concentrate on a handful, not an armful, of high payoff activities. As you have heard, we have a number of new improvement ideas we are actively exploring. Relatively soon we are going to have to winnow down this list. We need your feedback in determining what the highest payoff opportunities are.
- Better utilize existing information. We have a multitude of information from the 2002 CFS, the 2002 Economic Census, and possibly even some of our current programs that can help inform data-driven decisions related to both coverage and methodological improvements.
- Ensure data requests align with business recordkeeping practices and accounting conventions; test new inquiries; avoid collecting data on "rare" activity; use check boxes to identify changing business practices—embrace simplicity; resist complexity.
- Develop a road map or plan that mitigates the risk associated with reduced resources and changing economic conditions.
- Jointly establish milestones, measure progress, and hold staff accountable for meeting agreed upon dates.

CONCLUDING REMARKS

I want to thank the Transportation Research Board for their excellent work in creating this CFS data user conference. For all of those on the organizing committee, I know how much work it takes to set up the conference logistics, review papers, and all that it takes to prepare for 2 days here. All of us greatly appreciate the efforts of the planners, authors, discussants, and all the participants that will make this conference a success and we look forward to the seeing the results.

Workshop Resource Papers

Scope and Industry Coverage of the 2007 Commodity Flow Survey

RONALD J. DUYCH
Bureau of Transportation Statistics

PREFACE

This resource paper provides information and background material to facilitate a workshop discussion on the scope and industry coverage of the 2007 Commodity Flow Survey (CFS). The workshop will be held at the CFS Conference scheduled for July 8–9, 2005, in Boston, Massachusetts.

The topic of the scope and industry coverage of the CFS for discussion purposes can be broken into five research areas:

1. Those commodities and industries that have been out-of-scope to the CFS;
2. Incomplete and inconsistent industry coverage that occurred in the 2002 CFS when the commodity classification scheme changed from the standard industrial classification (SIC) codes to the North American Industrial Classification System (NAICS) system;
3. The dramatic growth of auxiliaries and their role in our increasingly sophisticated transportation system;
4. The role of third-party logistic providers (3PLs) in transportation brought about by the adoption of supply chain logistics and the outsourcing by shippers of their transportation functions; and
5. The necessity to better understand how to capture and identify intermodal shipments in the CFS.

This paper will comment on all five of these research areas but will focus primarily on the last four because they are the most achievable given budget and technical considerations and are important to producing an accurate and comprehensive CFS.

BACKGROUND

The planned 2007 CFS will be the fourth survey in a program that began with the 1993 CFS and was followed by surveys in 1997 and 2002. The CFS is a component of the Economic Census undertaken every 5 years in those years ending in 2 and 7. The first CFS was delayed until 1993, after the 1992 Economic Census, because the primary sponsoring agency, the U.S. Department of Transportation's (USDOT) Bureau of Transportation Statistics (BTS), was not created until the passage of The Intermodal Surface Transportation Efficiency Act of 1991 on December 18, 1991.

With the passage of the Research and Innovative Technology Act on November 30, 2004, BTS now resides within the USDOT's Research and Innovative Technology Administration.

Since its 1993 inception, the CFS has been undertaken through a partnership between the Census Bureau and the BTS, with BTS providing 80% of the funding and the Census Bureau the remaining 20%. Industry coverage of the CFS remained the same during the first three surveys, with CFS coverage extending to business establishments with paid employees located in the United States and classified as mining, manufacturing, wholesale trade, and selected retail industries (electronic shopping and mail-order houses). Establishments classified in services, transportation, construction, and most retail industries as well as farms, fisheries, most government-owned establishments, and imports have been excluded and are not in-scope to the CFS. These industries are either not contained in the Census Bureau's Business Register or are not primarily shippers of commodities and their inclusion would exceed the resources available for conducting the 2007 CFS.

CURRENT ISSUES AND PAPER FORMAT

The current issues involving the scope and industry coverage of the CFS can be categorized as historical in nature, having occurred out of changes in transportation since the CFS was first implemented, or because of changes in the coding scheme used to classify commodities. The conversion from the SIC codes used in the 1993 and 1997 editions of the CFS, to the NAICS system used in the 2002 CFS has had a major impact on many commodity sectors.

Historically industries that were out-of-scope in the CFS included industries such as oil and gas exploration, household goods, and several others as shown in Table 1. These industries were excluded for multiple reasons and include industries not on the Census Bureau's NAICS-based establishment list known as the Business Register [previously called the Standard Statistical Establishment List when establishments were classified by SIC codes], and therefore not surveyed in the Economic Census. The Economic Census along with confidential administrative records forms the foundation of the Business Register and is also an integral part of the CFS weighting process.

In all likelihood the industries that historically have been out-of-scope in the CFS, with some possible exceptions, will not be included in the 2007 survey. Budget constraints and technical concerns present formidable obstacles that must be overcome before expanding the scope and industry coverage of the CFS beyond the Business Register can be considered. The focus of this paper and the discussion that it is intended to stimulate is meant to ensure that the measurement of the CFS commodity groups and industries surveyed to capture these flows represent complete and accurate coverage.

The following five areas will be discussed in the forthcoming sections.

- The first discussion area will document the historical industries and commodity groups that have been out-of-scope in the CFS. The remaining four sections discuss areas where research has uncovered issues with the current scope and industry coverage of the CFS that need to be addressed to satisfy the goal of complete and accurate coverage of the establishments and industries to be surveyed in the upcoming 2007 CFS.
- The second research area concerns the effect of the change-over from SIC codes to the NAICS codes used in the 2002 CFS and planned for the 2007 CFS.
- The third research area involves the treatment of auxiliary establishments in the CFS. Now known as "enterprise support" establishments, these are establishments that operate in support of a corporation's central function.
- The fourth area discussed in this document is the need to research and better understand the importance of 3PLs. These companies supply warehousing and transportation services under

contract to companies who previously would have had these services provided by in-house staff, corporate facilities, and equipment.

- The fifth research area reviewed is intermodal transportation and its coverage in the CFS. What type of intermodal traffic is being captured in the CFS and issues with its identification will be discussed.

There may be considerable overlap among the last four research areas discussed in this document. Research in the four nonhistorical CFS gap areas of SIC to NAICS conversion, auxiliaries, 3PLs, and the identification of intermodal traffic may have significant consequences on the scope and industry coverage, the construction of the sample frame, and the design of the questionnaire for the 2007 CFS. Research in these areas for the 2007 CFS is required by the major changes that have occurred in transportation since the CFS was first implemented in the 1990s as well as the conversion in commodity classification schemes and shortcomings in capturing and identifying intermodal shipments.

HISTORICAL CFS COVERAGE GAPS

The CFS does not represent a comprehensive picture of total national freight shipments, and it is very unlikely that a single survey ever will. The basis for the CFS sample frame is the Census Bureau's Business Register, which is constructed from the Economic Census held every 5 years in years ending in 2 and 7 and supplemented with the Census Bureau's Annual Survey of Manufacturers and Company Organization Survey. In addition, administrative records from other federal agencies are used to further expand and enhance the data contained within the Business Register. The sources of data contained in the Business Register produce a database that is detailed but restricted and highly confidential.

The Business Register is the universe from which the CFS sample frame is drawn and is the basis for the processes of weighting establishment responses in the CFS. Not all establishments in the United States are contained in the Business Register. Table 1 shows some establishments and their industries or categories that are not contained in the Business Register.

The expansion of the scope and industry coverage of the CFS to include establishments in Table 1 would require using multiple sample frames that would extend beyond the Census Bureau's Business Register. The use of multiple sample frames creates problems of duplication between establishment lists and difficulties in developing a weighting process for survey responses.

The technical concerns of going beyond the Business Register to expand the scope of the CFS are formidable but so are the budget ramifications of such efforts. The institution of a dual or multiframe CFS would create the need of expenditures far exceeding what any budget projections for the 2007 CFS could accommodate. The time required to identify, obtain, and review frames also creates a prohibitive situation for realistically considering this course for the 2007 CFS. Consequently the expansion of the CFS beyond what is currently in-scope would have to be carefully considered and for the purposes of this discussion it is recommended that the attention of the workshop participants focus on the other, more achievable, research areas presented in this paper.

TABLE 1 Major Coverage Gaps in the 2002 CFS by Mode

In Business Register	Shipping Establishment	Modes Used
No	Farms	For hire truck, private truck
No	Fisheries and hunting	For hire truck, private truck, air
No	Logging	For hire truck, private truck, rail, water
No	Producers and distributors of crude petroleum	Pipeline, water
Yes	Construction	For hire truck, private truck
Yes	Services	For hire truck, private truck, mail/parcel delivery
Yes	Publishing	For hire truck, private truck, mail/parcel delivery
Yes	Retail (other than electronic and mail order shopping)	Mail/parcel delivery
No	Government-military	All
No	Government-municipal solid waste	For hire truck, rail, water
Partial	Government-other	For hire truck, mail/parcel delivery
Yes	Households and business moving	For hire truck, mail/parcel delivery
No	Foreign establishments (imports)	All
Yes	Paper products	For hire truck, private truck, mail/parcel delivery
Yes	Warehousing	All
Yes	Central administrative offices	For hire truck, mail/parcel delivery

SOURCE: FHWA, Freight Analysis Framework 2 project, BTS compilation, 2005.

SIC–NAICS CONVERSION

A critical issue associated with the scope of the CFS is the comprehensiveness of the industry sectors used to construct the CFS sample frame. It is imperative that a strong attempt be made to include in the scope and industry coverage of the CFS all industry sectors and subsectors that ship hazardous materials (with some exceptions) or significant freight within a Standard Classification of Transported Goods (SCTG) code. The transition from SIC to NAICS was first enacted by the Census Bureau with the publication of data from the 1997 Economic Census (Table 2). This changeover for classifying businesses and reporting industry statistics is a supply based or production-oriented, economic concept that was implemented to reflect our changing economy. The adoption of NAICS did cause a break in the availability of many time series data. While data for over two-thirds of the four-digit SIC codes could be estimated from the NAICS system, many industry definitions are dramatically changed. The NAICS, by which establishments are assigned to industry sectors, is a supply-based, or production-oriented, classification system. Another way of describing the difference between NAICS and the previously used SIC codes is that in NAICS those establishments with a similar production process are classified in the same industry while in the SIC system establishments were classified based on their primary activity. As a result of the conversion to NAICS some establishments that were in a single SIC group moved into multiple NAICS sectors. Consequently, some of the establishments that ship certain commodities have been transferred into sectors that are out-of-scope, for example, retail or agriculture, and are no longer included in the CFS.

Two of the most prominent examples of the transition from SIC to NAICS in the 2002 CFS were SCTG 25: Logs and Other Wood in the Rough, and SCTG 29: Printed Products. Commodities covered under these two codes that were in scope for CFS under SIC moved to sectors that were not included in the scope of the 2002 CFS. Other examples include industries classified under SCTG 18: Fuel Oils and SCTG 19: Coal and Petroleum Products n.e.c.

In the 2002 CFS, a portion of establishments shipping fuel oil and propane classified as Petroleum Bulk Stations and Terminals (SIC 5171) fell out-of-scope because in NAICS they were classified as retail establishments under NAICS 454311: Heating Oil Dealers and NAICS 454312: Liquefied Petroleum Dealers. Research conducted by BTS indicates that the out-of-scope portion of heating oil and liquefied petroleum sold via retail establishments was fewer than 3% of the total industry sales. Although this was not a significant portion of total industry sales, it was significant in terms of the number of establishments. The out-of-scope establishments shipping fuel oil and liquefied petroleum gas represented 15% of the total. The fact that fuel oil and liquefied petroleum gas are classified as hazardous materials makes it essential that shipments of these commodities are fully captured in the 2007 CFS. This accuracy is particularly necessary given that the data on these commodities are used in analyzing the safety and security of hazardous materials transportation. Summary data extracted from the Census Bureau's SIC 1987 bridge to NAICS 1997 are included in Tables 4 and 5.

It is also important to note the major changes that occurred when the classification and definition of auxiliaries was modified in the 2002 NAICS manual. Due in part to this revision in the 2002 NAICS coding scheme, the number of establishments classified under Code 493: Warehousing and Storage increased from 6,497 in 1997 to 12,123 in 2002. Overall only 422 SIC categories or about 40% remained substantially unchanged during the transition to NAICS.

AUXILIARIES

Auxiliaries, now also known as "enterprise support" establishments, but hereafter referred to as auxiliaries, are establishments that provide management or support services to one or more establishments of the same enterprise. Locations that perform headquarters services, payroll services, or warehousing and transportation services are examples of auxiliaries. Given the focus of the CFS, we are specifically interested in those auxiliary establishments that perform warehousing and transportation functions because they often act as intermediaries between shippers (manufacturers and importers) and retail stores. Knowledge of shipments reported by manufacturers and importers must be complemented by shipments from the intermediaries (auxiliaries) to final retail store destinations to prevent value, tonnage, and ton-miles from being understated in the CFS.

TABLE 2 Wholesale Trade, Durable Goods—Comparison of the Number and the Value of Shipment of Establishments Within Wholesale Trade Division of 1987 SIC and Wholesale Trade Sector of 1997 NAICS

SIC	NAICS	Description	Establish-ments	Value of Shipments (1,000)	Paid Employees	Annual Payroll (1,000)
501		Motor vehicles and motor vehicle parts and supplies	48,055	561,792,495	554,976	15,692,967
	4211	Motor vehicle and motor vehicle parts and supplies wholesale	29,328	533,352,124	375,731	11,458,634
Percent change because of reclassification from SIC to NAICS			–39.0%	–5.1%	–32.3%	–27.0%
502		Furniture and home furnishings	18,603	82,707,554	191,703	6,431,522
	4212	Furniture and home furnishings wholesale	15,246	75,006,478	157,465	5,316,976
Percent change because of reclassification from SIC to NAICS			–18.0%	–9.3%	–17.9%	–17.3%
503		Lumber and other construction materials	22,427	118,764,214	258,227	8,669,637
	4213	Lumber and other construction materials wholesale	14,267	89,175,875	155,535	5,296,176
Percent change because of reclassification from SIC to NAICS			–36.4%	–24.9%	–39.8%	–38.9%
504		Professional and commercial equipment and supplies	49,675	387,497,154	759,613	34,778,167
	4214	Professional and commercial equipment and supplies wholesale	45,351	367,383,550	716,113	33,292,437
Percent change because of reclassification from SIC to NAICS			–8.7%	–5.2%	–5.7%	–4.3%
505		Metals and minerals, except petroleum	12,583	150,493,610	174,029	6,898,028
	4215	Metal and mineral (except petroleum) wholesale	12,583	150,493,610	174,029	6,898,028
Percent change because of reclassification from SIC to NAICS			0.0%	0.0%	0.0%	0.0%
506		Electrical goods	44,295	381,044,602	547,776	25,011,949
	4216	Electrical goods wholesale	38,234	357,691,888	475,766	22,524,717
Percent change because of reclassification from SIC to NAICS			–13.7%	–6.1%	–13.1%	–9.9%
507		Hardware and plumbing and heating equipment and supplies	26,926	109,016,098	279,585	9,935,530
	4217	Hardware, and plumbing and heating equipment and supplies wholesale	21,194	92,189,762	219,233	7,977,961
Percent change because of reclassification from SIC to NAICS			–21.3%	–15.4%	–21.6%	–19.7%
508		Machinery, equipment, and supplies	77,587	330,544,034	780,186	29,599,541
	4218	Machinery, equipment, and supplies wholesale	76,643	328,968,331	772,550	29,401,798
Percent change because of reclassification from SIC to NAICS			–1.2%	–0.5%	–1.0%	–0.7%
509		Miscellaneous durable goods	37,126	177,634,042	341,276	10,720,001
	4219	Miscellaneous durable goods wholesale	37,783	185,455,758	351,839	11,070,270
Percent change because of reclassification from SIC to NAICS			1.8%	4.4%	3.1%	3.3%

SOURCE: U.S. Census Bureau—BTS compilation, 2005

TABLE 3 Wholesale Trade, Nondurable Goods—Comparison of the Number and the Value of Shipment of Establishments Within Wholesale Trade Division of 1987 SIC and Wholesale Trade Sector of 1997 NAICS

SIC	NAICS	Description	Establish-ments	Value of Shipments (1,000)	Paid Employees	Annual Payroll (1,000)
51		Wholesale trade, nondurable goods	183,850	1,935,906,522	2,621,962	86,779,592
	422	Wholesale trade, nondurable goods	162,841	1,879,940,402	2,398,296	81678408
Percent change because of reclassification from SIC to NAICS			–11.4%	–2.9%	–8.5%	–5.9%
511		Paper and paper products	19,744	134,224,675	296,912	9,151,004
	4221	Paper and paper product wholesale	15,848	117,062,485	214,350	7730308
Percent change because of reclassification from SIC to NAICS			–19.7%	–12.8%	–27.8%	–15.5%
512		Drugs, drug proprietaries, and druggists' sundries	8,053	203,147,771	190,127	8,394,864
	4222	Drugs, and druggists' sundries wholesale	8,053	203,147,771	190,127	8394864
Percent change because of reclassification from SIC to NAICS			0.0%	0.0%	0.0%	0.0%
513		Apparel, piece goods, and notions	21,266	125,860,153	215,100	8,020,588
	4223	Apparel, piece goods, and notions wholesale	20,707	124,104,420	207,574	7759577
Percent change because of reclassification from SIC to NAICS			–2.6%	–1.4%	–3.5%	–3.3%
514		Groceries and related products	41,949	590,785,074	858,481	26,883,955
	4224	Grocery and related products wholesale	41,760	588,970,062	854,919	26778099
Percent change because of reclassification from SIC to NAICS			–0.5%	–0.3%	–0.4%	–0.4%
515		Farm-product raw materials	10,343	166,786,245	97,521	2,306,012
	4225	Farm-product raw material wholesale	10,343	166,786,245	97,521	2306012
Percent change because of reclassification from SIC to NAICS			0.0%	0.0%	0.0%	0.0%
516		Chemicals and allied products	15,920	128,923,496	165,768	7,241,315
	4226	Chemical and allied products wholesale	15,920	128,923,496	165,768	7241315
Percent change because of reclassification from SIC to NAICS			0.0%	0.0%	0.0%	0.0%
517		Petroleum and petroleum products	12,711	272,459,061	151,555	4,838,952
	4227	Petroleum and petroleum products wholesale	11,297	267,623,942	137,829	4479853
Percent change because of reclassification from SIC to NAICS			–11.1%	–1.8%	–9.1%	–7.4%
518		Beer, wine, and distilled alcoholic beverages	4,850	69,703,203	151,677	5,667,069
	4228	Beer, wine, and distilled alcoholic beverage wholesale	4,850	69,703,203	151,677	5667069
Percent change because of reclassification from SIC to NAICS			0.0%	0.0%	0.0%	0.0%
519		Miscellaneous nondurable goods	49,014	244,016,844	494,821	14,275,833
	4229	Miscellaneous nondurable goods wholesale	34,063	213,618,778	378,531	11321311
Percent change because of reclassification from SIC to NAICS			–30.5%	–12.5%	–23.5%	–20.7%

SOURCE: U.S. Census Bureau—BTS compilation, 2005.

TABLE 4 1997 Economic Census Statistics on Wholesale Petroleum and Petroleum Products Establishments: Bridge Between SIC 517: Petroleum and Petroleum Products and NAICS*

SIC	NAICS	Description	Number of Establishments	Sales/Receipts/ Revenue/Shipments ($1,000)
517		**Petroleum and petroleum products**	**12,711**	**272,459,061**
5171		**Petroleum bulk stations and terminals**	**9,104**	**181,554,365**
	422710	Petroleum bulk stations (except LP)	6,045	50,922,994
	422710	Petroleum bulk terminals (except LP)	1,225	120,085,945
	422710	Liquefied petroleum bulk stations and terminals	420	5,710,307
	454311	Heating oil dealers (selling for consumption-retail)	813	3,283,889
	454312	Liquefied petroleum dealers (selling for consumption-retail)	601	1,551,230
5172		**Petroleum and petroleum products, except bulk stations and terminals**	**3,607**	**90,904,696**
	422720	Petroleum and petroleum production wholesale (except bulk stations and terminals)	3,607	90,904,696

*Establishments in the shaded area (NAICS 454311, and 454312) which were considered as wholesale by SIC have moved to retail under NAICS, and consequently were not covered in 2002 CFS.
SOURCE: U.S. Census Bureau—BTS compilation, 2005.

BTS has conducted analyses that demonstrate the growing importance of distribution centers in the retail, multistore environment. For example, Wal-Mart, the nation's largest retailer with worldwide sales of $285 billion, reports in its 10-K Securities and Exchange Commission (SEC) filing that, at the end of 2004, it had 90 distribution centers for its Wal-Mart stores in the United States and that these distribution centers shipped 80% of the merchandise sold by those retail stores.

Similarly, the SEC filing of The Home Depot, Inc., the nation's second largest retailer with approximately $65 billion of sales, states "At the end of fiscal 2003, approximately 40% of the merchandise shipped to our stores was passed through the network of distribution centers and transit facilities." Clearly, if the shipments from distribution centers to stores are not gathered and included, CFS will significantly understate freight flows in the United States.

TABLE 5 1997 Economic Census Statistics on Wholesale Petroleum and Petroleum Products Establishments: Percent of the Total Number, and Sales of Establishments Classified in Wholesale Petroleum and Petroleum Products Under SIC 1987 (SIC 517) That Can Be Bridged to NAICS 4227 (Wholesale Petroleum and Petroleum Products)*

NAICS	SIC	Description	Number of Establishments	% of SIC Establishments Covered by NAICS ($1,000)	Sales/Receipts/ Revenue/Shipments Covered by NAICS ($1,000)	% of SIC
		WHOLESALE TRADE				
4227		**Petroleum and petroleum products wholesale**	**11,297**	**84.5%**	**267,623,942**	
422710		**Petroleum bulk stations and terminals**	**7,690**		**176,719,246**	**97.3%**
	5171	Petroleum bulk terminals	1,225		120,085,945	
	5171	Petroleum bulk stations, except LP (selling for resale) (wholesale)	6,045		50,922,994	
	5171	Liquid petroleum bulk stations and term (selling for resale) (wholesale)	420		5,710,307	
422720		**Petroleum and petroleum prod wholesale (except bulk stations and terminals)**	**3,607**	**100.0%**	**90,904,696**	**100.0%**
	5172	Petroleum and petroleum product wholesalers, except bulk stations and terminals	3,607		90,904,696	

*Text in bold indicates NAICS classification.
SOURCE: U.S. Census Bureau—BTS compilation, 2005

Not only will freight flows be understated, but modal data will be distorted. Wal-Mart, for example, indicates that general merchandise is transported to stores primarily through its private truck fleet; for-hire carriers are employed to transport the majority of perishable and grocery merchandise. The distribution centers of Wal-Mart and Home Depot together ship commodities of approximately $300 billion of value.

Definitions and classifications have changed in NAICS from the 1997 to 2002 editions. As pointed out earlier, in the 1997 NAICS there were 6,497 establishments listed under code 493: Warehousing and Storage. The 2002 edition of NAICS, which will be used in the 2007 CFS, shows 12,123 establishments under code 493: Warehousing and Storage. This major change in auxiliaries classified as Warehousing and Atorage establishments occurred largely because establishments without inventory were not previously included in this sector. Given that NAICS codes 4931 and 5511 represent auxiliaries sampled in the 2002 CFS, careful attention must be paid to this research area when planning for the 2007 CFS.

THIRD–PARTY LOGISTIC PROVIDERS

To obtain a complete understanding of commodity shipment levels and modal patterns for the CFS in 2007, research needs to be conducted on 3PLs. In recent years as corporations focus inward to concentrate on their core businesses, more companies are contracting out their logistic functions of warehousing and transportation to 3PLs, which perform many identical functions for an enterprise as the auxiliaries do, but that are not owned by the company. Therefore, if the CFS surveys all establishments owned by an enterprise in a given NAICS sector and the enterprise does not contract out its warehousing or shipping functions, one can be assured that all shipping activity has been captured. However, if a 3PL has been contracted by an enterprise to perform their transportation or warehousing functions, then an incomplete picture of the freight flows of that enterprise and the commodities it ships may be obtained. If the 3PL maintains the shipping records, they may have to be surveyed as well.

It is not known how much of the freight moved through distribution centers or auxiliaries operated or owned by 3PLs, was captured in the 2002 CFS. Some of the establishments that are operated by 3PLs may not be owned by either the enterprise or the third-party provider. In this instance, these establishments and the freight moving through them may have been out-of-scope in the 2002 CFS. In an attempt to prevent, or at least minimize this data gap from occurring in the 2007 CFS, research on how enterprises utilize 3PLs and record their freight shipments needs to be undertaken.

Evidence exists that the use of 3PLs by major corporations and enterprises is a growing phenomenon with many if not most large shippers using this service. Research involving 3PLs could include questions of the following type. How common is the use of 3PLs by enterprises that are in scope for the 2007 CFS? Where are the shipment records from the establishments operated by the 3PLs kept? Who owns the establishments and what are their NAICS classifications? Most 3PLs are likely to be classified as being in the transportation sector, which in the past has been out-of-scope in the CFS. Given the circumstances surrounding the transportation of freight by 3PLs for major shippers, research is required so that the CFS scope can be carefully adjusted to provide a more complete and accurate presentation of commodity flows in 2007.

INTERMODAL

Intermodal transportation is a term that in recent years has often been a source of confusion. This term has its origins in the railroad industry and generally relates to the transfer of a piece of transportation equipment containing a shipment of commodities between modes, usually truck and rail. The most common and classical examples involve piggyback trailers and shipping containers being transported on railroad flat cars [Trailer on Flat Car (TOFC) and Container on Flat Car (COFC)]. Recent use of the term has expanded to mean any transfer of freight between modes and even the transfer of passengers from one mode to another. For the purposes of discussing intermodal traffic in the CFS, we are focusing on the application of this term to TOFC and COFC traffic.

The CFS questionnaire in 1993 and 1997 asked the respondent whether a shipment was containerized. Because of the poor quality of the responses received in both the 1993 and 1997 CFS the question was dropped in the 2002 CFS. It was assumed however that TOFC and COFC intermodal traffic was being captured in the CFS and displayed in the modes of Truck and Rail and Parcel, U.S. Postal Service, or courier mode. An analysis of the 2002 CFS SCTG codes for the truck and rail mode shows that very few of what many CFS users thought were TOFC or COFC shipments are actually in fact such shipments. Many of the types of commodities shown in Table 6 and Table 7 would not be shipped in either a piggyback trailer or a container.

TABLE 6 Twenty Commodities Commonly Shipped by Truck and Rail Mode by Tonnage from the 2002 CFS

SCTG	Commodity Description
02	Cereal grains
03	Other agricultural products
06	Milled grain products and preparations, and bakery products
07	Other prepared foodstuffs and fats and oils
08	Alcoholic beverages
15	Coal
19	Coal and petroleum products, n.e.c.
20	Basic chemicals
22	Fertilizers
23	Chemical products and preparations, n.e.c.
24	Plastics and rubber
26	Wood products
27	Pulp, newsprint, paper, and paperboard
28	Paper or paperboard articles
31	Nonmetallic mineral products
32	Base metal in primary or semi finished forms and in finished basic shapes
34	Machinery
36	Motorized and other vehicles (including parts)
41	Waste and scrap
43	Mixed freight

NOTE: A rank ordering for the 20 commodities listed cannot be given because of the statistical variability of CFS tonnage estimates.
SOURCE: BTS—special compilation, 2005

TABLE 7 Twenty Commodities Commonly Shipped by Truck and Rail Mode by Value from the 2002 CFS

SCTG	Commodity Description
02	Cereal grains
03	Other agricultural products
06	Milled grain products and preparations, and bakery products
07	Other prepared foodstuffs and fats and oils
08	Alcoholic beverages
09	Tobacco products
14	Metallic ores and concentrates
20	Basic chemicals
23	Chemical products and preparations, n.e.c.
24	Plastics and rubber
26	Wood products
27	Pulp, newsprint, paper, and paperboard
28	Paper or paperboard articles
32	Base metal in primary or semi-finished forms and in finished basic shapes
34	Machinery
35	Electronic and other electrical equipment and components and office equipment
36	Motorized and other vehicles (including parts)
40	Miscellaneous manufactured products
41	Waste and scrap
43	Mixed freight

NOTE: A rank ordering for the 20 commodities listed cannot be given because of the statistical variability of CFS value estimates.
SOURCE: BTS—special compilation, 2005

The CFS in all likelihood does capture a significant part of TOFC and COFC intermodal traffic, but it cannot be identified as such from the processing of the responses. Intermodal traffic is included in the CFS modes of parcel, U.S. Postal Service, or courier, truck and rail, and truck. Given that TOFC and COFC traffic has increased since 1993 according to industry sources, the year of the first CFS by over 40%, and currently exceeds over 10 million trailers and containers annually, strong attention and thought should be applied to improving this data in the 2007 CFS.

Filling Gaps in the U.S. Commodity Flow Picture
Using the CFS with Other Data Sources

FRANK SOUTHWORTH
Oak Ridge National Laboratory

ABSTRACT

The U.S. Commodity Flow Surveys (CFS) provide a wealth of data on the movement of freight within the country, by mode of transportation, dollar value, annual tonnage, door-to-door shipment distance, and shipment size. Yet gaps in CFS coverage and a lack of spatial and commodity detail limit the value of this data for planning and policy supporting studies. This paper identifies these gaps and considers how supplementary data sources combined with statistical modeling techniques can be used to create a more complete picture of national and regional commodity flows. The first half of the paper describes current data gaps and available data resources. The second half of the paper discusses possible gap-filling solutions. Solutions fall into two categories: data synthesis using current data sources, and improved and expanded data collection methods.

INTRODUCTION

Following a 16-year hiatus in federal data collection the 1993, 1997, and 2002 U.S. CFS have filled a large gap in the U.S. freight data universe. As a result of these surveys we now have data on the annual volume of commodity movements taking place into, out of, and within each of our states, the District of Columbia, and our largest metropolitan areas, broken down by mode of transport. We also have data on the "door-to-door" travel distances and shipment sizes associated with these movements, as well as both their annual tonnages and dollar value. However, in trying to use these data in planning and policy supporting studies we run into two kinds of problems. First, the surveys do not cover all U.S. freight movements. Second, the surveys only support the representation of origin-to-destination (O-D) movements between quite large geographic regions, and even then the O-Ds they do produce are limited in the level of commodity detail the surveys can support (because of sample size), or are allowed by law to reveal.

Gaps in coverage, difficulties in determining where different data sets overlap, and the need to consult a number of different data dictionaries to develop an adequate representation of a region's freight movements can prove more than just frustrating. They also make it difficult to establish with confidence the accuracy and statistical robustness of the resulting estimates of these commodity movements. As recent reviews have pointed out, this is the data universe the freight analyst is faced with currently, whether dealing with nationwide, statewide, or metropolitan area-wide commodity flows (*1-4*).

The purpose of this paper is to describe succinctly these gaps in CFS coverage and to discuss ways that other data sources can be used to fill them. The principal beneficiaries of a successful gap filling activity are analysts and policy makers in federal, state, and metropolitan agencies where the volumes of goods moved are key inputs to investment decisions affecting future transportation infrastructures and services. In particular, the paper is used to explore the following questions:

1. What other data sets are available for filling the current gaps in CFS coverage and detail?
2. What data modeling techniques exist for combining CFS data with these other data sources?
3. What new and emerging ways of collecting freight movement data might we tap into?

Appendix C of the circular provides a summary of questions raised and directions for future research and development.

DATA PRODUCTS WE NEED

Before entering this discussion it's useful to define the nature of the data products we're looking for. For present purposes two data products are of most interest, whether developed on a fully national or individual statewide basis:

- A multidimensional freight flow matrix, the principal dimensions of which are freight traffic generating origins, destinations, modes and types of commodities moved; and based on this flow matrix:
- A series of traffic assignments showing how freight vehicles move over a region's roadways (highways, railways, waterways, pipelines) and through its various modal (e.g., truck village) and intermodal (seaport, airport, truck–water, truck–rail, and truck-pipeline) transfer terminals.

With these two products we can address a wide range of issues and studies commonly faced by regional planners: from the creation of time series statistics on total freight activity, to an analysis of competition and cooperation between modes, to an analysis of the economic, safety, and environmental impacts of site-specific transportation capacity expansion or service modification projects. Both products, of course, imply the availability of other supporting data sets, notably data to represent a transportation network, as well as coefficients that translate emissions, fuel, and travel times into suitable movement costs. Focusing on the uses to which the CFS can be put, it is with the quality of the supporting "freight movement data" that we are concerned here.

THE DATA CHALLENGE: WHAT THE DATA DO AND DON'T TELL US

For the purpose of generating detailed freight movement matrices, the O-D data sets available to most users can be grouped into three classes:

1. A nationwide multimodal commodity shipment survey of "door-to-door" movements (i.e., the U.S. CFS);

2. Nationwide and mode-specific freight carrier activity surveys reporting "station-to-station" freight movements, reporting tons moved by specific industries and types of transportation equipment; and

3. International trade and trans-border traffic flow surveys, reporting principally commodity specific, dollar valued trades.

Taken together these data represent a loosely connected patchwork quilt with a number of holes in its coverage. The reader is directed to references (*1-5*) for recent listings and discussions of currently available databases falling under each of these headings. The principal gaps in each type of data are summarized below.

CFS Strengths and Weaknesses

The CFS has a number of unique strengths. In particular

- It is fully national in scope;
- It covers all the major surface transportation modes (truck, rail, water, petroleum pipelines), as well as shipments of air freight;
- It identifies the true geographic O-D of each shipment (and therefore also provides estimates of "door-to-door" shipment distances);
- It collects data on both the weight and dollar value of all in-scope shipments;
- It has a time series in the form of the 1993, 1997, and 2002 surveys; and
- It is done in conjunction with the Economic Census, providing concurrency with other data sets.

On the debit side, in particular

- Not all commodities are covered by the CFS;
- The survey does not, in theory, capture imports;
- The spatial detail available to its mode-specific O-D matrices is limited to a small number of rather large geographic regions;
- The volume of "intermodal"[1] freight reported may be low, due at least in part to definitional issues;
- The shipment length detail available from non-geographically disaggregated products is very limited in its supporting commodity-level detail;
- The surveys have seen some content changes, and a 4 to 1 reduction in sample size between 1993 and 2002 that makes for some large coefficients of variation in reported estimates; and
- There are discrepancies in the estimates generated by the CFS and the U.S. Army Corps of Engineers' waterborne commerce data, the latter based on industry-wide carrier reporting that produces larger ton and ton-mileage figures.

Coverage Issues: Commodities.

All three (1993, 1997, 2002) CFS surveys sampled business establishments in mining, manufacturing, wholesale trade, and selected retail industries (*6*). The surveys also cover selected auxiliary establishments, such as the warehouses of in-scope multi-unit and retail companies. The surveys do not cover establishments classified as farms, forestry, fisheries, construction, transportation (including household goods carriers), governments (including military and mail shipments), foreign establishments, services, and most establishments in retail. As a result, the CFS has been conservatively estimated to cover less than 75% of all the freight tons moved annually in the United States (*7*). A common trait of many of the missing shipments is the dominance of highly localized, essentially truck-only movements. This applies to most retail, and also to a good deal of activity in the construction and personal delivery services industries, both significant sources of short range truck miles of travel according to the 1997 and 2002 U.S. Vehicle Inventory and Use Surveys (VIUS) (*8*).

Coverage Issues: Imports

Being a survey of U.S. shippers the CFS also does not capture imports. Here a difficulty in adding imported tonnages to the CFS arises because a shipment that may be classified as imported cargo in the international trading arena also finds itself being redefined as an internal shipment within the CFS once a change in ownership of the goods has taken place. This may occur at a warehouse or other storage facility located close to the U.S. port of debarkation. How often this occurs is unclear. The CFS does include goods exported by U.S. shippers. However, the estimated volume and value of these shipments is often at odds with the reporting of such exports in the U.S. Foreign Trade data.

O-D Detail

Perhaps the single most significant improvement to the CFS for planning purposes would be the addition of geographic detail. This is asking a good deal. Even for the current 114 CFS regions, once shipments are broken down by both O-D region the level of both commodity and modal detail that the Census Bureau can release becomes limited, because of lack of robustness in sample-based estimates, or to the need to protect the confidential nature of shipper activities. These constraints restrict the survey's O-D matrices to very general one- or two-digit commodity classes. At the two-digit level this represents only 42 commodity groupings. And not even the 1993 survey, based on a sample of some 200,000 shipping establishments, was able to produce complete commodity and mode-specific O-D matrices at the level of rather broad two-digit commodity codes. With the 1997 CFS sample reduced to some 100,000 establishments, and with the 2002 CFS further reduced to a sample of 50,000 shippers, data on the O-D–commodity-mode (ODCM) combinations most useful to transportation planning agencies proves quite sparse.

Intermodal Shipments

The survey also faces a difficult task in representing the volume of intermodal freight movements, i.e., movements of freight from origin to destination using two or more end-on modes. One reason for this is probably definitional, and related to ownership of cargo. Another may be that some survey respondents didn't know how their product reached its final destination, only the mode it left their establishment in. This is, anyway, a generic problem that has to be faced when tracking any product through its freight movement supply chain. For example, a grain shipment will typically change hands when it reaches a grain elevator. The farmer sells to the elevator, who in turn sells to a customer “down river.” In contrast, a coal shipment may belong to the mine owner all the way to the utility at which it is consumed, possibly via truck-rail, truck–water, or rail–water transfer. Parcel shipments are also recognized as a separate category of freight in the CFS, and these too probably involve more intermodal activity than is captured.

Shipment Lengths

To date the CFS program has provided limited additional assistance to the spatial analyst. In particular, by limiting distance reporting to a small number of rather broad distance intervals, and for rather broad aggregations within commodity classes, a good deal of information that could be used to inform spatial interaction modeling is not available. Out-of-scope CFS truck and rail freight movements, notably of imported goods, pose a similar problem, with the Transborder Surface Freight Data and the Port Import and Export Reporting Service data sets each providing their respective U.S.–Canadian, U.S.–Mexican, and U.S.–transoceanic movements at only the state level (and often for a destination address associated with the business office of the receiver, rather than with the true destination of cargo delivery).

Time-Series Issues

As a time series of freight movement activity, the reduction in CFS sample size from roughly 200,000 establishments in 1993 and to 50,000 in 2002 is of great concern. As a result, some of the more detailed O-D matrix elements available in 1993 are no longer statistically robust enough to be reported in 2002. The CFS has also seen a number of changes in its design and content that can affect trend analyses. One of these changes was the move from the original system of 89 National Transportation Analysis Regions, to a 1997 regionalization based around the nation’s 56 most populous metropolitan areas, small states, and remainder-of-state regions. The 1993 CFS used Standard Transportation Commodity Codes, the 1997 and 2002 surveys moved to Standard Classification of Transported Goods (SCTG). Between 1997 and 2002 the boundaries of some of these metropolitan areas also changed. Of more concern was the impact of the North American Industrial Classification System (NAICS) re-classification of the underlying business establishments from which the survey was drawn. This resulted in a loss of data for both the lumber and printed matter industries, as covered in the 1993 and 1997 surveys (i.e., they fell out of scope). In 1997 the request for data on containerized shipments was dropped, while the method of asking for information on hazardous materials shipments was improved, ensuring among other things that petroleum shipments were captured in this category. The difficulty of creating petroleum pipeline O-Ds from the 1993 survey data led to this aspect of the survey being down-played in the 1997 and 2002 data creation efforts. Some of these changes may prove important when trying to construct temporal trends in ton, dollar, ton-mile, or vehicle-mile statistics; or if trying to use the 1993 and 1997 data sets to fill gaps in the 2002 data.

If such trend information is important then we might also ask whether a different approach to the survey design is in order. In particular, would a continuously sampled CFS help? This last issue is discussed further in (*4*) where the potential strengths as well as the practical challenges of continuous sampling are outlined, with evidence for some success in past freight surveys.

MODE-SPECIFIC CARRIER SURVEYS

In contrast to the CFS, the nation's carrier-based surveys are mode specific. Given the CFS problems of scope and detail discussed above, these surveys offer a natural option for enhancing the freight data picture.

True Versus Line-Haul O-Ds

Via the Surface Transportation Board's Rail Carload Waybill sample[2] and data contained in the Army Corps of Engineers' Waterborne Commerce database[3] it is possible to estimate, respectively, the volume of freight moving over specific sections of a railroad's track and over specific reaches of the nation's navigable waterways, and with a little work to associate these shipments with specific station-to-station or dock-to-dock routings over the U.S. rail and water networks. Both of these annual data collection efforts cover all of the commodities moved by their respective modes as well as details of the types of railcars and barges used to move different types of freight. The waterborne commerce data is an approximately 100% sample of inland barge, intra-coastal and Great Lakes movements. The rail waybills are a much smaller sample, but emphasizing the larger and heavier unit train movements. (They do not capture export shipments carried on the Canadian railroads operating inside the United States, however).

What is missing for the purpose of O-D estimation is the true origin and destination of these movements. This poses a problem, since a significant amount of the freight transported by rail and water involves truck draying to or from a rail or barge terminal, sometimes involving travel distances that place either or both the origin and destination of the freight outside the CFS region or State associated with the railcar or barge line-haul movement. This lack of true O-D information also applies to our current surveys of air freight carriers, again requiring that we somehow infer the true origin and destination of the truck drays that are involved in the vast majority of these low weight but high valued shipments. The most readily accessible forms of this air freight data, the Office of Airline Information's Air Freight Statistics, also provides only total tons of freight (and mail) transported, without commodity breakdowns.

More Data Needed on Trucks

The principal source of data on U.S. truck activity, the VIUS[4] contains no O-D data per se. This makes the CFS the only source of nationwide data on O-D truck movements, and the CFS captures no details on the type of truck used other than its for-hire versus private ownership status. What the VIUS offers is considerable detail, and a time series back to 1963, on the types of trucks used to haul freight of different types. In doing so it has a number of strengths and weaknesses of its own. On the credit side, it offers operator estimates of each vehicle's annual activity broken down by commodity carried and typical operating range (in distance intervals). The vehicle characteristics data is especially rich, including data on a vehicle's body type, length, axle configuration, empty and loaded operating weigh and mileage, ownership, fuel use, and hazardous cargo transport. On the debit side the commodity detail is quite limited: 33 classes prior to 2002, expanded in the 2002 VIUS to 51 classes based on the two-digit SCTG codes used by the CFS. The 5-year interval between each VIUS also means that we often have to wait a few years to see the effects of any important changes in trucking practices.

Too Many Commodity Classification Schemes

Direct comparison to, or combination of, information from the above carrier surveys with CFS data, or with each other, is further complicated because each has its own commodity coding scheme. While in 1997 and 2002 the CFS commodities are classified according to the SCTG, the rail waybills use STCC, and the waterborne commerce data is based on yet another commodity classification scheme. U.S. Foreign Trade Statistics are based on yet a fourth scheme: the Harmonized Commodity and Coding System, and the United Nations uses yet another scheme (the Standard International Trade Classification) for reporting international trades. The VIUS, which previously used standard industrial classification (SIC) commodity/industry codes, was converted to SCTG codes in the 2002 survey, making it compatible with the CFS. Finally, the coding scheme associated with the economic activity data sets commonly used to support both forecasting and geographic disaggregation of CFS-based commodity flows is the Census Bureau's NAICS (which replaced the SIC codes used by the Economic Census prior to 2000).

To combine data from two or more of these data sources means using a suitable "cross-walk" between the different commodity/industrial sector coding schemes. While such cross-walks already exist, there is necessarily some degree of lost accuracy in the resulting merger, especially at the more aggregate levels of some of the more diversified commodity grouping.

U.S. Trade Data: Movements of Imported and Exported Cargo

A number of federal agencies are involved in the collection, processing and dissemination of international trade and transportation data. U.S. merchandise trade statistics are processed and released by the Foreign Trade Division of the U.S. Census Bureau. International merchandise trade data are captured from administrative documents required by the Departments of Commerce and Treasury. The U.S. Customs Service collects these documents at the port of entry or exit unless the information is filed electronically using the Automated Broker Interface on imports or the Automated Export Reporting Program on exports.

Census also releases overall trade and transportation statistics that include data elements on the value, commodity, weight, country of O-D, and U.S. port used. Many agencies obtain special extractions and tabulations from Census and then perform additional quality assurance reviews and analyses to meet the needs of their own customers. These include: data on North American land trades (by truck, rail, mail, and pipeline) released and disseminated by the Bureau of Transportation Statistics (BTS) as the Transborder Surface Freight Data)[5]; data on U.S. international maritime trade [released to the Maritime Administration (MARAD) and Army Corps of Engineers and disseminated in multiple formats][6]; and data on U.S. transportation-related goods and overall trade data (released to Bureau of Economic Affairs/Department of Commerce and disseminated in multiple formats, including balance of payments information). A popular private sector product based on Customs data is PIERS (Port Import Export Reporting Service)[7].

Despite this wealth of information, current U.S. merchandise trade data pose a number of problems when we try to construct O-D freight movement matrices from them. In particular (*4*)

- Current reporting requirements mean that shipping weight is currently only collected for imports, and not for exports.
- In many cases the reported port of entry or exit is not the actual seaport but the port of duty filing, and electronic filing has increased the number of such filings.

- Data on the domestic O-D of international trade is often reported incorrectly. It is not uncommon for origin of movement series respondents, who may be intermediaries in the goods movement process, to erroneously report either a headquarters location or to specify the location of the U.S. port of exit as the point of an export's origin. The impact is greatest on the allocation of non-manufactured exports, where intermediaries are more common, notably farm products, minerals, and other bulk commodities.
- Because of current reporting requirements merchandise trade statistics do not distinguish goods moved by intermodal combination. Export mode is defined here as simply being the mode used when the U.S. international border is crossed. On the import side, the mode of transportation is defined as the last mode used when the freight arrives at the U.S. port of clearance or entry.

Besides these and a few other issues of content, it is worth noting that anyone unfamiliar with foreign trade data can spend a good deal of time searching the web for specific products. Annual imports and exports by U.S. seaport, foreign port of O-D, and commodity class can be obtained. Finding and getting access to them, and associating them with other data sets covering seaport activities can be a challenge. One-site access to at least the most detailed O-D data tables would be beneficial to all users.

DATA SYNTHESIS TECHNIQUES: SOME PROMISING DIRECTIONS

To the extent that there is any "standard practice" in freight modeling, it involves finding ways to combine data from different sources to support estimation and forecasting of freight movement volumes. References (*9-17*) identify some recent studies that exemplify the sort of data integration problems we face at the metropolitan/regional (*9-12*), statewide (*13-15*) and national/transborder (*16, 17*) scales. Of note, such efforts include a number of state departments of transportation-(DOT) based projects that make use of the multi-sourced TRANSEARCH® database; a proprietary product developed by Reebie Associates[8] that offers one approach to what can seem a rather daunting data integration challenge.

Two complimentary lines of attack are suggested for getting the most information out of not only the CFS but also our other freight movement data sets. One employs various mathematical and statistical modeling techniques to merge data from these different sources. The second (see Section 4) involves the use of new surveys/survey design options as well as a set of rapidly evolving non-survey data collection techniques. Given the gaps in CFS coverage highlighted earlier, two important data synthesis activities are discussed below:

- Filling gaps in CFS (in- and out-of-scope) commodity flow data and
- Creating spatially detailed commodity and vehicle/vessel based freight flow matrices for highways, railways, and waterways.

Filling Gaps in CFS O-D Tables

If we think of our data problem as one of filling in a multidimensional ODCM matrix, a strong candidate for bringing the various elements of such a matrix together is log-linear modeling (*18–21*). Mechanically, this approach can be linked to a series of matrix adjustments using the technique of iterative proportional fitting (IPF) which ensures that the values reported or estimated for each cell in the ODCM matrix sum to known or estimated and more aggregate marginal totals. And given forecasts of these aggregate marginal activity totals, the same log-linear models can be used to project the matrix of ODCM cell values into the future.

An example model may help. In this example we assume a single commodity class for brevity and solve for the other three (O-D-M) dimensions. When a region is referred to as generating freight traffic we refer to it as region "*i*," and when it receives this traffic as region "*j*." Individual modes are as designated "*m*." The data product we seek is a fully filled in matrix of freight flows, measured in annual tons moved, $\{F_{ijm}\}$, broken down across each of these three dimensions. For this given commodity we can estimate the following multiplicative model of the tons shipped from region *i* to region *j* by mode *m*, which we would solve in natural log form as:

$$\ln F_{ijm} = \theta + \lambda_i^O + \lambda_j^D + \lambda_m^M + \lambda_{im}^{OM} + \lambda_{jm}^{DM} + \lambda_{ij}^{OD} + \lambda_{ijm}^{ODM} \quad (1)$$

Here the various λ's, often termed the model effects, are a set of model estimated parameters that will return the original cell estimates. For example, the λ_{ij}^{OD} effect returns the impacts of O-D separation on the resulting cell estimate, while λ_i^O represents the size of origin effect. Given a completely filled in flows matrix Equation 1 will reproduce the cell estimates exactly. We are interested in how such a model performs with missing data.

In the CFS we will be missing a large number of $\{ijm\}$ cells: as well as some $\{ij\}$ and other two-dimensional cells. Setting such cell values initially = 1.0 (log = 0.0) and applying Equation 1 we can obtain an estimate of each cell's missing value from the reported cell values. Such estimates are often termed minimum information estimates. Better yet, we can introduce entirely new data into the problem. An appealing feature of this sort of IPF modeling, as described in (*19*), are the many possibilities for treating missing data elements. In particular, we can combine data from the CFS matrix with data from other sources, such as the railcar Waybills (suitably modified to match CFS regions and commodity classes). Here this Waybills data can be used as a second estimate or "data model" of the rail flows in each commodity class. We can do this in a number of ways. We can replace CFS-based missing cell data with waybill estimated values and then use IPF to bring the full matrix (in all four dimensions) back into compliance with the original CFS flow margins. We can also treat the railcar waybill flows as though they were a separate dimension or set of commodity specific tables in the rail portion of the CFS flows matrix, and fill in the missing valued cells using a combined CFS and waybills-inclusive log-linear model.

We can carry out the same operations on those parts of the commodity flow matrix involved with water and air freight transportation, for example using Corps of Engineers and Office of Airline Information data respectively, in place of the railcar Waybill information as the second "model" of these flows. Finally, we can carry out a similar operation for truck shipments: but in this case we will have to substitute an actual model of flows in place of a second data set. An additional possibility here is to incorporate a set of travel distance intervals as yet another dimension into the log-linear modeling solution, broken down by commodity class. The CFS reports tons moved in the mileage ranges less than 50, 50 to 99, 100 to 249, 250 to 499, 500 to 749, 750 to 999, 1,000 to 1,499, 1,500 to 2,000, and over 2,000 mi.

Constructing Spatially Detailed O-D Matrices

Construction of Spatially Detailed Truck Trip Matrices

This is perhaps the most pressing data need at the present time for both statewide and metropolitan area planning. Without the equivalent of a spatially explicit railcar waybill or inland waterways vessel manifest, truck trip data falls to the CFS, the VIUS, route specific truck traffic counts, and any local trucking survey data that may (very occasionally) become available. CFS data is best suited to statewide analyses involving long-haul freight activity. It can however offer some useful regional control totals, as well as value-to-weight statistics, for use in metropolitan area studies. The following discussion assumes a statewide or similar (e.g., multistate corridor) type of planning application.

One way to create O-D flows at a level of spatial detail finer than CFS regions is to combine CFS data with county-based economic activity data. This involves allocating CFS-based freight activity across counties based on the volume (earnings, employment levels) of freight-generating industrial activity reported by Economic Census. One way to accomplish this is to pass this economic activity data through a regional input-output model that translated dollars of industrial activity into dollars of commodities produced and consumed (*22–25*). Recent Census releases of annual employment and payroll data at the five-digit zip code now allow modelers to consider using different within-CFS regional aggregations of this data that may in some cases (some commodities) be better suited to the problem of generating spatially explicit freight flows than counties, especially if the next step is to turn these flows into trucks and assign them to the highway network.

Input–Output (I-O) software, including well supported and commercially available codes, is now generally available, allowing analysts to associate the inputs and outputs of different commodities with specific industries. Where data on the annual production of a specific commodity class already exists (e.g., bushels of grain produced in a county) the principal use of I-O modeling is in the consumption, and therefore destination, end of an O-D movement. What is required is a commodity-to-industry conversion table that associates the amount of each commodity required to produce a unit of that industry's output. We can represent this as follows. Let $u(g,n)$ refer to the sales of commodity g to industry n. Given this data for all $n = 1, 2, \ldots N$ industries in a region/county/zip code area, we can compute $Q(g)$, the total quantity of commodity g consumed as:

$$Q(g) = u(g,1) + \ldots + u(g,N) + e(g) \tag{2}$$

In practice further adjustments to the data are usually needed. Getting county-level or zip code area tonnages from this approach then means using CFS provided ton-per-dollar shipped statistics. Getting the number of truck trips from these tonnage values also requires suitable data on average truck payloads, statistics that need to include the percentage of empty as well as fully or partially loaded truck trips (e.g. backhauls) involved: see (*15*) and (*16*) for example applications, both using VIUS truck weight data.

Given such a spatially disaggregated set of commodity specific productions (Os) and consumption totals (Ds), the next step is to link the two to create a set of O-D flows. One way to do this is to calibrate a set of commodity specific spatial interaction models, subject to CFS region-to-region control totals. Doing so, however, requires additional information, such as the average distance shipped (per ton) or a distribution of trips over different distance intervals. Again, CFS can provide both of these types of statistic, but for limited commodity or regional breakdowns.

Given a set of truck flows there is then the issue of validating the resulting estimates. About the only truck activity data available for this purpose in most regions is average annual daily truck traffic counts, including the counts states report to the FHWA as part of their Highway Performance Monitoring System (HPMS) submissions[9]. If these counts can be grouped into appropriate O-D specific traffic corridors then some level of aggregate truck trip comparisons might be attempted. However, for this approach to be meaningful requires both a reliable set of representative truck counting sites and counters with the ability to identify trucks of different sizes and axle configurations. It also requires a reliable means of assigning commodities to truckloads. A more careful look at how truck count data might be used is in order. A number of mathematical programming models have been developed that allow modelers to combine O-D and link count data to estimated truck movement matrices (*26–28*). Applications to date have focused largely on the intra-urban scale. For example, see (*27*), which describes a flexible approach that lets the analyst place greater reliance on either the O-D or count data, as warranted.

Construction of Rail, Water, and Intermodal Trip Matrices

This usually occurs as part of long-haul freight studies. Here the CFS can play an important role in support of the modally more comprehensive (in commodity coverage terms) Waybills and Waterborne Commerce data sets. Among other things the creation of rail or waterway inter-regional O-Ds means dealing with the issue of truck drays to and from these line-haul modes. Here the CFS offers some annual region-to-region control totals. It might also be used in the future to provide distributions of estimated truck–rail, truck–water, and also truck–airport drayage distances. Data on such draying activity is of considerable importance to states where a significant volume of intermodal freight crosses their borders. Truck draying costs, and hence distances, can also be important in determining the geographic size of the shipper "market area" served by water and rail modes, and by specific airports. These drayage distances are currently estimated for the CFS by Oak Ridge National Laboratory. How well they reproduce actual drays needs to be established. The key unknown here is the true location of the rail or river loading/unloading dock for each shipment. Perhaps one or more of the data collection approaches discussed below can help with this question.

NEW APPROACHES TO DATA COLLECTION: SOME PROMISING DIRECTIONS

With funds for freight data collection limited at all levels of government we need to take full advantage of any opportunities that come along to fill in the freight movement picture. The following are three areas worth exploring:

Linking CFS, VIUS, and Other Trucking Data

As identified in (*4*) the principal options for sampling truck trip activity patterns, besides a shipper survey such as the CFS, are (*a*) vehicle-based sampling, (*b*) vehicle–driver intercept surveys, and (*c*) vehicle tracking surveys. Each approach has its pros and cons. The costs of options (*b*) and (*c*) render them unrealistic as nationwide sampling options. Vehicle identification number-based sampling is currently used by the VIUS, and here a number of interesting options may be worth researching, such as

- The collection of truck activity diaries for a sub-sampling of vehicles covered by the VIUS, focusing on vehicle routing, backhauling, repositioning, and operating speed aspects of the freight pickup and delivery operation.
- A tie-in between a more continuous truck trip dairy-based survey and the currently quinquennial VIUS, making use of the latter's sample frame[10]. Diaries might be collected on an annual, rotational basis, possibly with different types of vehicles or commodities selected each year for diary completion. The characteristics of the vehicles reported in these travel diaries might then be tied to operating characteristics in the larger, 5-year vehicle sample.
- A more radical approach might include a redesigned CFS, such as an alternating bi-annual shipper (CFS) and motor vehicle (VIUS) survey program, with the latter including some form of truck trip diary sampling.

It is also worth exploring how we might tie local truck/commercial vehicle activity surveys, as occasionally collected by metropolitan planning organizations, to such developments in data collection.

The International Trade Data System: How (and When) Can We Use It?

International freight data collection in the United States is currently undergoing a major change. In the near future international trade data will flow to the U.S. Department of Transportation (USDOT) and some 104 other federal agencies (at last count) via the International Trade Data System (ITDS)[11]. ITDS is a federal government information technology initiative (Initiative IT06) to coordinate, standardize, and ultimately simplify our federal border clearance and other international trade and transportation processes. It will enhance and replace the Automated Commercial System currently used by the Bureau of Customs and Border Protection within the Department of Homeland Security. Traders will submit standard electronic data for imports or exports only once, with the ITDS system serving as a "single window" system through which trade transactions data can flow between private traders and over 100 federal agencies involved in international trade.

The ITDS is a timely response to the federal government's need for much greater visibility of incoming foreign cargos in this time of heightened concern over terrorist actions. To this end each federal agency submits a list of data elements it deems key to its operations. The USDOT is developing five different portals into the ITDS: one each for MARAD, Federal Motor Carrier Safety Administration (FMCSA), NHTSA, FAA, and a BTS-supported and -maintained portal that will supply freight movement data to other modal administrations within the USDOT (FHWA, FRA). It remains to be seen how and when this data becomes available for general use;

and raises the question of how such data might be used to study complete end-to-end movements of freight through international supply chains.

Freight Informatics: Cargo Tracking, Supply Chains and Electronic Manifests

Freight Informatics

The growing presence of many different kinds of real time information gathering technology means that future traffic data collection is going to make greater use of non-survey-based approaches. Hopefully these methods will prove less expensive than traditional survey methods in the not too distant future. Existing technologies include the following (*4*):

- Active roadway sensors: fiber optic sensors, inductive loop detectors, magnetic sensors, piezoelectric sensors, pneumatic road tube, weigh-in-motion sensors;
- Passive roadway sensors: Infrared sensors, microwave radar, passive acoustic array sensors, ultrasonic sensors, video image sensors;
- On-board sensors: Bar-code scanners, microchip-based smart cards, radio frequency identification devices (RFID) and remote intelligent communication (RIC), smart active labels (SALs), satellite/Global Positioning System (GPS)-based vehicle tracking; and
- Wide-area sensors: IKONOS satellite imagery, light detection and ranging, small plane, helicopter, and uninhabited autonomous vehicles/micro aerial vehicles (UAVs/MAVs).

An RFID system typically consists of a tag or label containing data storage, an antenna to communicate with the tag, and a controller to manage the communication between the antenna and the computer. An RFID tag can be embedded in a package or placed on a person. Combined with RIC technology, radio frequency-based wireless reporting can be used to track the location, condition, and content of goods at every stage in a product's supply chain, and do it in near real time. This includes the emerging technology of SALs which use RFID tags containing an internally powered microchip linked to an antenna for wireless reception and transmission purposes. A read/write mode, suitably powered RFID can be used as a dynamic, electronic cargo manifest. The potential for increased cargo security alone is going to bring this sort of "smart tag" technology into the mainstream for freight and inventory management.

A now widely used tracking technology is GPS. The commercial component of the GPS is a worldwide radio-navigation system formed by linking together 24 orbiting satellites and their network of ground stations. Vehicle or cargo tracking down to a few meters is already possible, with further spatial refinements (down to centimeters) under development. Useful tracking of vehicles might base sampling on high volume highway corridors or high volume freight gateways, possibly on an annual, rotational basis. The use of MAVs seems likely to bring down the price of surveillance at major traffic intersections or along major traffic corridors. Here the potential for freight flow analysis would be in the combination of traffic count data from these aerial devices with O-D survey data, possibly as joint inputs to the "link O-D models" mentioned in the section above.

The potential for largely automated freight data collection seems obvious, given enough time and resources currently being used to develop informatics technologies. Less obvious, and in need of study, is the use to which this information can be put by public agencies. We also need to ask what other information technology is out there and what other uses we can find for those listed above.

Tracking Freight Supply Chains

Hopefully the ITDS program will eliminate the current weaknesses in our import–export data. If it can do so, then we might also ask whether a similar effort might not be used to collect the physical O-D movements of domestic cargos. A logical next step would then be to combine both of these domestic and international data sets. Perhaps the ultimate expression of this idea is the creation of a universal electronic manifest. Such a manifest would provide essential information for shippers and carriers to manage inventories and logistics, as well as meet the documentation needs of domestic and international trade, hazardous cargo movement, and the growing security needs of domestic transportation. It would modernize existing paper-based waybills and allow cargo tracking across all modes.

Unless or until the ITDS or another UEM-based system evolves, we will have to rely heavily on data integration tools such as those discussed above to piece together this goods movement puzzle. Recent U.S. experiments with an electronic supply chain manifest suggest that information technology applications that benefit private freight movement agents will advance quickly. A goal for the FMCSA Commercial Vehicle Information Systems and Networks (CVISN) program (*29*) is the integration and automatic processing of multisourced data on the carrier, vehicle, driver, and cargo (including oversize and hazard designations) associated with domestic or international commercial truck trips. A CVISN link to the ITDS/Automated Commercial Environment has also been discussed. This sort of data integration activity across federally supported programs needs further exploration. What other useful connections can we find if we look at the whole breadth of the government's information technology-based (IT) information gathering activities?

Similar experiments with different freight IT applications are ongoing in Europe (*30*). How might these different technologies be used to develop aggregate statistics for use by participating public agencies, notably in the estimation of hourly, daily, and (through aggregation) seasonal O-D freight flow volumes? There is also considerable potential, some of it already being tapped, for monitoring and measuring the travel speeds and en route delays associated with location specific truck, rail, and waterborne commerce movements.

Such possibilities also suggest an alternative approach to survey design: the use of "supply chain surveys," involving a mixed sample drawn from a mix of establishment types, i.e., from shippers, carriers, distributors, terminal operators, receivers, and also freight forwarders. More than one U.S. experiment in the tracking of complete O-D supply chains is currently underway, again making use of recent developments in electronic reporting. An experiment with this type of data collection was also recently tried in Europe, based on more traditional data collection methods involving both face to face and telephone interviews with the different supply chain participants (*31*). This sort of survey offers a greater understanding of the full logistics costs involved in moving freight from source to destination. However, it also requires a potentially complex, multistage sample design. Both a benefit (and an added complication) of such a survey is the collection of data on who is responsible for moving the freight at each stage, and what institutional arrangements (e.g., between carrier and shipper, shipper and broker, shipper and receiver) have been made to facilitate this. The European experience identifies a number of problems with achieving good response rates, and with higher costs per successful response than with traditional shipper surveys (*31*). In particular, the probability of getting a complete description of complex, multi-actor supply chains proved to be low. How such an approach might stand up under different sampling designs and under mandatory reporting requirements is currently unclear[12].

SUMMARY

The patchwork quilt that is today's freight data coverage requires a good deal of effort to create even a base case set of commodity flows for use in regional planning studies. Spatial disaggregation of flows requires that the CFS be combined with other data sets. Combined commodity and spatial disaggregation within the CFS is especially limited in this context. A number of quantitative methods can be used to help us get the most out of both the CFS and other federally supported freight movement data sets. New ways of collecting freight data, and especially data on truck movements, offer the promise of not only greater coverage but also greater understanding of how freight really moves through the transportation system. Directions for future research and development identified in the paper include

- The use of iterative proportional fitting and log-linear modeling techniques, including spatial interaction modeling, to produce maximum likelihood estimates of empty cell values in commodity and mode specific freight movement matrices;
- The use of economic activity data and spatial input–output models to disaggregate truck trip commodity flow matrices;
- The tracking of complete source-to-final market product supply chains, following the ownership as well as the physical and geographic aspects of en route cargo transfers;
- The measurement of truck draying distances associated with intermodal shipments;
- The use of both CFS and foreign trade data to develop spatially detailed estimates of the true origins, destinations, and modes used when transporting imported and exported goods, converted to both tons and dollar valued trades;
- The development of functional linkages between the CFS, VIUS (possibly expanded to include truck trip diaries), and perhaps also local area truck trip activity surveys;
- The gradual incorporation of vehicle counting and vehicle tracking information into the estimation and validation of truck-based commodity flows; and
- The rationalization of foreign trade data statistics in support of mode and O-D tracking.

The need for better freight movement data is considerable, and the costs of using poor quality data to plan future investments are likely to rise along with the pressure such movements place on our transportation infrastructures. Perhaps one final research task ought to be an objective assessment of such costs versus the benefits of providing better data to public agency planners and decision-makers.

REFERENCES

1. *Special Report 276: A Concept for a National Freight Data Program*. Transportation Research Board of the National Academies, Washington, D.C., 2003. http://gulliver.trb.org/news/blurb_detail.asp?id=1730
2. *Conference Synthesis: Data Needs in the Changing World of Logistics and Freight Transportation* (A. H. Meyburg and J. R. Mbwana, eds). New York State Department of Transportation, Albany, 2002. http://www.dot.state.ny.us/tss/conference/synthesis/pfd.
3. Southworth, F. *The National Intermodal Transportation Data Base: Passenger and Goods Movement Components*. Bureau of Transportation Statistics, U.S. Department of Transportation, 1999.
4. Southworth, F. *A Preliminary Roadmap for the American Freight Data Program*. Bureau of Transportation Statistics, U.S. Department of Transportation, 2003.
5. Mani, A., and J. Prozzi. *State-of-Practice in Freight Data: A Review of Available Freight Data in the U.S.* Project Report 0-4713-P2. Center for Transportation Research, University of Texas at Austin, 2004.
6. 2002 Commodity Flow Survey. Report EC02TCF-US. Census Bureau, U.S. Department of Commerce, Suitland, Md., 2004. http://www.census.gov/econ/www/cfsnew.html.
7. Capelle, R. B., et al. *Freight USA: Highlights from the 1997 Commodity Flow Survey and Other Sources*. Report prepared by Oak Ridge National Laboratory for the Bureau of Transportation Statistics, U.S. Department of Transportation, 2000.
8. 2002 United States Economic Census: Vehicle Inventory and Use Survey. EC02TV-US. Census Bureau, U.S. Department of Commerce, 2002. http://www.census.gov/svsd/ www/02vehinv.html.
9. Gordon, P., and Q. Pan. Assembling and Processing Freight Shipment Data: Developing a GIS-Based Origin Destination Matrix for Southern Freight Flows. University of Southern California, Los Angeles, 2001. http://www.metrans.org/Research/Final_Report/99-25_Final.htm.
10. Clarke, M., and L. Cheng. Initial Development of an Innovative Commodity Flow and Truck Model for the Greater Los Angeles Basin. International Chinese Transportation Professionals Association Annual Conference, Washington, D.C., 2005. http://www.ictpaweb.org/publication/DC/fu-2005conpaper/1-1-4- cargotechnicalpaper.doc.
11. VPRA Associates and Cambridge Systematics Inc. Goods Movement Truck Count Study. Report prepared for the Southern California Association of Governments, Los Angeles, 2002.
12. Liu, N. L., and P. Vilain. Estimating Commodity Inflows to a Substate Region Using Input-Output Data: Commodity Flow Survey Accuracy Tests. *Journal of Transportation and Statistics*, Vol. 7, No. 1, 2004, pp. 23–37.
13. Brogan, J. J., S. C. Brich, and M. J. Demetsky. Application of a Statewide Intermodal Freight Planning Methodology. Report VTRC 02-R5. Virginia Transportation Research Council Charlottesville, 2001. http://www.virginiadot.org/vtrc/main/online_reports/pdf/02-r5.pdf.
14. Parsons-Brinckerhoff-QD, Inc. Developing a Framework to Assess Goods Movement in New Jersey. Technical Appendix. Report prepared for the New Jersey Department of Transportation, Trenton, 2002. http://www.state.nj.us/transportation/works/freight/ documents/GMFtechApp.pdf.
15. Beagan, D., and L. Grezenback. *Freight Impacts on Ohio's Roadways.* Report FHWA/OH-2002/026. Ohio Department of Transportation. Columbus, 2002. http://www.dot.state.oh.us/ planning/Freight/freight_impacts_report.htm.
16. Battelle Team, Reebie Associates, Wilbur Smith Associates, and Global Insight. *Freight Analysis Framework: Derivation of the FAF Database and Forecast.* Report prepared for the FHWA, U.S. Department of Transportation, 2002. http://ops.fhwa.dot.gov/freight/ documents/WebMeth/index.htm#1998.
17. Measurement of Ton-Miles and Value-Miles of International Trade Traffic Carried by Highways for Each State. *International Trade Traffic Study* (TEA-21, Section 5115). Bureau of Transportation Statistics, U.S. Department of Transportation, 2003.

18. Willekens, F. J. Log-Linear Modeling of Spatial Interaction. *Papers of the Regional Science Association*, Vol. 52, 1983, pp. 87–205.
19. Southworth, F., and B. E. Peterson. Disaggregation within National Vehicle Miles of Travel and Fuel Use Forecasts in the United States. In *Spatial Energy Analysis* (L. Lundqvist, L.-G. Mattsson, and E. A. Eriksson, eds), 1990.
20. Agresti, A. *Categorical Data Analysis*. Wiley, N.Y., 1990.
21. Wrigley, N. *Categorical Data Analysis for Geographers and Environmental Scientists*. Blackburn Press, U.K., 1985.
22. Sorratini, A. J., and R. L. Smith, Jr. Development of a Statewide Truck Trip Forecasting Model Based on Commodity Flows and Input–Output Coefficients. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1707,* TRB, National Research Council, Washington, D.C., 2000, pp. 49–55.
23. Liu, N. L., and P. Vilain. Estimating Commodity Inflows to a Substate Region Using Input-Output Data: Commodity Flow Survey Accuracy Tests. *Journal of Transportation and Statistics*, Vol. 7, No. 1, 2004, pp. 23–37.
24. Jackson, R. W., et al. A Method for Constructing Commodity by Industry Flow Matrices. Research Paper 2004-5. Regional Research Institute, West Virginia University, Morgantown, 2004. http://www.rri.wvu.edu/pdffiles/jacksonetal2004-5wp.pdf.
25. Canning, P., and Z. Wang. A Flexible Modeling Framework to Estimate Interregional Trade Patterns and Input–Output Accounts. *Journal of Regional Science*, Vol. 45, No. 3, 2004.
26. Van Zuylen, J. H., and L. G. Willumsen. The Most Likely Trip Matrix Estimated from Traffic Counts. *Transportation Research 14B,* 1980, pp. 281–293.
27. List, G. F., and M. A. Turnquist. Estimating Truck Travel Patterns in Urban Areas. In *Transportation Research Record 1430,* TRB, National Research Council, Washington, D.C., 1994, pp. 1–9.
28. Sherali, H. D., R. Sivanandan, and A. G. Hobeika. A Linear Programming Approach for Synthesizing Origin–Destination Trip Tables from Link Traffic Volumes. *Transportation Research 28B,* 1994, pp. 213–223.
29. Commercial Vehicle Information Systems and Networks (CVISN). Recommendations for Primary Identifiers. Baseline Version V1.0. White Paper. Federal Motor Carrier Safety Administration. Washington, D.C., 2002.
30. Sustainable Urban and Regional Freight Flows (SURFF). EU/DG XIII Transport Telematics Applications Program, TR 1053, European Union, 2000. http://www.euroweb.net/ surff/about.htm.
31. Rizet, C., M. Guilbauly, J. C. van Merijeren, E. Bjister, and M. Houee. *Tracking Along the Transport Chain Via the Shippers Survey*. International Conference on Transport Survey Quality and Innovation, 2003. http://www.its.usyd.edu.au/.

NOTES

1. I.e., shipments requiring end-on transfers between two different modes to reach their destination.
2. http://www.stb.dot.gov/stb/industry/econ_waybill.html.
3. http://www.iwr.usace.army.mil/ndc/.
4. http://www.census.gov/svsd/www/products.html.
5. http://www.bts.gov/transborder/.
6. http://www.marad.dot.gov/Marad_Statistics/index.html.
7. http://www.piers.com/default2.asp.
8. http://www.reebie.com/images/transearch.asp. This database was also used in the FHWA's recent effort to develop a nationwide set of freight movement maps and supporting commodity flow matrices and truck-to-highway assignments; see (*16*).
9. HPMS. http://www.fhwa.dot.gov/policy/ohpi/hpms/.

10. For the 2002 VIUS the Census Bureau purchased a sample frame generated by R.L. Polk and Co.
11. http://www.itds.treas.gov/sitemap.html.
12. Even if the average response rate is 75% at each stage in the supply chain (a little higher than the rate experienced by the CFS) a three-interview chain would yield only a 42% successful completion rate, assuming independence of responses at each stage.

Shipment Characteristics in the Commodity Flow Survey
Can One Describe an Elephant?

BRUCE LAMBERT
U. S. Army Corps of Engineers

Due to concerns over the future use and conditions of the nation's infrastructure, a national discussion on freight policy has emerged, driven in part by a perceived lack of understanding freight movements by federal, state, and local researchers and planners. This growing awareness of freight transportation has led to several conferences and a review of freight data and analytical programs on the national level (*1*). The "Data Needs in the Changing World of Logistics and Freight Transportation" findings recommended a new data architecture that examined the reason for freight movement, through comparable geographical and functional aggregations, with data that was kept current, developed jointly between the public and private sectors and used the latest technologies to track shipments. These findings are echoed in the other reports: current public freight data programs do not provide sufficient information for answering advanced inquiries into freight transportation. But simply describing the nature of freight shipments remains a difficult task.

John Godfrey Saxe's Poem "Describing the Elephant" transforms a Hindu folk story about six blind men who come upon an elephant into the familiar poem and moral tale. Each blind man describes what part of the elephant he first touched. The elephant's parts were described (in order) as a wall (its side), a spear (its tusk), a snake (its trunk), a tree (its legs), a fan (its ears), and finally a rope (its tail). Understanding transportation can be as difficult as the blind man describing an elephant, because of the complexities of the nature of transportation. For example, an export shipment leaves a plant in the Atlanta, Georgia, area by truck, goes to a railhead to move intermodally by rail to the West Coast. That shipment involved two or more railroads, and the product arrives at Long Beach, California, where the container is drayed into the port, loaded on a vessel, and moved to Japan. During that movement, the product changed mode several times and may have also changed ownership. The shipment passed through several states and urban areas during its journey. The rail traffic moved on a stack train with other shipments, related only to each other by the rail shipment itself. There were costs associated with the shipment, such as rates, etc., but also economic benefits.

The Commodity Flow Survey (CFS) serves as a source for understanding the gross movement of freight in the United States. There currently exists no database within the public dominion that links geography, modes, and commodities within a common collection and analytical framework that will satisfy every user's expectations. Given a renewed focus on freight movements and the data and analytical tools necessary to assist decision makers, the current CFS may not provide enough information on shipment characteristics, but would serve as one instrument in developing new freight data architecture. The questions regarding shipment characteristics and the CFS involve: What are shipment characteristics and which shipment characteristics are useful for both planning and policy purposes? What shipment characteristics are in the CFS survey, and what additional items are necessary to either maintain or improve the CFS in 2007?

SHIPMENT CHARACTERISTICS FOR PLANNING AND POLICY PURPOSES

Describing transportation (or prioritizing data needs) is also complicated by the different uses for the data once it is collected (*2*). Like the six blind men, shipment information can satisfy different uses or applications, but the parts may appear inconsistent without understanding the whole. The act of describing shipment characteristics becomes important for five distinct applications:

1. To determine the economic value of transportation activities,
2. To estimate infrastructure planning and maintenance needs,
3. To provide market research,
4. To examine transportation related policies, and
5. To assist in evaluating various security options.

These five represent a brief list, but do provide a starting point for discussing the importance of understanding shipment characteristics and specific user needs across many different applications.

The Economic Value of Transportation Activities

When examining the economic value of transportation, users tend to focus on the cargo's value as a method of linking freight transportation in a manner consistent with other activities. For other analysis, freight information expressed in taxes, revenue activities, jobs or interdependencies provides meaningful insights. Information on cargo value, especially when linked to other data elements, often becomes the most important shipping characteristic to the non-transportation community. In many ways, knowledge about the actual road infrastructure is unimportant until linked to the relevant economic planning and development activities that freight transportation (infrastructure) supports.

Based on personal experience at the Port of Long Beach in the early 1990s, the port's official traffic information reported metric revenue tons (MRT)—a measure of either the cargo's weight or value, depending on which would generate the higher tariff rate. Reporters struggled with the concept that a measure would depend upon the physical dimensions of the cargo but it was not comparable to the MRT reported by other ports, which had different tariff structures. Tonnage proved difficult to explain because most reporters did not understand the scale associated with cargo movements. Only when freight activity was presented in dollars did reporters generally grasp the magnitude of the port's traffic volumes.

Research into the economic value of transportation activities may ask questions concerning the following.

- Are these characteristics linked to other databases to assess economic relationships?
- What is the economic importance of this activity? How many jobs does this create or support? How do industries relate with other sectors through transportation movements?
- What does this activity provide the local, regional, or national economy regarding revenues or taxes? Does this provide local sustainability?
- What are the major markets the local area serves?
- What are the most important industries (commodities)?

- What economic trends does shipment information provide?

Estimating Infrastructure Planning and Maintenance Needs

Given the large investment in infrastructure already made by federal, state, and local governments and the private sector (railways, harbors and terminals, locks and dams, roads, airports, pipelines, terminals, etc.), investment decisions require understanding a different set of shipment information. This information may come not only from commodity information but from other data sources. When examining system usage, commodity and shipment information becomes dependent upon understanding operational information. For example, how many units move through a certain point and what modes were they carried on, and what are the nature, costs, and timing of these connections? In one sense, shipment characteristics become a proxy for capturing information on the actual physical movements, but must be linked to other data elements to transform basic commodity information into actual units moving on the system. Other information may be included in the research, such as frequency of shipments, mode, or some linkages to other databases to better understand operational demands and infrastructure requirements.

Infrastructure planning and maintenance may ask questions concerning the following.

- What are the ultimate origins and destinations (O-Ds) of the cargo moving over the nation's infrastructure? How do these destinations (firms) consume the cargo?
- Do volume levels change by time of day or year?
- What cargo passes through the local area that did not originate within the local area (pass through cargo)? What does this mean for the local infrastructure and for the cargo itself? What would it mean if the cargo had to go via other routes?
- When will traffic volumes reach levels that require new investments?
- Can traffic switch to other areas or modes, delaying or changing investment needs? Does one want to retain traffic or have the goods flow through other regions?
- What factors (rising fuel costs, new regulatory mandates, new technologies, etc.) may place future pressures on the system?
- Can commodity information be translated into actual physical movements, such as number of trucks, barges, railcars, etc.? Can this be translated into other economic information?
- How often do shipments move to and from a certain facility, location, or region?
- Can one balance freight demands with other users for the same infrastructure segments?
- What operational improvements are necessary within a certain area, such as electronic tolling, metering on-off ramps, or scheduling traffic movements, to improve system utilization?
- How are current facilities being utilized?

Provide Market Research

Market research tends to focus on investment needs in equipment or location, trying to capture cargo, or understanding the direction of the market. As such, shipment characteristics, when compared against internal company statistics, become measures of market share while potentially discovering new markets. Market research can also help private transportation providers make more efficient decisions. In the public sector market research becomes a method to examine potential shifts in transportation services from proposed or pending legislation (regulation) or to understand if changing transportation patterns may result in different policy or planning efforts.

Market research may ask questions concerning the following:

- What forces are driving shippers? How does a company competitively respond?
- What is my competition doing? (This is done at many levels, ranging from national economic development down to a local business development.)
- What equipment is needed to carry the cargo?
- What are the shippers and carriers associated with a particular shipment?
- What are the prevalent transportation costs, rates, and service needs for particular shipments?
- What backhaul cargos exist? Can a company better control the movement of empty equipment?

Examining Transportation Policies

As with infrastructure planning, transportation policies look at changing infrastructure needs, but this may also include changing regulatory oversight. For example, changing truck size and weight limits may lead to more cargo switching between trucks or railroads. In addition, transportation policies may look at the distribution of transportation services or access within a region, the impact of transportation activities on a region, or the emissions associated with freight traffic. Policy questions seek to understand the nature of government's actions (or inaction) and the potential response by various groups to see if the policy will produce the desire outcome.

Transportation policy may ask these questions:

- Can traffic be switched from one mode to another? To what extent can or do policies impact mode choice?
- What potential changes can be made to the system to improve its operational nature?
- What amount of fuel taxes are collected in area?
- What relationship exists between industries and changes in infrastructure and transportation services?
- How does either infrastructure investment or transportation operations support economic growth?
- How responsive are shippers and carriers to regulatory changes?

Evaluating Various Security Options

The increased focus on security resulted in a growing awareness that the transportation system can be both the means of destruction and the target itself. When examining freight statistics, the question centers on developing methods to understand in real time the commodities moving on the nation's infrastructure (all modes), the relationship between infrastructure and economics, and understanding methods to quickly respond to incidents.

Security research may ask these questions:

- Can system vulnerability be assessed?
- Can system redundancy be evaluated to construct response plans?
- Can one develop in-transit visibility of the cargo information and routing choice?
- What are the commodities actually moving on the system, and where are they moving?
- Who are the actual shippers and receivers of these cargos?
- What modes tend to carry what commodities?
- What is the average shipment size for certain commodities?
- What are the population groups alongside a given transportation corridor or segment?
- What is the level of risk or exposure associated with freight shipments?

PRIORITIZING SHIPMENT CHARACTERISTICS NEEDS

Several reports highlighted the need for better freight data, but the specific shipment information was not included (*3*). Other reports outlined the most important shipment characteristics, many of which were included in the proposed elements listed in *TRB Special Report 276: A Concept for a National Freight Data Program*. This report listed the following elements as the basis for a national program: O-D; commodity characteristics, weight, and value; modes of shipment; routing and time of day; and vehicle/vessel type and configuration. These elements would serve as the basis for a national freight data program (*4*). A review of state transportation data will highlight the need of surveying carriers and firms regarding O-Ds, but generally state based research appears limited in its ability to understand traffic passing through to other markets.

When examining federal databases focusing on commodity movements, most data collection efforts contain the following: a geographic element (either an O-D or gateway facility), a time dimension (daily, monthly, or annual), and some commodity detail (at various levels of detail). By implication, some information is also available on at least one mode (rail, highway, etc.) depending upon the collection mechanism. Information on transportation activities may also include some information on equipment type, time of day, route, and carrier. International statistics will include value information, but generally domestic shipments do not report shipment value. With the exception of the rail waybill, rates (costs) are generally collected from sources separate from the shipment data collection.

A Hypothetical Database

Other parts of the conference will discuss collection methods, data quality, etc., but the nature of shipment characteristics provide the critical interface between the data collected and the ability to translate information into data. This section will examine a hypothetical database to further highlight the users' expectations on freight data, followed by a discussion on prioritizing shipment characteristics needs and the role of the researcher.

For this paper, shipment characteristics are defined as anything that describes the nature and movement of a product, including not only the shipment itself, but the transportation, supply chain, and equipment associated with that product shipment. This represents a very broad definition, but the ideal database would describe every element related to each individual shipment (*5*). After outlining the desired needs, the shipment characteristics reported by the CFS will be compared to this mix of anticipated data and analytical needs. For example, one hypothetical database (excluding the current CFS) would record and estimate the following characteristics for each shipment:

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

The list contains a mix of data elements, some of which could be directly calculated from existing data programs. Other elements would come from modeling efforts, revenue files, or operational databases. Thus, the ideal freight transportation data and analytical framework involves not only improving data collection efforts but also economic models.

The Role of the Researcher

The list in the previous section shows how complicated it is to describe transportation and freight movements, just as if one had asked blind men to describe the elephant. Unlike the blind-men, if these data elements represented the universe and were readily available one could answer the majority of current freight research needs. Shipment characteristics become one means to understand various dimensions of the freight industry. Some studies may give higher importance to various shipment characteristics. Economic research requires the value of the cargo; infrastructure planners desire frequency and weight estimates; market researchers desire the commodity type, equipment needs, and market size; policy research needs vary by operational considerations, equipment use, and infrastructure needs; and security concerns include company detail, volume, routing, modal use, and commodity detail. The ability to mix information purely related to the physical transportation activity itself with information on the shipment itself becomes more important given the lack of a single database covering all transportation activities. This leads researchers to use existing databases and make assumptions to extend the reported shipment information into other applications, especially uses beyond the original data program's design.

Unlike our example of blind men, most researchers have a clearer set of objectives when examining transportation activities. While the use of good information is critical, the researcher's ability to use the information provides an indication of the usefulness of any data program. First, the researcher must determine how important the information is for the completion of the research effort, especially given limited budgets and time to complete the study. Second, the researcher must determine "how good is good enough?" as the incremental gains from developing and using new data and models theoretically diminishes over times, but the gains occur at an increasing cost. Third, the researcher may or may not be experienced in using transportation data, which may result in either the use of an inferior database or the incorrect assumptions on shipments. The researcher may ultimately define the ability of the CFS (or any related database) to be useful in satisfying the original intent of the data collection effort—understanding the nature of freight shipments.

SHIPMENT CHARACTERISTICS CURRENTLY IN THE CFS

While it is important to discuss users' needs and expectations on freight data, the collection and reporting of shipment characteristics information is the most important function of the CFS. As such, the CFS's success depends upon its ability to provide relevant information on freight movements. This section discuss how the CFS defines a shipment, what information is actually collected (released) and how the CFS compares to the shipment characteristics in the hypothetical database described earlier.

CFS Shipment Definition

The nature of the shipment and the implications of observed activities become the basis for all other transportation elements within the CFS, as the commodity's characteristics determine the shipment's volume, frequency, transportation costs and rates, and mode.

The CFS defines a shipment as **"**A shipment is a single movement of goods, commodities, products etc., from your location to a customer or to another location of your company" (Figure 1). Full or partial truckloads are counted as a single shipment only if all commodities on the truck are destined for the same location. If a truck makes multiple deliveries on a route, each stop is counted as one shipment. Interoffice memos, payroll checks, or business correspondence are not considered shipments. Shipments such as refuse, scrap paper, waste, or recyclable materials are not considered shipments unless the establishment is in the business of selling or providing these materials. This definition of each unique shipment between a shipper and a customer implies the observation of shipments moving on the system, and potentially provides meaningful insights into the transportation system not readily captured in other data sets.

What we mean by a "shipment":

For the purposes of this survey, a shipment is a single movement of goods, commodities, products, etc. from your location to a customer or to another location of your company.

"Commodities" refer to items that your location produces, sells, or distributes, *not* to items that are considered by-products of your location's operation.

What we don't mean by a "shipment":

Do *not* include as shipments items such as inter-office memos, payroll checks, business correspondence, etc.

Do *not* include as shipments items such as refuse, scrap paper, waste, and recyclable materials **unless** your location is in the business of selling or providing these materials to others.

A special note about "shipments":

A full, or partial, truckload should be counted as a single shipment only if all the commodities on the truck are destined for one location.

If a truck makes multiple deliveries on a route, **please count each stop as one shipment.**

Item E: Sampling Instructions

If you reported in item D that you had 40 or fewer shipments for the week, complete Item F (Shipment Characteristics) for **ALL** of your shipments covered by the one-week reporting period.

If you reported in item D that you had more than 40 shipments for the week, follow the instructions in item E in order to select a sample of shipments on which to report in item F.

FIGURE 1 Definitions of shipment from the 2002 CFS; Instruction Guide, CFS-1100, U.S. Department of Commerce.

Actual Shipment Information in the CFS

The 2002 CFS Instruction Guide includes a copy of the actual survey (Figure 2*a-d*), as well as instructions on completing the survey. The Survey Form is included in Figure2*a-d*. (The major shipment characteristics remained unchanged between the 1997 and 2002 surveys.) Based upon the shipment characteristics surveyed, the CFS actually records the following information presented in Table 1.

The CFS does a very good job providing a format for collecting information that survey respondents can complete based on the various codes used in the survey (these items are indicated by italics). The many items collected provide additional information to assist CFS data checks but are not reported in the final reports. Several characteristics could be inferred from the actual business establishment itself. Items A–E and G involve basic housekeeping information, including the business address, total number of shipments, business operations.

When researchers mention the CFS, it is generally in relationship to the final outputs, mostly the summarized reports and the table files. Generally the following elements are reported

TABLE 1 Shipment Characteristics in the CFS

Item	Source	Uses Lookup Codes	Reported in Summary Tables
Total number of shipments	Item D	No	No
Selection rate	Item E	No	No
Shipment identification number	Item F	No	No
Shipment date (month, day)	Item F	No	No
Shipment value	Item F	No	Yes
Shipment weight in pounds	Item F	No	Yes
Commodity code from standard classification of transported goods (SCTG) list	Item F	Yes	Yes
Commodity Description at the five-digit level	Item F	No	No
Hazmat flag [United Nations(UN) or North American (NA) number]	Item F	Yes	Yes
U.S. destination (city, state, zip code): gateway for export shipment	Item F	No	Yes
Modes of transport	Item F	Yes	Yes
Foreign destination (exports only: city, country)	Item F	No	Yes
Export mode	Item F	Yes	No

in every table: tons, value, ton-miles, and average distance shipped. These shipment characteristics are reported in various tables summarized against some combination of mode, commodity, shipment size, and geographic detail. When compared to the survey form, several shipment characteristics are aggregated or dropped entirely (elements in bold are not reported in the summary tables).

For example, Table 2 presents the typical CFS file format. Value, tons, ton-miles and average miles per shipment are reported across the top of the table. A descriptor, in this case mode, is listed on the left. Only two of the four key variables, tons and value, are summarized by the actual survey responses. Both ton-miles and average miles shipped are developed using the shipment origin, destination, and mode of transportation. Bureau of Transportation Statistics (BTS) employees produced mileage calculations for the 2002 CFS, working on-site at the Census Bureau.

Comparison of the CFS to the Hypothetical Ideal

The CFS's definition of shipment characteristics is narrower than the definition presented earlier. The CFS list only includes primary data collected via its surveys, and of the 19 items listed in the hypothetical ideal, the CFS only reported on eight items with various levels of detail. The CFS provides good information on hazmat flags for hazard cargos, value, tons, mode, commodity, export flows and mode, and CFS regional O-D flows, but not all survey information is published.

OMB No. 0607-0889: Approval Expires 10/31/2004

FORM CFS-1000 (12-11-2001)

2002 COMMODITY FLOW SURVEY
CENSUS OF TRANSPORTATION

U.S. DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. CENSUS BUREAU

Reporting period:

Please return by:

RETURN TO

U.S. CENSUS BUREAU
1201 East 10th Street
Jeffersonville IN 47132-0001

(Please correct any error in name, address, and ZIP Code)

BEFORE COMPLETING YOUR REPORT, please read the accompanying Instruction Guide. Visit our website at **www.census.gov/CFS** or telephone us at **1-800-772-7851** if you would like additional information.

This survey requests a limited amount of data on a sample of your outbound shipments. Its purpose is to develop information on the characteristics of freight flows in the United States. This information is essential for understanding transportation markets, investment needs, and the economic, safety, energy, and environmental consequences of transportation.

If book figures are not available for the information we are requesting, estimates are acceptable.

Thank you for your timely completion of this report.

Item A Is the establishment name shown in the mailing address correct?

1 ☐ Yes
2 ☐ No — *Enter correct name.*

Item B *Mark (X) the ONE box which best describes this establishment during the one-week reporting period shown above.*

1 ☐ In operation
2 ☐ Temporarily or seasonally inactive
3 ☐ Ceased operation — *Give date*

Month	Day	Year

Item C Is this establishment's physical location the same as the address shown above? (PO boxes or rural routes are not physical locations.)

1 ☐ Yes
2 ☐ No — *Enter physical location below.*

Number and street		
City, town, village, etc.	State	ZIP Code

NOTE — The rest of this questionnaire requests information about shipments (or deliveries) from the establishment located at the address in the mailing label.

If you entered a different address in Item C — *Please complete the form for shipments originating from the location listed in Item C.*

Item D **TOTAL NUMBER OF SHIPMENTS** — Please enter the total number of outbound shipments (or deliveries), including customer pick-up, for the one-week reporting period shown above. If book figures are not available, please provide your best estimate.

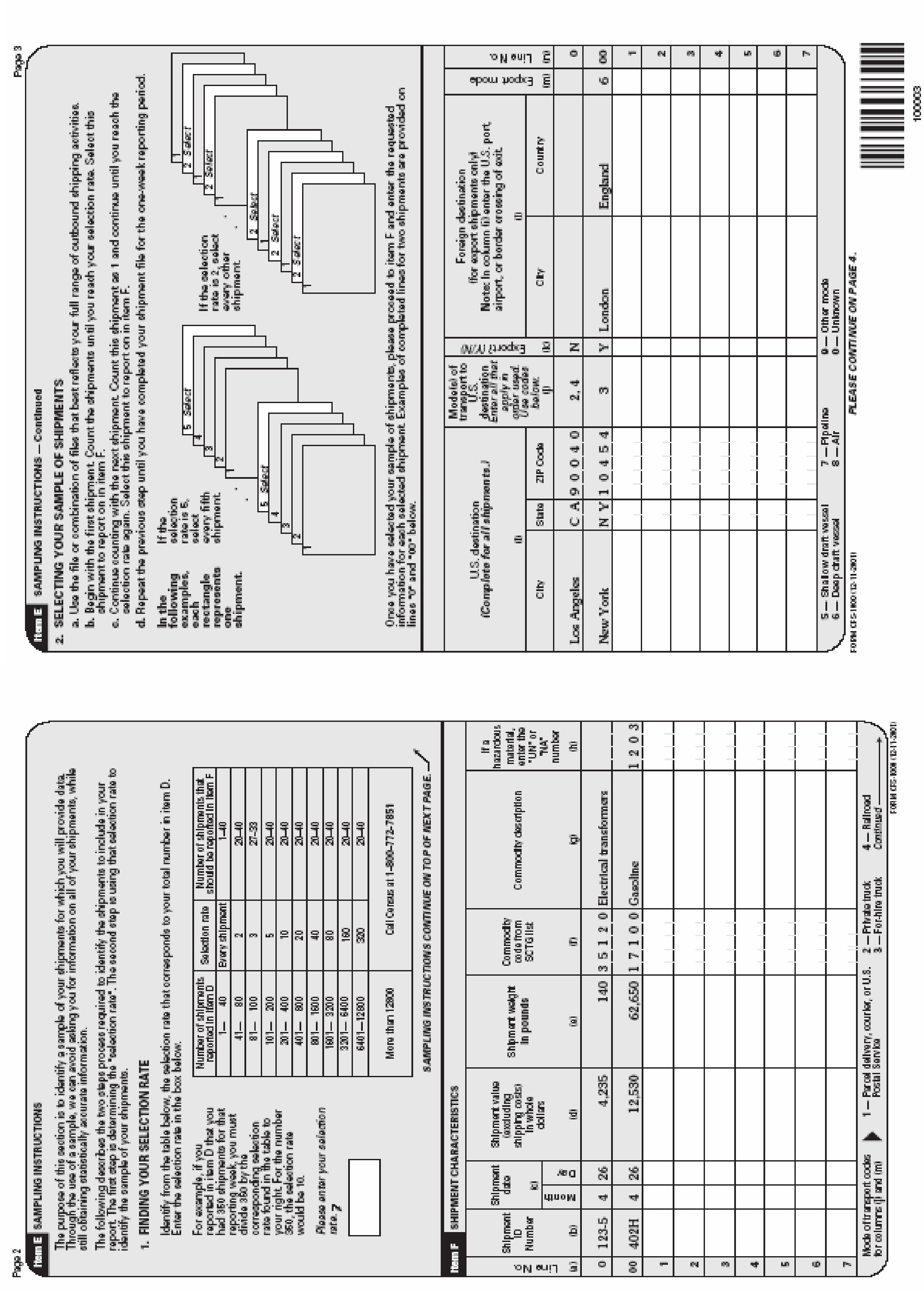
This number should reflect **ALL** shipments (not just those listed in Item F) and deliveries leaving this location during the one-week reporting period. *Please see Instruction Guide for a definition of "shipment."*

DO NOT PROCEED UNTIL YOU HAVE COMPLETED ITEM D.

YOUR RESPONSE IS REQUIRED BY LAW. Title 13, United States Code, requires businesses and other organizations that receive this questionnaire to answer the questions and return the report to the U.S. Census Bureau. By the same law, **YOUR REPORT IS CONFIDENTIAL.** It may be seen only by persons sworn to uphold the confidentiality of Census Bureau information and may be used only for statistical purposes. Further, copies retained in respondents' files are immune from legal process.

100001

FIGURE 2*a* 2002 CFS Instruction Guide.



Page 2

Item E SAMPLING INSTRUCTIONS

The purpose of this section is to identify a sample of your shipments for which you will provide data. Through the use of a sample, we can avoid asking you for information on all of your shipments, while still obtaining statistically accurate information.

The following describes the two steps process required to identify the shipments to include in your report. The first step is determining the "selection rate". The second step is using that selection rate to identify the sample of your shipments.

1. FINDING YOUR SELECTION RATE

Identify from the table below, the selection rate that corresponds to your total number in item D. Enter the selection rate in the box below.

For example, if you reported in item D that you had 350 shipments for that reporting week, you must divide 350 by the corresponding selection rate found in the table to your right. For the number 350, the selection rate would be 10.

Please enter your selection rate.

Number of shipments reported in Item D	Selection rate	Number of shipments that should be reported in Item F
1— 40	Every shipment	1–40
41— 80	2	20–40
81— 100	3	27–33
101— 200	5	20–40
201— 400	10	20–40
401— 800	20	20–40
801— 1600	40	20–40
1601— 3200	80	20–40
3201— 6400	160	20–40
6401—12800	320	20–40
More than 12800	Call Census at 1-800-772-7851	

SAMPLING INSTRUCTIONS CONTINUE ON TOP OF NEXT PAGE.

Item F SHIPMENT CHARACTERISTICS

Line No. (a)	Shipment ID Number (b)	Shipment date (c) Month	Day	Shipment value (excluding shipping costs) in whole dollars (d)	Shipment weight in pounds (e)	Commodity code from SCTG list (f)	Commodity description (g)	If a hazardous material, enter the "UN" or "NA" number (h)
0	123-5	4	26	4,235	140	3 5 1 2 0	Electrical transformers	
00	402H	4	26	12,530	62,650	1 7 1 0 0	Gasoline	1 2 0 3
1								
2								
3								
4								
5								
6								
7								

Mode of transport codes for columns (j) and (m) ▶ 1 — Parcel delivery, courier, or U.S. Postal Service 2 — Private truck 3 — For-hire truck 4 — Railroad Continued →

FORM CFS-1000 (12-11-2001)

Page 3

Item E SAMPLING INSTRUCTIONS — Continued

2. SELECTING YOUR SAMPLE OF SHIPMENTS

a. Use the file or combination of files that best reflects your full range of outbound shipping activities.

b. Begin with the first shipment. Count the shipments until you reach your selection rate. Select this shipment to report on in item F.

c. Continue counting with the next shipment. Count this shipment as 1 and continue until you reach the selection rate again. Select this shipment to report on in item F.

d. Repeat the previous step until you have completed your shipment file for the one-week reporting period.

In the following examples, each rectangle represents one shipment.

If the selection rate is 5, select every fifth shipment.

If the selection rate is 2, select every other shipment.

Once you have selected your sample of shipments, please proceed to item F and enter the requested information for each selected shipment. Examples of completed lines for two shipments are provided on lines "0" and "00" below.

U.S. destination (*Complete for all shipments.*) (i) City	State	ZIP Code	Mode(s) of transport to U.S. destination *Enter all that apply in order used. Use codes below.* (j)	Export? (Y/N) (k)	Foreign destination (for export shipments only) Note: In column (l) enter the U.S. port, airport, or border crossing of exit. (l) City	Country	Export mode (m)	Line No. (n)
Los Angeles	C A	9 0 0 4 0	2, 4	N				0
New York	N Y	1 0 4 5 4	3	Y	London	England	6	00
								1
								2
								3
								4
								5
								6
								7

5 — Shallow draft vessel 6 — Deep draft vessel 7 — Pipeline 8 — Air 9 — Other mode 0 — Unknown

FORM CFS-1000 (12-11-2001) *PLEASE CONTINUE ON PAGE 4.*

100003

FIGURE 2*b* 2002 CFS Instruction Guide.

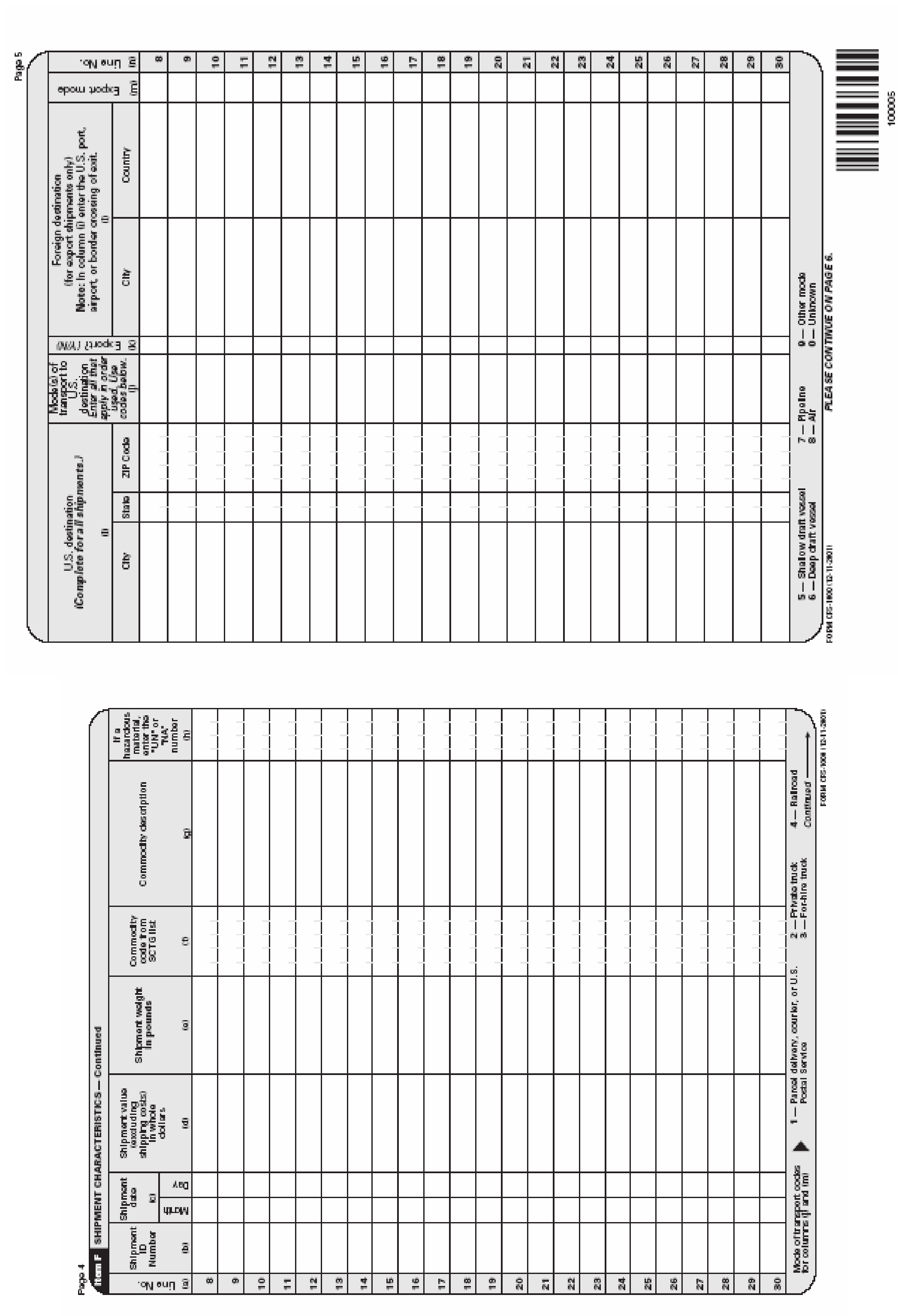
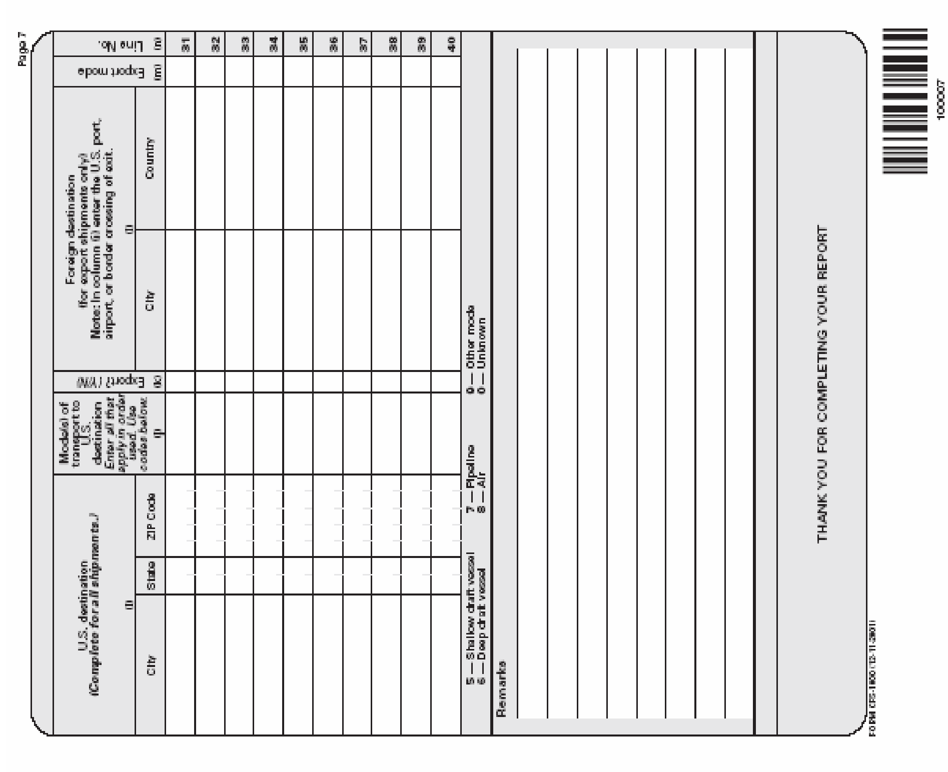


FIGURE 2*c* 2002 CFS Instruction Guide.



Page 7

U.S. destination (Complete for all shipments.) (i) City	State	ZIP Code	Mode(s) of transport to U.S. destination Enter all that apply in order used. Use codes below. (j)	Export? (Y/N) (k)	Foreign destination (for export shipments only) Note: In column (l) enter the U.S. port, airport, or border crossing of exit. (l) City	Country	Export mode (m)	Line No. (n)
								31
								32
								33
								34
								35
								36
								37
								38
								39
								40

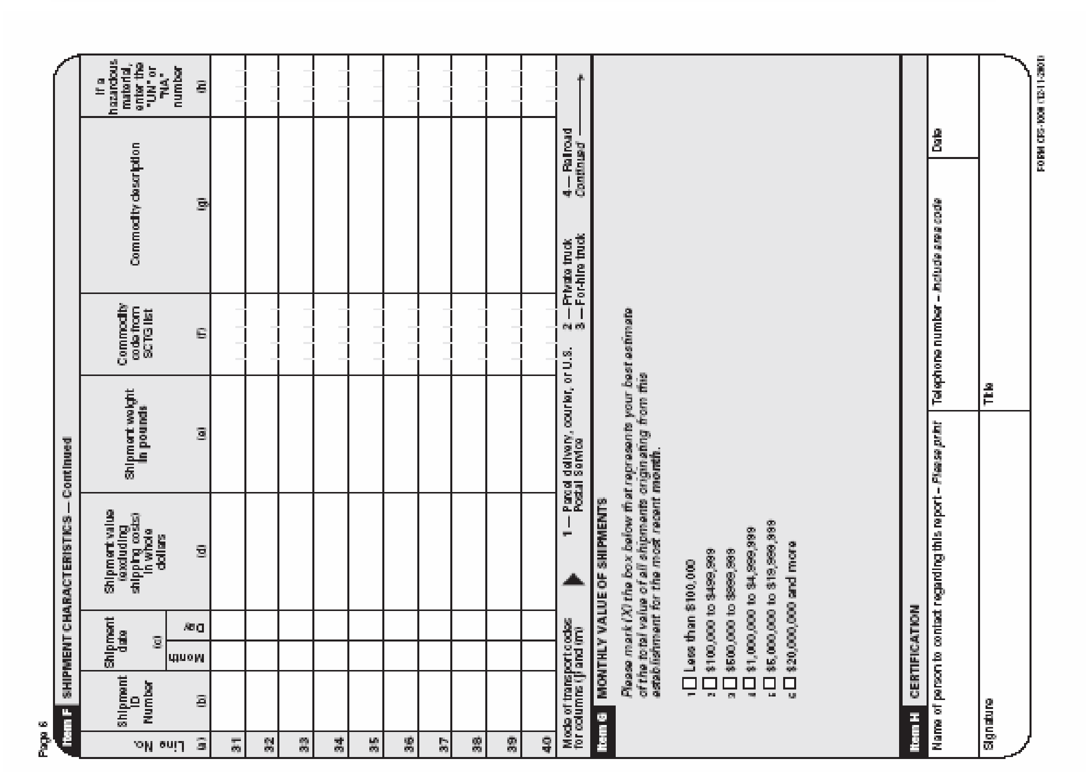
5 — Shallow draft vessel 6 — Deep draft vessel 7 — Pipeline 8 — Air 9 — Other mode 0 — Unknown

Remarks

THANK YOU FOR COMPLETING YOUR REPORT

FORM CFS-1000 (12-11-2001)

100007



Page 6

Item F SHIPMENT CHARACTERISTICS — Continued

Line No. (a)	Shipment ID Number (b)	Shipment date (c) Month	Day	Shipment value (excluding shipping costs) in whole dollars (d)	Shipment weight in pounds (e)	Commodity code from SCTG list (f)	Commodity description (g)	If a hazardous material, enter the "UN" or "NA" number (h)
31								
32								
33								
34								
35								
36								
37								
38								
39								
40								

Mode of transport codes for columns (j) and (m) ▶ 1 — Parcel delivery, courier, or U.S. Postal Service 2 — Private truck 3 — For-hire truck 4 — Railroad Continued →

Item G MONTHLY VALUE OF SHIPMENTS

Please mark (X) the box below that represents your best estimate of the total value of all shipments originating from this establishment for the most recent **month.**

1 ☐ Less than $100,000
2 ☐ $100,000 to $499,999
3 ☐ $500,000 to $999,999
4 ☐ $1,000,000 to $4,999,999
5 ☐ $5,000,000 to $19,999,999
6 ☐ $20,000,000 and more

Item H CERTIFICATION

Name of person to contact regarding this report – *Please print*	Telephone number – *Include area code*	Date
Signature	Title	

FORM CFS-1000 (12-11-2001)

FIGURE 2*d* 2002 CFS Instruction Guide.

TABLE 2 Shipment Characteristics by Mode of Transportation for the United States: 2002

Mode of Transportation	Value		Tons		Ton-Miles (1)		Average Miles per Shipment
	2002 (Million$)	Percent of Total	2002 (Thousand)	Percent of Total	2002 (Millions)	Percent of Total	
All modes	8,397,210	100	11,667,919	100	3,137,898	100	546
Single modes	7,049,383	83.9	11,086,660	95	2,867938	91.4	240
Truck (2)	6,235,001	74.3	7,842,836	67.2	1,255,908	40	173
For-hire truck	3,757,114	44.7	3,657,333	31.3	959,610	30.6	523
Private truck	2,445,288	29.1	4,149,658	35.6	291,114	9.3	64
Rail	310,884	3.7	1,873,884	16.1	1,261,612	40.2	807
Water	89,344	1.1	681,227	5.8	282,659	9	568
Shallow draft	57,467	0.7	458,577	3.9	211,501	6.7	450
Great Lakes	843	—	38,041	0.3	13,808	0.4	339
Deep draft	31,034	0.4	184,610	1.6	57,350	1.8	664
Air (including truck and air)	264,959	3.2	3,760	—	5,835	0.2	1,919
Pipeline (3)	149,195	1.8	684,953	5.9	S	S	S
Multiple modes	1,079,185	12.9	216,686	1.9	225,715	7.2	895
Parcel, USPS, courier	987,746	11.8	216,686	0.2	19,004	0.6	894
Truck and rail	69,929	0.8	25,513	0.4	45,525	1.5	1,413
Truck and water	14,359	0.2	42,984	0.2	32,413	1	1,950
Rail and water	3,329	—	23,299	0.9	114,986	3.7	957
Other multiple modes	3,822	—	105,107	0.2	13,788	0.4	S
Other and unknown modes	268,642	3.2	19,782	3.1	44,245	1.4	130

USPS = U.S. Postal Service

The inability to produce or display all the information collected in the CFS comes from two main items: concerns over statistical quality and maintaining the confidential nature of the survey. Statistically, values within a cell are suppressed if certain statistical criteria are not met, which may limit the actual useable numbers of cells within a table. Information is also suppressed to ensure the survey data (the microdata) remains unpublished to satisfy Title 13 law, which protects confidential business information collected for statistical purposes.

Regarding the hypothetical list, some of these elements are not appropriate for the CFS to release, such as business name, although there may be regulatory or marketing needs that would find the information useful. Furthermore, the hypothetical list includes some elements that are modeled after the data is collected. These additional models could be developed by either the Census/BTS team or could occur after the CFS is released, but in both cases the final data should recognize that these models are outside of the scope of the CFS. The CFS does cover (to some degree) most of the items requested in the *TRB Special Report 276*, although not at the same rate or detail requested, primarily because of these same limitations.

The gap between the three lists (CFS, *TRB Special Report 276*, and the hypothetical list) highlights the frustration planners and researchers express regarding finding a consistent data source for examining freight transportation research. Others attempt to bridge the gap by developing (or purchasing) proprietary models that disaggregate commodity flow information. The role of the private sector in closing the gap between publicly available data and analytical tools further suggests a long term strategy for freight data may involve some modeling work based on the currently available data programs. Some thoughts on bridging the perceived gap that exist between the hypothetical model and the CFS are discussed in the next section.

WHAT ADDITIONAL ITEMS ARE NECESSARY TO IMPROVE THE CFS IN 2007?

For any surveyed shipments, the 2007 CFS should not capture less information than in either 1997 and 2002 CFS. The ability to directly compare the current survey to prior surveys potentially captures information on the changing nature of transportation. Consistent survey forms should reflect changes in shipper needs over time, which directly translates into changing market locations, equipment needs (for example, the deployment of specialized equipment, such as refrigerated vehicles) or modal changes within certain industries or corridors.

Regarding the needs of the various users identified earlier, most of the information in the CFS could be very useful, providing information on many of the various shipment characteristics needed by transportation practitioners. However, the shipment information reported in the CFS only goes so far in providing information on specific shipment characteristics, and primarily at a national/regional level. Assuming the 2007 CFS remains funded, there are certain incremental improvements related to shipment characteristics that may be generated from the CFS. These include questions on geography and routing, commodity information, seasonal patterns, the survey itself, new reports based on the current CFS form, and finally, a discussion on managing expectations.

Geography and Routing

One of the consistent cries against the CFS is the inability of the Census to disclose information below the reported CFS regional data. While these groupings provide good general coverage to a region, most researchers want information on/at a lower geographic level, such as a county, zip code or traffic analysis zone. The CFS is unable to release the microdata (the actual firm survey information), limiting the ability to assign traffic to a specific location. While this coverage is adequate for non-Federal users, there needs to be some mechanism for federal users to access the data to calibrate or examine existing data or models. (Most mandated federal data efforts have safeguards in place to prevent unlawful disclosure.) Most users recognize the data dilemma associated with the use of confidential data, but infrastructure planning needs may be better served by either suppressing or aggregating commodity information to provide more geographic detail below the current CFS region structure.

The current survey prevents modeling freight routings simply because no route information is collected beyond the location of the shipper and the city and state of the receiver. This information is used to estimate ton-miles and distance shipped, which implies some routing to make the estimates but the routing information is not published. There could be some sample routes listed for certain commodities or markets presented in a map format that serves to hide some of the location information. For some modes, it may be possible to identify some routing information by simply adding a new item that asks “does your facility possess modal specific terminal(s), such as inland water dock or a rail siding?” If the respondent answers yes, other information could be used to assign that shipment to a particular network. (Like all microdata elements, the summary information would be available, but not the detailed firm level information.)

A new series of tables based on concentric circles for a certain industry, such as a matrix of average shipment size versus banded shipment distances in 50-mi increments, may be useful for researchers. For some commodities, there may be different markets and corresponding differences in mileage. If the average shipment mileage is 500 mi but there are two distinct markets, one with an average shipment distance of 200 mi and a second of 800 mi, the use of the 500-mi average would misrepresent the shipments of this particular commodity.

There are some concerns about developing national databases that report localized shipment information but these may be unfounded. Given the scale of U.S. transportation, the ability to provide specific shipment information for a local geography could be provided if some other shipment characteristics could be masked. Further, a more tailored local set of traffic generation tools would provide a mechanism for states and local planners to examine traffic routing assignments with more geographic detail. It may also be relevant for the CFS collection effort to be tied to state CFS programs, which could allow states some access for their specific use.

Commodity Information

Commodity information plays a critical role in the CFS as the mechanism to determine what traffic moves, and serves in some regards as a proxy for business or industry sectors. The survey may complicate the ability to use all survey information because of many different factors that may result in some inconsistencies reported in the summarized data.

First, the survey may be limited somewhat by a respondent's ability to provide usable answers, as commodities move in many sizes, shapes, and quantities. For example, how does one describe the shipments of diverse commodities, such as oil, cattle, grain, and shoes? Each item's accounting reflects its industry standard to describe shipment size. Barrels of oil, head of cattle, bushels of grain, or pairs of shoes, are common descriptions of these specific movements which are normally not reported as value or tons. Some better descriptions or conversion elements may help survey respondents complete the survey, including additional material posted on a website to discuss various conversion ratios from commodity or industry specific counting mechanisms into tonnages.

Second, the scope of the CFS survey of establishments excludes some key shipments, such as certain bulk cargos (*6*), (grain shipments from farm to market), and import shipments. These gaps may provide misleading indications concerning the nature of transportation. For example, can domestic shipment patterns of import competitive industries be adequately explained without understanding import traffic? Did import sensitive industries structurally change from 1993–2007? If so, did the domestic transportation pattern of both domestic and imports across the U.S. change over the same time period? The current CFS cannot adequately address the potential changes that may have resulted from imports displacing domestic shipments.

The survey could also include more information on the physical nature of the shipment. By adding a new column entitled “shipment unit,” the survey respondent will be given a choice of several options, such as containerized, palletized, refrigerated, bulk shipment, etc. The ability to link certain traffic patterns with actual equipment may provide additional information for traffic generation models but better aligning certain commodities with certain equipment needs at a modal level. This information does not necessary need to be for detailed regional averages, as even some aggregated national averages would provide useful insights.

Seasonal Patterns

While the CFS is collected at several different times during the year, it may be useful to consider the development of seasonal traffic patterns. If a firm ships primarily agricultural goods, there will be a clear spike in traffic during certain seasons reported in the data. Over the course of the year there are many consistent shipment patterns that are important to understand system use demands.

Time may also review supply chain patterns over the survey time period. For example, Firm A (toy manufacturer) could receive some shipments reported by Firm B (plastics shipper) in Firm B's CFS survey. In the next quarter, Firm A would report its shipments (finished toys) in the CFS sample. The linkage would provide some understanding of the nature of supply chains within certain industries, in addition to serving as a crude proxy for the economic analysis for transportation activities, although this effort may overburden the current CFS workload.

Survey Information

As with any database, the CFS is only as good as the initial survey size, the response rate, and the actual quality of the survey responses. While the statistical reliability estimates are important, a discussion on the data inputting and scrubbing would be helpful, especially if no one can readily assume that respondents possessed sufficient knowledge to complete the survey. For example, how many calls did Census receive on the CFS and what were standard inquiries regarding filling out the forms? It would be useful to know the follow-up rate by Census employees for reported surveys, and if these patterns varied by industry or region. (This implies certain additional errors may be embedded in the reported shipment characteristics.) Furthermore, some understanding of the statistical extrapolation techniques, including the reliance upon other databases, would provide users with more confidence in the CFS.

There are many different analyses on the CFS survey instrument itself that would be meaningful in understanding both the use and possible improvement of the CFS. For example, it would be interesting to know how many people checked Item C. If so, how were these reconciled to the existing survey? How much of the geography mirrors headquarters effects, where one headquarters coordinates the transportation of satellite offices, but the product actually moves from other locations, especially given the growth of third party logistics firms. Furthermore, how are shipments reported in Item D and G when compared to the actual survey reported in Item F? One would assume that the average value and weight from D and G would compare favorably with the Item F responses.

New Shipment Characteristics Available from Existing Survey Documents

There are several shipment characteristics that could be inferred values from the CFS. These include business establishment by type, geography (zip code, city, and state), and time of year. The CFS could be linked to the Economic Census to develop estimates of shipment patterns from or into a certain plant type or industry type, but also general inbound shipments into certain facilities. This may generate useful information for state and local planners who must estimate initial trip generation data by commodity or industry sector. Since there exists no bridge between industry type and commodity detail in the CFS and one would have to be developed based on other database information or sources. One concern is that if more tables were developed from the CFS, the usefulness of these new tables may be nullified if it is feared more information would provide a mechanism to reverse engineer the data to estimate the confidential microdata.

Reporting the Correct Total for U.S. Transportation

The CFS, by limiting its survey to only domestic shipments, may provide misleading indications concerning the nature of transportation shipments. For example, when comparing Table 2 to the table in Figure 3 from the Waterborne Commerce of the United States, National Summary (listed in the Appendix), both the definitions and values are strikingly different. The CFS reported 681 million tons on the waterway system, while the Waterborne Commerce reported 2,335 million tons. Average haul in the Waterborne Commerce equaled 308.9 mi while the CFS estimated the average miles per shipment at 568 mi. The CFS reported 282 billion ton-miles while the Waterborne Commerce reported 721 billion ton-miles. Some of the differences could be explained by CFS respondents not knowing the true mode used in the shipment and that gross comparisons of average shipment are more relevant at detailed commodity level. Also water handles tremendous tonnages of bulk and import cargos, items outside the current CFS survey scope. By excluding these items from the CFS, the resulting summarized tables report large discrepancy of tons, ton-miles, and in some cases, average shipment length. This implies that even the reported findings may be suspect in describing the value, tons, ton-miles of cargo shipments, resulting in users assuming the CFS contains more information than actually captured by the survey itself.

To correct this, the CFS should look at the use of the detailed sets for various modal databases which should then be shared with the appropriate federal data or statistical agency to double check the conversions. If irreconcilable differences exist, at least some mechanism exists to report the differences between the two reported data sets. Once the data set is finalized, the now calibrated CFS may be improved by more closely integrating the information with other databases on both missing elements and comparative freight statistics. By doing so, the CFS enhanced may provide information on the time sensitiveness of the shipment, the type of conveyance (which is done through the modal subgroups), piece counts and size, the mixing of commodities within the same or multiple shipments, and transportation costs. Data integration and quality checks against existing databases could also provide better confidence in developing the shipment characteristics reported in the CFS summary tables, this may be to painful process given current funding and staff resources.

Federal Use of the CFS

While the federal government's data and analytical needs appear consistent with other entities, the unique role of the federal government as both a collector and analytical user of information suggest there should exist two versions of the CFS—one for a general audience and a second for federal government purposes. Currently access to the CFS microdata may be granted, but the general inability of other federal agencies to use the CFS to calibrate against other non-regulatory data and analytical needs limits the ultimate usefulness of the CFS. Today, federal agencies must develop or purchase additional databases or models to reestimate information

Table 1-4: TOTAL WATERBORNE COMMERCE, 1984-2003
TON-MILES, SHORT TONS AND AVERAGE HAUL+++ BY TYPE OF TRAFFIC

Year	Grand Total	Foreign			Domestic				
		Total	Great Lakes	Coastal	Total	Coastwise	Lakewise	Internal	Intraport
Ton-Miles**** (millions)									
1984	966,478.2	78,758.2	*32,249.4	**46,508.8	887,720.0	593,923.1	49,784.4	242,855.4	1,157.1
1985	964,911.2	71,941.2	*27,171.8	**44,769.4	892,970.0	610,976.5	48,184.0	232,707.5	1,102.0
1986	944.910.5	71.509.4	*24.235.3	**47.274.2	873.401.0	580.888.7	43.198.2	248.116.9	1.197.3
1987	968,403.7	72,988.2	*22,003.3	**50,985.0	895,415.5	586,818.4	50,076.7	257,336.4	1,183.9
1988	970,685.0	80,656.6	*24,448.4	**56,208.2	890,028.4	561,594.9	58,159.5	269,035.7	1,238.4
1989	911,232.2	95,682.3	*27,118.9	**68,563.4	815,549.9	483,888.6	58,307.6	272,157.4	1,196.3
1990	+932,151.4	+98,607.6	*24,330.8	**+74,276.9	833,543.8	479,133.6	60,929.9	292,393.3	1,087.0
1991	+938,671.3	+90,272.3	*17,102.6	**+73,169.7	848,399.0	502,133.0	55,339.1	289,959.0	968.0
1992	+949,344.7	+92,660.1	*21,027.6	**+71,632.5	856,684.6	502,311.0	55,784.6	297,638.7	950.3
1993	+882.605.8	+92.947.9	*26.106.8	**+66.841.1	789.657.9	448.404.2	56.438.1	283.893.6	921.9
1994	909,405.0	94,485.8	*28,299.6	**66,186.2	814,919.2	457,600.7	58,263.4	297,762.4	1,292.7
1995	914,362.5	106,634.9	*31,656.5	**74,978.4	807,727.7	440,345.1	59,703.8	306,329.1	1,349.6
1996	870,193.4	105,506.9	*34,259.9	**71,247.0	764,686.5	408,086.1	58,335.3	296,790.6	1,474.5
1997	813,800.1	106,390.2	*32,535.9	**73,854.3	707,409.9	349,843.0	62,165.9	294,023.0	1,378.1
1998	780,449.8	107,704.5	*34,136.7	**73,567.8	672,795.3	314,863.9	61,654.3	294,896.4	1,380.7
1999	764,447.5	108,585.9	*34,846.7	**73,739.2	655,861.5	292,730.0	57,045.2	304,724.1	1,362.2
2000	763,421.1	117,621.8	*38,623.3	**78,998.5	645,799.3	283,871.6	57,879.1	302,558.4	1,490.2
2001	736.930.0	115.243.8	*36.580.1	**78.663.8	621.686.2	274.558.8	50.853.5	294.860.9	1.413.0
2002	721,421.7	109,341.1	*34,921.3	**74,419.9	612,080.5	263,688.2	53,652.9	293,410.3	1,329.0
2003	**714,289.5**	**108,294.1**	***32.921.5**	****75,372.7**	**606,146.3**	**278,918.7**	**47,539.4**	**378,352.3**	**1,335.9**
Tons (millions)									
1984	1,832.6	803.3	58.8	744.6	1,029.3	307.7	98.0	542.5	81.1
1985	1.785.0	774.3	51.3	723.0	1.010.7	309.8	92.0	534.7	74.3
1986	1,870.5	837.2	45.8	791.4	1,033.2	308.0	87.4	560.5	77.4
1987	1,962.8	891.0	45.9	845.1	1,071.8	323.5	96.5	569.8	82.0
1988	2,082.9	976.2	52.5	923.7	1,106.6	325.2	109.7	588.1	83.7
1989	2.135.2	1.037.9	54.8	983.1	1.097.3	302.0	109.1	606.0	80.2
1990	2,159.3	1,041.6	50.5	991.1	1,117.8	298.6	110.2	622.5	86.4
1991	2,087.6	1,013.6	41.8	971.8	1,074.0	294.5	103.4	600.4	75.6
1992	2.127.9	1.037.5	45.5	992.0	1.090.4	285.1	107.4	621.0	76.8
1993	2,123.3	1,060.0	43.6	1,016.4	1,063.2	271.7	109.8	607.3	74.4
1994	2,208.8	1,115.7	50.1	1,065.6	1,093.1	277.0	114.8	618.4	82.9
1995	2,233.5	1,147.4	51.9	1,095.5	1,086.2	266.6	116.1	620.3	83.1
1996	2.276.7	1.183.4	56.4	1.127.0	1.093.4	267.4	114.9	++622.1	++89.0
1997	2,326.9	1,220.6	57.7	1,162.9	1,106.3	263.1	122.7	630.6	89.8
1998	2,332.3	1,245.4	62.4	1,183.0	1,086.9	249.6	122.2	625.0	90.1
1999	2,316.7	1,260.8	62.4	1,198.3	1,055.9	228.8	113.9	624.6	88.60
2000	2,419.1	1,354.8	64.0	1,290.7	1,064.3	226.9	114.4	628.4	94.6
2001	2,387.4	1,350.8	62.5	1,288.3	1,036.6	223.6	100.0	619.8	93.2
2002	2,335.2	1,319.3	59.6	1,259.7	1,015.9	216.4	101.5	608.0	90.0
2003	**2,387.9**	**1,378.1**	**56.3**	**1,321.8**	**1,009.7**	**223.5**	**89.8**	**609.6**	**86.9**

** Based on distance transported on Great Lakes and St. Lawrence River to International Boundary at St. Regis, Quebec, Canada
** Based on distance transported on United States Waterways from entrance channels to ports and waterways.
**** Excludes intra-territory traffic for which ton-miles were not compiled.
+* An error in the routing network caused the coastal port foreign ton-mile totals to be slightly inflated for calendar years 1990-1993. The corrections are reflected in these revised totals. Ton-miles for individual ports were not affected.
++ Beginning in 1996, fish was excluded.
+++ "Average Haul" is the average distance a ton of cargo travels by water. It is computed by dividing the number of ton-miles by the number of tons.

1-6 TOTALS U.S. ARMY CORPS OF ENGINEERS

FIGURE 3*a* Waterborne commerce of the United States, calendar year 2003, Part 5—National Summaries.

Table 1-4: TOTAL WATERBORNE COMMERCE, 1984-2003
TON-MILES, SHORT TONS AND AVERAGE HAUL+++ BY TYPE OF TRAFFIC
(continued)

		Foreign			Domestic				
Year	Grand Total	Total	Great Lakes	Coastal	Total	Coastwise	Lakewise	Internal	Intraport
Average Haul*** (miles)									
1984	527.4	98.4	548.7	62.5	862.5	1,930.5	508.0	448.0	14.3
1985	540.5	92.9	529.5	61.9	883.5	1,972.2	523.8	435.2	14.8
1986	505.2	85.4	528.8	59.7	845.3	1,885.9	494.5	442.7	15.5
1987	493.3	81.9	479.1	60.3	835.4	1,813.9	519.0	451.6	14.4
1988	464.9	82.6	465.8	60.8	804.3	1,727.0	530.3	457.5	14.8
1989	426.8	92.3	494.5	69.7	743.2	1,602.1	534.5	449.1	14.9
1990	+431.7	+94.7	482.2	+74.9	745.7	1,604.4	553.1	469.7	12.6
1991	+449.6	+89.1	409.2	+75.3	789.9	1,704.8	535.0	483.0	12.8
1992	+446.1	+89.3	462.2	+72.2	785.7	1,761.7	519.4	479.3	12.4
1993	+415.7	+87.7	598.2	+65.8	742.7	1,650.3	513.8	467.5	12.4
1994	411.7	84.7	564.5	62.1	745.5	1,651.8	507.6	481.5	15.6
1995	409.4	92.9	610.4	68.4	743.6	1,651.6	514.1	493.8	16.2
1996	382.2	89.2	607.9	63.2	699.4	1,526.2	507.8	477.1	16.6
1997	349.7	87.2	563.5	63.5	639.5	1,329.5	506.5	466.3	15.3
1998	334.7	86.5	546.8	62.2	619.0	1,261.3	504.7	471.8	15.3
1999	330.0	86.1	558.2	61.5	621.1	1,279.4	500.9	487.9	15.4
2000	315.6	86.8	603.0	61.2	606.8	1,250.9	506.1	481.4	15.8
2001	308.7	85.3	584.9	61.1	599.7	1,227.9	508.5	475.7	15.2
2002	308.9	82.9	585.8	59.1	602.5	1,218.5	528.8	482.6	14.8
2003	**299.2**	**78.6**	**585.0**	**57.0**	**600.3**	**1,248.2**	**529.5**	**456.6**	**15.4**

*** Excludes intra-territory traffic for which ton-miles were not compiled.
\+ An error in the routing network caused the coastal port foreign ton-mile totals to be slightly inflated for calendar years 1990-1993. The corrections are reflected in these revised totals. Ton-miles for individual ports were not affected.
+++ "Average Haul" is the average distance a ton of cargo travels by water. It is computed by dividing the number of ton-miles by the number of tons.

FIGURE 3*b* Waterborne commerce of the United States, calendar year 2003, Part 5—National Summaries.

already captured in the CFS surveys, which suggests not only duplicate efforts and a waste of government funds, but a potential loss of interagency use (and cross promotion) of important transportation information. Most federal freight transportation data collection programs provide safeguards to protect confidential business and individual data information, but ways to share the detailed information with other federal users. The CFS should examine ways to do the same.

Managing Expectations

Even with greater computers and processing capacities, the reality is that no database currently captures and compiles the data needs identified earlier into a single, user friendly, accessible interface. The large development and maintenance costs would be prohibitive, but also the actual mass may result in a database too large to use.

Any CFS improvement should balance (or at least acknowledge) the costs associated with collecting and publishing data statistics, and not just the benefits from a better data program. For example, how much time will respondents spend filling out the survey if the data expand? Will the recommended improvements be carried out based on current funding and staffing? If the survey occurs every five years, do we really need to generate sufficient detail to answer every question at the exclusion of other databases? Would the moneys available for CFS improvements be better spent on corresponding specialized surveys or data collection efforts?

The CFS is only collected every 5 years, making some very specific data less relevant when compared to other data elements that are updated on a more current cycle. The CFS should focus on developing the data that can sustain the long lags between survey collections, given some indication of the magnitude of the U.S. freight system and the supporting linkages to economic activity at a national level.

Does this mean the CFS should not be linked to other data sets? No, for when the CFS is combined with other information available from other government agency collection efforts, the CFS has the potential to generate a very detailed database on actual business activities, even with the masking of confidential data. However, data integration could lead to reporting data that could be considered confidential once combined with other elements, which may put additional limits on the actual shipment information released.

Regarding the future of the CFS, every five years the CFS should answer the following question: "if the CFS were rebuilt from scratch, what would it look like?" Since the first CFS in 1993, new data sets and technologies hold great promise as a means to improve the quality of the CFS data. The CFS should also look at receiving electronic submissions, including allowing users to complete the survey on the internet, which may speed up processing while providing some additional data quality on the shipment characteristics reported by the shipper. The addition of electronic submissions may help offset some data errors from such a large and diverse set of respondents.

CONCLUSION

The ultimate need for data is to generate confidence in the decision maker's process. But for data to be useful, it must provide the correct information in the correct form for comprehension. The question is what shipment characteristics are the most relevant and are these being adequately distributed to users of the information?

Shipment characteristics may be useful in accurately describing transportation activities or may result in more blind men trying to describe an elephant. The conflicting freight totals appear to like the men describing the elephant, arguing over what is the correct value. (One word of caution: The story of the blind men implies they spent more time arguing about the elephant than they spent actually touching the elephant.) Transportation contains many striking differences, especially based upon various user perspectives, data collection methodologies, etc.

The linchpin of the commodity flow survey is its findings on certain key shipping characteristics. These values serve as the basis for meaningful discussions by decision makers, but users must be aware of the limitations of the CFS as a survey and its lack of complete coverage. (There is always the potential for analysts and planners to complain there is either not enough information or too much information to make any decision.)

Today and in the near future no one database can definitely capture the total freight activity in America for all modes through primary data collection efforts. Unlike the six blind men, we are not looking for differences, but similarities, in describing how freight moves, why freight moves, and where freight moves. The CFS is uniquely qualified to provide one mechanism to describe the national freight transportation system, which may support other related research efforts. There is a need for the CFS, linked to other data and analytical programs, to reexamine the most shipment characteristics used (and presented) for the broadest set of users to determine the critical questions that must be answered.

However, a word of caution is needed. Another story says six blind elephants were asked to describe a man. The first elephant put his leg on man and claimed he was flat. Each elephant then took his turn and all agreed with the first elephant's findings. The CFS (or any other freight database) should not be relied upon to more than it can within a reliable modeling framework but the CFS should also not be viewed as a single enterprise and discard it after a period of time.

QUESTIONS TO CONSIDER

The following questions are presented to provide some additional discussion topics during the shipment characteristics session. The questions are listed by general topic area.

The Commodity Flow Survey Itself

- If the CFS were rebuilt from scratch, what elements would be collected and used based on information collected from shippers?
- Is it fair to the CFS to make it the single database that will provide all the questions to be answered?
- Is the CFS primarily a federal tool or a local database? What characteristics would be needed if the tool were better refined to serve federal needs? To serve local needs?
- Without the full coverage of all shipments, does even reporting the data imply some misleading indicators regarding total freight movement?
- How much will improvements cost? What funding requirements are necessary to make these changes? Can the CFS charge for specialized data reports?
- Can the CFS develop a framework to incorporate user feedback to either improve current data programs or to develop new data collection efforts?
- What is the appropriate level of public and private sector data collection or integration efforts? Should all transportation data be provided for free or should some market mechanisms exist to provide information not available in current publicly available data programs?
- What is the correct approach for concurrent publicly and privately available data and model development for transportation analysts?

Shipment Characteristics

- What are the important shipment characteristics for the broadest class of user needs?
- What characteristics are reported most often in various documents—trade journals and publications, economic research, regulatory analysis, planning documents? Can CFS tables be constructed to show these common elements (including maps) about freight shipments?
- Should the CFS only focus on surveying intra-city goods movement and allow other databases or integration efforts to include bulks, etc., to generate a more comprehensive picture of freight shipments?
- What comprehensive commodity databases could be generated by combining CFS at the sample level with Waterborne Commerce statistical data, Highway Performance Monitoring System, county business patterns, e-stats, foreign trade, Rail Waybill, etc. and still maintain a confidential "rolled-up" database?

- What shipment characteristics are the most critical to improve in the current CFS structure?
- Should more efforts be geared towards developing correct national totals across modes?
- Can the detailed microdata be released after an appropriate time lag, such as releasing the 1993 CFS now and the 1997 CFS after the 2007 CFS is published? Can the microdata be released without the specific company information?

The Survey Itself (Data Collection and Data Quality)

- Can the CFS receive electronic transmission of a full database in lieu of the paper filing? Does this change what can be collected or released? What does this imply for electronic filing standards?
- Are the survey form and instructions adequate for most people to complete the survey? Does the CFS mislead survey data users by misreporting the shipment information to the survey respondent?
- What institutional knowledge about the CFS shipment characteristics remains within the minds of the CFS processors and should it be shared with the user community?
- Does the CFS adequately report survey processing so users understand the data quality issues separate from statistical checks?
- What level of data quality can the transportation accept?
- What non-survey information could be integrated into the CFS reports to estimate the gaps outside of the CFS's current scope? (i.e., can we move the CFS from only a survey into an analytical tool?)

New Shipment Characteristics from the CFS

- By surveying business establishments, what other information can be extended on the company level, such as number of employees, etc., for developing trip generation models by industry establishment?
- Can seasonal frequencies be developed and/or released?
- Can industry or regional averages concerning shipment patterns be generated from the other elements in the CFS survey?
- Can the inferred information be presented in the reported tables?
- Are the use of average miles shipped the most meaningful statistic for developing routing models?

Should an "Industry Series" be developed for the CFS? Can a series linking to other economic data sets be developed that addresses how economic data can be linked with the CFS data?

NOTES

1. TRB has participated in many of these efforts. A briefing listing includes *TRB Special Report 276: A Concept for a National Freight Data Program*; *TRB Special Report 277: Measuring Personal Travel*

and Goods Movement—A review of the bureau of transportation statistics' surveys and data needs in the changing world of logistics and freight transportation.

2. There is an implied assumption that the users listed here are not simply planners at state Departments of Transportation, but include academics, private sector companies, consultants, reporters, economic researchers, etc.
3. American Association of State Highway Transportation Officials (AASHTO), "Freight Bottom-Line Report"
4. *TRB Special Report 276: A Concept for a National Freight Data Program*, Transportation Research Board of the National Academies, Washington, D.C., 2003, http://books.nap.edu/html/SR276/SR276.pdf.
5. Garrido, R. A. *Insights on Freight and Commercial Vehicle Data Needs* provides a similar listing, based on historical usage by various nations. The file can be accessed at http://www.its.usyd.edu.au/conferences/international_conference_on_transport_survey_quality_and_innovation%20(new)/South_Africa_Papers/Garrido%20Freight%20Data-reviewed-2.doc
6. Bulk cargos generally are homogenous commodities that move in large quantities, such as grains or petroleum products.

Intermodal Shipments, Warehousing, and Third Parties
A Special Measurement Issue

ROLF SCHMITT
Federal Highway Administration

ABSTRACT

The Commodity Flow Survey (CFS) is designed to capture shipments moving by more than one mode, but appears to significantly under represent intermodal shipments. Possible reasons include gaps in coverage of shippers, misclassification of shipments because the shipper outsourced supply chain management or the carrier changed modes, and disparities between presumed definitions of intermodal shipments and actual definitions based on record-keeping practices. This paper suggests alternate strategies for resolving these problems. Many of the gaps in coverage are also explored in the resource paper on the CFS scope.

THE PROBLEM

The CFS is supposed to be an excellent source of data on intermodal shipments because it is designed to capture ultimate origins and destinations rather than terminal-to-terminal moves of shipments by all modes and intermodal combinations. Shippers are surveyed since they are more likely than carriers to know the final destination and all modes used for shipments by multiple carriers, intermodal combinations, or shipper-owned equipment. The CFS includes a separate mode for mail and parcel delivery services since shippers are not likely to know the individual modes used by the U.S. Postal Service (USPS) and other intermodal small package carriers.

The 2002 CFS measured 218.7 million tons of goods moving 226.7 billion ton miles valued at $1.1 trillion in shipments using multiple modes or USPS and parcel carriers. These intermodal shipments represent 2% of tons, 7% of ton miles, and 13% of the value of all shipments in the 2002 CFS. The truck-rail combination alone accounted for 243 million tons of goods moving 46 billion ton miles valued at $70 billion in shipments.

While the 2002 CFS captured large volumes of intermodal shipments, including some unexpected combinations, comparisons with Rail Waybill statistics and other analyses suggest that a majority of intermodal shipments are missing. Seven sources of missing shipments are:

1. **Shipments by retail and service establishments, central administrative offices, governments, and households.** A significant quantity of mail and parcels are shipped from domestic sources not in the CFS scope. While mail order houses are included in the CFS, most retail establishments are excluded because the majority of their sales are hauled away in the consumer's private vehicle. Thanks to toll-free phone numbers and the Internet, many businesses are doing significant mail order business in addition to over-the-counter sales. Service establishments, central administrative offices (typically corporate headquarters), governments, and households also ship mail and parcels. The quantity of shipments per shipper is very small,

so these types of shippers are not included in the CFS. However, these shippers include over 100 million households, about 100,000 government entities, and thousands of others, so their total shipments of mail and parcels is significant.

2. **Imports and in-transits.** Significant quantities of imports move in containers on railcars from deep water ports or on railcars crossing land borders. Some foreign shipments entering the country are destine for another country, and are actually in-transits rather than imports. Imports and in-transits are not measured by the CFS unless they are reshipped by a domestic establishment after arriving in a port of entry. Foreign establishments who originate the imports and in-transits are beyond the reach of the Census Bureau.

3. **Warehousing.** Some intermodal shipments originate in warehouses which are either out of scope or covered poorly. Warehousing includes establishments in sub-sector 493 of the North American Industrial Classification System (NAICS) and warehousing auxiliaries. NAICS sub-sector 493 includes warehousing and storage services provided on a for-hire basis to other companies, and the auxiliaries are in-house warehousing and storage facilities (such as the distribution center for a grocery chain). Establishments in NAICS sub-sector 493 were not included in the CFS because they are not shippers and are not responsible for knowing where merchandise stored in their facilities are sent. Warehousing auxiliaries are in scope because they are part of the shipper's company and more likely to know the intended destination of their shipments; however, preliminary investigations indicate that many warehousing auxiliaries were missed or misclassified in the lists of establishments from which the CFS sample was drawn.

4. **Third parties.** Many domestic intermodal shipments are probably being reported as truck only because the move was arranged by a third party and the shipper sees only trucks picking up the goods. When the CFS was launched in 1993, most shippers managed their own logistics and were likely to know all modes used for their shipments. Many shippers, including very large companies, have outsourced their logistics to third parties, and are less likely to know or care about all the modes used.

5. **Carrier uses of other carriers.** Third parties are not the only reason for misclassification of mode by the shipper. For-hire motor carriers often put their trailers on flatcars for longer distance moves, often without the knowledge of the shipper.

6. **The pipeline problem.** Pipeline operators have even more control over commodity movements than other carriers, mixing and redirecting flows of crude oil between origins and destinations in ways that do not correspond with the concept of a shipment. The CFS excludes crude petroleum moves by pipeline, which in turn excludes intermodal moves involving the pipeline. This is a relatively small problem for coverage of intermodal shipments, but a big hole in the coverage of total flows in the CFS.

7. **Definitions.** What exactly is an intermodal shipment? Popular conceptions, terms used by industry and government, and the operational definition used in the CFS are not necessarily the same.

CAN CHANGES IN SCOPE FIX UNDERREPORTING OF INTERMODAL SHIPMENTS?

The first source of missing shipments can be captured by expanding the scope of the CFS, which is discussed in another resource paper. Retail and service establishments, central administrative offices, and governments are already covered in the quinquennial Economic Census and Census of Governments. Households are sampled annually, but would require a modified questionnaire. As discussed in the workshop on scope, there may be more cost-effective alternatives to adding these types of shippers to the CFS.

Imports cannot be estimated through a simple expansion of the CFS scope since foreign shippers are beyond reach of Census surveys, and warehousing and third parties cannot be added to the CFS scope without changes to the questionnaire or processing that would limit potential double counting and other problems. These sources of missing shipments, problems related to the definition of intermodal shipments, and strategies or improving the representation of intermodal shipments in the 2007 CFS are explored in this paper.

Shipments involving pipelines cannot be added by merely including pipeline operators in the CFS scope. Unless an effective surrogate for shipment can be found, pipeline flows must be estimated through a model that distributes flows among production and consumption areas.

IMPORTS AND IN-TRANSITS

Imports and in-transits can be captured in several ways:

1. International brokers can be sampled as shippers and given CFS questionnaires modified to collect shipment information for imports. This assumes that nearly all imports are handled by international brokers, and that the brokers can be identified, sampled, and the sample expanded to a universe in the same way that domestic shippers are handled. This also assumes that the questionnaire can be modified to exclude imports that are already in scope because of change in ownership at the port of entry. This approach may not capture in-transits.
2. The CFS questionnaire can include a screening question on imports for a follow-up survey. Shippers would be asked if they imported goods, and then given a subsequent questionnaire to measure how much of what from where by what mode. This approach has a minimal increase in respondent burden since only establishments that import receive the follow-up questionnaire, and the follow-up questionnaire can be customized to deal specifically with imports. This approach assumes that most imports are received by domestic shippers already in scope, and that imports already captured by changing ownership at the port of entry are not double counted. This approach would not capture in-transits, and would probably result in imports being reported for a delayed time period than domestic shipments.
3. The CFS questionnaire can be modified to include a section for reporting imports. Shippers would be asked if they imported goods, and then directed to a section asking about import shipments received. This approach does not get the import data out of synch with the rest of the CFS, but adds complexity and bulk to the CFS questionnaire. As in the second approach, most imports are assumed to be received by domestic shippers already in scope, and ways must be found to avoid double counting imports already captured by changing ownership at the port of entry. This approach would not capture in-transits.
4. Import documents can be sampled for a parallel survey. Import and export documents were sampled in the mid 1970s and the brokers or shippers responsible for the document were asked about the domestic leg of the shipment. This could be done in synch with the reporting period for domestic shipments in the CFS, and might capture in-transits as well as imports. This approach imports requires procedures or changes to the CFS to exclude imports already captured in the CFS by changing ownership at the port of entry.

5. Shipment characteristics measured in the CFS can be incorporated into the manifests to be used in the International Trade Data System. If shipment weight, interior origin, and modes used between port of entry and interior destination are added to import documents, then the only remaining challenge is to assure that the CFS excludes imports already captured in the CFS by changing ownership at the port of entry.

Which of these approaches is the most cost-effective? Which will provide the greatest quality? Which will minimize respondent burden?

WAREHOUSING AND THIRD PARTIES

Warehousing auxiliaries were included in the 2002 CFS to capture the part of retail other than mail order establishments where most shipping takes place. These are usually company-owned distribution centers from which truck loads of goods are sent to individual stores. The major shortcoming for these establishments in 2002 appears to be inadequate coverage or misclassification of warehousing auxiliaries in the establishment lists used for CFS samples. Greater attention to warehousing auxiliaries while compiling the establishment lists will presumably correct the problem.

Warehousing and storage establishments in NAICS sub-sector 493 are a different matter because shipments from those locations are presumably controlled by shippers covered elsewhere in the CFS. Does the CFS questionnaire adequately direct the shipper to include shipments from public warehouses and give true origins for those shipments? Should shipments from public warehouses be excluded from the CFS questionnaire for shippers and a parallel survey of establishments in NAICS sub-sector 493 is conducted? In a parallel survey, how would shipments be sampled and expanded to a universe?

The warehousing issue is closely related to the third party issue. If the shipper has outsourced the supply chain, then only the third party logistics firm knows what modes of transportation and intermediate warehousing and storage facilities were used. Does the answer lie in a parallel survey, similar to or built on the parallel survey alternative for for-hire warehousing and storage establishments? Or should third parties be treated like leasees in the Vehicle Inventory and Use Survey (VIUS)? If someone other than the owner of a sampled vehicle in the VIUS operated the vehicle in the reporting year, then the party who knows how the vehicle was used is supposed to answer the questionnaire. If a shipper outsources some or all of its shipments to a third party, should the shipper be directed to forward the questionnaire to the third party? How would the instructions direct shippers who use more than one third party arranger of transportation.

THE CONTAINERIZATION QUESTION

Supplemental surveys or redirection of the CFS questionnaire to for-hire warehousing establishments and third parties are intended to assure that all shipments are captured and that intermodal shipments are properly classified. If a shortfall in the reporting of intermodal shipments has more to do with misclassification of shipments than with capturing all shipments, then an additional question for each shipment on the CFS questionnaire could be developed for use in confirming or reclassifying the shipment. The 1993 CFS questionnaire included a column for “"Containerized? yes/no” for each recorded shipment. The instructions said that containerized “means that the shipment left your establishment in an intermodal container or stackable tank without permanently attached wheels. These containers typically vary from 20 to 53 ft in length, and are carried on truck chassis, trains, and ships.” The responses to this field were so inconsistent that the containerization item was dropped. Can a more effective question be developed to identify shipments that are packaged to move intermodally?

WHAT IS INTERMODAL?

While the concept of an intermodal shipment as one which travels by more than one mode of transportation seems obvious, the concept loses its precision when commonly perceived supply chains collide with shipping documents. Grain elevators at river ports provide a good example. These are clearly intermodal facilities, with grain shipments arriving from farms and other grain elevators by truck and rail, and leaving by barge or ship. These grain elevators are a transfer point and temporary storage facility on an intermodal supply chain that stretches from farms to bakeries and other food processors. However, the shipments to the grain elevator over land are separate from the shipments from the elevator by water, in part because the grain has been commingled before being sent onwards and in part because the moves to and from were covered by separate shipping documents. This intermodal facility is actually a transfer point between single mode shipments.

What is intermodal? The most obvious intermodal shipment is illustrated by cargo moving in a container from a vessel or a train onto a truck chassis without changing ownership and under one shipping document. The most obvious example of two shipments by single modes is the shipment of raw materials by one mode to a manufacturing facility where the materials are transformed into a finished product and shipped out by another mode under separate ownership and separate shipping documents. What about cargo that arrives at a shipper’s warehouse by one mode and leaves by another mode after being repackaged? There is a change in mode and the movements are covered by separate shipping documents, yet there was no transformation of change in ownership. Does the distinction of two single mode shipments and one intermodal shipment depend on ownership, transformation of material, repackaging, or the kind of container used, or is it just a function of documentation?

If the constraints of record-keeping practices are put aside for the moment, what changes must happen before an intermodal movement becomes two single mode movements? Ownership of the material? Physical transformation of the material? Repackaging for different destinations? Repackaging for the same destination? Co-mingling bulk material? Passing through Customs between the international and domestic transportation systems? How should these elements be combined into a more precise concept of intermodal that conforms to popular understanding?

The realities of recordkeeping practices must enter the equation following agreement on a more precise concept of intermodal. How do the ways in which shipment records are kept affect the operational definition of intermodal? Are there ways to bring an operational definition of intermodal into closer alignment with the concept of intermodal?

The current definition of “shipment” in the CFS documentation attempts to balance common perceptions with record-keeping practices: “a shipment is a single movement of goods, commodities, or products from an establishment to a single customer or to another establishment owned or operated by the same company as the originating establishment (e.g., a warehouse, distribution center, or retail or wholesale outlet). Full or partial truckloads are counted as a single shipment only if all commodities on the truck are destined for the same location. If a truck makes multiple deliveries on a route, each stop is counted as one shipment." Applying the published explanation for trucks to other modes, changes in ownership, repackaging, or multiple destinations effectively turn an intermodal movement into multiple single-mode shipments.

The definitional issue is best resolved by asking why we care. Freight policy statements often proclaim the importance of seamless, intermodal transportation, which raises the question of how much seamless, intermodal transportation exists. Since most bulk and liquid shipments move through systems designed to link appropriate forms of transportation, the better question may be how many shipments of manufactured goods move in containers that can be carried by truck, rail, or water? "Intermodal" becomes "containerized," which would force the CFS designers to find a better way of asking the "containerized" question.

The definitional issue is underlain by an emerging difference in the interests of CFS users. The traditional focus is on geography: from where to where do commodities move. The CFS is designed to answer this question by identifying the locations of the shipper and the consignee for each shipment. A new focus is on supply chains, which may involve multiple shippers and consignees in multiple regions. For example, the U.S. Department of Agriculture wants to track beef from hoof to hamburger so that a health problem at either end can be traced to its source or to the affected region. Supply chains are more likely to involve multiple modes than shipments among one pair of players in the supply chain. The CFS measures aggregate flows among regions. Input-output models and other analytical techniques must be combined to turn shipments among regions into supply chain flows among industries.

CONCLUSION

The CFS seriously underreports intermodal shipments, in some instances because the shipment was out of scope and in other instances because the shipment was misclassified. Some of the underreporting can be fixed by changes in scope, more effective treatment of imports, and a better question for containerization; however, misclassification by shippers is inevitable when third parties are involved or carriers decide to hitch a ride on other carriers without telling the shipper.

Some Suggestions for Improving CFS Data Products

NATHAN ERLBAUM
New York State Department of Transportation

JOSE HOLGUÍN-VERAS
Rensselaer Polytechnic Institute

INTRODUCTION

The Commodity Flow Survey (CFS) collects data on the movement of goods in the United States and the District of Columbia. It provides information on commodities shipped, their value, weight, and mode of transportation (air, motor carrier, rail, water and pipeline and intermodal combinations). To a certain extent, the origin and destination (O-D) of shipments of mining, manufacturing, wholesale trade and selected retail establishments are covered.

Currently, the CFS is part of the Economic Census response that is mandatory under Title 13 of the U.S. Code. The establishments selected are asked to provide data on shipments sent during 1 week for each quarter. According to federal law governing Census Bureau reports, the data collected from the private sector cannot be disclosed in any way or form that permits identification of individual firms or establishments. This requirement translates into the practice of not disclosing cells in the tabulations that do not reach a minimum threshold of the number of companies. The net result of doing this is to reduce the level of geographic detail provided by the CFS. The lack of appropriate geographic detail is widely acknowledged to be a major weakness of the CFS, along with industry scope, commodity detail, periodicity, and timeliness. All of which has been made worse by the decisions to reduce the CFS sample size from 200,000 establishments in 1993 to 100,000 establishments in 1997 and to 50,000 establishments in 2002. This reduction in sample size had a much greater impact on data reliability with the resultant suppressions than it had on data disclosure.

Analysts and researchers in both the public and private arena use data from the CFS. However, CFS data are often inadequate because of gaps in shipment and industry coverage, the lack of geographic and commodity flow detail at state and local levels, lack of international flows, and the inability to capture rapid changes in economic and global trade cycles. On average, the CFS data (1993, 1997) were released 3 years after it was collected. As a result, the CFS must be supplemented with data from other sources or models to support analysis and mapping of spatial commodity flow patterns.

The purpose of this discussion is to examine some alternative data collection, packaging, and products that may improve the usefulness of the CFS by understanding the data issues that transportation professionals expect the CFS to fulfill. In this way it may be possible to both alter what is collected and how data should be packaged to better serve the CFS end user.

The paper starts with a brief summary of the Saratoga Freight Data Needs Conference because its findings provide a rather compact definition of freight data needs as perceived by a wide spectrum of users. Then, the paper focuses on identifying the expectations of CFS users (section III). Subsequent sections discuss ways to improve the CFS data products from different perspectives: within the current CFS structure (section IV), by adopting combinations of selective statistical aggregation and improving linkages to other data products (section V), by taking advantage of and fostering freight transportation research (section VI), and finally by defining the kind of data products the freight transportation modeling community would like to have.

For the most part, the authors have focused on data needs for the most demanding of users, referred to here as "power users," which are the ones that make intensive use of the CFS for both modeling and policy analyses. This group is comprised of professionals and researchers from consulting companies, state and metropolitan planning organizations (MPOs), software developers, and universities. Although numerically a relatively small group of individuals, power users do have considerable influence in transportation planning decisions because they do the technical work, developing the software tools and the analytical techniques that support the analytical process of freight transportation planning. The decision to focus primarily on power users seems justified by the fact that, since they are the most demanding of users, satisfying their needs is likely to meet the expectations of sporadic users. Furthermore, should sporadic users become power users in the future, enhancing the usefulness of the CFS will enable them to have at their disposal the kind of data products that they would need. It is also important to acknowledge that the CFS has other uses, especially at the federal level that may not be discussed in this paper. For example, the Bureau of Economic Analysis uses of commodity detail for their U.S. and regional input-output accounts.

SARATOGA FREIGHT DATA NEEDS CONFERENCE

A conference entitled "Data Needs in the Changing World of Logistics and Freight Transportation" was held November 14–15, 2001, in Saratoga Springs, New York. The conference was sponsored by the New York State Department of Transportation (NYSDOT), Transportation Research Board, Bureau of Transportation Statistics (BTS), FHWA, AASHTO, and North East Association of Transportation Officials.

The main objective of the conference was "to provide transportation officials with a broader understanding of data issues associated with the changing focus of the global competitive markets and its implication on the existing transportation infrastructure, trade corridors and market areas." The conference focused on four main themes which are very germane to the discussion here:

- Understanding the underlying reasons for freight movements in a complex world where supply chains and trade areas constitute the context in which freight is generated;
- Identifying the purpose for which the data are to be used;
- Ensuring that future data collection efforts will be useful to a broad spectrum of users; and
- Taking actions to develop a consistent framework for future data collection.

The conference produced a remarkable consensus on a number of crucial issues related to these themes; two which are extremely pertinent to any discussion of CFS data products are

- The recognition that freight flows are regional, national, and global in nature and involve freight corridors and trade and market areas and

- The need to understand the underlying causes of freight flows before deciding on what additional information is to be collected.

From that consensus, recommendations were made which stressed the need to

- Understand freight data gaps,
- Collect additional local O-D data,
- Integrate modal data collection activities,
- Develop and use innovative technologies for tracking freight, and
- Determine the true impact of congestion on economic development.

The subsequent discussion will focus on the first three. In doing this, it is important to highlight that the appropriateness of the suggestions made in this paper depend on the goals pursued. If the ultimate goal is to improve CFS data products from the standpoint of what is actually created and distributed, then the main focus should be on suggestions pertaining to timeliness and relevancy of the tabulations. However, if the ultimate goal is to identify how to add value to the current freight data collection program so that it could answer the types of questions raised in the Saratoga conference, then the transportation community should focus on other options. Otherwise, the CFS will continue to be the CFS as it is today, that is, inadequate for many state transportation planners.

Since a key objective of the CFS is to estimate and understand commodity flows, it is important to examine the interrelationships of at least the following items, and then assess how the CFS might be structured and compiled differently.

- Costs and impacts of commodity flows on regions and local areas,
- Relationship between supply chains,
- Flows and firm location behavior, and
- Cost and benefits of international trade.

In the following sections, the authors discuss a number of ideas to increase the short-term usefulness of the CFS, within the current structure; present alternative concepts for the freight data collection program; and discuss long-term freight transportation modeling needs. At the end of the paper, the set of critical questions are included.

USER EXPECTATIONS

This section discusses some of the questions CFS power users ask themselves when using the CFS.

Industry Sector

Why can't we identify movements by industry sector [two-digit North American Industrial Classification System (NAICS)] in addition to just commodity type [two-digit Standard Classification of Transported Goods (SCTG)]? This is important since it is the economic flow of commerce that drives the movement of goods to market. Input-output models tend to be industry based; summarization by industry type would create new tables in the CFS that would aggregate data at a different level. The issue of industry clusters and market areas may become much more apparent and geographic detail below the state level may be possible.

Origins and Destinations

O-D has typically been defined as: destination states from the state of origin. The CFS is basically a shipment survey from establishments. In other words what is sent out is tracked. In the 1997 CFS a new table appeared summarizing all origin states for a destination state, almost the reverse flow. Unfortunately, this is not a complete state to state table. In the outbound direction the non-disclosed data is readily available by state, but in the reverse direction it is not available. What is lacking is a survey of shipments received by establishments. Why can't the establishment being sampled also respond to questions regarding packages received, including origin and the delivery mode of transport?

Trade Corridors

Key trade corridors are defined by route(s) and specific destination locations anchored by metropolitan areas. The cost to ship is often dependent upon the reverse flow from the destination and how much in or out of balance the directional movements are. Do we understand anything about state-to-state or regional flow imbalance as to how it influences modal choice and price? The CFS reports value of the shipment. Why can't it also report the shipment cost along with the value? Theoretically it should be easy to ascertain when the package was shipped and when it arrived. Why can't the survey then report the in-transit duration time? Such survey modifications would provide very important pieces of information.

A careful examination of destination states from states of origins from the 1993 and 1997 CFS will easily show major state trading blocks in which intra-regional shipments occur. Many states have significant proportions of their originating trade with one or more adjacent states. This relationship forms a new geographic aggregation that may not be reflected by traditional Census groupings. Why can't movements be classified in terms of which regional trading areas they are within, part of or between? This is an important aspect of beginning to identify the concept of trade corridors. Within this context the movement could be to or from or within a metropolitan area (MA).

Modes and Domestic and International Totals

Have we included all modes, or modal combinations for both domestic and international flows? The regional import-export component is such that domestic shipments capture only a portion of the entire supply chain movement. The mode by which goods penetrate inland from ports is missing from the CFS. International movements are also lacking. Why can't the Transborder Surface Freight Data (TSFD), which identifies province to state moves on a monthly and annual basis, be combined with the CFS to provide information on those movements during the time period that the CFS is being conducted? After all, the Census Bureau processes these data for BTS, so coordination should be easy. The Census also processes total imports and exports to foreign countries from states. Why can't these be combined to help the account total, and then enumerate any the CFS domestic movement that is an export? Equally important is information about waterborne movements, both domestic and international. The U.S. Army Corp of Engineers collects these data as activities in ports and on navigable waters. Why can't they be included during the period of time the CFS is being conducted? (The integration of the various data sets available is further discussed later in the paper.)

Time Series

How do flows vary over time? The 1993 and 1997 CFS are already in CD format. The 2007 CFS should also combine the data from the three previous CFSs (1993, 1997, 2002) and produce a set of time series numbers based on state to state flows assembled along with these and other data on differing levels of detail.

Measurement Units

Traditionally the CFS has focused on value, tons, and ton-miles. Within the computation of ton miles is the measure of distance, indicating what proportion of value and tons go how far. The CFS provides these data by commodity. Why not provide them by industry shipped and industries received and then collapse them into the categories that stratify the vehicle choice for the transportation component? Is it low value-to-weight ratio or high value-to-weight ratio? Does it have to be there tomorrow or can it take its time? It would be useful to contrast these aspects against shipment cost and distance and flow balance.

One of the big problems with the CFS is that it is often used to try to estimate total flows without having a good set of account totals at the network level. The CFS represents movements of commodities, whereas traffic counts typically capture movements of vehicles, and are readily available and very detailed. Unfortunately, there is very little detail on network flow as this is typically derived from origin and destination information, and assumptions are required for empty back haul and local movements. Equally important, other sources of modal data that could establish or serve as an account total are not included or well linked as part of a comprehensive and integrated effort of the Census Bureau to improve upon the data it reports. A proposal about how to address this issue is discussed later in the paper.

IMPROVING DATA PRODUCTS WITHIN CURRENT CFS STRUCTURE

This section discusses some ideas that are expected to add value to the CFS and should be relatively easy to implement because they only involve changes in the presentation format, and the way in which the raw data is processed.

Linkages to Previous CFS Data

One of the most important improvements would be to undertake and link flow pairs at the state to state, state to MA, and MA to MA level over time. In this way the analyst can compare changes on flow pairs over time for tons, value, ton-miles, and distance. The real question is whether in-transit travel time and shipment cost can be ascertained. These are two very useful pieces of information that illustrate the immediacy of the shipment versus the cost, or potential for modal tradeoffs. It would then enable the analyst to track performance over time. Having knowledge about these data items as part of a time series may help in understanding why intermodal facilities really work, as well as to provide the answer to the key question: "If you build (or improve it) will they come?"

Data Formats

- Microdata are a "must have" for the power user, as the experience with the Vehicle Inventory and Use Survey (VIUS) indicates. The 1997 VIUS CD has a user interface, but it also has a database containing microdata that enables a user to construct tables with greater dimensions and for variables not specifically analyzed in the national reports or standard tables. CFS data could be made available in a similar format.
- Database formats such as MS Access, SAS, SPSS, dBase, etc., could be prepared as an easy end user option for technical analysts. Previous CFS data products should be reprocessed to this format for consistency.
- O-D matrices should be provided along with tables that compare change over time for value, tons, ton-miles, distance, and shipment cost. Some assessment or table should be provided to examine modal shift over time. Sub state flow to and from MAs should be used to identify corridors and growth corridors. The interstate essentially connects all MAs; figuring out corridor flows should not be that difficult. In fact even at a state level it becomes pretty obvious. Once corridor flows are established, then flow imbalances can be ascertained, maybe even by mode.
- Linkages to waterborne, air, port and customs import/export data should be developed and these data sets combined with the CFS, as they can provide the account totals lacking in the CFS.

Tables and Data Sets

- Intercity O-D patterns. If the CFS sample is drawn on an MA and non-MA basis as opposed to a state basis then aggregate tonnage, value, and distance could easily be developed from MA to MA, MA to rest of same state, or MA to rest of another state. Although this may indeed increase the overall sample size, this may prove to be a worthwhile improvement because it would provide a more meaningful geographic basis for the analyses.
- Interstate O-D patterns. If the focus of the CFS is based upon sampling MA and non-MA areas the aggregation up to the state is possible, as long as provisions are made to disaggregate multistate jurisdictions.
- Trading area, blocks, states, and regions. These delineations are easily made at the state and regional level the way the CFS is presently constituted. State-to-state movements can easily be depicted graphically. Values can be computed for each state and the most significant trading partners then identified and then trading areas become obvious. If the CFS is structured as an MA-based survey then this type of aggregation can be easily undertaken, and then trade corridor measures of tonnage and value, could be made available.
- Movements within large MAs. Clearly in multistate labor market areas or very large MAs some delineation along jurisdictional boundaries can be made. For example, the central MA for New York–New Jersey–Connecticutt–Pennsylvania could easily be separated in the New York and New Jersey portions. The New York portion given its size could easily be separated into the New York City primary MA and the Nassau–Suffolk MA providing that the sample was drawn to reflect state boundaries for very large MAs. Knowing that goods moved to the New

York–New Jersey–Connecticutt–Pennsylvania central MA is not as useful as knowing into which portion it was destined for. The number of establishments in this type of jurisdictional disaggregating is larger than in most other MAs and even some states.

- Flow tables should be readily available for both directions for tons, value, distance and mode. Industry at the two-digit level (NAICS) should be created and commodity should be collapsed into larger aggregations or into groups that reflect how these commodities might be shipped, since that categorization determines mode and cost. By approaching the problem this way it may be possible to provide the three dimensional tables of O-D, mode, and commodity that are absent from the current CFS configuration.
- As an alternative to geographic aggregation, the authors suggest that more aggregate categories be used to identify the commodity type. As a result of using more aggregate codes the identification of a individual company is going to be more difficult, which will mitigate the concerns of the private sector, while enabling the Census Bureau and BTS to provide more geographic detail. For modeling purposes, the amount of information lost is minimal, since most behavioral models use two digits NAICS or standard industrial classification (SIC) codes and almost never use more than two-digit codes (Holguín-Veras, 2002). As an additional benefit, a more aggregate definition of the commodity type reduces the burden on respondents.
- Ports. Waterway ports airports and rail terminal yards should be uniquely identified if and where possible. Clearly container shipments may come from any one of these facilities. Just simply knowing to which MA or MA portion such movements travel, would be an improvement upon what is now available. When containers arrive in ports, their origin and origin mode and possibly their final destination and destination mode may also be known. While shippers may not know this, the CFS could have a port component to ascertain this portion of the flow. Once these containers enter a port, they create a domestic movement by truck, water, or rail that is not being measured.
- Cost of shipment by mode. If the value of a shipment can be ascertained, the postage or shipping cost can also be determined, along with possibly the mode for all or part of its shipment. This information could be collected and weighted by tonnage and distance to produce O-D cost per mile to ship tables by mode.
- Flow balance. In an MA-level survey it will be possible to easily determine the flow split for all O-D pairs by mode or in the aggregate.
- Enhancing the characterization of commodity flows and the linkages with land use. The current CFS predominantly focuses on transportation flows, with almost no attention being paid to the characteristics of the establishments that generate these flows. In the context where the sample size of the CFS has been reduced, there is more pressure to make the CFS as useful as possible despite greater variability. One way to accomplish this is to add a short section to the CFS that gathers the set of establishment characteristics most likely to be explanatory variables of commodity generation (e.g., area of company, number of employees), or tie it to some type of county business pattern characterization. Among other things, these data would enable the estimation of trip generation functions that could be used to fill the gaps in the CFS and provide a wealth of information that would benefit both state and local freight transportation planning.

STATISTICAL AGGREGATION AND LINKAGES TO OTHER DATA PRODUCTS

This section starts with a summary of a number of papers that have produced findings of relevance to the purposes of this section and then focuses on the discussion of how to improve linkages to other data products. Among other things, this will enable the reader to get an idea about the state of the art of transportation modeling, what works, and what does not. This is central to the paper's objective because it leads to the idea that the CFS should be reformed according to what research indicates are underlying relationships and the context for interregional flow, rather than maintained for the sake of format consistency.

Understanding how transportation infrastructure planning for highways, railroads, and port facilities and regional development evolves at the metropolitan planning level is critical in defining what type of commodity flow data are required. Equally important is the understanding that there are adjacent state relationships in the movement of goods and that these adjacencies form interregional commodity flows. Thus within the context of a national survey there must be the recognition that commodity flow may be determined by interregional relationships and hence the survey sample domain may be better served by a design that leverages those relationships. Also, in this context, a survey of domestic commodity movement absent some recognition of the impact of international trade literally will miss the boat.

ON THE AGGREGATION OF COMMODITY TYPES

Celik and Guldmann (2002) produced some suggestions about the underlying determinants that could improve the way the CFS is structured and the data products that can be reported out. In their work they examine the concept of adjacency to determine if having a common physical border has an effect on commodity flow between states. They also look at imports and exports that may be included in the CFS from the custom districts where the commodity enters or leaves the United States. States with custom districts are typically coastal (oceans or Great Lakes) or border with Canada or Mexico. They also examine a somewhat condensed list of commodity categories compared to the entire list in the CFS. Thus if one takes an economic view of looking at commodity flows, it should be possible to increase the level of geographic detail through commodity aggregation.

- Classifying shipments based on aggregated commodity type at the two-digit NAICS level may eliminate confidentiality issues.
- Aggregating states based on adjacency (dominant state trading partners) when known shipping relationships are apparent will define new geographic areas reflecting trade.
- Identifying when shipments go through a customs port to determine that the domestic move is part of an international movement. Providing a flag on the flow could reveal new movement patterns to explain TSFD.
- Aggregating commodity codes based on how they might be grouped for shipment or uses a particular mode. This is typically what happens anyway.
- Aggregating will enable the enhancement of geographic detail below state for ports, cities, and MAs in the development of flow data.
- Continuing to provide the existing tables for the purposes of historical time series.

Das and Andriamananjara (2004) aggregated commodity types in their report to the Office of Economics, U.S. International Trade Commission. In this paper seven composite clusters of commodity types are used:

- Agriculture,
- Natural resources,
- Food manufacturing,
- Light manufacturing,
- Heavy manufacturing,
- High-tech manufacturing, and
- Services.

In two different papers (Holguín-Veras 2002, and Holguín-Veras and Wang 2004) super groups of commodity types were successfully used to model the vehicle choice process and the decision to use electronic toll collection (ETC) as the payment method. In the first case, the data used came from an O-D sample in Guatemala City; and in the second case from a sample collected in the New York–New Jersey area.

The aggregation used in Holguín-Veras (2002) was based on SCTG codes and consisted of

- Cereal grains (02);
- Agricultural products (03);
- Prepared foods (07);
- A super group of agriculture and meat products (SCTG 01–09 except 02, 03 and 07);
- Monumental or building stone (10);
- Natural sands (12);
- A super group of minerals (SCTG 10–15 except 10 and 12);
- A super group of fuel and oils (SCTG 16–19);
- Chemical products and preparations (23);
- A super group of chemicals (SCTG 21–24 except 23);
- A super group of forest and wood products (SCTG 25–29);
- Non-metallic minerals (31); and
- A super group of manufactures and electronic products (SCTG 30–40 except 31).

Furthermore, this research found that the binary variables representing the type of inter sectorial flow (noted below) linking the activities at the origin and at the destinations were statistically more important than the commodity types:

- Retail to retail,
- Retail to wholesale,
- Retail to other,
- Wholesale to retail,
- Wholesale to wholesale,
- Wholesale to other,
- Other to retail,
- Other to wholesale, and

- Other to other.

Holguín-Veras and Wang (2004) estimated discrete choice models to represent trucking companies' decision to use or not to use the ETC system used in the New York–New Jersey area (E-ZPass). It was found that the following super groups of commodity types had a statistically significant role in the choice process:

- General merchandise,
- Cars,
- Building materials, and
- Food.

Furthermore, this research found that this choice process is also a function of company size, the type of market the trucking company is in (e.g., independent owner operators, full truck load) which provides backing for the suggestion to gather company specific attributes.

These categories greatly simplify the commodity categories that need to be analyzed, while still facilitating behavioral modeling, without violating confidentiality. In this context, O-D by mode and aggregate commodity may be possible. Pursuing this path might facilitate a better mapping of industry, commodity, and trading cluster at varying levels of geography. Combining this understanding with value, tons, distance and cost to ship would provide new data attributes for consideration.

Ogden (1992) provides a general discussion of the determinants of freight transportation demand. Niles (2003) built on Ogden's discussion focusing on the aspect of the cargo moving within cities to understand what influences the movement of trucks. The essence of Niles's discussion is that there are really three basic demand factors associated with freight movement:

1. What is moved,
2. The places between which goods are moved, and
3. The business-related processes that govern both the movement and its timing.

Based on the identified three sets of demand factors, Niles (2003) merged the comprehensive Standard Transportation Commodity Code (STCC) list of commodities in the CFS (the same could be done for SCTG) with everyday experience based on those commodity categories used in the VIUS to simplify the categorization of the most significant types of freight found in intra-urban transport and suggested the use of six major categories of commodity types:

- Homes,
- Factories,
- Warehouses,
- Retail and service,
- Intermodal, and
- Other.

It is becoming much more evident that given sample size limitations, collapsed tables could be constructed that would increase the level of geographic detail by decreasing the specificity of both the commodity type and the land use to which it travels and the activity at that land use.

CREATING LINKAGES TO OTHER DATA PRODUCTS

It is clear that, because of the huge complexity of modern logistics, it is extremely important to gather logistic information about supply chain patterns because this is the point where business decisions influence the characteristics of the freight movements. Although the CFS as presently constructed does not address this demand component, a proposal to integrate the CFS with elements of a carrier survey to assess logistic patterns is discussed later in the paper. However, there are some intermediate steps that could be taken to improve the overall usefulness of the U.S. data collection program.

In a shipper survey it is easy to ascertain the commodity of the item and some general concept of the land use or industry that is at the origin. By including a survey component that also samples the destinations for these shipments, and asks about the industry and the land use to which the item is being shipped, a much more realistic picture from a practical standpoint can be obtained for the first two demand factors. This concept also suggests that a tighter integration between the VIUS and the CFS is essential to improve the data that is collected.

The fact that both the CFS and the next VIUS will be conducted approximately during the same time window provides a good opportunity for coordination between these important surveys. The VIUS captures resident-based truck movements; it provides zones of operation (e.g., 50 mi, 100 mi) but it fails to ask do you cross a state boundary and into what state(s) did you cross? This is especially important if the range of operation is within an MA and the MA is in a multistate jurisdiction. For example, a 50-mi radius of operation in northern New Jersey could take the vehicle into New York, Pennsylvania, or Connecticut. The VIUS tells only about vehicles registered in the state and not about the vehicles that visit the state. It has very detailed operational characteristics (e.g., weight and body type information). By simply ascertaining the states of operation, as well as the MA, the VIUS data could be used to construct tables to reflect incoming vehicles from adjacent states. These data then could be used to better understand the CFS flows since both surveys are conducted at the same time. It's important to remember that registration of vehicles may be more related to tax policies in each state and less to do with bases for operation.

Improving the linkages between the CFS and VIUS may prove instrumental in improving the overall usefulness of both programs. Doing this would necessitate the definitions of

- An appropriate set of truck types (e.g., light, medium, medium heavy, and heavy);
- Company focus (e.g., truckload, less than truckload, intermodal, package express);
- Ownership types [e.g., private carrier, common carrier, third party logistic provider (3PL)]; and
- Geographic focus (e.g., long-distance trucking, intra-urban).

The integration of common concepts from the VIUS to the CFS affords the opportunity to enhance the usefulness of the CFS. Equally important are linkages to other products such as the Annual Survey of Manufacturing and the Economic Census: Manufacturers.

Recently in New York State, a business location analysis tool was made available to the MPOs by the NYSDOT. This tool enables the MPOs to examine the composition of businesses establishments as geographic point locations by SIC or NAICS two-digit categorization, sales, and number of employees within their existing traffic analysis zone (TAZ) structure. Since these data reflect the actual locations (more or less) of where workers work and in what industry, the

MPO is better able to understand how to restructure their TAZ boundaries. These business establishments then become a surrogate for land use type in their travel model.

The CFS as an establishment survey currently asks about shipments from the establishment. It could just as easily ask similar questions concerning the shipments received at that establishment. It could define business by their industry type and collapse the commodities to the industry-based groups discussed above, while at the same time maintaining the existing STCC/SCTG categories for the traditional tables. By integrating VIUS metropolitan categorization with state and MA information, (something that is easily done based on the address of the establishment) the shipment customer destinations and the origins for shipments received could be known, thus additional geographic categorization as well as shipment information can be obtained. While standardized tables have severe restrictions, the tables that could be generated based upon the suggested levels of aggregation above would have many more observations and greater geographic depth. The resultant data products would much more in line with how analysts try to model freight–truck flow in urban contexts.

The bottom line here is that the CFS should not endure simply because that is the way it is. Rather it should evolve to reflect what research indicates, and even what the CFS data indicates, to accommodate and reflect how shipping decisions are reflected onto the network of travel. In so doing, derivative products that use the CFS in part or entirety such as, Reebie TransSearch data or the FHWA Freight Analysis Framework activities will also benefit.

THE ROLE OF RESEARCH IN ENHANCING CFS DATA PRODUCTS

The usefulness of data for transportation demand modeling is to a great extent a function of geographic detail, timeliness, and the availability of the set of relevant attributes that enable the modeler to capture the specific dynamics of the particular process being studied. Because of the unique features of transportation modeling, data are needed to model both supply and demand.

In this context, since the main focus of the CFS is on freight demand, not much effort is spent on collecting data about the specifics of supply, i.e., how the logistic and freight industry go about transporting the commodities (e.g., logistic chains, distribution patterns). This is a fundamental weakness of the overall freight data collection program of which the CFS is part. However, as a demand-centered data collection exercise, the CFS does a fairly good job of capturing the general attributes of commodity flows (e.g., shipment size, commodity type), though it is far from being a timely product and, as is widely acknowledged, the CFS does not have sufficient geographic detail to become useful as an aid for transportation planning at either the state or the MPO level. These are major deficiencies that prevent taking full advantage of the potential of the CFS as a vital component of a freight data collection program.

Fulfilling the objectives of a complex data collection program such as the CFS requires a major paradigm shift from its current statistical focus to a program that is closely intertwined with (and that supports) the development of cutting edge freight transportation modeling techniques. This change is needed so that the CFS provides the foundation for the holistic analyses of the freight transportation the U.S. needs. A more meaningful integration between the CFS and freight transportation research would: (*a*) facilitate the analytical integration of data coming from the various data collection programs; and (*b*) enhance the CFS's ability to estimate the impact of truck traffic in the U.S. network by properly considering the role played by commercial vehicle empty trucks. These areas are discussed next.

FREIGHT MODELS AS DATA INTEGRATORS

Currently, freight data is collected in the United States in a piecemeal fashion. A recent report (BTS, 2001) provides a very useful example of the fragmented nature of the current data collection programs:

> A shipment of electronic equipment moving from overseas to a U.S. retail outlet arrives in the Port of Long Beach, California, via containership. The container is transferred to a railcar and travels by train to Chicago, where the load is broken into separate shipments. The electronic equipment is placed on a large truck with other shipments and driven to a distribution center in Indianapolis, where the truck's cargo is unloaded and the electronic equipment is separated out. The shipment is then placed in a delivery van and driven to its final destination in Fort Wayne, Indiana. (BTS, 2001)

This implies that three different data programs could potentially collect data for the very same shipment:

1. The US Waterborne Commerce Data (the water portion of the trip);
2. The Waybill Sample (the rail component); and
3. The CFS (the final leg of delivery from the warehouse to the end user), as discussed in the BTS report above and shown in Figure 1.

Among other things, this data collection structure masks the underlying production–consumption relations because the distribution center that is the origin of the commodity flow captured by the CFS is nothing more than an intermediate stop in the logistic chain. In this context, O-Ds of the physical trips do not equate with the production and consumption places as indicated in Figure 2. This presents a major challenge to freight transportation planning because the lack of integration among the various freight data programs is not conducive to a transparent identification of the production and consumption patterns.

In this context, freight transportation research has an important role to play in the development of analytical techniques aimed at the integration of disparate data streams. Among other things freight transportation models, particularly freight O-D synthesis, are bound to play a significant role as data integrators by weaving together data from the Waterborne Commerce, the Waybill Sample, and the CFS.

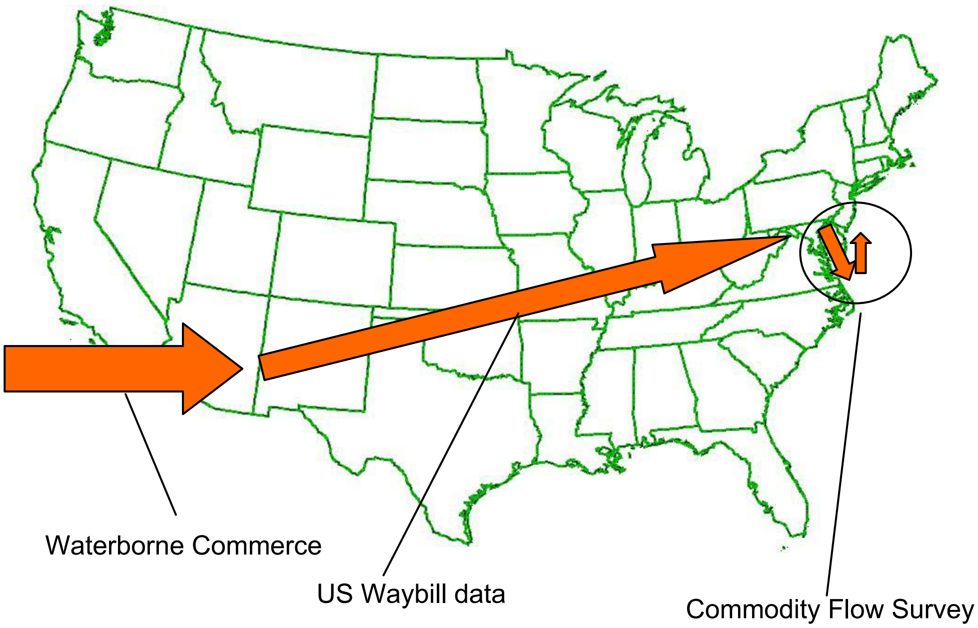


FIGURE 1 A typical intermodal shipment.

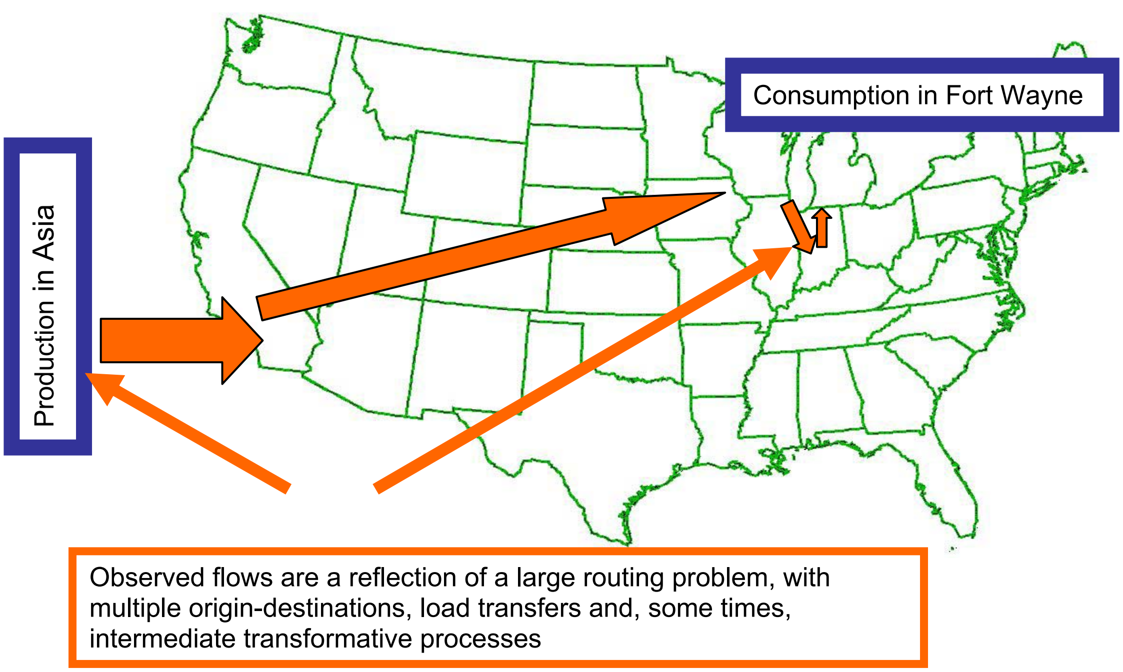


FIGURE 2 Production and consumption relations.

FREIGHT MODELS AS TOOLS TO COMPLEMENT THE CFS: COMMERCIAL VEHICLE EMPTY TRIPS

The CFS, because of its shipper-based nature, is able to capture the economic characteristics of the cargoes transported (e.g., shipment size, commodity type). However, because of its orientation, the CFS is unable to gather data about the vehicle flows associated with the transportation of the cargoes. This is extremely important for transportation planning purposes because the vehicle flows are the ones that directly impact the transportation network in terms of capacity, infrastructure deterioration, and environmental impacts.

Although data about vehicle flows is needed for all freight transportation modes, the trucking case deserves special discussion because of the significant role played by commercial vehicle empty trips that may represent 30% to 50% of the truck traffic in different corridors (Holguín-Veras and Thorson, 2003a; Wilbur Smith, 1969). Nationwide, the importance of empty trips can be appreciated by noting that according to the VIUS, 56.09% of truck miles correspond to empty trips (Table 1). The mean value becomes 27.47% if the companies that reported 100% empty mileage (most likely corresponding to trucks used for personal purposes) are not included. In all, something in between 24% and 33% of truck miles correspond to empty trips. As it may be expected, neglecting this number would lead to significant underestimation of the number of truck miles in the nation's highways.

As illustrated in Figure 3, the trucking industry undertakes complex tours to transport the commodities. A typical trucking industry practice is that two out of three moves must be profitable, i.e., with a paying load. As a result, the commodity flow from Chicago to Indianapolis (and from there to Fort Wayne) may indeed be transported in tours comprised of a number of stops to pick up and deliver, or by a full truck load movement. In either case empty trips are likely to be generated. This clearly indicates that, to fully quantify the network impact of the commodity flows, the corresponding empty trips must be estimated.

Since the CFS focuses on commodity flows, the estimates of truck traffic produced using the CFS are only able to describe loaded trucks (because the empty truck traffic is not a direct function of the commodity flows). Furthermore, research has shown (Holguín-Veras and Thorson 2003a and 2003b) that attempting to use simple expansion procedures to compensate for the missing empty trips leads to significant errors in the estimates of directional truck traffic that consistently reach 30% on average and, in some cases almost 300%.

TABLE 1 Percentages of Truck Miles Traveling Empty

Mean Value of Empty Mileage (%)	All Trucks	Straight Trucks Not Pulling Trailer	Straight Trucks Pulling Trailer	Truck Trailer Pulling a Trailer
Including 100% empty mileage	56.09%	58.54%	32.79%	24.15%
Not including 100% empty mileage	27.47%	27.55%	32.51%	23.83%

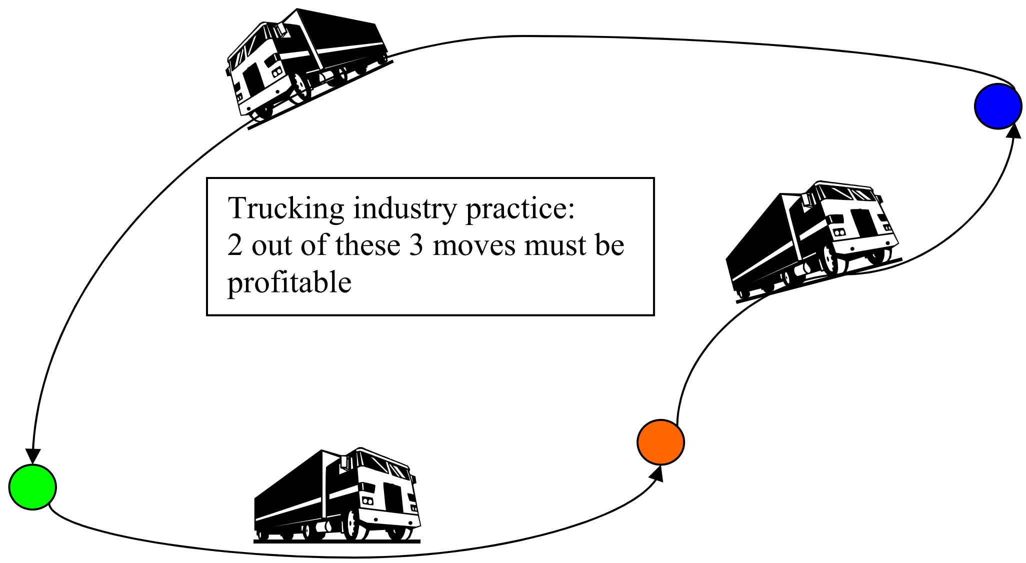


FIGURE 3 Typical empty trip patterns.

Fortunately, in spite of the complex probabilistic nature of the problem, the empty trips could be reasonably well estimated using approximations based on trip chain models (see Holguín-Veras and Thorson, 2003a and b). These models produce estimates of the empty trips by producing probabilistic estimates of trip chains using commodity flow data.

Although there have been significant developments in the area of empty trip models that enable analysts to infer empty trips from commodity flow matrices, more research is needed before such models could provide reliable estimates of empty truck trips for the U.S. network. Supporting such research, and more importantly, ensuring that federal, state, and local estimates of truck traffic properly model empty trips must be a priority.

CFS DATA PRODUCTS NEEDED FOR TRANSPORTATION MODELING

Developing a full picture of commodity flows and their impact on the transportation network requires data about: (*a*) characteristics of the cargoes transported; and (*b*) the logistic chains by which the cargoes reach their final destinations. The former is required to develop a good understanding of the underlying patterns of freight demand and the associated decision making processes, such as commodity generation, mode choice, vehicle choice among others; while the latter provides information about the supply side, i.e., how the freight and logistic industries organize themselves to transport the commodities.

The need to have basic information about both demand and supply translates into a major data collection challenge because the shippers, who know the details of the cargoes transported, usually do not know the particulars of the transportation and logistic process by which the cargoes reach their destinations; while the carriers, that know the details of the transportation and the overall logistic chains, are usually unaware of the characteristics of the cargoes they transport. As a result, the only way to get data about both the supply and demand sides is to incorporate elements of both shipper and carrier based surveys.

The lack of information about logistic chains is a major problem because of the complexity of modern logistics. Table 2 and Figure 4 show the itinerary followed by an Internet purchase made by the first author that required three truck trips (1–2, 4–5, and 5–6) and two air flights (2–3 and 3–4). It shall be clear that modeling such complex flows require data about the actual logistic patterns followed by the private sector; knowing that a shipment, (ultimately a commodity flow) from Great Falls, Montana, to Clifton Park, New York, took place is simply not enough.

In its current form, the CFS offers a partial though important view of freight flows because its shipper based focus enable the CFS to gather a rich array of data pertaining to shipment characteristics, such as: shipment size, frequency of shipment and commodity type. These data have proven to be extremely important for econometric modeling of decisions pertaining to mode choice, vehicle choice, routing patterns, and the like. The reason is that the commodity type is a proxy for the opportunity cost of the cargoes, which has been found to be an important explanatory variable. However, because of its shipper focus, the CFS is very limited in its ability to depict current logistics and/or transportation processes.

Two alternative approaches are worth discussing: a comprehensive longitudinal (tracking) survey and a combination of a shipper and a carrier based survey. The main features of these surveys are discussed next.

TABLE 2 Itinerary of a Typical Shipment

Node	Time/date	Location	Action
	26-Jan-05		
#6	4:25 P.M.	COLONIE, LATHAM, NY, US	DELIVERY
	9:20 A.M.	COLONIE, LATHAM, NY, US	OUT FOR DELIVERY
#5	5:16 A.M.	COLONIE, LATHAM, NY, US	ARRIVAL SCAN
#4	12:21 A.M.	E. SYRACUSE, NY, US	DEPARTURE SCAN
	21-Jan-05		
#3	6:51 P.M.	MINNEAPOLIS, MN, US	ARRIVAL SCAN
	20-Jan-05		
	2:00 P.M.	BILLINGS, MT, US	DEPARTURE SCAN
#2	7:55 A.M.	BILLINGS, MT, US	ARRIVAL SCAN
	19-Jan-05		
	9:39 P.M.	GREAT FALLS, MT, US	DEPARTURE SCAN
	7:07 P.M.	GREAT FALLS, MT, US	ORIGIN SCAN
#1	6:20 P.M.	US	BILLING INFORMATION RECEIVED

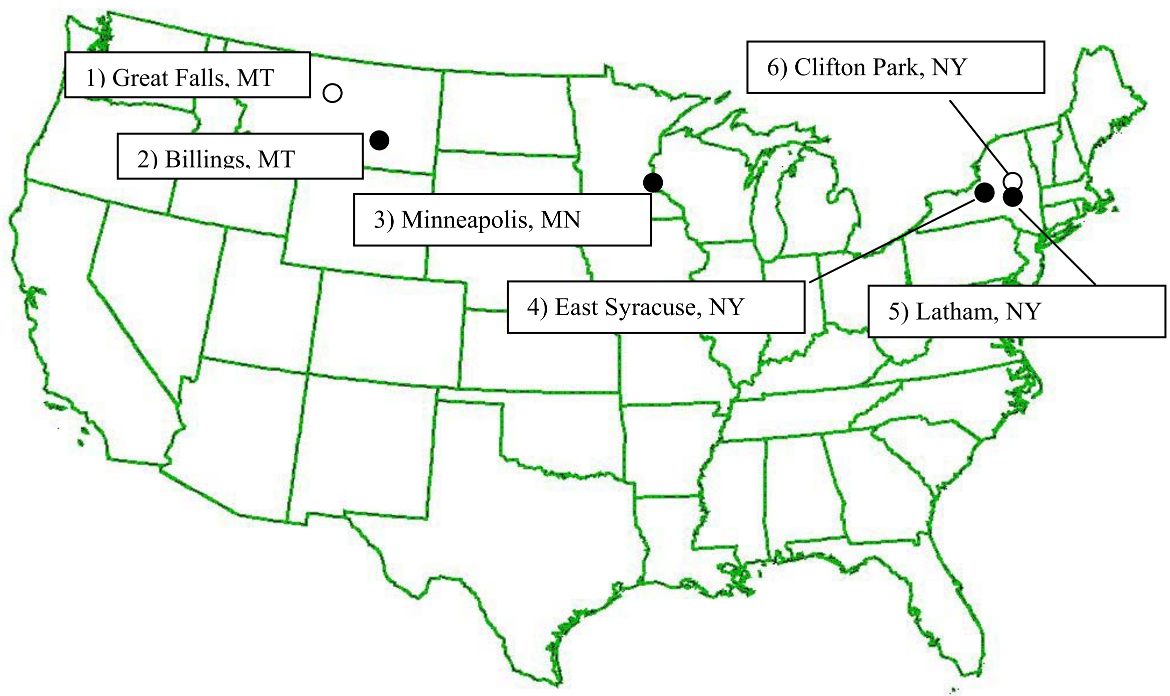


FIGURE 4 Logistic patterns.

LONGITUDINAL (SHIPMENT TRACKING) SURVEY

The main feature of this survey is that it would track individual shipments from the place of origination to the place of final consumption. In this context, the shipment characteristics would be provider by the shipper; while the carrier and logistic data would be provided by the actual carriers doing to the transportation, or any transformative process. This concept has been already successfully tested in France as part of a project funded by the European Commission entitled MYSTIC (Methodological Framework for Modeling European Passenger and Freight Transport) (Rizet, 2003). As part of MYSTIC, a shipper survey involving longitudinal tracking of shipments through the various stages of the logistic chain was successfully conducted. The sample was conducted by means of a two-stage sampling plan involving the sequential selection of establishments (among those with more than 10 employees) and shipments (with quotas to ensure a minimum data). Data gathered included information about the attributes of the shipment, who is the customer, who organizes the transportation of the cargoes, who actually perform the transportation, and which modes and routes are actually used.

The advantages of a longitudinal survey are that it: reveals actual behavior of all involved in the logistic chain; and is conceptually simple to understand. Among the disadvantages, one must highlight its cost, response burden on the private sector, and equally important, that it would require a significantly large sample to be able to provide consistent estimates of total flows by commodity type and mode.

COMBINATION OF A SHIPPER- AND A CARRIER-BASED SURVEY

This alternative would require development of a new survey, though it could be two separate but complementary surveys, with elements of both a shipper and a carrier survey. The former component would be a CFS like survey that would gather data about shipment characteristics and would provide control totals of the amount of cargo transported in a given year. The latter component would be a smaller longitudinal survey that would gather, in the manner described before, data about the various steps in the logistic process. Statistical modeling would be required to link up both surveys and produce a comprehensive set of estimates.

This alternative is likely to be the most cost efficient because it would build on the existent CFS know-how and would provide invaluable information to complement the process of imputation of logistic patterns currently done by the Oak Ridge National Laboratory. Although this alternative would require significant statistical modeling, these challenges could be overcome if there are resources and the will to move forward.

KEY QUESTIONS TO BE ADDRESSED

Having put forward a preliminary idea about how to improve the usefulness of the CFS and the overall freight data collection program, the authors would like to highlight a set of questions for discussion.

- Can movements by industry sector (two-digit NAICS) in addition to commodity type (two-digit SCTG) be identified in the CFS?
- Can a superset of industry and commodity categories be constructed to enhance geographic data detail?
- Can the sampled establishments be asked to provide data on shipments received (what, how, where did they come from)?
- Can data about shipment cost (outbound and inbound), in addition to value be collected?
- Can regional trading areas be identified using the CFS data?
- Can flows and summary tabulations be aggregated based on regional trading areas?
- How does state to state or regional flow imbalance influence modal choice and price?
- Can the CFS be integrated with the transborder data (Canadian–Mexican province to state, monthly, and annual flows) to provide information on those movements during the time period that the CFS is being conducted?
- Can Census combine data on total imports and exports to foreign countries with the CFS to improve the accuracy of account total, and enumerate any CFS domestic movements that are exports or imports?
- Can tabulations using the 1993, 1997, 2002, and 2007 CFS be produced to examine how commodity flows, value, tons, ton-miles, modal choice, etc., vary over time?
- Can traditional CFS tabulations be broken down so they provide industry shipped and industry received summaries? Can they be collapsed into the categories that stratify the land use or vehicle choice for the transportation component?

- Can travel time to destination or time from origin and shipping cost be determined separately from the value of the goods shipped?
- Can the CFS data be provided in a format that will enable users to be able to create their own tabulations and analyses, as the VIUS does?
- Can the CFS focus on state and MAs and their major state subcomponents (e.g., New York City), so that the resulting breakdown can provide state-specific portions as well as possible indication of major internal flows (New York City, Nassau–Suffolk, and Westchester–Rockland–Putnam)?
- Can better waterborne, air, port, and customs import–export data be included, nationally, regionally, or at state or MA levels to have a complete picture?
- Can the characterization of commodity flows and linkages with land use patterns be enhanced so that commodity generation could be studied?
- Can a proactive program that enables the integration of the timing and the data streams provided by the different data collection activities (e.g., Waterborne Commerce, Waybill Sample, CFS, and TSFD) be developed?
- Can a pilot test of replacing the CFS with a longitudinal (tracking) survey be explored?
- Can a pilot test of combining the CFS, as a shipper based survey, with elements of a carrier-based survey to gather both commodity flow data and information about logistic chains be explored?

REFERENCES

Bureau of Transportation Statistics. 2001. Accessed at http://www.bts.gov/publications/transportation_statistics_annual_report/2001/html/chapter_02_box_06_one_intermodal_shipment_024.html.

Celik, H. M., and J. M. Guldmann. Spatial Interaction Modeling of Interregional Commodity Flows. Presented at 42nd European Congress of the Regional Science Association, August 2002, Dortmund, Germany.

Das, G. G., and S. Andriamananjara. Hub-and-Spokes Free-Trade Agreements in the Presence of Technology Spill Overs: An Application to the Western Hemisphere. Working paper 2004-09-A. Office of Economics, U.S. International Trade Commission, 2004.

Holguín-Veras, J. Revealed Preference Analysis of the Commercial Vehicle Choice Process, *Journal of Transportation Engineering*, Vol. 128, No. 4, 2002, pp. 336–346.

Holguín-Veras, J., and E. Thorson. Modeling Commercial Vehicle Empty Trips with a First Order Trip Chain Model. *Transportation Research Part B*, Vol. 37, No. 2, 2003a, pp. 129–148.

Holguín-Veras, J., and E. Thorson. Practical Implications of Modeling Commercial Vehicle Empty Trips. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1833*, Transportation Research Board of the National Academies, Washington, D.C., 2003b, pp. 87–94.

Holguín-Veras, J., and Q. Wang. On the Attitudinal Factors Explaining the Use of Electronic Toll Collection Technologies by Freight Carriers. Submitted to *Transportation Research Part C: Emerging Technologies* (CD ROM). Presented at XIII Pan American Conference of Traffic and Transportation Engineering, Albany, New York, 2004.

Data Needs in the Changing World of Logistics and Freight Transportation. New York State Department of Transportation, Saratoga Springs, November 2001.

Niles, J. Trucks, Traffic and Timely Transport: A Regional Freight Logistics Profile. Mineta Transportation Institute. June 2003.

Ogden, K. W. *Urban Goods Movement*, Ashgate Publishing Limited, England, 1992.

Rizet, C. Energy Consumed in Freight Transport: Estimates from Shipper Surveys. Presented at 2003 Association for European Transport, Strasbourg, France, 2003.

Wilbur Smith and Associates. *Motor Trucks in the Metropolis*. Automobile Manufacturers Association, August 1969.

Commodity Flow Survey
Improving Methods to Enhance Data Quality and Usefulness

JOHANNA ZMUD
NuStats Partners, LP

INTRODUCTION AND BACKGROUND

This resource paper provides a foundation for a workshop discussion on potential survey methods improvements for the 2007 Commodity Flow Survey (CFS). It presents background information on the CFS, such as survey objectives, data users and uses, and 2002 CFS survey methods. The paper offers brief discussions on methods issues and challenges and presents survey design considerations that would enhance data quality and usefulness.

CFS SURVEY GOALS

The CFS provides information on the flow of goods in the United States, specifically data on shipments originating from manufacturing, mining, wholesale, auxiliary warehouses, and selected retail establishments in the 50 states and the District of Columbia. While it is an establishment survey, the survey's focus is on the characteristics of shipments—neither on establishments nor on more detailed carrier information. The survey's goals are to estimate the characteristics associated with the ultimate origin and destination of shipments, the distances traveled by shipments of goods, the commodities shipped, the modes of transportation used to transport shipments, and the volume of shipments measured by weight and value. To collect data to meet the survey goals, a three-stage sample design is used (i.e., establishments, reporting weeks, and shipments) stratified by geography, industry, and size of establishment. During post-processing, CFS data are aggregated at the level of states and Bureau of Economic Analysis regions. While such aggregation is necessary so that the data meet certain statistical reliability and respondent confidentiality criteria, it limits the usability of the data for analyzing intrastate commodity and vehicle flows, as well as analyzing flows between metropolitan areas.

USERS AND USES OF CFS DATA

CFS data have a range of users and of uses. According to a 2003 Letter Report on the Commodity Flow Survey of the Committee to Review the Bureau of Transportation Statistics (BTS) Survey Programs (*1*), data from the CFS are used by various government agencies for policy and program purposes; by academic researchers; and by consulting companies, whose clients can range from businesses, state departments of transportation (DOTs), federal agencies, and associations (*Special Report 277*, 2003). Uses of the data include: analyzing trends in goods movement over time, conducting economic analyses, forecasting future demand for good movement and associated infrastructure and equipment needs, establishing benchmarks for estimating national accounts, and analyzing and mapping spatial patterns of commodity and vehicle flow. Primarily, CFS data are used at the national or state levels, where it is most reliable. In recent years there has been increased interest in CFS data from state DOTs and metropolitan planning organizations that reflect their growing interests in freight issues. The aggregated data, however, are of less utility for these sets of users, who require data at finer levels of geographic detail than is currently provided.

CFS HISTORY

Data on the flow of goods in the United States have been collected since the 1960s. The predecessor to the CFS, the Commodity Transportation Survey (CTS) was conducted by the Census Bureau between 1963 and 1983. These surveys produced measures on the flows of goods and materials according to mode of transportation. The early implementations covered only shipments of more than 25 mi made by manufacturing establishments, and they excluded Alaska and Hawaii (Davie, 2003). The 1977 implementation covered the entire United States, and included shipments of 25 mi or less. The last CTS was conducted in 1983, but data were not published because its small sample size had resulted in estimates with substantial bias.

National transportation policy and program requirements in the early 1990s resurrected interest by the DOTs and the Census Bureau in restoring a commodity-based survey to be conducted as part of the Economic Census, under the authority granted by Title 13, United States Code. Title 13 directs the Census Bureau to conduct economic censuses every 5 years, for years ending in “2” and “7.” The title’s language also makes response mandatory, establishes penalties for nonparticipation, and requires that the Census Bureau maintain the confidentiality of the information provided by respondents. The first of the series, however, was conducted in 1993, rather than 1992. Consequently, a CFS has been conducted by the Census Bureau under primary sponsorship from the BTS in 1993, 1997, and 2002 (Table 1).

OVERVIEW OF CFS METHODS

The design of the CFS has remained virtually static between 1993 and 2002. The Letter Report on the Commodity Flow Survey recommended that the 2007 CFS would benefit greatly from modifications that would update the 2002 CFS methods to use available funds more effectively, increase the sample size, and improve the overall usability of the data. An overview of 2002 CFS methods is presented to provide a basic knowledge foundation for discussions of potential improvements.

TABLE 1 Survey Costs and Sample Sizes: 1993–2002

	1993	1997	2002
Survey costs	15 million	19 million	13 million
Sample sizes	200,000	100,000	50,000

Source: Letter Report on the Commodity Flow Survey, Transportation Research Board, 2003

SAMPLE DESIGN

The CFS has historically relied on sample sizes of 200,000 (1993) and 100,000 (1997). The 2002 CFS sample was significantly smaller—comprised of 50,000 establishments drawn from a universe of approximately 760,000 U.S. establishments. While nearly a 50% reduction from 1997, the sample design used for the 2002 CFS was still extremely complex because of the fact that the population of shipments is extremely large and variable, and shipping records are kept in a geographically wide-ranging universe of establishments (Black et al., 2000). An overview the 2002 CFS sample design is presented below.

First Stage

The first-stage sample frame consisted of a subset of establishment records from the Business Register maintained by the U.S. Census Bureau that lists all known establishments located in the United States or its territories. This list includes establishments classified in mining (except oil and gas extraction), manufacturing, wholesale, or electronic shopping and mail order retail industries, as well as auxiliary establishments (e.g., warehouses and central administrative offices) with shipping activity. The list does not include establishments classified in forestry, fishing, utilities, construction, transportation, services, all other retail industries, farms, and government-owned entities (except government-owned liquor stores). The sampling frame was stratified by geography, and industry and primary stratum were formed using geographic-by-industry combinations. Because the 2002 sample was about half the size of the 1997 sample, certainty components were used to ensure coverage of less frequent types of shipments (e.g., air, water, rail, and hazardous materials). It was decided to identify those establishments which made the bulk of these types of shipments in 1997 and to select them with certainty. This design strategy was used to reduce the sampling variability of the estimates.

Second Stage

The frame for the second stage of sampling consisted of 52 weeks from January 6, 2002, to January 4, 2003. Each establishment selected into the 2002 CFS sample was systematically assigned to report for four reporting weeks, one in each quarter of the reference year. An establishment's assigned reporting week was in the same relative position for each quarter.

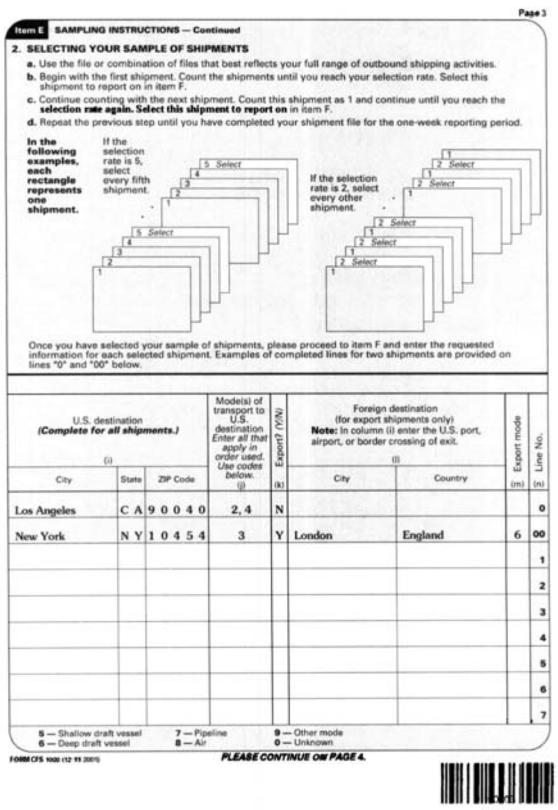
Third Stage

Each respondent was asked to construct a sampling frame consisting of all shipments made by the establishment during each of the four reporting weeks. Each respondent was instructed to count or estimate the total number of shipments comprising the sampling frame and to record this number on the questionnaire. For each assigned reporting week, if an establishment made more than 40 shipments during that week, the respondent was asked to select a systematic sample of the establishment's shipments and to provide information for only those sampled shipments. If an establishment made 40 or fewer shipments during that week, the respondent was asked to provide information on all of the establishment's shipments made during that week.

DATA COLLECTION

The 2002 CFS was a mail survey. The questionnaire, itself, was seven pages in length, designed with a row and column orientation (Figure 1). The statement "your response is required by law" was on the front page, along with a statement of confidentiality, and contact information for support. Included in these seven pages were instructions for selecting the sample of shipments. A separate eight-page instruction booklet accompanied the questionnaire. Each establishment was mailed a questionnaire, instruction guide, and a commodity coding manual for each of its four reporting weeks. These packets were mailed once every quarter.

For each reported shipment, respondents were asked to provide information on shipment identification number, the date on which the shipment was made, value, weight, commodity code, commodity description, mode(s) of transportation, U.S. destination or port of exit for exports, an indication of export shipment, and the United Nations or North America number for hazardous materials shipments. For a shipment that included more than one commodity, the respondent was instructed to report the commodity that made up the greatest percentage of the shipment's weight. For an export shipment, the respondent provided the mode of export and the foreign destination city and country.



Page 3

Item E SAMPLING INSTRUCTIONS — Continued

2. SELECTING YOUR SAMPLE OF SHIPMENTS

a. Use the file or combination of files that best reflects your full range of outbound shipping activities.

b. Begin with the first shipment. Count the shipments until you reach your selection rate. Select this shipment to report on in item F.

c. Continue counting with the next shipment. Count this shipment as 1 and continue until you reach the **selection rate again. Select this shipment to report on** in item F.

d. Repeat the previous step until you have completed your shipment file for the one-week reporting period.

In the following examples, each rectangle represents one shipment.

If the selection rate is 5, select every fifth shipment.

If the selection rate is 2, select every other shipment.

Once you have selected your sample of shipments, please proceed to item F and enter the requested information for each selected shipment. Examples of completed lines for two shipments are provided on lines "0" and "00" below.

U.S. destination (*Complete for all shipments.*) (i) City	State	ZIP Code	Mode(s) of transport to U.S. destination *Enter all that apply in order used. Use codes below.* (j)	Export? (Y/N) (k)	Foreign destination (for export shipments only) **Note:** In column (i) enter the U.S. port, airport, or border crossing of exit. (l) City	Country	Export mode (m)	Line No. (n)
Los Angeles	C A	9 0 0 4 0	2, 4	N				0
New York	N Y	1 0 4 5 4	3	Y	London	England	6	00
								1
								2
								3
								4
								5
								6
								7

5 — Shallow draft vessel 7 — Pipeline 9 — Other mode
6 — Deep draft vessel 8 — Air 0 — Unknown

FORM CFS 1000 (12-11-2001)

PLEASE CONTINUE ON PAGE 4.

100003

FIGURE 1 Excerpt of 2002 CFS questionnaire.

DATA PROCESSING

An automated editing system was introduced in 1997 that enabled Census staff to identify and correct problematic reports (*2*). The edit check program checked for inconsistencies, range errors, critical omissions, and assigned a series of flags for missing or incorrect data. Respondents who reported shipments having typically large value or weight when compared to their rest of the reported shipments were contacted to verify the reported numbers. This was done to reduce estimation problems caused by large and infrequent shipments.

STABILITY OF 1993, 1997, AND 2002 CFS METHODS

The design of the CFS has remained fairly stable over the past 10-year period (Table 2). While this serves a purpose in ensuring stable measures of key estimates for trend purposes, it fails to fully utilize innovations in methods and advances in data collection technology over the same 10-year period. This does not mean that zero changes have been in the CFS design. Some changes in the CFS design were made between the 1993 and 1997 CFS. Some specific examples of these changes included

- Reducing sample size 200,000 establishments in 1993 to 100,000 establishments in 1997 to allow for more intensive follow-up of problem reporters and ensure higher quality, more timely and accurate data;
- Changing the measure of size estimate used in first-stage sampling from an employment or payroll based measure to an estimate of annual total value of shipments to more efficiently and reliably stratify the establishments represented in the sampling frame;
- Reducing the number of primary strata from 18,000 to 3,400 through the use of clustering to offset increases in sampling variability because of the decrease in sample size;
- Shortening the reporting period from 2 weeks to 1 week to reduce respondent burden;
- Improving the respondents' third-stage sampling instructions, along with an example and illustration of shipment sampling;
- Identifying and including all large shipments made during the year for all selected establishments; and
- Identifying establishment to be sampled with certainty based on prior CFS reports.

According to the Census, these changes were made to improve timeliness of data products, reduce respondent burden, and lessen the influence of large and frequent shipments (Black, 1997). Post-evaluation of these modifications (via statistical measures) demonstrated improvement over the 1993 sample design.

TABLE 2 Comparability of 1993, 1997, and 2002 CFS Methods

	1993	1997	2002
Sample size	200,000 selected from universe of about 790,000 establishments	100,000 selected from universe of about 770,000 establishments	50,000 selected from universe of about 760,000 establishments
Basic sample design	Three-stage design (establishment, reporting period, shipment), stratified by industry and geography	Same as 1993	Same as 1993
Industry coverage	Based on 1987 SIC, manufacturing (not printing), mining (not mining services and oil and gas extraction), wholesale, retail catalog and mail order houses, auxiliaries	Same as 1993	Based on 1997 NAICS, manufacturing (not prepress services), mining (not support activities and oil and gas extraction), wholesale, retail electronic shopping and mail order houses, auxiliaries
Data collection mode	Self-administered, mail questionnaires	Same as 1993	Same as 1993
Data entry mode	Key from paper	Same as 1993	Optical scanning with key from image
Reporting period	2 weeks in each of the four calendar quarters of reference year	1 week in each of the four calendar quarters of reference year	Same as 1997
Shipment sample	Respondents constructed frame and drew sample, reporting characteristics for each sampled outbound shipment	Same as 1993	Same as 1993
Shipment data items	Total value, total weight, commodity that contributes most to shipment's weight, mode(s) of transport, origin, destination, containerized (Y/N), hazardous materials (Y/N), export, mode of export and destination	Same as 1993	Total value, total weight, commodity that contributes most to shipment's weight, mode(s) of transport, origin, destination, hazardous materials (UN/NA), export, mode of export and destination
Commodity classification	STCC	SCTG	Same as 1997
Modes of transport	For-hire truck, private truck, rail, air, inland water, deep sea water, pipeline, parcel, USPS, or courier, other, unknown	Same as 1993	Same as 1993

NAICS: North American Industrial Classification System; UN/NA: United Nations or North America; STCC: Standard Transportation Commodity Classification; SCTG: Standard Classification of Transported Goods; USPS: U.S. Postal Service.

Evaluations after the 1997 and 2002 CFS surveys confirmed the need for further research and improvements in the design of the CFS. In fact, it was uncertainties about the availability and level of funding that severely limited opportunities for research on and implementation of methodological improvements for the 2002 CFS (*Special Report 277*, 2003).

2007 CFS ISSUES, CHALLENGES, AND DESIGN CONSIDERATIONS

The Letter Report on the Commodity Flow Survey recommended that BTS and the Census Bureau proceed with planning for the 2007 CFS. It was suggested that this planning effort should explore opportunities for methodological improvements, particularly those with a potential to reduce survey costs through more efficient sampling and data collection. Equally important were changes that would mitigate respondent burden to improve response rates, as well as improvements that would enable a larger sample size at finer levels of geography to provide data more useful to a wider range of users. This section of the paper provides an overview of current issues, challenges, and design considerations in fulfilling this recommendation.

SAMPLING METHODS: ISSUES AND CHALLENGES

Sampling for establishment surveys is fundamentally different than for household surveys (Plewes et al., 1988). The following are examples of the types of issues that differentiate establishment survey samples.

- Establishment surveys rely on list frames, which are subject to problems associated with errors of commission and omission as well as misclassification.
- Selection of establishment samples must be current with respect to economy dynamics—both in terms of the constant cycle of business births and deaths and in terms of economic cycles (i.e., recession periods or industry changes).
- Sample rotation should be considered to ease the burden on businesses that have been participating in the survey for multiple iterations.
- Total universe coverage is impeded by changes in ways of doing business.
- The distribution of establishment populations are typically quite skewed, with a few large firms commonly dominating totals for most characteristics of interest.

Such issues affect the frame development and maintenance, sample design, and estimation practices. Establishment surveys differ from household surveys in another important regard related to sampling. While innovations in sampling, such as dual-frame sampling, are being field tested to more efficiently sample households, innovations in establishment sampling have not been as evident. Perhaps this is because establishment surveys tend to rely on list frames, which lend themselves less than other possible frames to innovative techniques.

SAMPLING METHODS: CONSIDERATIONS

Census and BTS are considering a larger sample size for the 2007 than the 50,000 establishments surveyed in 2002. Obviously, the survey budget will be a determinant in the number of establishments that will be sampled as well as the reporting requirements that will be established. In addition, the final sample design will need to reflect other data needs and methods-related considerations, such as

- The optimal sample size and design based on survey objectives (i.e., desired precision levels for specific shipment characteristics, modes, trend changes, etc.).
- The level of geographic detail the data will provide—avoiding disadvantageous breakpoints—such as sample size not large enough to determine commodity flows at the requisite level of geographic detail.
- The various trade-offs for first- and second-stage sampling. For example, the advantage of sampling 100,000 establishments canvassed twice per year versus 50,000 canvassed four times per year.
- The trade-offs between having a larger sample size and tailoring the survey experience to the respondent to increase response rates. Sample size makes an enormous difference in the types of response-inducing procedures that can be employed.

In addition to considerations of sample size adjustments, Census and BTS need to prioritize potential sample design changes on the basis of their contributions to improving sampling efficiency or minimizing statistical error of the estimates. The menu of potential sample improvements includes

- Exploring the possibility of a CFS pre-canvass operation that might be used to identify non-shipping auxiliary establishments, to better understand reporting arrangements, and to identify shippers that use less common modes or those that ship hazardous materials or exports;
- Investigating the inclusion and thresholds for selecting establishments with certainty;
- Improving the measure of size estimates critical to first-stage sampling, particularly for auxiliary establishments;
- Examining ways of providing finer levels of geographic and commodity detail, while maintaining the statutory obligation for confidentiality of individual establishments (*3*);
- Changing the way shipments are sampled, such as stratifying shipments by size and including all large shipments over a long period (1 month, 1 year) rather than the current 1-week reference period; or
- Exploring ways to minimize respondent errors during third-stage sampling, such as implementing alternative shipment selection procedures.

INSTRUMENT DESIGN: ISSUES AND CHALLENGES

Some changes were made to the CFS questionnaire between the 1993 and 1997 CFS. For example, the respondent instructions for the third-stage sampling were revised to include a written example of the sampling procedure with a diagram. Also an illustration was designed to depict how information about each sampled shipment should be recorded by providing examples of a multiple-mode shipment as well as an export shipment of hazardous materials. However, these changes were predicated on Census Bureau and BTS staff experience gained from the 1993 CFS and not from pretesting of the 1997 questionnaire (Black et al., 2000). Nor was the effectiveness of the changes evaluated with respondent debriefing subsequent to the 1997 CFS. Pre-existing issues with the design of the questionnaire that would affect adequacy of response or usefulness of instructions have never been explored with respondents, themselves.

As Dillman (2000) notes, establishment surveys, more than other types, are frequently designed in ways that make them very difficult to complete. He cites "tradition" as the cause, more than anything else. Mandatory, government surveys, like the CFS, have often been thought of as "forms" that had to be made as short and precise as possible, rather than as queries to be read, fully comprehended, and thoughtfully answered. For some of these same reasons, pretesting of mandatory establishment surveys has been uncommon. Willimack et al. (2002) outlined various ways to improve the design and testing of establishment surveys, such as cognitive interviews, focus groups, site visits, record-keeping studies, and consultation with subject area specialists and other stakeholders. They also recommend documenting respondents' feedback and conducting ongoing quality evaluations to diagnose questionnaire problems. Finally they recommend the use of experimental comparisons of the original and revised items.

Information on problem reports, missing data, and nonresponse from both the 1997 and 2002 CFS suggest that it is time to pretest the questionnaire and instructions. At the same time, it is important to assess ways in which the burden on respondents may be minimized, above the changes made to the 1997 questionnaire, as this was a clear recommendation from the Letter Report on the Commodity Flow Survey. Respondent burden is significant because it is associated with nonresponse (discussed in a subsequent section of this paper). At the same time, the relationship between privacy/confidentiality and nonresponse is also significant. As Prewitt (2004) discussed, "even though the census is mandatory, 71% of the public [in a poll] said that an intrusive census questions should go unanswered and another 13% were unsure, leaving only 16% saying that a question considered invasive should be answered." He places blame for this situation on a "new urgency" about privacy and confidentiality resulting from unimpeded access to individual's personal information brought about by new technologies and data mining. Pretesting the CFS materials would provide insight into the best ways to collect data that may be viewed as confidential business information, and therefore, left unanswered.

INSTRUMENT TESTING: CONSIDERATIONS

BTS's priorities for the 2007 CFS include doing a general content review of the questionnaire and cognitive testing of the questionnaire and the accompanying instructions for possible improvements. The agency is interested in analyzing how respondents understand and respond to the questions on the survey instrument, and whether or not there are better ways of capturing the information of interest. Dillman (2000) suggested conducting cognitive interviews on-site when pretesting establishment survey questionnaires. This option should be considered for the 2007 CFS as it would be helpful in ascertaining how businesses are organized and determining what questions should be asked of businesses, of whom within the business, and when in the business year. Important research questions are

- Identifying in terms of job title or position, who is the person within the business organization most likely to know or be able to find the answers to survey questions;
- Identifying which person in the organization (establishment) has the authority to grant permission for reporting;
- Determining whether establishments keep records at the level that is required for answering the items on the questionnaire;
- Examining whether survey questions be structured to conform to the business' record-keeping practices, including its fiscal year; and
- Determining if there particular times of the year when requested data are more readily available?

Of course, the benefit to be derived from such on-site cognitive interviews would need to weighed against the additional burden this might place on business organizations. Another instrument evaluation option that Willimack et al. (2002) identified was consultation with stakeholders. Given the recommendations in the Letter Report on the Commodity Flow Survey, this would be an important consideration for the 2007 CFS. Stakeholder panels, comprised of CFS data users, could be used to take account of user demands for both data consistency to support survey-to-survey trend analysis and the need to minimize data elements to mitigate respondent burden. Such data user panels might also address how to incorporate new content (or not). Users have called for data on such things as transportation costs and service characteristics for tracking service quality and modeling mode choice. Again, incorporation of new content requires the need to balance user needs with minimizing respondent burden.

It might also be important to establish "pretest" panels comprised of potential business respondents or other subject area experts. An important discussion for a respondent panel would be strategy development for meeting the potentially conflicting requirements for finer levels of geographic and commodity detail with the need to protect the confidentiality of individual level data. Another potential research activity with this type of panel would be to conduct ethnography of the CFS "interview" to examine how responses are affected by respondents' understandings of the larger survey context. This research activity would address such issues as: Why do establishments think they are being interviewed, and what do they think will be done with the answers. And, how do these considerations impact their CFS reports?

MODE OF COLLECTION: ISSUES AND CHALLENGES

Transportation survey researchers are currently using a plethora of survey modes for data collection. There are at least five data collection modes in common use today—face-to-face (or intercept), telephone, mail, Internet, and Global Positioning System (GPS). But this list of survey modes only begins to scratch the surface with the options available for capturing data. There are not only various survey modes to consider, but also sub-modes to consider as well. Telephone and face-to-face surveys can be conducted either by paper-and-pencil or computer. Even computer-based applications have their options: desktop computers, laptops, handheld-devices, or Pen Tablet PCs. While setting standards for the selection of a particular survey mode might be welcomed, the reality is that the selection of a survey mode is dependent on the survey situation. And, the survey situation can best be defined as encompassing the target population, eligible respondents, data needs, and budget available.

Yet, for the 2002 CFS and for prior enumerations in 1993 and 1997, all responses have been collected via paper questionnaires. As of October 21, 2003, however, the Government Paperwork Elimination Act required federal agencies to allow entities the option to submit information or transact with the agency electronically, when practicable. In addition, previous CFS survey respondents have inquired about an electronic reporting option. Preliminary attempts were made to design an electronic questionnaire for the 1997 CFS, but the product proved too burdensome and did not advance past the test stage. Because of the timing constraints involved in initiating the 2002 CFS, the schedule did not allow for further exploration of this option.

MODE OF COLLECTION: CONSIDERATIONS

New data collection methods could bring important benefits for the 2007 CFS by offering the potential to reduce respondent burden, to increase respondent's willingness to participate, thereby increasing response rates (effectively increasing the sample size), and with built-in edit checks increasing data quality. While technological development and new modes of data collection provide opportunities for more effective and efficient data collection, it is important to understand the cost-quality-usefulness trade-offs that new methods bring to a survey situation. For example, electronic reporting options developed for the 2002 Economic Census allowed businesses to extract data directly from their own spreadsheets and import it into survey software (*Special Report 277*, 2003). Some federal agencies have begun to implement electronic options in their establishment surveys and have reported significant benefits (Fox et al., 2004; Hak et al., 2003). They point out such advantages as

- Reduced respondent burden and potential increase in response rates;
- Improved data accuracy/completeness;
- Quicker dissemination and receipt of survey forms and improved control of survey responses;
- Ease of making changes to questionnaire versions and added flexibility in tailoring instruments; and
- Decreased data keying and capture costs (since the respondent enters the data).

The advantages of electronic reporting, however, depend in part on the extent to which establishments are capable of and willing to use an electronic option. Thus, the cost-benefits of such applications need to be researched. The benefits do not always justify the costs and resources required to maintain a dual system. Other noted disadvantages have included

- Increased concerns of respondents about privacy and confidentiality of submitted information that require additional security protections for survey entry and transmittal of information;
- Limiting respondent ability to answer questionnaire sections "out of order" or to get an overall sense of survey contents, while facilitating the correct navigation through the questionnaire; and
- Providing the opportunity for "mode effects" resulting from the use of both paper and electronic response options.

Thus, the move to new methods or new technologies might be justified on the basis of higher data quality and/or faster processing from collection to final data release. However, these benefits should be weighed against generally greater costs and higher risks, especially in initial applications in a particular survey context.

In the 1990s, many household surveys gravitated toward survey designs that are best described as "mixed-mode" in an effort to combat nonresponse. Dillman (2002) pointed out that the "future of surveying is far more likely to evolve toward the use of mixed-modes for different survey situations." Simply defined, a mixed-mode survey is one that uses two or more methods to collect data for a single data set. The Internet is one of the newest modes be offered to respondents as a way of allowing them to select the mode they prefer. We understand that the Census Bureau and BTS are engaged in research to determine the efficacy of providing the option of a web-based questionnaire for the 2007 CFS, in addition to the mail option. This strategy would enable the CFS to reach different establishments in different ways. For example, establishments equipped to provide data electronically through electronic data interchange systems could provide CFS data by Internet data entry, and at the same time, the mixed-mode design would not preclude mail for others. Fox et al. (2004) highlighted the benefits Internet-based questionnaires offered as a part of a mixed-mode design. These benefits were enhancing response rates, improving data quality, and improving timeliness of reporting. At the same time, they caution that potential for cost savings also exists, although in some cases offering an additional data collection mode might actually increase costs.

The use of new technologies, such as GPS and radio frequency identification, for passive, real-time data capture of shipment movements has been discussed for several years. While the technologies exist, the stumbling blocks to implementation have been cost, distribution of units, and retrieval of units. The actual data on movements can be relayed from the units in real-time to a master storage unit concurrent with the shipping process. But calculations of the volume of units that would need to be purchased, distributed and retrieved for future use amounted to costs greater than the perceived value of the technology application. It is theorized that perhaps the purchase and implementation of the units could be handled in a decentralized manner at the level of the shipper (rather than in a centralized manner by the survey organization). However, this scenario is further complicated by the perceived reluctance of shippers to allow access to such proprietary, real-time inventory control information. Thus, it is unlikely that such technological innovations would be considered for CFS 2007.

NONRESPONSE: ISSUES AND CHALLENGES

The CFS is conducted every 5 years as part of the economic census. As such, reporting is mandatory and penalties are established for an establishment's failure to comply. In spite of the mandatory reporting authority, the CFS still suffers from nonresponse. Four levels of nonresponse occur in the CFS—establishment, quarter (reporting week), questionnaire item, and shipment. In other words, not all sampled establishments complete the questionnaire. Furthermore, establishments responding in the first quarter do not always respond to at least one of the remaining quarters. Of responding establishments, item nonresponse also occurs. That is, respondents do not always provide information for every sampled shipment, or omit responses about requisite shipment characteristics. More detail about nonresponse in the 2002 CFS is presented below.

UNIT NONRESPONSE

Each establishment selected into the CFS sample was mailed a questionnaire for each of its four reporting weeks. Of the approximate 50,000 establishments sampled in 2002, less than 32,000 responded (i.e., provided usable shipment data) for at least one quarter. Approximately 17% of the sampled establishments were determined to be ineligible for the survey (e.g., out-of-scope industries, non-shipping establishments, etc.). However, almost one-quarter of the sampled establishments that were assumed to be eligible did not respond or provided data that was too problematic to be included in the final results. As expected, differential response patterns were detected. For example, specific industry sectors (e.g., manufacturing and mining) and larger establishments exhibited higher response rates. Also, response rates varied significantly based on geographic boundaries. These differences bring into question how data collection methods can be modified to focus on more problematic response areas and increase overall establishment response rates for the CFS. Measures were taken to compensate for establishment level non-response, for example, an industry-level adjustment weight was applied. This weight utilized information from other surveys and censuses conducted by the Census Bureau to account for establishments from which no usable response was received. It also adjusted for changes in the population of establishments between the time the first-stage sampling frame was constructed and the year in which the data were collected.

ITEM NONRESPONSE

For a given establishment, the respondent was requested to provide the following information about each of the establishment's reported shipments: shipment identification number, the date on which the shipment was made, value, weight, commodity, mode(s) of transportation, domestic destination or port of exit, an indication of whether the shipment was an export, and the United Nations or North America (UN/NA) number for hazardous material shipments. For an export shipment, the respondent was also instructed to provide the mode of export and the foreign destination city and country. A respondent may not have been able to provide value, weight, or a destination for one of the sampled shipments. If this data item could not be imputed, then this shipment did not contribute to tabulations and was deemed unusable. A shipment is deemed usable only if it has valid entries for value, weight, and origin and destination zip codes. Similar to unit nonresponse, to account for these unusable shipments, a shipment nonresponse weight was also applied.

NONRESPONSE: CONSIDERATIONS

Issues and considerations discussed under Instrument Design and Mode of Collection identify potential strategies to combat or at least mitigate nonresponse. While such strategies are important, equally important is the need to conduct research in two areas: (*a*) research into the level of burden that firms are willing to tolerate and ways to mitigate burden and (*b*) evaluation of CFS nonresponders. Both of these activities will identify the type and level of bias that is present in survey results, help focus future methods research efforts on specific problems, and serve as a useful guide for reducing survey nonresponse in the future.

A FUTURE CFS DESIGN CONSIDERATION: CONTINUOUS DATA COLLECTION

The Letter Report on the Commodity Flow Survey noted a prominent limitation with the CFS, with its 5-year cycle, was its inability to capture rapid changes in economic cycles. The lack of coverage of the intervening four years means that time trends in freight activity, such as the effects of emerging from a period of recession or severe drought, cannot be studied satisfactorily using CFS data alone. In addition, since freight moving industries were deregulated and computerization allowed tighter inventory control, there has been dramatic changes in how freight is shipped (Loudon, 2000). There is more less-than-truckload shipping and more use of parcel delivery services, and there is more backhaul shipping—to name just a few. A CFS survey design needs to have sufficient currency to capture such rapid changes in shipment methods.

Thus, a longer term issue than the 2007 CFS, but one that warrants serious consideration in the near future, is whether to improve on the timeliness of the CFS by transitioning to an annual or continuous survey, as is being done with the American Community Survey (ACS). The ACS is a nationwide survey that collects socioeconomic and housing information and is the planned replacement for the long form in the 2010 Census. In the past, the long form (LF) data were collected, processed, and tabulated once each decade. Forms were sent to a sample of about one of six households so that reliable estimates were published for small areas such as tracts and block groups. The ACS, when fully implemented, will collect LF data throughout the decade. Although statistics from any individual year of the ACS are not assumed to produce reliable estimates for small area estimates, multi-year (3- and 5-year averages) are assumed to produce reliable, useful, and timely statistics to replace the LF.

Any type of ACS-like change could not take place until after the 2007 CFS, but a change of this magnitude would require several years of planning. Obviously, there are many operational constraints associated with switching the periodicity or enumeration schedule of the CFS (e.g., no longer parallel tracking with economic census). Among other questions that must be addressed are

- How often are CFS data for specific characteristics needed and at what level of geography?
- How does this impact the ability to monitor freight shipment trends?
- How would the sample design and size need to change for a continuous enumeration?

CONCLUSIONS

The CFS methodology can and should be improved. Evaluation of the methodological changes implemented prior to the 1997 CFS indicated that these were successful in providing a more efficient sample design, lessening respondent burden, and improving timeliness of data products. In the past 10 years, survey methodology has benefited from many research-based improvements in sample design, instrument design, data collection, and nonresponse mitigation. For various reasons, the CFS has not taken advantage of such methodological improvements. It is time to do so. This paper has discussed current issues, challenges, and survey considerations that are relevant to the CFS. It underscores the need for a continuous research and evaluation plan that would enable BTS and the Census Bureau to stay abreast of methodological enhancements, as well as real changes in freight transportation, that impact CFS data quality and usefulness. Such a research plan could provide an increased understanding of the response process and respondent burden to increase response rates, alleviate privacy and confidentiality concerns, improve data quality, etc., for establishment surveys. This is an area in which not enough research is currently taking place. The issues discussed in this paper also draw attention to the value that ongoing dialogue with data users and stakeholders provides. Not only do these persons bring substantive expertise on tough decisions such as whether or not to change the way shipments are sampled, but also their input helps to ensure methods employed are providing data of most importance and utility.

REFERENCES

Black, J. Can Respondents Construct a Frame and Draw a Sample? Experiences from the 1993 Commodity Flow Survey. Proc., Survey Research Methods Section, American Statistical Association, Alexandria, Va., 1997, pp. 216–221.

Black, J., W. Davie, and J. Jonas. An Evaluation of Sample Design Changes for the 1997 Commodity Flow Survey. Presented at the 2000 Conference of the American Statistical Association.

Davie, W. Profile of the Commodity Flow Survey. Presented at International Trade Traffic Study Workshop, Bureau of Transportation Statistics, November 21, 2003.

Dillman, D. Mail and Internet Surveys: The Tailored Design Method. John Wiley & Sons, New York, 2000.

Dillman, D. Navigating the Rapids of Change: Some Observations on Survey Methodology in the Early Twenty-First Century. *Public Opinion Quarterly*, Vol. 66, No, 3, 2002, pp. 473–494.

Fox, J. E., W. Mockovak, S. K. Fisher, and C. Rho. Usability Issues Associated With Converting Establishment Surveys to Web-Based Data Collection. Statistical Policy Working Paper #38, 2004.

Hak, T., A. E. Anderson, and D. K. Willimack. Determinants of Web Reporting: A Qualitative Study of Mode Selection. Proc., Federal Committee on Statistical Methodology Research, Office of Management and Budget, Washington, D.C., 2003.

Loudon, W. Surveys of Freight Shippers and Carriers: Lessons Learned. Presented at ITE District 6 Meeting, San Diego, Calif., 2000.

Plewes, T., K. Copeland, C. Corby, R. Fecso, S. Freedman, M. Gonzalez, C. Konschnik, S. Slowinski, A. Tupek, P. Waite, and G. Werking. Statistical Working Paper 15: Measurement of Quality in Establishment Surveys. Federal Committee on Statistical Methodology, Office of Management and Management, 1988.

Prewitt, K. What If We Give A Census and No One Comes? *Science*, Vol. 304, 2004, pp. 1452–1453.

Special Report 277: Measuring Personal Travel and Goods Movement. Transportation Research Board of the National Academies, Washington, D.C., 2003.

Willimack, D. K., E. Nichols, and S. Sudman. Understanding Unit and Item Nonresponse in Business Surveys. In *Survey Nonresponse* (Groves et al., eds.), Wiley, New York, 2002.

NOTES

1. This committee was convened by the Transportation Research Board and the Committee on National Statistics in response to a request from BTS.
2. A primary reason for the reduction in sample size for the 1997 survey was to enable staff to identify and follow-up with problem reporters.
3. Although the 1997 CFS, with a sample size of 100,000 establishments, collected potentially useful local-level data, these microdata cannot be made available to the public because their release could compromise the confidentiality of data providers.

Workshop Summaries

Scope, Comparability, Shipment Characteristics, and Special Measurement Issues

CATHERINE LAWSON
MICHAEL BRONZINI
ARNIM MEYBURG
PAUL BINGHAM

Compiled by
KATHLEEN HANCOCK

INTRODUCTION

The first four workshops of this conference focused on the scope, data comparability across years and with other data sources, shipment characteristics, and special measurement issues of the Commodity Flow Survey (CFS).

Scope workshop participants initiated their discussion with a delineation of five areas of concern:

- Limitations of the sampling frame as drawn from the Census Bureau's Business Register result in out-of-scope establishments being selected.
- Changing from Standard Industrial Classification (SIC) linked classification of commodities to North American Industry Classification System (NAICS) has impacted the sampling frame.
- Growth of auxiliaries has and will impact the CFS.
- Growth of third-party logistics (3PL) firms has and will impact the CFS.
- Different understandings of intermodalism are impacting how CFS surveys are completed and interpreted.

Comparability workshop participants were presented with the following themes:

- Gaps in CFS coverage,
- Lack of consistent time series,
- Supplementary data sources,
- Interpolation and imputation methods, and
- Improved data collection methods.

Shipment characteristics workshop participants considered what the ideal database would consist of and whether any of this information could be acquired from linkages to other data sources.

Special Measurement workshop participants were asked to consider six specific issues:

- Intermodal shipments and shipments by third parties,

- Imports and in-transits,
- Warehousing,
- Third parties,
- Containerization, and
- Pipeline shipments.

Because of the overlap in responses from participants to the issues raised in these workshops, this summary provides an overview of the key points that were raised across all four sessions.

DATA GAPS

Several data gaps were identified in the resource papers and by the participants. These gaps result from commodities and/or services being out of scope; ambiguity, uncertainty, or inability of the current survey to capture the desired information; and the inability of Census to release information because of statistical rigor or private sector confidentiality issues.

Out of Scope

The sampling frame for the CFS is drawn from the Census Bureau's Business Register that has been classified by industry type. Any business not included in this list is automatically out of scope including fisheries and logging companies. Printed materials moved from manufacturing to information businesses were also rendered out of scope with the change from SIC to NAICS.

The CFS sampling frame is based only on firms in the United States. As a result, no data are collected from firms outside of the United States, resulting in no import information. It is also not in scope to survey goods on incoming ships. Commodities may be recognized once they are part of the flow of goods from distribution centers in the United States but not before.

Survey Limitations

Participants indicated that a single survey would be too cumbersome and costly to obtain the "ideal" database for commodity flow. Even establishing a set of common factors would be difficult since different users have different needs. Some discussion focused on a need to understand and evaluate information obtained from the current survey before deciding on changes. Participants indicated that developing a matrix of data elements by CFS users would be useful.

Because the CFS is a shipper survey, data about other integral activities are not captured. These include carriers, intermodal facilities, 3PL firms, and auxiliaries (businesses that provide management or support services). The growth of these operations will compound the impact of related data gaps in the CFS.

Pipelines make up a unique mode of transportation as defined by operating characteristics. The concept of shipments is not directly applicable. However, other sources of data may be available to capture this information.

The practice of pre-canvassing could help to clarify eligible firms and terminology. Some confusion may be occurring because respondents are not using the term "intermodal" and "containerization." Another area of potential confusion is whether drayage is considered a part of a shipment.

One method that was suggested to address the commodity code issue was to get some easily obtainable industry information for each shipment. Presently, respondents are asked only for the commodity code of the shipment. Commodity flow related studies that are conducted by the Bureau of Economic Analysis and other agencies, such as production of economic input-output tables, are based upon the classification of the producing or shipping industry, not the commodity type actually produced. It would be relatively easy to modify the CFS to ask for the shipping and receiving industries related to each shipment. This information would be very useful for producing cross-walk tables.

Another consideration was to improve the coverage and accuracy of the truck-rail intermodal data reported in the CFS. This would require effort to investigate the true origin and destination (O-D) of intermodal movements that involve drayage. One approach would be to ask the carriers or 3PLs involved in the movements to provide the data, rather than the shippers. Another approach would be to investigate distribution center locations and movements from a distribution center to a final destination, e.g., movements of automobiles from the factory via rail to a distribution center, then by truck to the automobile dealerships. Many thought that other areas of "low hanging fruit" to improve CFS coverage and accuracy through relatively low cost additional surveys or analysis were available.

Confidentiality Issues

Participants were interested in understanding the relationship of reduced detail, either for commodity code or for geographic location, on resulting commodity flow information. Is the level of detail (the three- or four-digit level) necessary, or could the data be collected at the two-digit level since the data are most often used at the two-digit level? For users who require more than two-digit level data, can the data be masked to protect confidentiality? One consideration was the possibility that more firms would respond if the burden of reporting four-digit commodities were removed. Another comment indicated that more shipment data could be released on a geographic basis if commodity type was dropped. Many participants had questions about the tradeoffs between level of detail and resulting usable information, raising more questions than ideas.

DATA COMPATIBILITY

Two types of compatibility were addressed, internally in the CFS across time and externally across other data sources.

CFS Over Time

The CFS was conducted in 1993, 1997, and 2002. Changes in the survey included the commodity and industry classification systems, the geography of the regions for which estimates are made, and changes in some of the specific questions asked about the sampled shipments. Participants observed that there were sound reasons to make these changes, which resulted in compatibility issues across time. Some of the difficulties imposed by these changes could be avoided with a good, well-documented cross-walk between the various classification systems used, but this has not been provided or widely distributed for all data sources. Many users indicated a desire for such cross-walks.

Compatibility Across Data Sources

The CFS is a survey of shipments, while other surveys that are used to add detail not in the CFS, such as the Rail Waybill Survey and the U.S. Army Corps of Engineers Waterborne Commerce Data, are carrier surveys. This basic difference is the major source of the incompatibility across surveys. The other significant sources of incompatibility include differing commodity coding schemes, commodities that are out-of-scope for the CFS, lack of consistency in dealing with imports and exports, difficulties in matching up "intermodal" shipments, and differing regional geographies.

Several participants observed that the basic source of the compatibility problem is that the extant surveys are produced by different agencies to meet their own needs, with little thought as to how these surveys can be fit together to produce a comprehensive picture of freight activity. An overriding need exists to develop complete data system architecture and integrated data collection program that breaks down these agency silos. Specifically, several things could be coordinated by federal agencies, like using the same definitions for commodity codes and vehicle types. Common definitions would allow improved linkages between data sets.

SUPPLEMENTING THE CFS WITH OTHER DATA SETS

One of the most discussed issues in the four workshops was the need to combine the CFS with other data sources to obtain necessary information. Comments from participants indicate that the CFS, by itself, addresses a narrow set of needs but must be linked with other data to be useful to a broader audience.

Issues related to linkages from the CFS to other data sets include

- What elements allow the merging of data sets?
- What methodologies are necessary to ensure that shipments are not double counted and over-estimated?
- Could O-D data be used to link data with other databases? This is not to say that commodity types are not important, but would O-D provide a better link?
- What data could be used to capture pass-through traffic?
- Could Census and Bureau of Transportation Statistics partner with other agencies to obtain complimentary information, i.e., Department of Homeland Security for Hazmat?

Much interest was expressed in how commodity flow data could be extracted from the data files already created and maintained by shippers for their own business systems. One way to make this easier (than the current version of CFS) for the respondents is to make use of data dumps of existing company data, rather than asking them to reformat the data or fill in survey records. Other countries such as Sweden have used data-dumping successfully to gain access to shipper data. Of course, many companies are not willing to share shipping data. They have legitimate questions such as how would the data be scrubbed and what is in it for my company?

Statistics Canada addressed such concerns by having a third-party collect the data, then scrub it and provide it back to the contributing industrial partners. Atlanta has had success in getting industry to come to the table by explaining that the projects they need will be programmed only if their needs are made known. Making extensive use of company databases would likely require public–private partnerships.

Considerable interest was also expressed in developing a new motor carrier roadside survey, since CFS is currently the only large source of truck commodity flow data, but lacks much of what could be obtained from a roadside survey. It was also recognized that this would be a costly undertaking.

ADDITIONAL SUGGESTIONS

Some additional suggestions made by the participants included the following:

- Investigate the known disparity between waterborne commerce data and the CFS, to reveal opportunities to either capture more of these data, or better fuse the two data sets.
- Establish a data center at a university transportation center to promote cooperation between metropolity planning organizations (MPOs), industry, departments of transportation (DOT), and others.
- Collect commodity flow data via a standardized local survey, conducted by state DOTs or MPOs, then integrate the data to produce the national data set. This would put CFS on the same model as the Highway Performance Monitoring System (HPMS). Given some of the problems with HPMS data, some stated this might be a bad idea.
- Explore whether add-ons to CFS be used in the future.
- Certain public health issues have arisen that resulted in animal tagging, which might provide an opportunity to add some new agricultural data to the CFS.
- Archived intelligent transportation system data could be used to augment the CFS, especially for truck flow data. This might prove to be an important source of verification data for the loaded network product discussed below.
- Incorporate advanced technologies into the data collection process, such as radio frequency identification tags for tracking shipments and Internet-enabled interfaces for completing surveys.
- Consider European surveys and data collection activities.
- A new CFS data product that would aid analysts would be a loaded network, based on micro data. It appeared to many that such a product could be generated without compromising statistical accuracy and privacy concerns. This would save analysts the trouble, expense, and inherent inaccuracies of producing such a product by disaggregating the published national or region-to-region data.
- There was interest among many participants in producing a guide to the CFS for planners since few MPOs seem to be using CFS directly. As one participant stated: “CFS at the regional level only makes a nice paragraph in the long range plan.” Several pointed out that the CFS has been used indirectly through value-added data products, and that at the MPO level it is necessary to perform some data integration for many applications. The proposed guide would explain CFS strengths and weaknesses, and how it can be used with other public data in a state or metropolitan context.

- It was observed that no one who has completed a CFS survey—shippers, producers, etc.—was at this workshop. At a minimum we need outreach to the industry associations to bring them to the table in planning the CFS. We need to go to the industry meetings, rather than expect them to attend our workshops.

Improving CFS Data Products

THOMAS E. ZABELSKY
U.S. Census Bureau

INTRODUCTION

The primary purpose of this workshop was to discuss the development of data products and dissemination tools that meet the needs of the Commodity Flow Survey's (CFS) principal data users, including approaches for improving the utility and accessibility of products, customer services, and methods of dissemination for the 2007 survey.

The workshop participants focused their discussions on the following broad areas of inquiry:

- What are the most important changes that would improve the current CFS data products?
- How would you prioritize geographic detail, commodity detail, modal detail, and industry detail in future CFS data products? What is the minimum acceptable level of detail in each of these when showing data for commodity flows?
- What sorts of value-added or interpretive analyses, if any, should accompany the release of the CFS data?
- How important is the availability of micro-data, and what sorts of disclosure avoidance techniques could be applied to protect the confidentiality of individual responses?

HIGHLIGHTS

The participants identified the following important data, products, and services for the 2007 CFS:

- Detailed origin–destination (O-D) tables emphasizing more geographic and industry data, reverse geographic flows, and additional aggregations;
- Fully loaded network product that would identify freight corridors and trade and market areas, and provide a better understanding of the economic relationships, connections, and other causal factors influencing freight flows;
- Trip generation and trip distribution variables to facilitate modeling of missing data and freight flows between surveys;
- Increased access to the CFS micro data to provide users with additional detail and greater flexibility in their use of CFS data; and
- Value-added products like the FHWA's Freight Analysis Framework that would provide interpretive analyses, and projections and forecasts for policy makers and others.

SUMMARY OF WORKSHOP DISCUSSIONS

The following is a summary of the views expressed.

Geographic Detail

Detailed O-D data are important to improve freight analysis and forecasting, and many transportation investment decisions made at the state and local level. Leveraging the detailed O-D information that is uniquely part of the CFS data set was the area that many felt provided the best opportunity for improving the CFS data products.

A repeated concern is that insufficient geographic data are produced in the CFS. In part, this is because of limited resources that constrain sample sizes and force trade offs between geography, commodity, and industry detail; and the need to protect the confidentiality of individual responses that can lead to the suppression of some data cells. One could assume it would be possible to increase the level of geographic detail by aggregating commodities. Presented with this choice, most of those voicing an opinion gave priority to producing more geographic data at the expense of commodity detail.

Commodity Detail

Discussion focused on the minimum acceptable level of commodity aggregation. The immediate reaction of some was that two-digit Standard Classification of Transported Goods (SCTG) detail would be sufficient. Others, however, expressed the concern that two-digit commodity aggregations may not be adequate for some commodity groupings. Because of the different shipping characteristics of some of these products (either mode choice, equipment used, or the like) some three- or four-digit commodities would be needed.

Nonetheless, trade offs of this sort provide the opportunity to publish additional geographic information.

Industry Detail

Many felt that data aggregated by industry would be extremely useful. It would enhance CFS users' ability to identify industry clusters and market areas. Input–output models tend to be industry based and considerable economic data are available by industry. Together, these data could further the understanding of the underlying forces that generate freight movements.

It was suggested that two-digit North American Industrial Classification System (NAICS) -based tables be produced showing detailed O-D by commodity by mode, where SCTG commodities codes could be rolled up to allow the most detailed geographic data possible.

Geographic Aggregation

In 2002, the published CFS geographic levels were defined by a combination of the 50 states, the District of Columbia, and the largest metropolitan areas (MAs) based on their population size in Census 2000. MAs not among the largest were collapsed with other MAs and non-MA areas within a particular state to form a "remainder-of-state" area.

It was pointed out that in some instances the remainder-of-state areas were not always useful aggregations. Connecticut was sited as an example where the commodities, methods, and patterns of shipping differ widely across the areas covered by the remainder-of-state. It was suggested by one participant that the Bureau of Transportation Stastitics (BTS) and Census contact individual states to get their views on alternative aggregations that better reflect their state's overall economic activity.

It was also suggested that Bureau of Economic Analysis (BEA) Areas could be used as an alternative in publishing sub-state geography. This is a sampling as well as publication issue. In 1993, National Transportation Analysis Regions (NTARs) were the primary sub-state geography in the CFS. NTARS are consolidated BEA Analysis Areas. The sample was stratified by kind of business (standard industrial classification) and NTAR, and NTAR data were published. While the same stratification criteria were used for 1997, only limited NTAR data were published. There was more interest in data published by political geography and that sub-state data be additive to the state level. Consequently, the Census Bureau and BTS published the 1997 CFS data by MA and, to ensure these data were additive to the state, included only those portions of the MA in the state. For 2002, the Census Bureau stratified the sample by kind of business (NAICS) and MA and remainder of state, and dropped the publication of all NTAR data.

The CFS provides data on state-to-state flows. It was suggested that aggregating adjacent states where dominant trading relationships exist could be used to define new geographic areas within the CFS to reflect regional trading blocks.

It was also suggested that O-D tables be available not only by state (or other geographic area) of origin but also by destination. These "reverse tables" are now produced by state, but because of reliability issues, only limited data are provided. It was felt that a larger sample size and further aggregation of commodity and other detail might allow the publication of more data and possibly of tables produced at a sub-state level.

The limited size of the CFS sample contributes to the lack of flow data below the state level. This limits the users ability to assess the relationship between freight movements and business patterns and flows along major corridors. Some noted the importance of also measuring movements within some very large multijurisdictional MAs. The CFS was never intended to provide detailed coverage of local freight movements. Some thought limiting the measurement of sub-state flows to only the very largest MAs, and rolling commodity and modal detail up no more than two-digit SCTG for truck and all modes of transportation combined, might allow the publication of some of these data.

Additional Data Content

Many suggested that a descriptive set of establishment characteristic variables be collected, such as payroll, employment, or other measure of size or activity that would allow the estimation of trip generation functions that would provide modelers and others the basis for estimating missing data or producing estimates of freight flows between surveys.

It was also suggested that Census and BTS explore the collection of data on shipping costs that would allow the estimation of O-D costs per mile by various modes of transportation.

Network Flows

There was discussion on the need to develop a national data architecture that would help guide data collection at all levels of government and private industry. Freight flows are regional, national, and global in nature. Agreement on identification of national networks is important and a beneficial product would be the CFS data loaded onto that network. The data would illustrate the importance of market areas, trade corridors and regional connectivity. Showing the flow of dollars in addition to traffic along the network is important to multiple jurisdictions in making network improvements.

The need for improved data on intermodal transfers was discussed. It was suggested that marine ports, airports, and rail terminals be surveyed, reasoning that they would be better able to provide information (mode, destination) on intermodal freight movements than the shipper. An obvious concern would be ensuring that this portion of the flow not be duplicated from the shipper's report.

Data Products

Attendees indicated a preference that standard database formats be used and, to the extent possible, historic CFS data be reformatted to be compatible. While many would like to see tables that compare CFS data over time, there was recognition that changes in industry and product classification, geography, and survey design make this much more difficult. The feeling was that it was more important to introduce survey improvements that would allow a more accurate depiction of current freight flows than to retain a consistency with the past to maintain a comparable time series.

Many felt products need to be flexible and users given the ability to produce their own tables.

There was brief discussion on a suggestion to link the CFS results to other data products like the Transborder Surface Freight Data, U.S. Army Corp of Engineers Waterborne Commerce Data, Rail Waybill data, and Vehicle Inventory and Use Survey. Some felt it more appropriate that users attempt to establish those links.

The importance of freight data researchers and modeling in integrating these various data sources was briefly discussed. For example, the importance of vehicle flows associated with freight movements was mentioned as an important area that cannot be measured directly in the CFS. As much as 30% of truck miles on the nation's highways represent empty trips. Trip chain models using CFS data provide some basis for approximating these empty trips.

The impact of international trade on domestic commodity movements was discussed. There was particular interest in movements originating from Canada and Mexico. It was pointed out that the New York State Department of Transportation has acquired some commodity flow data from Ontario. Some wondered if it was possible to link the CFS with any of these foreign sources of data.

It was noted that following the 1993 survey, BTS issued special interpretive data products that showed the volume of freight originating, ending, and flowing within and through each state. These were well received and the suggestion was made that BTS produce similar reports in the future.

Access to Micro Data

A question was asked about whether confidentiality rules could be relaxed that would allow users more access to the micro data. Those rules are stringent and firm, and exceptions cannot be made. It was pointed out that the Census Bureau and BTS would be exploring alternatives to their current disclosure protection methodology, including adding noise to the data, to allow much more of the data being released without jeopardizing the confidentiality of individual responses. It was noted, however, that many more cell suppressions occurred in the 2002 CFS for reasons of data reliability than avoiding disclosure.

Exploring Survey Methods to Enhance CFS Data Quality and Usefulness

JOY SHARP
Bureau of Transportation Statistics

RESOURCE PAPER OVERVIEW

The workshop commenced with an overview highlighting current survey objectives and methods, introducing research areas and plans for the 2007 Commodity Flow Survey (CFS), and proposing additional considerations for 2007 and beyond. Key points follow.

CFS Objectives

The CFS provides information on the flow of goods in the United States. It captures characteristics of shipments originating from select types of business establishments in the 50 states and the District of Columbia. Shipment characteristics include

- Origin and destination (O-D),
- Distances traveled,
- Commodities shipped,
- Modes of transport, and
- Volume measured by weight and value.

2002 CFS Methods

Approximately 50,000 establishments were selected from the Census Business Register for the 2002 CFS sample.

The design of the sample consists of three stages:

- Stage 1: Establishments (stratified by geography, industry, size);
- Stage 2: Reporting weeks (1 week in each quarter of the year); and
- Stage 3: Shipments (20 to 40 per week; respondent enumerated and drawn).

Data Collection Methods

Questionnaire is a self-administered, mail-out/mail-back survey consisting of a 7-page booklet. Other materials sent in mailout include

- Eight-page instruction booklet and
- Commodity coding manual.

These materials are mailed to sampled establishments each quarter.

1. Data processing
    - Completed forms are optically scanned and keyed from image;
    - Data are processed through an automated editing system; and
    - Callbacks are made for verification and clarification.
2. Sample weighting: data are weighted to represent an annual sample of shipments; different weights applied include
    - Shipment weight,
    - Shipment nonresponse weight,
    - Quarter weight,
    - Quarter nonresponse weight,
    - Establishment-level adjustment weight, and
    - Industry-level adjustment weight.

Current Methods Refinement Goals

In general, planning for the 2007 CFS includes the following goals:

- Reduce survey costs through more efficient sampling and data collection.
- Achieve larger sample size at finer levels of geography to improve utility for a wider range of users.
- Mitigate respondent burden to improve response rates.

The following are some considerations for achieving these goals. Many are currently being considered or are included in ongoing research and improvement efforts for the 2007 CFS.

1. Refine sample size, design and coverage. Efforts are needed to
    - Improve coverage of shipments through reevaluation of sample frame and sample selection criteria (industry, certainty, size);
    - Increase shipment detail at finer levels of geographic detail;
    - Reduce sample loss;
    - Consider trade-offs between first- and second-stage sampling; and
    - Consider trade-offs between first-stage sample size and implementing response enhancement strategies.
2. Survey design: Efforts to improve survey design focus on
    - Pretesting questionnaire and instructions;
    - Overcoming problem reports and missing data;
    - Determining "respondent burden" points; and
    - Overcoming nonresponse issues.
3. Data collection: One current consideration for the 2007 CFS is the introduction of an electronic Internet-based reporting option. This option could allow for
    - Efficiencies in data capture/processing;
    - Quicker dissemination/receipt of survey forms;
    - Decreased data keying and capture costs; and

- Improved control over survey responses.

4. Pilot test: The importance of a pilot (field) test cannot be underestimated. Changes made to the questionnaire and data collection methodology require evaluation prior to full scale implementation to
    - Ascertain impact of questionnaire changes on response (unit, item);
    - Determine cost-quality-usefulness trade-offs of electronic reporting; and
    - Identify "mode-effects" resulting from use of both paper and electronic reporting options.
5. Concluding remarks and future considerations (beyond 2007): CFS methods can be improved for the 2007 CFS. Priority areas include
    - Better coverage of shipments;
    - More detail on shipment characteristics;
    - Lessening respondent burden; and
    - Improving timeliness.
6. Changes beyond the 2007 CFS include
    - Passive data capture options, such as Global Positioning System (GPS) and radio frequency identification (RFID), and
    - Continuous data collection to improve data currency and timeliness.

WORKSHOP DISCUSSION

Workshop participants were seated at five tables and each table was presented a methodological theme with related questions which had been posted prior to the conference as follows:

- Table 1: Sampling Methods. How can sample design be improved? How can shipment sampling instructions provided to the respondent be improved?
- Table 2: Instrument Design. Is the current questionnaire able to respond to industry changes? Are there existing issues with the questionnaire that may result in poor response or burden?
- Table 3: Mode of Collection. Are multiple data collection instruments needed? Should other modes of data collection be explored?
- Table 4: Nonresponse. How can we improve the response process to ensure questionnaire reaches the appropriate respondent(s)? What additional measures should be implemented to improve overall establishment response?
- Table 5: Future Design Considerations. Should the CFS be conducted more frequently? Should establishments continue to report quarterly?

Key observations and ideas from each topic are provided below.

1. Sampling methods:
    - The three-stage sample design is very complex. Changes could be made at each stage to improve overall design.
    - At the first stage, more effort could be made to identify eligible establishments in universe. A precanvass of establishments, similar to the one currently being considered

by project staff, would be valuable in improving frame knowledge and making first-stage sampling more efficient.

- Using previous survey enumerations, an evaluation of each of the three stages could determine which stage contributes most to sampling efficiencies/deficiencies. This evaluation would help direct resources to the area for the largest direct improvements in the sampling process.
- Use of a web-based (or electronic form) might be very beneficial in the third stage, since it could provide assistance in the shipment sampling process and further improvement of sampling instructions.

2. Instrument design:
   - Participants noted specific items on the CFS questionnaire that could be problematic. These included destination city and mode since they may not be known for certain shipments, especially if the shipment is sent via an intermodal mode. Shippers may know the initial mode (e.g., truck) but are often unaware of changes (e.g., transfer to rail). In addition, shipments sometimes are re-routed to other locations without the shipper's knowledge.
   - Education of respondents could be a key to improving response quality. This could require sampling fewer shippers and would allow for better follow-up contacts with those shippers.
   - The questionnaire formatting might be improved to assist the respondent completing the forms. The current design of the form consists of a spreadsheet over two pages. This presentation is somewhat disjointed and may mislead the respondent in initially thinking that only the information on the left is requested.
   - Some duplication of information requested appears to exist. Streamline data reporting might ease the response process. Two specific areas are zip code and city/state, and commodity code and description. While both pairs of items are requested, this is, in part, for quality control purposes. Thus, the elimination of what appears to be duplicative items might ease response burden, but it could introduce additional data quality issues.
3. Mode of collection:
   - A web-based option could be preferred by many survey respondents and be more effective in mitigating third-stage sampling error. Regardless, maintaining a paper form option is important.
   - Several newer technological options (GPS, RFID) could also be useful to alleviate response burden, as well as to validate or adjust data.
   - The establishments sampled in the CFS differ greatly based on industry type, size, shipping volumes, etc.; yet, the CFS uses only one form for all establishments. Tailoring questionnaires to conform more to the special characteristics/practices of different establishment types could yield better data and improve the response process.
   - Better form design could result from a better understanding of record-keeping practices. Cognitive testing would be beneficial in learning more about the response process.
4. Nonresponse:
   - While approximately 51,000 establishments are included in the sample, only about 31,000 actually contribute to the results. Nearly 9,000 were determined to be ineligible, while the remainder consisted mostly of nonresponders.

- A precanvass operation would be beneficial in pre-identifying ineligible establishments.
- Cognitive testing could also be a useful tool in understanding why establishments do not respond. Better response avoidance techniques could be developed based on these results.
- Sampling with replacement or some other sampling processes that would factor in the target number of establishments and likely response rate could reduce the nonresponse rate.
- An awareness campaign prior to conducting the 2007 CFS would be beneficial in creating awareness and demonstrating the importance of the survey. Cover letters with endorsements of important associations, etc., might also lead to a stronger sense of credibility and importance of the CFS, thus improving response.

5. Future design considerations:
    - Most data users agreed that they would like data on a more frequent basis than the current 5-year cycle. A continuous survey is preferred, but financial and operational constraints are recognized.
    - Supplemental data and data collection efforts could be used to enhance the CFS. This might include efforts such as roadside data collections, fusing data from outside groups (e.g., chambers of commerce, trade associations), and employing existing electronic data options (e.g., free pass, weigh stations, manifest data).
    - Partnerships could be established with state data collection efforts to share costs and improve geographic detail.

Applications

National Transportation Applications

How TRANSEARCH Uses the CFS

Paul Ciannavei
Global Insight, Inc.

TRANSEARCH is a proprietary nationwide database of freight traffic flows that provides coverage of highway, rail, air, water, and intermodal activity. The database has many conceptual similarities with the Commodity Flow Survey (CFS), but has been produced on an annual basis for the last 20 years. TRANSEARCH does, however, utilize several discrete inputs from the CFS. This presentation explores how and why specific elements of the CFS are used:

- Commodity $/ton values,
- For-hire/private trucking mode share split,
- Selected origin-to-destination (O-D) truck flow volumes,
- Truck length-of-haul profiles, and
- Identification of commodities moving via air mode.

In addition, much of the CFS information is reviewed and used in the annual TRANSEARCH development quality control process.

A more general discussion will outline why additional information from the CFS is not integrated into TRANSEARCH, and a brief discussion of the background and history of the proprietary database is also included.

TRANSEARCH BACKGROUND AND HISTORY

TRANSEARCH was first created over 20 years ago and had very direct ties to the 1977 Census of Transportation (COT). One of the key features provided by TRANSEARCH at the time was that it estimated much of the data that was unable to be released in the COT. The major user group for this information at that time was the railroad industry, which continues to this day as one of the significant markets for the data.

The commercial applications of the information prompted an annual updating of TRANSEARCH which continues to this day. However, as the 1977 COT became dated, the commercial product was forced to follow a path of Darwinesque evolution to adequately reflect changes in the freight marketplace. This process has included finding, creating, and integrating additional sources of data and adopting and adapting econometric modeling techniques. The product has also undergone transformations to widen its appeal to a broader group of users. This has included development of greater detail in the area of geographic market flow coverage (down to the county-level), and increased industry/commodity scope to include and distinguish warehouse and distribution center activity, truck drayage of rail/highway and air intermodal activity, and inclusion of significant truck movements of non-manufactured (agriculture, minerals) commodities.

Today, the client-base for TRANSEARCH includes

- Railroads (applications include market planning, strategic analysis—particularly for mergers, operations research, model diversion assessment);
- Motor carriers (with similar applications as used by the railroads);
- Other freight carriers (including United Parcel Service and FedEx, plus others);
- Other commercial entities (investment banks, real estate development, energy); and
- Public-sector planners (FHWA, state departments of transportation and metropolitan planning organizations).

In addition to the CFS, TRANSEARCH is constructing from a wide range of public and private sources of information. These sources include, but are not limited to

- Department of Commerce (DOC) annual survey/census of manufacturers;
- DOC import and export information;
- Surface Transportation Board (STB) Annual Waybill Sample;
- U.S. Army Corps of Engineers (USACE) Waterborne Commerce and Port data;
- FAA enplanement and flow data;
- Energy Information Administration coal transportation data;
- U.S. Department of Agriculture Census of Agriculture;
- Bureau of Economic Analysis industrial input/output tables;
- Industrial trade organizations; and
- Proprietary data exchange with motor carriers and railroads.

APPLICATION OF COMMODITY FLOW SURVEY DATA

TRANSEARCH primarily utilizes data elements from the CFS that are not available through other information sources. These areas are discussed individually below.

Commodity $/Ton Values

Commodity $/ton values are calculated from the CFS. Because TRANSEARCH uses the Annual Survey of Manufacturers (ASM) to establish annual production levels by industry/commodity, conversion of the dollar value of output reported in the ASM to tons is necessary. From the CFS a wide range of these conversion factors can be calculated with multidimensional aspects, such as commodity values by mode of transport and geographic market area. This type of information is not readily available from any other source. The $/ton values maintained for TRANSEARCH production are updated annually for the intervening non-CFS years using inflation-based factors derived from sources such as the Producer Price Index.

For-Hire/Private Trucking Mode Share Split

The TRANSEARCH development process starts by establishing industrial production levels by market area, and then nets out traffic moving by rail, water, air, and pipeline modes, leaving a preliminary estimate of trucking volumes. The truck volumes are separated into three sub-modes: for-hire truckload; for-hire less than truckload; and private trucking. Commercial users of the data find these distinctions of the trucking industry useful, and there are also benefits to the public-sector users, as these distinct segments employ some very different operational patterns.

The CFS provides essential information on the relative shares of truck traffic, by geographic market areas, and commodity types which is not available elsewhere.

Selected Origin-to-Destination Truck Flow Volumes

Although TRANSEARCH uses a robust sampling of actual truck movements reported by the nation's leading motor carriers, trucking origin-to-destination flow volumes from the CFS are fed into the development process. Even though the sample of the trucking industry obtained through the proprietary Data Exchange Program typically covers about 75 million individual shipments, this sampling is biased toward the larger longer-haul national and regional carriers. Consequently, CFS information on the very significant level of shorter-haul and local trucking activity is a very valuable input. This type of information on a nationwide basis can not be found elsewhere.

Truck Length-of-Haul Profiles

Supply chain and distribution practices vary from industry to industry. The information on average length of haul for truck movements by commodity is very useful in developing commodity-specific flows in TRANSEARCH. Typically, lower value items move shorter distances, as transportation can represent a significant portion of total costs. In addition, product characteristics can dictate shipping patterns, such as for wet cement. Also, shipper's selection of private versus for-hire transportation options can be discerned from the CFS data. Again, this type of information on national patterns is not available from other sources.

Identification of Commodities Moving via Air Mode

Third-party service providers (those which do not operate aircraft) play a significant role in the air freight industry. The ensuing commercial relationships inhibit the industry's willingness to reveal potentially sensitive information about clients and their products. Consequently, the CFS is, again, the only source where some information is available to actually identify the types of goods moving via air.

Quality Control

A key input in the annual quality control process that TRANSEARCH uses each year is a comparison with the CFS. The proprietary database does have some distinct differences in coverage and detail that make direct comparisons difficult, such as the unique identification of traffic moving from warehouse and distribution centers. However, comparisons with the CFS are useful to determine if freight movement volumes of specific commodities, and between broader market areas, are verified by the only other source of similar flow information. In addition, when confronted with questions from clients about the volume of specific movements, the CFS results are consulted to help verify or refute questions.

OTHER COMMENTS

One of the primary considerations in the annual development of TRANSEARCH is maintaining year-to-year compatibility of the data for the long-term base of clients. While changes in methodologies, sources, and coverage have been made over the years to better capture a true picture of the freight transportation markets, adoption of these enhancements receives critical consideration. Consequently there are many areas in which CFS information is not utilized as it is not believed to provide improvements over existing sources and procedures. Rail and water information, for example, from the CFS has not been judged to provide a better picture of this activity than that which is available through the STB Waybill Sample, our proprietary data exchange with the railroads, and the various data sets available from USACE on water activity.

Modal Choice in Product Shipments

Analysis of Nonpublic Census Microdata

GALE BOYD
MARIANNE MINTZ
KI YOUNG PARK
ANANT VYAS
Argonne National Laboratory

This paper explores how micro data within the Census itself can be used to supplement the Commodity Flow Survey (CFS). By establishing research using the raw CFS and other economic data collected at Census we gain a more comprehensive understanding of the uses of CFS data

Modal choice of product shipments has changed in the last 20 to 25 years. This has implications for the shipping sector, transportation patterns, energy use, and pollution. Aggregate data show large shifts in mode with higher value products more likely to use premium shipment modes, like air freight. These issues are important for a variety of forecasting, e.g., the National Energy Modeling System uses forecasts of industry growth to estimate energy demand from commercial freight traffic. However, the aggregate data is insufficient to isolate the combined effects of the shipper specific, i.e., the establishment the shipment originated from, and shipment specific, i.e., the shipment characteristics in terms of size, distance, and value of the item shipped, on mode choice. This paper presents preliminary results from a project that examines the raw, non-public shipments data from the CFS linked to the corresponding non-public establishment data from the Census of Manufacturing (CM).

Through a cooperative research program with the Census Center for Economic Studies (CES), the raw data from the CFS and CM have been linked (*1*). Initial focus has been on linking the detailed shipments records from the 1997 CFS to the 1997 CM. The non-public data from the 2002 Economic Census is only available to the CES after all publication series are completed, so the newer published 2002 CFS and CM are not included in this study. Most of the records in the CFS are from the manufacturing sector. Linking of the CFS to the Census of Mining and the Census of Wholesale and Retail Trade will be performed after initial analysis of the CFS to CM link is completed. If permission to use the 2002 data is obtained from CES the project will be expanded to encompass that data as well.

The raw data for the CFS are detailed shipment specific records for those establishments in the CFS sample. Each data element represents an individual shipment from that establishment, categorized by Standard Classification of Transported Goods (SCTG), mode, destination, and other economic characteristics. For any establishment in the CFS sample, multiple shipment records are common. The multiple data records representing each establishments' shipment data have been linked to the establishment level data from the CM and the Standard Statistical Establishment List (SSEL), also called the Business Register. The CM provides detailed data on the operations of the establishment for 1997; the SSEL provides detailed data on the location and firm ownership of the specific establishment. In particular the SSEL provides the establishment classification based on the North American Industrial Classification System (NAICS) and on the standard industrial classification (SIC) system. The data from the CFS has been linked to the CM and SSEL by using the appropriate permanent plant identifiers in the respective data bases.

Initial analysis of the linked micro data has focused on examination of the distribution of value of shipments, tons, ton miles, and number of shipments cross tabulated by NAICS. The most useful tabulations have been at the three- and four-digit level (*2*).

To better understand the relationship between SCTG and the NAICS a measure of commodity level diversity in shipments is constructed for each industry, where commodity refer to the SCTG shipment taxonomy and industry refers to the NAICS establishment taxonomy. The index we choose is the Herfindahl–Hirschman Index (HHI). This index is a commonly accepted measure of market concentration and is commonly use for evaluation of mergers in antitrust cases. For our purposes it is calculated by squaring the shipment share of each commodity and then summing the resulting numbers. For example, for an industry consisting of four commodities with shares of 30%, 30%, 20% and 20%, the HHI is 2600 (302 + 302 + 202 + 202 = 2600). The HHI takes into account the relative size and distribution of the commodity shipments in an industry and approaches zero when an industry ships a large number of different commodities of relatively equal size. The HHI increases both as the number of different commodities in the industry decreases and as the disparity in size between those commodity types increases. For anti-trust, an HHI between 1000 and 1800 points is considered to be moderately concentrated, and those in which the HHI is in excess of 1800 points are considered to be concentrated. However, it is not clear if those ranges are relevant to our application. The HHI does provide a useful measure of relative concentration.

By using the tabulations of the distribution of value of shipments, tons, ton miles, and number of shipments cross distribution of value of shipments, tons, ton miles, and number of shipments cross tabulated by NAICS and the HHI we hope to establish a meaningful level of NAICS aggregation of the CFS using this industry level taxonomy. The final goal of the project is to estimate a multinomial logit model of mode choice in industry. This shipment specific model choice model will determine the relative influences of shipper specific characteristics, like product mix, size, and industry /regional effects compared to shipment specific influences like length of haul, value ($/ton), and regional destination. This paper will present a preliminary examination of our applications of the micro-level CFS data to date.

In an earlier study using a bridge file between SCTG and SIC, provided by the Census, we constructed tabulations, by mode for various commodities within two-digit SIC. Two sectors that stand out in terms of the CFS data on $/ton growth are SIC36 and SIC 38, where the cumulative growth between 1977 and 1997 exceeds 250%. The data suggest substantial change in the value and weight relationship over 20 years. This is not at entirely surprising. What are more interesting are the detailed results for these two sectors. The 1997 CFS reveals large differences in the shipment characteristics in terms of $/ton for shipments that are sent by air and parcel compared to all modes (Table 1). This is also not surprising, since higher value, lighter weight products would be more likely to be shipped via these premium modes. However, this comparison illustrates how the detailed information on shipment characteristics may be used in a mode choice model. In particular we hypothesize that the value per ton shipped is an important determinant of the mode choice. This "light-weighting" of goods is a potentially large influence in future economic growth, so would have implication for forecasting mode choice and resulting energy consumption.

TABLE 1 Comparison of $/Ton by Mode and Detailed SCTG (Source: 1997 CFS)

Five-Digit SCTG in SIC 36	**All Modes**	**Air**	**Parcel**
Electric motors, generators, generating sets, rotary converters	$8.09	$40.02	$43.02
Electric cooking appliances and other electrothermic or electromechanical...	$7.08	—	$16.22
Line telephone or telegraph apparatus	$63.62	$106.51	$90.58
Electronic entertainment products, except parts	$15.62	$56.95	$37.86
Prepared unrecorded or prerecorded media	$29.97	$59.05	$93.73
Transmission apparatus for radio or television broadcasting, radio...	$102.87	$233.46	$194.63
Electronic components and parts	$60.75	$480.94	$275.32
Other electronic and electrical equipment	$10.62	$78.70	$56.76
Average	$17.55	$151.38	$89.65
Five-Digit SCTG in SIC 38			
Optical elements, instruments, and apparatus, except photographic and...	$86.25	$116.27	$151.89
Photographic and photocopying machines	$33.83	$89.57	$60.16
Surveying, hydrographic, oceanographic, hydrological, geophysical, drawing...	$44.67	—	$121.50
Instruments, apparatus, and appliances for medical, surgical, dental, or...	$48.25	$245.19	$118.49
Meters and other instruments and apparatus for measuring, checking, testing...	$69.47	$153.97	$145.90
Average	$53.72	$200.86	$121.30

Some of the questions that the above discussion suggests include the following:

- Is mode choice product specific or industry specific?
- Are there common characteristics of mode choice across industries, e.g. $/ton or length of haul?
- Are there shipper/shipment specific characteristics which are structural, like location or destination, which the aggregate data do not reveal?

These questions can only be adequately addressed with the micro-level data. If these are important determinants in the changes that have occurred in the past 20 years, and may be expected to continue, then this would be a significant improvement over using static mode choice.

The general form of the multinomial logit model is:

$$\text{Prob[choice } k \text{]} = \exp[\beta_k' X] \,/\, \textstyle\sum(j=0,J)\exp[\beta_{j,}' X]$$

It is common in transportation modeling that the *X* is the characteristic of the mode, but in our case *X* includes the cross sectional characteristics of the shipper and shipment. The various shipper and shipment specific effects that are under investigation, including which data may represent these effects are discussed in the next several paragraphs.

For example, the above discussion might lead to the notion that the mode choice is highly industry specific. However, examining the product specific data presented for SIC 36 and 38 raises the possibility that mode choice is a function of the value-to-weight relationship. Small high-value products get the fastest mode because of the time value inherent in the shipment. This approach might do well for the aforementioned SIC, but does that explain air freight shipments of flowers and fruit? The short economic life of these commodities is more likely the answer. Therefore a product-specific effect is expected, but the value argument may explain a lot of observed produce-specific differences. Shipper-specific preferences or structural effects may also explain mode choice.

Shipper-specific effects may include size, with larger establishments more likely to have private truck for captive shipments. Product mix may also play a role. Establishments with diverse mix might use diverse modes or tend to ship some products by the same mode as their primary output. This would explain why some fraction of a particular product goes by a less conventional mode. Structural effects like location close to a rail spur, waterway, or other type of shipment hub might influence a shipper's choice. These effects would be hard to capture without a detailed analysis of location, but could be included in a shipper specific variable. Of course, such a variable could not be released, but could be tested to see how robust the other variables are to such a specification.

Explanatory variables to be examined for the basic multinomial logit include but need not be limited to

- Shipment specific effects as given by CFS:
  - Length of haul;
  - Value;
  - $/ton;
  - Regional dummies for shipment destination; and
  - Product level (SCTG) dummies.
- Shipper specific effect as given by the LRD:
  - Product mix measured by specialization ratio;
  - Size measured by annual value of shipments, employment, etc.;
  - Industry dummies (or industry specific equations); and
  - Regional dummies for plant location.

The number of parameters in the multinomial logit model can get quite large, depending on the dimension of X and the number of alternatives, J+1. To reduce this somewhat our focus is on the basic mode choice. The CFS identifies mixed modes. These are modes that involve transshipments, e.g. a truck delivery or pickup from or to a rail or water depot. Often these mixed modes which include rail or water will be predominantly via that mode, with the truck portion providing a small fraction of the total transportation service. Therefore, rail–truck will be combined with rail; water–truck, as well as water–rail will be combined with water. Parcel, which is a mixed mode frequently with an air component, has grown in some high-value sectors. We will test the choice difference between parcel and air to see if there are any significant differences.

NOTES

1. This specific project is underway at the Chicago Research Data Center (www.chicagordc.org). For more information about the CES research programs see www.ces.census.gov/ces.php/home.
2. There are 512 SCTG codes; 1,168 NAICS codes; and 1,004 SIC codes. The five-digit NAICS corresponds roughly to the four-digit SIC. Data for about two thirds of the four-digit SICs can be directly derived from the five-digit NAICS. However there is not a direct correspondence between the most detailed SCTG and either SIC or NAICS.

Freight Analysis Framework and the Commodity Flow Survey

ROLF SCHMITT
TIANJIA TANG
Federal Highway Administration

T. RANDALL CURLEE
Oak Ridge National Laboratory

SUMMARY

The Freight Analysis Framework (FAF) estimates commodity flows and related freight transportation activity over the national highway network, waterways, and rail system among states, sub-state regions, and major international gateways. The end product of the FAF provides not only trade and economic pattern data but also congestion/capacity data associated with all the transportation networks carrying out all the trade activities. To make the FAF a more effective tool for measuring and analyzing the changing world of freight transportation, FHWA is depending heavily on the 2002 Commodity Flow Survey (CFS) to update the FAF and assure that the next generation of FAF methods and products is both transparent and reproducible. The CFS will be the cornerstone of the FAF Origin and Destination Database of commodity flows among the 106 to 114 CFS regions plus major international gateways, benchmarked every 5 years and updated annually with provisional estimates. The 2002 benchmark Origin and Destination Database will include forecasts every 5 years from 2010 to 2035; the 2007 benchmark Origin and Destination Database will include forecasts through 2040.

WHY WE CARE

The FAF provides basic information on the flow of commodities among regions and along major intercity transportation links. This information is essential for understanding key trends and issues such as

- Patterns of merchandise trade with domestic and international partners and the economic growth potential associated with that trade;
- Volumes of traffic passing through a location between distant origins and destinations (O-Ds), indicating the effects of external traffic on local transportation facilities and the importance of local facilities to distant places;
- Markets served by different modes of transportation and intermodal combinations;
- Locations exposed to risks of hazardous materials incidents and other safety aspects of freight transportation;
- Energy use and environmental consequences of freight transportation;
- Efficiency and productivity of logistical systems supporting the nation's economy;

- Likely impacts of transportation policies on efficiency, economic productivity, safety; and
- Growth in freight transportation activity throughout the United States, and the pressures created by that growth on the nation's transportation systems.

The understanding of all issues listed above is detrimental in the decision making process of developing transportation plans, programs, or projects to move goods and people efficiently and mitigate congestions.

As Abraham Lincoln said on the subject of federal funding for internal improvements:

> Statistics will save us from doing what we do, in the wrong places. ... The surplus, that which is produced in one place to be consumed in another; the capacity of each locality for producing a greater surplus; the natural means of transportation, and their susceptibility for improvement; the hindrances, delays, and losses of life and property during transportation, and the causes of each, would be among the most valuable statistics in this connection (*1*).

WHAT THE TRANSPORTATION COMMUNITY WANTS

Over the past 3 years, the Bureau of Transportation Statistics (BTS) and FHWA have convened two expert panels of the Transportation Research Board (TRB) to discuss requirements and strategies for freight data (*2*), the design of the CFS (*3*), and future directions for the FAF (*4*). In its report on a conceptual national freight program, the first panel called for data on

- O-D;
- Commodity characteristics, weight, and value;
- Modes of shipment;
- Routing and time of day; and
- Vehicle/vessel type and configuration.

When the panel was reconvened to review the FAF, it emphasized their desire that data be available at a level of geographic detail to support project design and planning. The panel also expressed the wish that freight activity statistics be based on enhanced data collection programs rather than on estimates from models. A separate expert panel reviewed the CFS and recommended continuation and enhancement of the survey as a basic data source.

Many customers of the FAF and related data programs may not need the level of geographic detail desired by the FAF review panel, but some customers want additional data elements such as circuity, delay, reliability, shipper and carrier costs, and other freight network performance measures. Customers have also asked for national indicators such as the value, tons, and ton miles by modes of transportation and type of commodity for all freight shipments, revenue and vehicle miles of trucks and trains, and travel times. Few customers have expressed a willingness to do without updates to the FAF data in the years required to implement a national freight program.

THE PLAN

The improved FAF will be benchmarked to the CFS and Economic Census, the Census of Agriculture, and other comprehensive freight data programs for the census years ending in 2 and 7 to provide the most complete, detailed, and statistically sound basis for estimates and forecasts, and the FAF should be updated annually with provisional estimates. FAF methods and outputs should be as transparent and reproducible as possible to maximize the FAF's utility and credibility. FAF data products should not exceed the geographic and commodity detail for which data quality is assured. Data quality of FAF benchmark O-D flow estimates should reflect the statistical standards developed for components of the Economic Census.

FHWA is especially dependent on the CFS for the planned Origin-Destination Database, which will include value, tons, and ton-miles of commodities by mode, commodity type, hazard class for hazardous materials, place of origin, and place of destination.

- For benchmark years ending in 2 and 7 and forecast years ending in 0 and 5: modes are truck, rail, water, pipeline, air and truck–air, truck-rail, truck–water, rail–water, other; commodities are two-digit Standard Classification of Transported Goods (SCTG) with selected three-digit breakouts; hazard classes are one-digit U.S. Department of Transportation (USDOT) codes; and places are CFS regions and major international gateways and regions.
- For provisional current year estimates: modes are truck, rail, water, pipeline, intermodal; commodities are two-digit SCTG; hazard classes are one-digit DOT; and places are CFS regions and major international gateways and regions.

CONCLUSION

The FAF has been a very successful product for FHWA and the freight community. FAF outputs have been accepted with relatively little scrutiny because they shed new light on the world of freight transportation. To continue this success, the FAF must continue to evolve and respond to user experiences, especially as problems with underlying methods and data are uncovered.

The planned FAF improvements are designed to improve the quality of published information, minimize competition with private vendors and dependence on proprietary data, and focus FHWA resources on improvements in the completeness and timeliness of the FAF and on the development of complementary analytical tools. The FAF will continue to be a national policy analysis tool and serve as a starting point for understanding state and local freight activity. The FAF will provide a framework—rather than become a substitute—for local data collection and analysis to support small-area planning and project design.

A very aggressive program of enhancements to the quality, timeliness, and relevance of the FAF is proposed. The Office of Freight Management and Operations looks forward to working with its partners in FHWA, USDOT, Census, and the private sector to assure that the FAF continues to meet the needs of the transportation community.

REFERENCES

1. Lincoln, A. Internal Improvements: Speech to the House of Representatives, June 28, 1848. In *Congressional Globe*, 30th Cong., 1st Session, 1848, pp. 709–711.
2. *Special Report 276: A Concept for a National Freight Program*. Transportation Research Board of the National Academies, Washington, D.C., 2003.
3. *Special Report 277: Measuring Personal Travel and Goods Movement: A Review of the Bureau of Transportation Statistics Surveys.* Transportation Research Board of the National Academies, Washington, D.C., 2003.
4. *Letter Report on the Freight Analysis Framework*. Transportation Research Board of the National Academies, Washington, D.C., February 10, 2004.

Applications

Applying the CFS to State and Local Transportation Issues

Synthesizing Truck Origin–Destination Table for Local Transportation Analysis Zones

A Nashville Metropolitan Area Case Study

HO-LING HWANG
SHIH-MIAO CHIN
Oak Ridge National Laboratory

LEE D. HAN
FANG YUAN
University of Tennessee

BACKGROUND

The average length for truck shipments originating in Tennessee grew over 27% between 1997 and 2002. In the same period of time, the tonnage of freight carried by trucks originating in Tennessee increased about 20%. Clearly, there are increasing freight demands on the existing highway system. To prepare for future transportation challenges, the Tennessee Department of Transportation (TDOT) is heavily engaged in comprehensive and inclusive transportation planning efforts. This planning process will enable the elected officials, decision makers, as well as transportation analysts to have a comprehensive understanding of current and future transportation conditions. As a result, they will be able to develop alternatives to addressing the state's transportation demands and develop strategic programs to guide future transportation investment decisions.

One of the key obstacles in transportation planning is the lack of origin and destination (O-D) trip matrices. This is particularly true in freight transportation planning. The Commodity Flow Survey (CFS) is the only nationwide freight movement data collection effort in the past two decades. CFS provides freight movement information at the national, state, and some selected metropolitan areas (MA) levels. To adapt CFS in freight studies at the metropolitan planning organization (MPO) level, however, additional efforts would be needed.

Reebie Associates, a consulting firm specializing in freight transportation and distribution, has developed a CFS value-added freight movement database, TRANSEARCH. As stated by the company, over 100 proprietary data sources, commercial as well as public—including the CFS, were collected and compiled in this database. To facilitate its freight transportation process, TDOT has purchased the Tennessee portion of TRANSEARCH. However, the state-level TRANSEARCH freight database cannot be readily used by the MPOs in Tennessee.

OBJECTIVE

The objective of this study is to explore the feasibility of adapting freight transportation data that was prepared at the state level to traffic studies for the MPO. This study synthesized a truck O-D matrix for a transportation analysis zones (TAZs) defined by the local MPO so that the assigned truck traffic is consistent with those observed on the roadway network.

BASE STUDY SITE

Nashville MA in Tennessee was selected for this study because there is significant truck traffic through this region. Additionally, the research team has a good working relationship with state and local transportation agencies. These positive relationships facilitated the acquisition of critical data from those agencies.

Data from the Nashville MPO

The digital roadway network database and the digital TAZ database, prepared and used by the planning office, were utilized in this study to synthesize the truck O-D matrix for the Nashville MPO. By doing so, results from this study can be transferred directly to the Nashville-area MPO. The study area for this project consists of 1,440 zones and covers five counties (Sumner, Wilson, Rutherford, Davidson, and Williamson). Small parts of the Robertson and Maury counties are also included in Nashville's TAZ.

The digital highway network prepared and used by the Nashville-area MPO for its general transportation planning work is quite extensive. It includes 18,702 links and 14,545 nodes. The high-occupancy vehicle (HOV) lanes on Interstates are coded as separated links. Note that these HOV links are excluded from the synthesizing truck O-D matrix process.

Data from TDOT

The most up-to-date annual average daily traffic (AADT) data (2004) was obtained from the Tennessee Roadway Information Management System (TRIMS) at TDOT. The geo-referenced information provided by TRIMS enabled the project team to separate passenger vehicles from single-unit trucks and multi-unit trucks. Multi-unit truck traffic was then geographically transferred to the Nashville-area MPO's highway network. As in many data collection systems, the truck AADT information is not perfect. Some of the data gaps are clearly visible when examining the map. Nevertheless, TRIMS provides adequate truck traffic information for this study.

Extracting Regional Data from State-Level Truck Information

A cookie cutter procedure developed by the Oak Ridge National Laboratory/University of Tennessee (ORNL/UT) research team for extracting a sub-state region from the state-level data is described in the following steps:

- Identify all links within a selected region.

- Identify all routes which include at least one link in the region.
- Identify all O-D pairs with routes that include at least one link in the region.
- Identify origin nodes and destination nodes for each O-D pair in the region that meets the following conditions:
  - O-D pairs with route passing through the region;
  - O-D pairs with route originated from, but terminated outside, the region;
  - O-D pairs with route originated from outside the region but terminated in the region;
  - O-D pairs with route originated from and terminated in the given region; and
  - Ignore the intermediate points for routes going in and out of a region multiple times. This is possible if a region is not convex or its roadways are winding.
- Transfer the O-D with route that includes at least one link within the region to the new origin node and destination node for the given region.
- Consolidate all selected O-D pairs to the new set of origin and destination nodes for the region.

Quality of the Preliminary O-D Matrix

To evaluate the quality of this O-D matrix for the Nashville area, 168 link segments in TRIMS were selected. These selected highway segments are either Interstates or freeways and expressways within the study region. These 168 link segments can be considered as "checking" locations to be matched with observed truck traffic volumes. A Mean Sum of Absolute Errors (MSAE) was then used as a goodness-of-fit measure to quantitatively assess the O-D matrix. Mathematically, MSAE can be expressed as the following:

$$\text{MSAE} = \frac{\sum_{i=1}^{N} \dfrac{\left|\text{Estimated}_i - \text{Observed}_i\right|}{\text{Observed}_i}}{N}$$

The MSAE between the estimated link traffic (assigned based on the new O-D truck matrix) and the observed multi-unit truck traffic (as reported in TRIMS) is approximately 0.47. In other words, the average discrepancy between the estimated and observed traffic volumes is 47% among these 168 locations.

Calibration

The preliminary truck O-D matrix was calibrated based on the above-mentioned 168 Interstate and freeway–expressway locations to create an adjusted truck O-D matrix. The estimated truck flow pattern from this adjusted truck O-D matrix is found to be very similar to the observed multi-unit truck flow pattern. The MSAE between the estimated link traffic assigned based on the synthesized O-D matrix and the observed multi-unit truck traffic reported in TRIMS is about 0.09. In other words, the average error of 47% induced by the newly "cut" O-D matrix was reduced to 9% by the calibration process.

CONCLUSION

This study demonstrated that data collected nationally can not be used directly by the local metropolitan planning agencies. Instead, additional calibration efforts are needed to adjust the preliminary O-D matrix that was abstracted from state-level flow to its local traffic conditions. After the calibration process, the adjusted truck origin and destination matrix can yield an assigned truck-traffic, on major freeways and expressways, with an acceptable level of error (less than 10%).

The calibration efforts for this study were concentrated on truck traffic on Interstates, freeways and expressways. This effort leads to a small discrepancy between the assigned and observed truck traffic volumes, on the average about 9%. With additional efforts, a more “balanced” O-D truck matrix can be formulated by placing more observed traffic “checking” points on arterials and other surface streets. It is expected that error rate would be reduced with a balanced O-D truck matrix.

Using National Freight Information to Study the Interactions Between Heavy Trucks and Passenger Cars Under Traffic Incidents or Emergencies

SHIH-MIAO CHIN
HO-LING HWANG
Oak Ridge National Laboratory

LEE D. HAN
FANG YUAN
University of Tennessee

BACKGROUND

Our nation's economy depends heavily on transportation infrastructure. Most businesses and industries depend on effective freight transportation systems to reach state, regional, national, and global markets. As shown in the most recent Commodity Flow Survey (CFS) results released by the Bureau of Transportation Statistics (BTS), the majority of our goods are moved by truck, accounting for more than $6 trillion in value and about 8 billion tons in weight in 2002. Specifically, trucks carry three-quarters of the value of the total freight shipped in the United States and two-thirds of the weight, according to this survey.

With the ever-increasing demand for freight and people movement on our relatively steady highway infrastructure, traffic congestion and delay have become major factors that impede the economic growth and degrade our quality of life. The degradation in the nation's mobility is further aggravated when traffic incidents and emergency conditions occur. Re-routing heavy trucks during such events requires thorough consideration of numerous constraints because of the height, width, clearance, weight, turning radius, etc., of these vehicles as well as the availability of suitable alternative roadways.

OBJECTIVE

The objective of this paper is to explore the feasibility of using national-level freight movement information in tandem with traffic simulation software to study the interactions between heavy vehicles and passenger cars during traffic incidents or emergency conditions. This research was conducted using a case study approach.

TRAFFIC SIMULATION TOOL: DYNASMART-P

The Traffic Estimation and Prediction System for Planning (DYNASMART-P) is one of the latest FHWA-sponsored traffic engineering analysis tools that can be used to help evaluate intelligent transportation system (ITS) traffic management and information strategies. It was designed for traffic engineers to model the effects of various ITS components (*1*). The DYNASMART-P is the first traffic simulation software that enables traffic engineers to use a separate truck origin and destination (O-D) table as the input information for truck traffic.

Study Region: Knox County, Tennessee

Knox County in East Tennessee was selected as the geographic area for this study mainly because of its significant truck traffic (e.g., Interstate 40, Interstate 75, as well as truck traffic from Interstate 81 traveling south or west). This area is also home to the research team. Team members not only possessed a good working knowledge of the traffic operations in and around Knox County, but also have good working relationships with state and local transportation agencies. These relationships facilitated acquisition of critical data required for applying DYNASMART-P to this study.

Data Sources

For this study, the most up-to-date annual average daily traffic (AADT) data for the study region was obtained from the Tennessee Roadway Information Management System (TRIMS) at Tennessee Department of Transportation (TDOT). Information from the TRIMS enabled the research team to determine traffic volumes for passenger vehicles, single-unit trucks, and multi-unit trucks separately for peak-hour periods. These peak-hour traffic patterns established the base-lines for the calibration process in this study.

The traffic analysis zone used in this study mostly follows the boundary of those traffic analysis zones (*2*) formulated by the Knoxville and Knox County Metropolitan Planning Commission [later renamed as the Knoxville Regional Transportation Planning Organization (TPO)]. Furthermore, the major street network used in this case study is based on the example file, prepared by the University of Maryland, as input for the DYNASMART-P. This geo-coded link-node network has detailed geographic representation in which on-ramps and off-ramps, as well as ramps of interchanges, are modeled and depicted.

All signal timing information for all signalized intersections in the Knoxville area, which is needed to run the DYNASMART-P, was provided by the Traffic Signal Group, Traffic Engineering Section in the City of Knoxville. Traffic signal related information was also coded by the University of Maryland.

Based on data prepared for the GeoFreight (*3*) system, annual freight (in tons) that originated, terminated, and traveled through Knox County can be identified. Utilizing additional information from the Zip Code Business Patterns (*4*), this freight information can then be “mapped” onto zip code areas and further “shared” to the transportation analysis zone (TAZ). Through this process, a truck O-D matrix for Knox County is established. The data reveals that freight flow in Knox County is mostly composed of through traffic. This is consistent with the fact that regional weekday profiles of multi-unit trucks show little peaking characteristics during both morning and afternoon peak periods.

Another piece of information needed for this study is a passenger vehicle O-D matrix. The original passenger vehicle O-D trip matrix was prepared by the Knoxville Regional TPO in early

1990s. This passenger vehicle O-D matrix was later modified by the University of Maryland and used as an example for the DYNASMART-P. In this study, the DYNASMART-P version of the passenger vehicle O-D table was calibrated to traffic as reported in TRIMS.

Calibration

Under the current study, efforts were made to calibrate the passenger vehicle and truck O-D trip matrices so that all assigned link-traffic would match TRIMS-based baseline traffic volumes. The basic framework of this calibration algorithm is to make minimal adjustments to the O-D trip matrix so that the total discrepancy between the assigned and the TRIMS-based traffic volumes, for all selected links, can be minimized. A goodness-of-fit measure, the Mean Sum of Absolute Errors (MSAE), is used to assess the calibrated O-D trip matrix. Mathematically, MSAE can be expressed as the following:

$$\text{MSAE} = \frac{\sum_{i=1}^{N} \frac{\left|\text{Estimated}_i - \text{Observed}_i\right|}{\text{Observed}_i}}{N}$$

The calculated MSAE for each of the adjusted O-D matrices is summarized in Table 1. Clearly, errors for the truck O-D matrices are relatively small when comparing to the passenger vehicle O-D matrix.

Simulation Scenario

These calibrated O-D matrices were then "loaded" onto the traffic network in the DYNASMART-P, using a constant rate for a period of 60 simulated minutes. The simulation was executed for 180 min. To evaluate different passenger vehicle and truck re-routing strategies during an emergency, eight incidents (with an assumed 50% reduction in capacity for 60 min) were simulated on Interstates links. The starting times for these eight simulated incidents are staggered during a period that is 30 to 60 min from the beginning of the simulation.

Two re-routing strategies were formulated and simulated in this process. The first strategy is to divert both passenger vehicle and truck traffic to surface streets as necessary for locations with capacity reduction. The second strategy involves a "truck preferential treatment" strategy which is intended to keep trucks on highways and divert passenger vehicles to local roadways.

TABLE 1 MSAE on Assigned Versus TRIM-Based Traffic Volumes

	Number of Check Segments	MSAE
Passenger vehicle O-D	194	29.5%
Single-unit truck O-D	122	7.3%
Multi-unit truck O-D	122	8.7%
Combined truck O-D	122	6.6%

Preliminary Results

From this study, it is found that DYNASMART-P simulated approximately the same number of vehicles under all four cases. The eight simulated incidents resulted in only small delays when averaged over all vehicles, delay increased from about 26 min per vehicle to 28 min per vehicle. However, average travel time for those incident-impacted vehicles is almost doubled. It should be noted that about 19% of all vehicles in the network were affected by these eight incidents.

This study also revealed that a truck-preferential-treatment strategy can improve overall traffic operations when highway capacity is reduced because of incidents. Under the same condition (i.e., with eight incidents on the network), when trucks are discouraged from using arterial and local streets, approximate 10% improvement in average travel time for all travelers can be observed. This improvement, however, translates into less than 2% of additional vehicles being "evacuated" from the network within a 3-h period. Although this improvement seems small, its magnitude could increase as traffic becomes more congested. The magnitude of improvements would also increase as the number of incidents increases, or as to the severity and durations of the incidents increased.

CONCLUSION

This case study demonstrated that national level freight movement data can be adapted to local traffic studies on trucks. Although the national freight movement information can not be applied directly in small geographic regions, it remains to be an important data sources for the metropolitan planning organization. With supplemental data from local agencies, a truck O-D table can be built to reflect local-level truck traffic volumes. Such a local truck O-D table can greatly improve transportation analysts' ability to study the increasingly congested urban highway traffic.

NOTES

1. Such as advanced traffic management systems, advanced traveler information systems, advanced public transportation systems, commercial vehicle operations, and emergency management systems.
2. A copy of the Knoxville and Knox County Metropolitan Planning Commission formulated TAZ boundary file can be downloaded from www.census.gov/geo/www/cob/tz2000.html.
3. Geofreight Intermodal Freight Display Tool, U.S. Department of Transportation, www.fhwa.dot.gov/freightplanning/geofreight.htm
4. ZIP Code Business Patterns CD-ROM: 2001, C1-E01-ZCBP-09-US1, U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau.

Using CFS Data to Guide Regional Transportation Policy and Investment

SCOTT DRUMM
Port of Portland

Commodity Flow Survey (CFS) data has played a significant role to help set the context for regional transportation policy and investment decisions in the Portland–Vancouver region. Data from the CFS has been a primary input into the region's Commodity Flow Forecast (1997) and the update in 2002. The CFS has also provided data that has helped answer business community questions about freight flows and engaged them in policy discussions regarding the Columbia River crossing as part of the Interstate 5 Trade Corridor project.

Both directly and indirectly, the CFS has been very helpful in helping us set the context for freight movement and to put freight issues on the regional transportation agenda. CFS data gives us the ability to frame the issues, convey the order of magnitude of freight's importance, and to identify areas where further data is needed. Ultimately, we would like to be able to use the data at a project level, but the CFS doesn't provide enough detail. That is to say, we would like to have the data at detail level sufficient to help make the case for a specific investment or to prioritize among competing investments. However, even at current levels of detail, the CFS has been useful. Due in part to CFS data in our Commodity Flow Forecast, we have secured $500,000 in regional funding for a freight data collection project that will provide us with some of the detail we need to make specific investment decisions, such as origin–destination and time of day data.

This presentation showed how and why our region has successfully used CFS data, identified where we have found gaps and problems, and suggested alternatives for making CFS data more accessible and more useful at a regional level.

Estimating Freight Flows for Metropolitan Area Highway Networks Using Secondary Data Sources

GENEVIEVE GIULIANO
PETER GORDON
JIYOUNG PARK
LANLAN WANG
University of Southern California

QISHENG PAN
Texas Southern University

INTRODUCTION

Trade and shipments between countries, regions and cities have been expanding dramatically in recent years. Yet, transportation planners and social scientists studying these phenomena have mostly relied on rules-of-thumb or infrequent and expensive shipper surveys to estimate freight flows. These two approaches are no longer adequate. We suggest that it is possible to estimate most of a metropolitan area's (MA) highway network truck shipments from secondary data sources, using these sources to generate relatively inexpensive and updateable link-specific estimates.

The major research steps involved are the following.

1. Utilize a regional input–output transactions table to estimate intraregional commodity-specific trip attractions and trip productions, and allocate these to small-area units.
2. Estimate commodity-specific interregional and international trip attractions and trip productions for those locations where airports, seaports, rail yards, or regional highway entry–exit points are located.
3. Create a regional commodity origin–destination (O-D) matrix using estimates from steps 1 and 2.
4. Load the O-D matrix onto a regional highway network with known passenger flows.

This paper discusses the first two steps of our approach. It is complicated by the fact that freight data from the most important data sources are described via various (often independent) classificatory systems and definitions. Much of our work has been devoted to reconciling data from these various sources.

In what follows, we describe a prototypical application of our approach to the Los Angeles metropolitan area [the five-county central metropolitan statistical area (CMSA)]. We describe the steps required to reconcile Commodity Flow Survey (CFS) data with other important data.

DATA SOURCES AND RECONCILIATION

Throughout, we rely on data sources that are widely available, suggesting that our approach is widely applicable. Our data sources are

1. A regional input-output data file from the Minnesota IMPLAN Group (*1*);
2. The CFS (*2*);
3. WISERTrade (*3*);
4. The California Department of Transportation's (Caltrans) Integrated Transportation Management Systems (ITMS) file (similar products may be available in other states);
5. A transportation analysis zone (TAZ) -level employment-by-place of jobs;
6. Data from the Waterborne Commerce of the United States (WCUS) (*4*); and
7. Data supplied by major airports in the region and complementary data from RAND (*5*).

Our approach requires the use of commodity- or sector-specific data from a variety of sources. This made it necessary for us to reconcile various classifications as best we could. We also developed the 47-sector system of "USC Sectors," a system that we could translate most sector classifications into. Table 1 summarizes the current state of convertibility.

TABLE 1 Data Source, Industrial Sectors, and Code Conversion Matrix (in current $)

DATA SOURCE	CODE	USC	SCTG	BEA	NAICS	IMPLAN (2001)	SIC	HS	SITC	WCUS
	USC		C, E	C, E	C, E	C, E	C, W	C, E	C, W	C, W
CFS 1997	SCTG	C, E		C, E	C, E	C, E	P	C, E	C, W	C, W
BEA 1997	BEA	C, E	C, E		A	A	P	A	P	P
Economic Census 1997	NAICS	C, E	C, E	A		A	C, W	C, E	P	P
IMPLAN 2001	IMPLAN	C, E	C, E	A	A		P	C, E	P	P
SCAG employment data 2000	SIC	C, W	P	P	C, W	P		P	P	P
WISERTrade 2001	HS	C, E	C, E	A	C, E	C, E	P		C,W	C,W
WISERTrade 2001	SITC	C, W	C, W	P	P	P	P	C, W		C, E
WCUS 2001	WCUS	C, W	C, W	P	P	P	P	C, W	C, E	

Notes: Highlighted cells are the conversion ratios used in this research project.
C: Created
A: Available
P: Possible to create
E: Bridge allocations are evenly distributed, where necessary, without any weights
W: Bridge allocations with plausible weights

Code Discriptions
USC: USC sectors created by USC SPPD Team
SCTG : Standard Classification of Transported Goods (http://www.bts.gov/cfs/sctg/welcome.htm)
BEA: Bureau of Economic Analysis (http://www.bea.doc.gov)
NAICS : North American Industry Classification System(http://www.census.gov/epcd/www/naics.html)
IMPLAN: IMPLAN 509-sector codes
SIC : Standard Industrial Classification (http://www.osha.gov/oshstats/sicser.html)
HS : Harmonized System (http://www.statcan.ca/trade/htdocs/hsinfo.html)
SITC: Standard International Trade Classification available from WISERTrade (http://www.wisertrade.org/home/index.jsp)
WCUS: Waterborne Commerce of the United States (http://www.iwr.usace.army.mil/ndc/data/datacomm.htm)

INTRAREGIONAL SHIPMENTS

To estimate truck shipments within a region, we must estimate commodity supply and demand within the region to and from local enterprises. We seek to generate a set of O-D matrices that describe commodity flows in and out of small areas for the nine aggregated Standard Classification of Transported Goods (SCTG) commodity sectors used for metro area CFS reports. This section describes the estimation of commodity flows associated with intraregional supply and demand.

To estimate the attractions and productions of commodities at each TAZ in the region, requires a regional input–output transactions table and small area employment data (introduced in Cho et al., 1999). The regional input–output model, with foreign shipments removed, provides the basis for estimating zone-to-zone shipments, once the regional coefficients are combined with small-area jobs-by sector data (Equations 1 and 2). In this case, employment data by sector by zone are from a data file provided by the Southern Association of Governments. The IMPLAN input–output transactions table provides the dollar values of inter-sector commodity flows that serve household consumption and the parts of final demand not associated with households.

We use TAZs which are approximately the size of census tracts. Two basic estimation equations are:

$$D_i^z = \sum_j a_{ij} X_j^z \tag{1}$$

where X_j^z is total regional output of commodity *j* in zone *z,* given base year employment in sector *j* and zone *z*;

a_{ij} is the *i*, *j*th element of **A**, the matrix of demand coefficients for the (open) input-output model; it represents the flow from *i* to *j* per unit output of *j*; and

D_i^z is the freight flows attracted from sector *i* to response to the demand in zone *z*.

The total attractions at the destination denoted by D_i^z represents the shipments of commodity *i* to zone *z* from transshipment zones (imports) and from other zones to accommodate regional demand.

Similarly, Equation 2 calculates the total supply of output *j* furnished by zone *z*,

$$O_j^z = \sum_i b_{ij} X_i^z \tag{2}$$

where X_i^z is total regional output of commodity *i* in zone *z,* given base year employment in sector *i* and zone *z*, b_{ij} is the *i*, *j*th element of B, the matrix of supply coefficients for the (open) input-output model. This is the flow from i to j per unit output of *i*. O_j^z is the freight flows produced in zone *z* to satisfy regional demands by sector j.

Total productions at origins denoted by O_j^z includes the shipments of commodity *j* to transshipment zones from zone *z* to accommodate non-local final demand (exports) and to other zones to accommodate local demand.

We can carry out these calculations for any number of commodity sectors. In this application, we aggregate to a smaller number of SCTG sectors because of data constraints imposed by the estimation procedures for international and interregional trip ends, as discussed in the following section.

INTERREGIONAL AND INTERNATIONAL SHIPMENTS

The CFS which collects data on the movement of goods within the United States every 5 years is a major data source to estimate interregional and international modal shares. As we noted, the CFS "inbound" and "outbound" flows include all flows that are shipped to and from local area establishments, even if they are shipped to or from a transshipment point for purposes of international and interregional trade; at the CMSA level, CFS modal data and sectoral data are limited, especially for inbound commodity flows. We, therefore, use some of the California, West-region and Pacific-division CFS data to estimate the modal distribution of some remainder of CMSA shipments.

The other key data source for our research is the IMPLAN data, purchased from MIG. It provides shipments data which can be reconciled with the CFS data in terms of the definitions of freight flows. The bridge table for IMPLAN and CFS sectoral classifications makes the aggregation of IMPLAN sectors into SCTG sectors possible. In general, it is the joint use of CFS and IMPLAN data that makes the estimation of interregional and international shipments by mode and by sector complete.

Spatial allocation of freight trip ends is also conducted. Interregional and international shipments to and from the CMSA pass through the region's two seaports, five airports, three rail yards, and six major highway entry–exit points. The allocations of inbound and outbound of truck and rail shipments are based on data from 1996 ITMS report from Caltrans. For the spatial

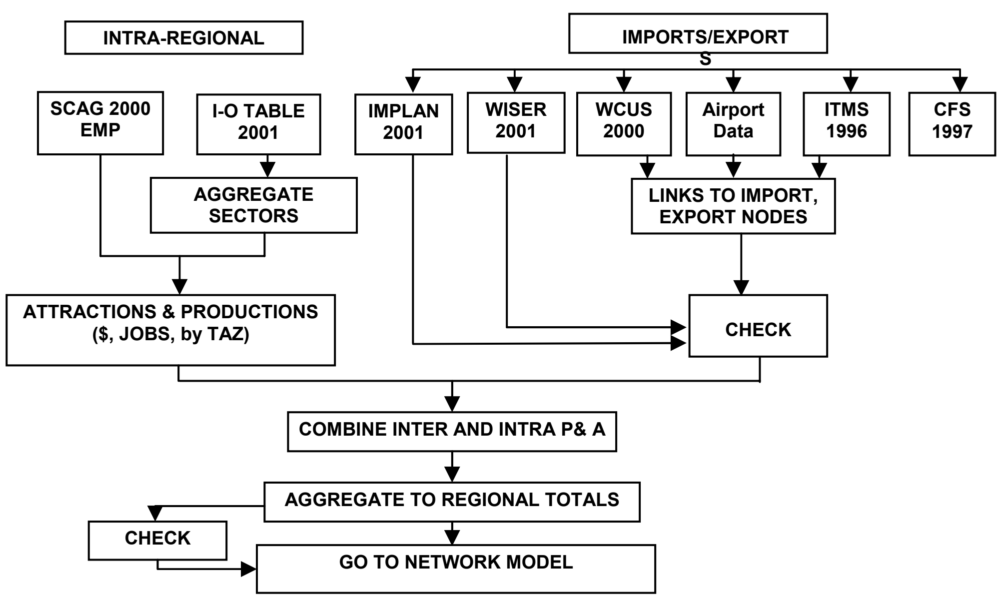


FIGURE 1 Freight flow model.

allocation of airborne and waterborne commodities, the freight shares for five airports were based on the airport statistics and seaports' data from WCUS or WISERTrade. A description of the detailed data processing is available at http://www-rcf.usc.edu/lanlanwa/NSF/CFS_USC_Draft.pdf. Figure 1 describes the basic data sources and logic behind this project:

NOTES

1. http://www.implan.com/index.php?page=index&Base_Session=b3f81c27d3c3a8efe3fbf9c347fd6bfe
2. http://www.bts.gov/programs/commodity_flow_survey/
3. http://www.wisertrade.org/home/index.jsp
4. http://www.iwr.usace.army.mil/ndc/data/dictionary/ddwcus.htm
5. http://ca.rand.org/stats/economics/airport.html.

Applications

Uses of the CFS Beyond Transportation

Economic Geography of Food Distribution in the United States

PATRICK CANNING
U. S. Department of Agriculture

The Economic Research Service (ERS) studies food markets. To measure food and ingredient supply chains, ERS researchers have developed a modeling framework for estimating a multiregional social accounting system. The framework uses information from the Economic Census, including the Commodity Flows Survey (CFS), plus U.S. Department of Agriculture (USDA) data and other statistical series to gain efficient and unbiased estimates of interregional trade. In turn, the modeling framework may be used to carry out diagnostic tests on the CFS. This poster presents some salient aspects of the research that pertain to the CFS.

INTRODUCTION

In the United States, there has not been a published multiregional input–output table in the past 25 years. Instead, a detailed national benchmark input–output account is published every 5 years along with numerous socioeconomic data for individual regional economies. If one were to use the published regional data to populate a multiregional social accounting matrix, imbalances and inconsistencies of regional "make" and "use" data would have to be reconciled to each other and to an interregional trade matrix. In theory, the CFS could be used for the latter purpose but researchers generally regard CFS data gaps to be prohibitive.

Increasingly, researchers are adopting efficient information processing rules and mathematical programming to adjust survey data used for data system accounts. These frameworks are designed to facilitate the fusion of inconsistent, unreliable and incomplete data. Since data systems have many known and rigid accounting identities, these can be introduced as linear constraints in the process of integrating survey data from independently enumerated sources.

How Does It Work?

Consider the example of a closed two-sector, two-region economy depicted in Figure 1. Suppose an economic census of this economy measured all elements of Figure 1-*a* in blue type. However, assume that only a small sample survey measures interregional commodity flows as depicted in Figure 1-*b*, with published data highlighted in yellow and green. Standard errors (VT) reported in the table show the published data to be considerably more reliable.

The published data in Figure 1-*b* represent subtotals, plus the row, column, and grand totals, and are linear combinations of interior data elements of this table. Denote G as a matrix containing all such linear adding up requirement. It is straightforward to obtain estimates of the "true" commodity flows, $\widetilde{T}$, that are blue by solving the objective function depicted in Figure 1-*c*. For this example, the results are reported in Figure 1-*d*.

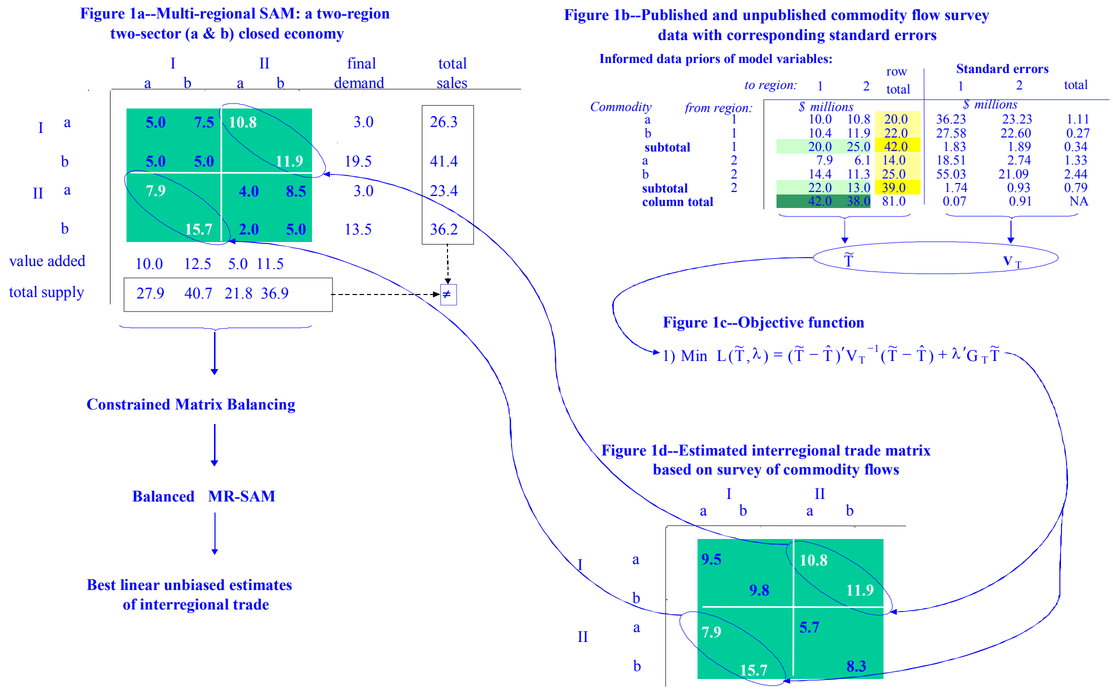


FIGURE 1*a-d* Hypothetical two-sector, two-region economy. [Best linear unbiased estimates (blue) of interregional trade using CFS: a mathematical programming approach.]

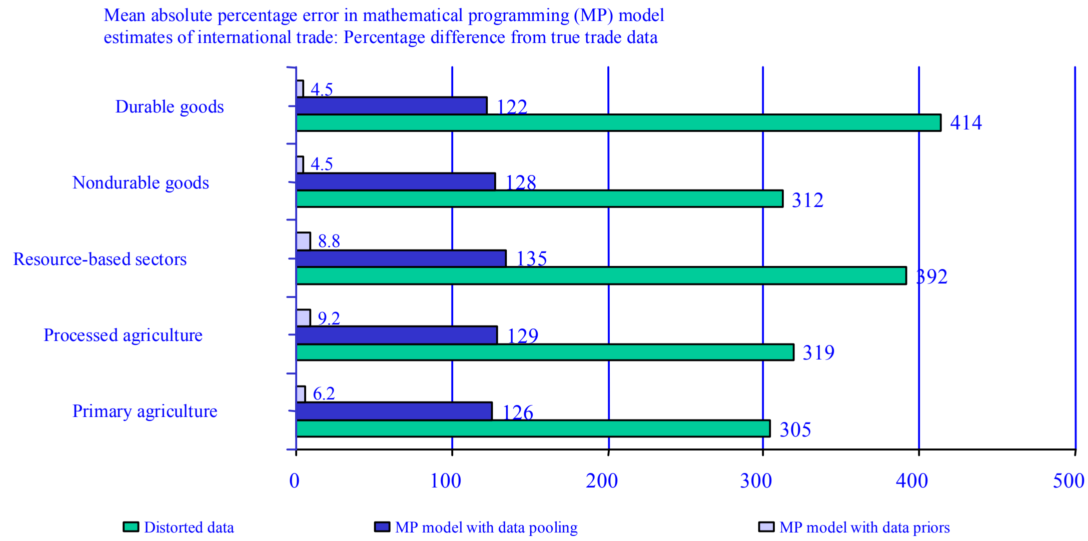
Do the Data in Figure 1-*d* Balance the Matrix in Figure 1-*a*?

By directly adding the interregional (off diagonal) elements of the trade matrix (Figure 1-*d*) to the MR-SAM (Figure 1-*a*), the resulting summation of any column would describe the total value of sector product either produced in (blue type) or shipped into (white type) that region. The corresponding row describes all in-region uses (blue type) plus out-shipments (white type). The row and column sums should be equal, but the trade matrix was estimated from a sample of shippers so is unlikely to balance the SAM. A constrained matrix balancing (CMB) framework, analogous to the equation in Fgure 1-*c*, would produce a balanced matrix and facilitate diagnostic testing of the CFS survey-data.

Figure 2 reports experimental results from a more generalized version of this framework, as applied to an actual global MR-SAM database. A discussion of the results is found in Canning and Wang (2005). The figure demonstrates the remarkable capacity of this framework to discover processes from sample data.

How USDA Uses the Data from This Framework

- Discover the general disposition of food and ingredient commodity flows in the United States between origins of movement or port of entry and destinations of use or port of exit.



Source: Canning and Wang (2005).

FIGURE 2 Experimental results: survey data facilitate efficient estimates of interregional trade.

- Facilitate applied economic geography research on food marketing systems.
- Evaluate USDA statistics vis-à-vis other USDA data and other federal data.
- Facilitate a cooperative research agreement with Professor Karen Polenske of Massachusetts Institute of Technology to work towards calibration of a multiregional input output table of the United States.
- Facilitate potential collaborations currently under negotiation:
  - A memorandum of understanding between USDA and the U.S. Army Corps of Engineers is being considered to explore data sharing and collaborative transportation research.
  - Negotiations are ongoing with the Census Bureau to examine the Economic Census within this framework.
  - Discussions and workshops with the Department of Energy, National Laboratory System has explored a leveraging of this research and research on critical infrastructure protection and related work.

CONCLUSIONS

A major obstacle in regional economic analysis and empirical economic geography is the lack of consistent, reliable regional data, especially data on interregional trade. This research seeks to integrate the information provided by the CFS into a mathematical programming model that employs an efficient information processing rule. In the process, an objective diagnostics framework is produced to evaluate the CFS in comparison to other data in the Federal statistical system. Preliminary results from this research to date include the following:

- The CMB framework employed demonstrates a remarkable capacity to discover interregional trade patterns from imprecisely measured initial estimates.
- The CMB framework provides a statistically objective use for higher statistical moments of survey data, such as measures of sample variance.
- Under reasonable assumptions about the CFS sample and enumeration, the CMB framework produces best linear unbiased estimates of interregional trade and the dual result of objective diagnostic testing procedures for the survey data.
- To fully leverage this framework for diagnostic testing and sensitivity analysis of the CFS, work with the 1997 and 2002 micro-data files is viewed as critical. To date, a USDA proposal to carry out this research at the Bureau of Census, Center for Economic Studies is under review.

REFERENCE

Canning, P., and Z. Wang. A Flexible Mathematical Programming Model to Estimate Interregional Input-Output Accounts. *Journal of Regional Science*, Vol. 45, No. 3, 2005, pp. 539–563.

Estimating Regional Trade Models in the Presence of Missing Data

RICHARD E. CHARD
Bureau of Economic Analysis

OBJECTIVE OF THE APPLICATION

The Commodity Flow Survey (CFS) is the only publicly available survey measuring trade flows among and between the states of the United States. This is an important data source for researchers interested in a variety of applications. However, the public use version of the CFS is not an ideal data source for two primary reasons: the degree and scope of data missing because of suppression; and, the use of the Standard Classification of Transported Goods for commodity classification.

In this paper, we explain how we overcame these problems to use the CFS in estimating Regional Purchase Coefficients (RPC). A RPC for a commodity is defined as the proportion of regional demand for a particular commodity that is met by regional production. Given that the CFS measures the value of shipments based on origin and destination by commodity, it would seem that the calculation of RPCs would be a simple task. However, given the limitations of the data collected in the CFS, the actual process becomes more complex.

APPLICATION CONTEXT

An Overview of the CFS Data

The data that we use for our estimation come from the 1993 and 1997 CFSs. The CFS is conducted jointly by the Bureau of Transportation Statistics and the Bureau of the Census. As noted, one difficulty for researchers using the CFS survey, including the most recent (1997) release, is the amount of data that is suppressed for various reasons (*1*). For example, the data suppression in the 1993 public use file affects 61,829 of the 86,700 entries (71.3%). The 1997 public use file has an even greater percentage of suppressed data. In fact, data suppression affects 87,640 of the 114,036 entries (76.9%) in the 1997 data. In our research, based on CFS data, we were interested not only in calculating the RPCs but also in developing and testing techniques to interpolate or impute the suppressed data.

Overview of Suppressed Data

The cell entries in the CFS are suppressed for three reasons: small values (code one); data validity (code two); and, confidentiality (code three). Generally speaking, the confidentiality suppression causes few problems, in that it accounts for a small number of cases overall (about 2,300 in 1993 and a little more than 300 in 1997). More than 75% of the suppressed data fall into the other two categories with a majority of that data suppressed because of small values (code one). Analysis that relies on listwise deletion, pairwise deletion or mean interpolation of the missing data is fraught with problems as several scholars have noted (Brown, 1983; Brown, 1994; Little and Rubin, 1987; Wothke, 2000; King et al., 2001). Thus, developing a method that consistently replaces the missing data would be of great value in estimating reliable and valid RPCs.

Process: How Was the CFS Data Used?

To calculate RPCs, measures of quantity, demand, and local demand met by local supply must be identified. The CFS is an excellent resource for this calculation in that the shipment valuation tables by commodity for state of origin and state of destination provide this data. However, because of the large amount of data missing because of suppressions, as discussed earlier, it was not possible to construct even the rudimentary estimates of RPC with the level of confidence desired from the public release version. For that reason, the missing data needed to be replaced with reliable estimates for the omitted values.

Combining CFS Data with Other Data Sources

To impute the missing shipment value data, new variables that could predict or at a minimum correlate with the variable containing missing values (shipment value data) needed to be identified. We relied on a variety of data collected from other sources including: the Bureau of Economic Analysis, the Census Bureau, and the Bureau of Labor Statistics to create an imputation data set. We examined a variety of methods of imputation and tested two of those methods.

The selection of an imputation method is somewhat limited by the sheer volume of missing data (i.e., data suppressions in the CFS). The ideal method would use an iterative maximum likelihood approach, however, given that there is three times as much missing data as data present, expectation maximization models that rely on maximum likelihood algorithms are likely to have difficulty converging (*2*). Because of this, we consulted a number of sources on missing data (King, Honaker, Joseph, and Scheve, 2001; Wothke, 2000; Allison, 2001; Franklin, 1989). We narrowed our methods to a program (AMELIA) developed by Gary King (see King, Honaker, Joseph, and Scheve, 2001 for a detailed discussion) that uses an Expectation Maximization with repeated samples technique known as EMis and a two-stage method (2SAIV), developed by Franklin (1989), that uses regression models as a starting point. We describe each method below along with a discussion of our results from these two different methods. Both methods have the desirable feature of basing estimates on existing data. Testing for reasonableness revealed a promising result for the imputed variable (2saiv value). Table 1 shows the comparison of sum, mean, median, and standard deviation for the data reported in the CFS for shipment value (value of shipment), the imputed 2SAIV variable (2saiv value) and the shipment value imputed with AMELIA (amelia value). Unfortunately, the data failed in this regard. The imputed values were tested and found to deviate too much from reasonable expectations to be valid. The table below shows some basic statistics for the public release data compared to the results of the two imputations.

TABLE 1 Comparison of Imputed Data (Values Are in Millions of Dollars)

Statistic	Value of Shipment	2SAIV Value	AMELIA (Value)
Number of Observations	26,396	114,036	114,036
Sum	26,301,668	28,626,331	38,012,931
Mean	996	153	333
Median	74	67	49
Standard Deviation	10,265	247	4,957

Results

The procedure outlined above produced what seem to be reasonable RPCs, in spite of the high degree of data missing because of suppressions in the CFS. The methods used to impute the CFS data seemed to improve the results. While there is no substitute to actually having access to the full data set, we present a reasonable procedure to increase the utility of the public use data set.

HOW USEFUL WAS THE CFS?

Without the CFS, even with the problems of suppressed data, the calculations of RPCs would be extremely difficult, if not impossible. Thus, the CFS provided an excellent, even if incomplete data source.

What Worked, What Did Not Work

The 2SAIV method, which is more forgiving of a high percentage of suppressed data than the maximum likelihood based AMELIA program, worked well. One contributing factor was probably that the highest percentage of suppressed data was for small values. If we had been willing to make an assumption on those values, replacing them through mean interpolation or some other system, the maximum likelihood approach may have worked better, however, the methods to replace that data would not have been the best available.

How Could the CFS Become More Useful for This Application?

Improvements in the CFS design could include a larger sample size to decrease the number of suppressions because of small values and data quality. Other additions to the data set that could assist those who wish to impute missing values could include reporting of means, medians and ranges for the values that are suppressed by category. This could be implemented without compromising the confidentiality of the data while simultaneously providing CFS users who wish to impute missing values with information that will help improve the accuracy of their estimates. Finally, the inclusion of a bridge from the SCTG to SIC and/or NAICS would greatly improve the utility of the CFS data.

REFERENCES

Allison, P. D. *Missing Data*. Sage Publications, Thousand Oaks, Calif., 2001.
Brown, C. H. Asymptotic Comparison of Missing Data Procedures for Estimating Factor Loadings. *Psychometrika*, Vol. 48, No. 2, 1983, pp. 269–291.
Brown, R. L. Efficacy of the Indirect Approach for Estimating Structural Equation Model with Missing Data: A Comparison of Five Methods. *Structural Equation Modeling: A Multidisciplinary Journal*, No. 1, 1994, pp. 287–316.
Franklin, C. Estimation Across Data Sets: Two-Stage Auxiliary Instrumental Variables Estimation (2SAIV). *Political Analysis,* Vol. 1, 1989, pp. 1–23.
King, G., J. Honaker, A. Joseph, and K. Scheve. Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation. *American Political Science Review*, Vol. 95, No. 1, 2001, pp. 49–69.
Little, R. J. A., and D. B. Rubin. *Statistical Analysis with Missing Data*. Wiley, New York, 1987.
Wothke, W. Longitudinal and Multi-Group Modeling with Missing Data. In *Modeling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches, and Specific Examples* (T. D. Little, K. U. Schnabel, and J. Baumert, eds.), Lawrence Erlbaum Associates, Mahwah, N.J., 2000.

NOTES

1. We should note that none of the data in the CFS is technically missing. In fact, data were collected for the entries that are read in as missing, but data were suppressed for one of three reasons: small values, data quality, and confidentiality concerns.
2. It may be possible to employ easily an iterative maximum likelihood approach to impute the missing data if the researcher were willing to make general assumptions and replace the small value missing data with zero, one, or some likely value between zero and one. Rather than doing this, we attempt to impute the data suppressed because of small values and then impose a recoding on the values, where imputed values greater than one are recoded to one and imputed values less than zero are recoded to zero, after the imputation.

Enhancing Interdependency Analysis with the CFS

Kenneth G. Crowther
Yacov Y. Haimes
University of Virginia

OBJECTIVE OF APPLICATION

Our modern era is characterized by a large-scale web of interconnected and interdependent economic and infrastructure sectors, coupled with threats of terrorism. This research demonstrates the value of integrating the Commodity Flow Survey (CFS) databases with the Inoperability Input–Output Model (IIM) to enable the analysis of interdependencies among connected economic and infrastructure sectors spanning various regions of the United States. The IIM estimates the cascading inoperability and economic losses that result from interdependencies within large-scale economic and infrastructure systems in a specified region (*1*). It has been applied in various studies for the Commission on High-Altitude Electromagnetic Pulse (*2*), the U.S. Department of Homeland Security (*3*), and the Defense Threat Reduction Agency. Derived from Bureau of Economic Analysis (BEA) data and a Nobel Prize-winning W. Leontief macroeconomic model, the IIM is a quick, inexpensive, holistic method for estimating economic impacts. The integration of the CFS databases with the IIM to enhance the capability and performance of the model is presented in three applications. First, we have used CFS databases to quantify direct disruptions resulting from an attack to a transportation asset (e.g., bridge), which prohibits or delays commodities from reaching their destinations. Second, we employ the CFS databases to enable interregional analyses and estimate how a disruption cascades across regions. Third, we plan future work that involves integrating CFS databases into a geographic information systems framework for spatial interdependency analysis.

TYPE OF APPLICATION (APPLICATION CONTEXT)

Okuyama and Chang (*4*) describe a long list of recent improvements that have been made to improve awareness of the impacts of natural or man-made disasters. The estimation process, however, is complicated by the high degree of interdependencies among infrastructure and economic sectors. Rinaldi et al. (*5*) describe many such interdependencies and recommend a general philosophy to discover and model them for planning and risk management purposes. Many methods have been employed to estimate the impacts of disasters, including IIM analysis. However, IIM analysis requires extensions in many cases to produce reasonable results that adequately represent the system of systems under study. The use of CFS and other databases enables certain extensions to the IIM for modeling transportation network interdependencies and economic impacts.

PROCESS: HOW WERE THE CFS DATA USED?

A transportation infrastructure asset can fail for many reasons, including congestion, closure for maintenance, and structural failure. The impacts will be felt in workforce commute, commodity flow, and business accessibility. Losses will accrue in each of these areas from delayed or absent workers, delayed delivery of commodities, and potential business demand reduction. An analysis of this scenario can be performed using publicly available databases through the U.S. Census Bureau in conjunction with other government agencies. Those used in this analysis include CFS to Destination, CFS from Origin, RIMS II Data, Journey to Work Data, Regional Employment Data, Regional Earnings Data, and Geographic Location Data. These databases are used to estimate the direct impacts to producing and consuming sectors of the Commonwealth of Virginia.

To illustrate the process, consider a scenario where both the Hampton Roads Bridge-Tunnel (HRBT) and Monitor–Merrimac Bridge-Tunnel (MMBT) are closed to traffic (Figure 1). HRBT is a section of Interstate 64 (I-64) located in southeastern Virginia near the mouth of the James River. The alternative route nearby is a portion of I-664, the MMBT. Our analysis will consider both to be disabled since if the HRBT is closed, the MMBT would have to handle the traffic and vice versa.

A major part of the analysis establishes the economic sectors that are primarily affected by the loss of commodity-trucking across the bridge using CFS data. The CFS provides a first-order estimate for the total number of trucks that originates or arrives in the region and the distribution of commodities across these trucks by weight or value. Trucks of origin that are delayed or stopped result in a demand reduction for the local region, while trucks of destination that are delayed or stopped result in a supply reduction for the local region. The CFS data, when supplemented with other databases provides insight into the direct effects to each sector of the economy resulting from a transportation outage. Figure 2 presents the 1997 CFS for trucking flows by two-digit sector for the Norfolk–Virginia Beach–Hampton region of the Commonwealth of Virginia.

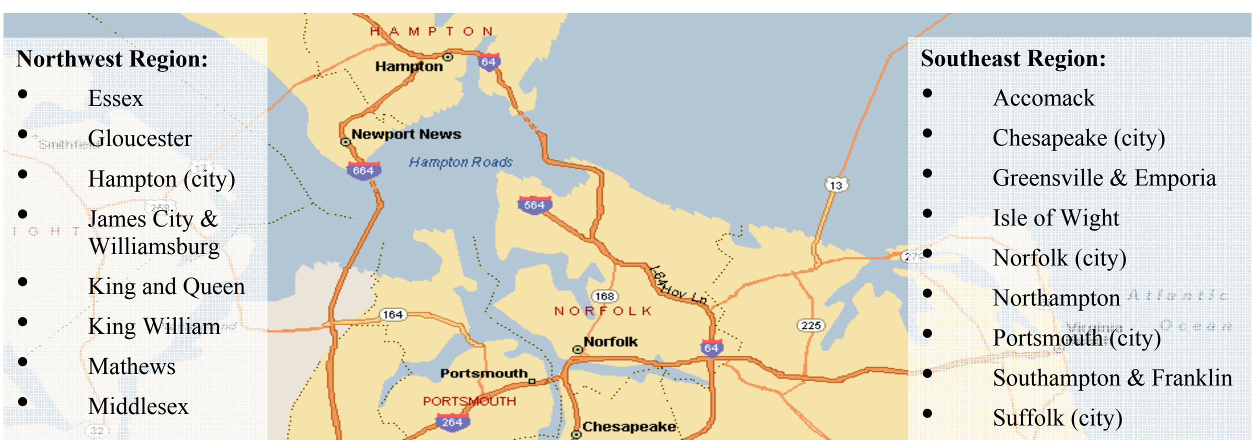


FIGURE 1 Map of region around the bridge-tunnels at the mouth of the James River.

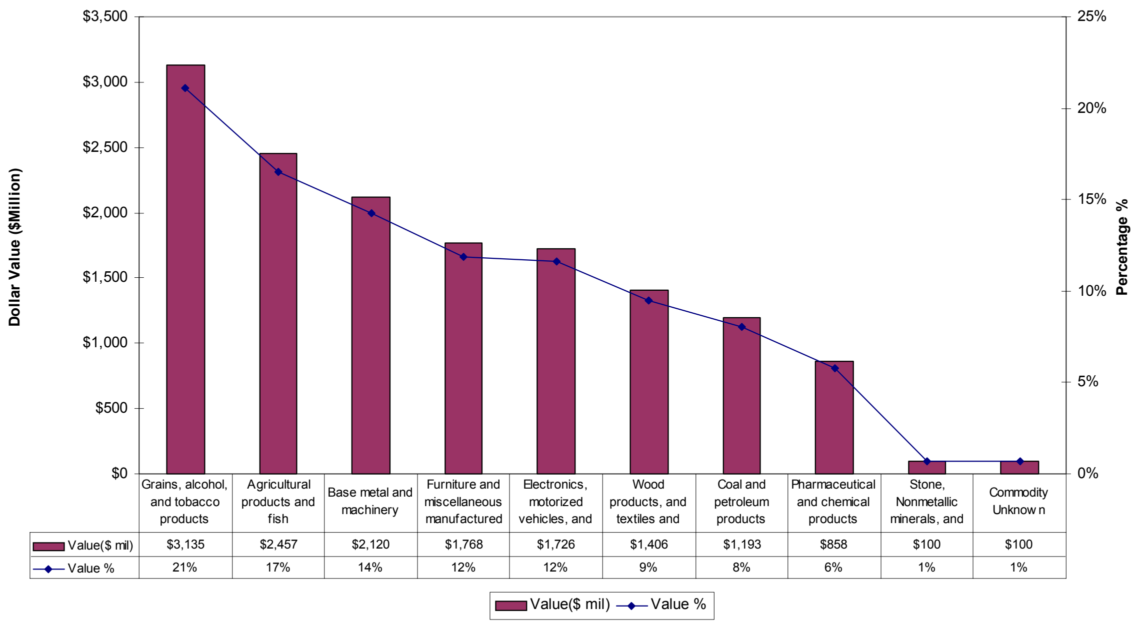


FIGURE 2 Truck flows (1997 CFS Table 6) for the Norfolk–Virginia Beach–Hampton, Virginia, region used to quantify a disruption to the local economy by transportation failure.

This commodity information is then mapped to North American Industry Classification System-classified industries based on consumption of commodities by industry and scaled down based on the nature of the closure and traffic of the transportation asset relative to the region's total traffic. The commodity distribution across sectors can be used to generate a vector that represents the total amount of direct loss for each sector of the total economy because of delayed commodities, which can be used as input to the IIM to estimate total and indirect economic impacts because of interdependencies. The information derived from Figure 2 also illustrates the value of the transportation assets to each sector because of the way it interconnects it to the entire local economy.

HOW WERE THE CFS DATA COMBINED WITH OTHER DATA SOURCES?

The TRANSEARCH database is available to Virginia Department of Transportation through Reebie and Associates. The 1998 TRANSEARCH database is built on approximately 50 million samples of Virginia commodity flows. It supplements this database for greater resolution of commodity flows by providing details of county-to-county commodity flows to the four-digit Standard Transportation Commodity Code classification across five different transportation modes. This database is used when higher resolution is desired.

Another stage of the analysis utilizes the Journey to Work database from Bureau of Labor Statistics and Local Earnings and Employment database from the BEA to estimate the impact to operation of loss of commuters who cross the bridge-tunnel complex. The Journey to Work data provides information about the counties that may use the HRBT or MMBT for commuting, and which residences and economic sectors will be most affected by the inability of commuters to arrive at work. We make the assumption that commuters traveling from the Northwest Region (as defined in Figure 1) to the Southeast Region, or vice-versa, will use this bridge-tunnel asset. The Local Earnings and Employment database provides a means of decomposing the workers across various sectors of the local economy that may be affected. By using this data we assume that the distribution of commuters across sectors is approximately the same as the distribution of workers across sectors. The reduction of workforce is assumed to create productivity losses to the individual sectors that will be affected. These databases aid in quantifying the direct input to the IIM, from which we determine the total economic effects to each economic sector in the region and total economic losses because of workers not able to cross the bridge-tunnel complex.

HOW USEFUL WAS CFS?

The CFS was useful mainly in understanding how the flow of trucking traffic is distributed across various commodity types, and the interregional connectivity between origins and destinations within these commodity types. Moreover, state and regional aggregations enabled better understanding of how certain transportation assets in specific regions compare in value to other regions across the nation because of the types and volumes of commodities they carry across distances.

WHAT WORKED, WHAT DID NOT?

Aggregations, in many cases, did not allow the resolution of analysis that was desired (hence the supplementation with an alternative database). Additionally, the commodity classifications are not supplemented with industry classifications that can easily be mapped to industry production and consumptions. Crude mappings between commodities and industries were used, which weakened the accuracy of the estimates of interdependencies and economic impacts.

HOW COULD CFS BECOME MORE USEFUL FOR THIS APPLICATION?

The single most useful addition to the CFS would be to add a single question to the survey questionnaire which requests information about NAICS industry classification of the shipper and receiver. This would be useful to produce (*a*) quantities that illustrate the overlap (or lack of overlap) between industry and commodity classifications by sector, and by region, and (*b*) tables that can be aggregated by industry classification in addition to commodity classification for applications that require the integration of CFS data with other economic databases.

A second useful addition to the CFS would be an improvement of the web site, where a user can customize some query of the total data and the web site would aid the user in finding the best resolution of detail for region and sector aggregation that still protects proprietary information of the companies being surveyed.

REFERENCES

1. Santos, J. R., and Y. Y. Haimes. Modeling the Demand Reduction Input–Output Inoperability Due to Terrorism of Interconnected Infrastructures. *Risk Analysis*, Vol. 24, No. 6, 2004, pp. 1437–1451.
2. Haimes, Y. Y., B. M. Horowitz, J. H. Lambert, J. R. Santos, K. G. Crowther, and C. Lian. Inoperability Input–Output Model for Interdependent Infrastructure Sectors: Case Study. *Journal of Infrastructure Systems,* Vol. 11, No. 2, 2005, pp. 80–92.
3. Crowther, K. G., and Y. Y. Haimes. Application of the IIM for Systemic Risk Assessment and Management of Interdependent Infrastructures: A Case Study. *INCOSE Systems Engineering*. Accepted for publication June 2005.
4. *Modeling Spatial and Economic Impacts of Disasters* (Y. Okuyama, and S. E. Chang, eds.). Spring-Verlag, New York, 2004.
5. Rinaldi, S. M., J. P. Peerenboom, and T. K. Kelly. Identifying, Understanding, and Analyzing Critical Infrastructure Interdependencies. *IEEE Control Systems Magazine,* December 2001, pp. 11–25.

Benchmarking the CFS to Waybill Data

SCOTT M. DENNIS
Bureau of Transportation Statistics

OBJECTIVE

This presentation compares railroad traffic estimates from the 2002 Commodity Flow Survey (CFS) to benchmark values found in the Surface Transportation Board's (STB's) Carload Waybill Sample (Waybill).

The CFS is an establishment-based survey of freight traffic conducted by the Bureau of Transportation Statistics and the Census Bureau. The 2002 CFS population estimates were developed from a 50,000 establishment stratified sample, and included all modes of freight transportation. The CFS sample frame excludes agricultural shipments from farm to first point of processing, imports from port of entry to first point of processing, crude petroleum, printed matter, commodities carried in non-revenue transportation, and shipments by government, service industries, utilities, most retailers, and households.

The Waybill is an annual carrier-based survey conducted by the STB. The Waybill includes railroad-only transportation, as well as the railroad portion of rail–truck and rail–water transportation presented in the CFS. The sample frame for the Waybill includes essentially all railroad traffic. The 2002 Waybill population estimates were developed from a 600,000 record stratified sample which included 22% of all railroad traffic.

The Waybill is regarded as a benchmark estimate of railroad industry traffic. Comparison of the CFS estimates of railroad traffic to the benchmark Waybill values illustrates the magnitude of in-scope issues, and identifies the differences in the railroad industry, rail–truck, rail–water, and rail–only estimates in the most recent data.

METHODOLOGY

The general methodology was to start with the benchmark Waybill estimates, eliminate traffic which would not be included in the CFS railroad estimates, and then add estimates of non-rail traffic that would be included in the CFS as part of rail–truck or rail–water multimodal shipments. Traffic was measured in ton-miles to avoid double counting tons or value when reshipped in the CFS, or double counting tons when rebilled in the Waybill.

The Waybill data were first adjusted for CFS scope issues including out-of-scope commodities, non-rail modes, and imports from point of entry to first point of processing. Crude Petroleum and Natural Gas [Standard Transportation Commodity Code (STCC) 13], Ordnance or Accessories (STCC 19), Printed Matter (STCC 27), Empty Containers (STCC 42), and Mail or Express Traffic (STCC 43) were all identified as out-of-scope for the CFS and eliminated from the Waybill. Freight Forwarder Traffic (STCC 44), which is primarily parcel shipments, was also eliminated, since this is considered a separate mode by CFS. Four separate categories of import traffic were also identified and eliminated from the Waybill. These included traffic originating in

Canada or Mexico, traffic listed as originating at Mexican border crossing points, intermodal traffic originating at U.S. ports, and finished automobiles originating at U.S. ports. After making these adjustments, the resulting estimate of railroad industry traffic should be comparable to rail ton-miles in Table 2a of CFS.

Estimates were then developed for rail–truck, rail–water, and rail-only traffic. Rail–truck traffic was defined as any remaining railroad trailer or container traffic, plus any remaining railroad finished automobile traffic. An originating and terminating truck mileage was applied to the tonnage of this traffic, and added to the rail ton-miles of this traffic, to develop an estimate of rail-truck ton-miles. Rail-water traffic was defined as any railroad coal, grain, or metallic ore traffic originating or terminating at transloading points. The average waterway length of haul for each of these commodities was applied to this traffic, and added to the rail ton-miles of this traffic, to develop an estimate of rail–water ton-miles. Any traffic remaining after the rail–truck and rail–water traffic was identified was considered to be rail-only. After making these adjustments, the resulting estimates should be comparable to the rail, rail–truck and rail–water ton-mile estimates in Table 1a of CFS.

RESULTS

The results of the analysis are summarized in Figures 1 through 4. The size of the graph in all four figures is proportional to the traffic involved.

Figure 1 illustrates the conversion from total railroad industry traffic as measured by the Waybill to in-scope railroad industry traffic as measured by CFS. The traffic eliminated from the Waybill amounts to only 13% of railroad industry traffic. Out-of-scope commodities and non-rail modes account for less than 2% of all railroad industry traffic. Import traffic accounted for almost 11% of railroad industry traffic, with most of this traffic coming not from overseas but from Canada. Comparing the in-scope railroad industry traffic in Figure 1 with the CFS in-scope

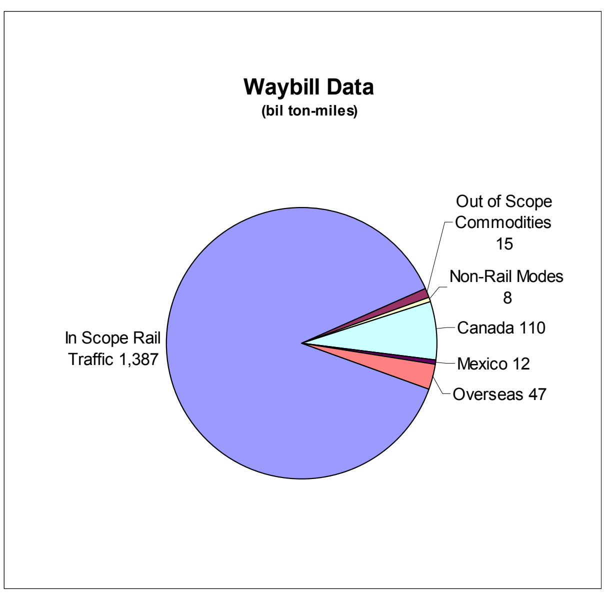


FIGURE 1 Railroad industry traffic.

railroad industry traffic in Figure 2 shows that the CFS underestimates total in-scope railroad industry traffic by only about 2%.

Figure 3 illustrates the division of Waybill traffic involving railroads into rail-only, rail–truck, and rail–water traffic. Figure 4 presents the same division of traffic based on CFS data. (The data in both figures include both rail ton-miles and estimates of truck and water ton-miles in multimodal shipments.) Comparing the traffic involving railroads in Figure 3 with the CFS traffic involving railroads in Figure 4 shows that the CFS underestimates traffic involving railroads by only about 3%. However, the composition of the traffic varies greatly between the Waybill and CFS data. Comparing the data in Figures 3 and 4 shows that the CFS underestimates rail-truck traffic by 80%, and rail-water traffic by 19%. The CFS therefore overestimates rail-only traffic by 15%.

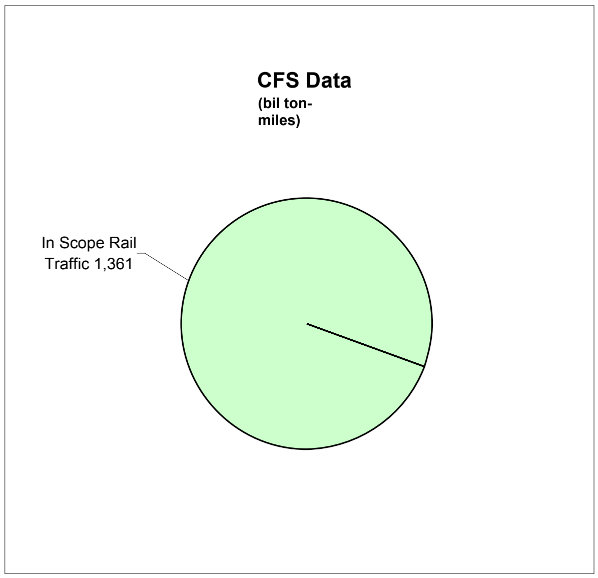


FIGURE 2 Railroad industry traffic.

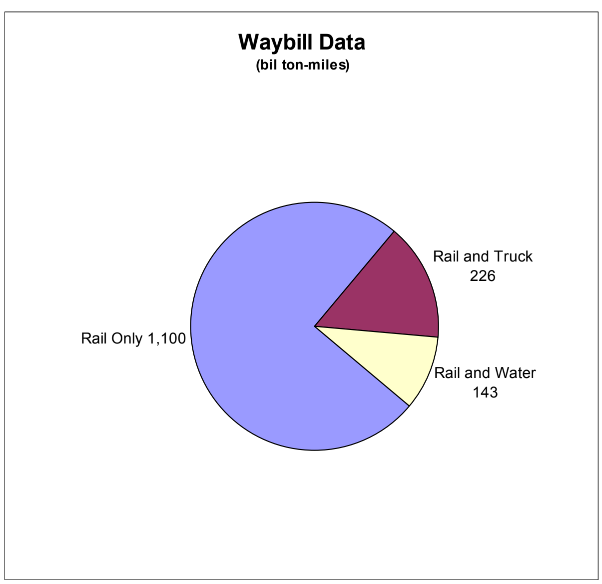


FIGURE 3 Traffic involving railroads.

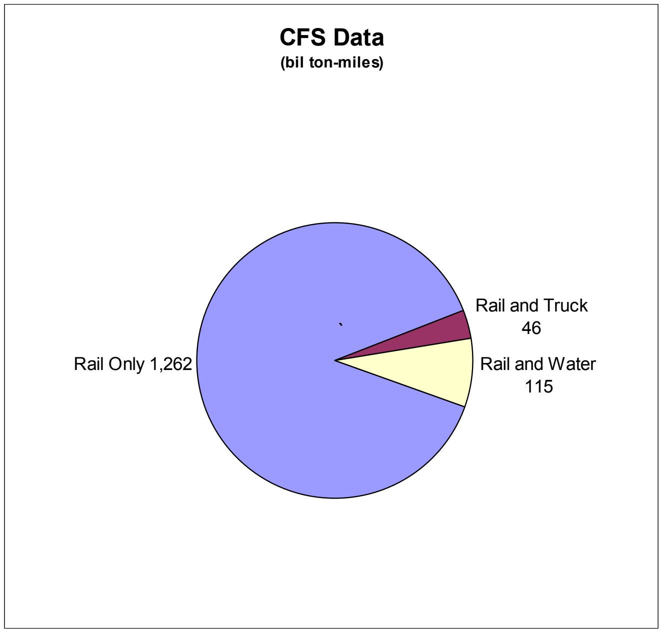


FIGURE 4 Traffic involving railroads.

IMPLICATIONS

Comparison of the CFS estimates to the benchmark Waybill values illustrates the magnitude of in-scope issues, and identifies the differences in the railroad industry, rail–truck, rail–water, and rail-only estimates in the most recent data.

The difference between the CFS estimates and the benchmark Waybill values for railroad industry traffic are due primarily to imports, not out-of-scope commodities or non-rail modes. Research for including more rail traffic in the CFS might best be focused on traffic from Canada, which accounts for the majority of the imports.

The CFS underestimates both rail–truck and rail–water traffic, and consequently overestimates rail-only traffic, compared to the Waybill benchmark. The underestimate of rail-truck traffic is very substantial. The volume of rail–truck traffic, along with its rapid growth, suggests that better identifying this traffic in the CFS should be an important research priority.

While this analysis does not benchmark railroad traffic by commodity, variations from benchmark values suggests potential modifications to the 2002 estimates, and possible improvement and refinements to the methodology and scope of the 2007 CFS.

Improving Estimation Process for Shipment Distances in the 2007 CFS

A Multimodal Approach and Integrated Interface

STEVE LEWIS
Bureau of Transportation Statistics

FELIX AMMAH-TAGOE
MacroSys Research

The Bureau of Transportation Statistics (BTS), in partnership with the Census Bureau, conducts the Commodity Flow Survey (CFS) every 5 years. As a shipper survey, the CFS collects a wealth of information on commodity flows, including mode of transportation, commodity types, value, and weight of goods. However, the survey does not collect information on the detailed actual routes used by carriers or the total miles traveled by the shipments. For each shipment, the survey respondents provide zip code of origin, zip code of destination, and mode sequence (e.g. truck–rail–truck). BTS turns this information into a multimodal route for each shipment in the survey and estimates the total mileages by mode of transportation.

For each of the past CFSs, this task was performed using routing models developed in FORTRAN and FoxPro by the Oak Ridge National Laboratory (ORNL). The logic that went into these routines is very good, but the implementation of the logic is now technologically outdated. For the 2007 CFS, BTS is now partnering with ORNL, ESRI, and MacroSys to translate this logic to a current generation programming platform within a geographic information system-based application environment.

This presentation covered our work plan and progress to date. It described the plans to improve the mileage estimation process and develop a new solver application tool that combines data preprocessing, mileage solver, and data post-processing in an integrated programming platform. It highlighted some of the system requirements involved in developing a database layer, an analytical layer, and a visualization layer. Also, the presentation described the combined multimodal network to be used in estimating distances for a given modal sequence and intermodal transfer.

Why Regional and State Data Are Not Enough
Potential Solutions

KAREN WHITE
Federal Highway Administration

INTRODUCTION

The Commodity Flow Survey (CFS) remains the first, best, and, frankly only, gateway to understanding freight movements in the United States. Without the creation of the CFS analysts were forced to rely upon anecdotal evidence, expert opinion and suspect data sets. The purpose of this paper is not to diminish the tremendous accomplishment of the CFS. Without the touchstone of the CFS many analysts (myself included) would just be poking in the dark and reading tea leaves.

In the CFS's current manifestation the U.S. Department of Transportation (DOT) participates as both a value added producer, creating the Freight Analytical Framework (FAF) and Geofreight; and a user, relying upon the CFS, FAF, and Geofreight as primary sources for analytical reports and recommendations to political decision makers. This paper will discuss the DOT uses of the CFS (and related products), deficiencies in the current CFS release formats and products, and solutions to improve the CFS products.

DEPARTMENT OF TRANSPORTATION'S USES OF THE CFS

The DOT has recognized the importance of freight data by investing in the CFS, creation of the FAF and GeoFreight and the recent creation of an internal Freight Data Working Group. The group is examining the freight data required to support the kinds of analysis being done by the various modes. The following summarizes recent and projected future transportation analysis using the CFS and related databases.

Western Uniformity Analysis

The Western Uniformity Analysis was a follow-on study to the Comprehensive Truck Size and Weight Study. The Western Governor's Association requested DOT examine a scenario where the states west of the Mississippi could join into agreements to allow larger and heavier trucks on expanded roadways (all the states in the study currently allow trucks larger or heavier than the federal limits). This analysis began by establishing the network that various trucks can travel and then measured the change to shipper's choices with expanded networks and/or more truck configuration alternatives. This study relied upon the CFS directly for commodity values and indirectly through the FAF detailed originations, destinations, and routing.

Positive Train Control (PTC)

The FRAundertook this analysis to determine the amount of traffic that rail intermodal could capture from highway if rail provided customers improved service resulting in lower transportation costs. The CFS and the county-to-county truck flows created in FAF were crucial in constructing this analysis. As a result of the PTC analysis, FRA has developed a data set that estimates the drayage from each county centroid to the nearest intermodal facility.

Interstate 81 Analysis

The Office of the Secretary (OST) undertook this analysis to estimate the reduction in truck through traffic and potential parallel diversion if I-81 in Virginia implemented an exclusive truck lane with a toll of nearly 37 cents per mile. The analysis expanded to analyze if shippers would abandon suppliers that are tied to the I-81 corridor while shifting to other sources that could demonstrate lower transportation and other logistics costs.

NCHRP 20-59(9) Guidelines to Maximize Transportation and Economic Resilience to Terrorist Attacks or Other Catastrophic Events

Although still in the planning stages it is reasonable to hypothesis that to estimate the economic impact on truck shipments, it will be necessary for the shipments to be flowed on the highway infrastructure. The analysis will need to connect each shipment to counties of origin and destination to determine the degree to which traffic can route around a catastrophic event. For example when the Oklahoma I-40 Bridge collapsed in 2002, FHWA was able to use the county-to-county FAF to estimate the impact on trucks that typically use that bridge on I-40. The analysis was able to re-route traffic and measure the total impact to the nation’s trucking. The re-routing of truck traffic around a catastrophic event requires a database with the origination and destination of each shipment moving along the impacted transportation center.

Strategic Multimodal Analysis

FHWA has initiated this project to create the tools to estimate the benefits and costs of alternative transportation investments. This analysis cuts across modes and investment alternatives to provide a holistic view of transportation investment within a region. The initial corridor analysis has begun and is developing the tools that rely upon county-to-county truck shipment data. The disaggregate shipment data allows estimation of truck rates for various configurations, tolling alternatives and estimation of alternative railroad or short-sea-shipping rates.

DEFICIENCIES OF CURRENT CFS PRODUCTS

All the above products and analyses require more geographic detail than the CFS provides and the commodity flows attached to the transportation network. The current CFS release is akin to knowing how much water flows through a state but masking which river the flow travels upon before the data is released.

The most successful product to arise from the CFS is the FAF. The wisdom of the FAF was not the creation of a large data matrix but the ability to translate that data into maps that combined freight data and highway networks to deliver a complete package for understanding relevant freight issues on capacity, network connectivity and mode choice.

The maps and discussion below highlight the difficulties mapping the CFS data and the difficulties mapping current truck shipments to the rail network as required for an examination of truck-to-rail diversion.

Routing Between Two "Rest-of-State" Zones

Figure 1 shows that routing on the Interstate System (red lines) is difficult using only the CFS regions and areas outside the regions. A shipment originating in Texas with a destination in Colorado could not be mapped to the Interstate system without further specification for both the origin and destination state.

Mapping Truck Shipments to the Railroad Network

Figure 2 shows the intermodal yards with red dots and the CFS zones with green outlines. A policy analyst confronts significant difficulty estimating the intermodal drayage of truck shipments located only by the state of origin and destination. DOT is increasingly looking to the CFS data to answer these and other rail–truck-related questions.

FIGURE 1 Interstate highway network and the 114 CFS zones.

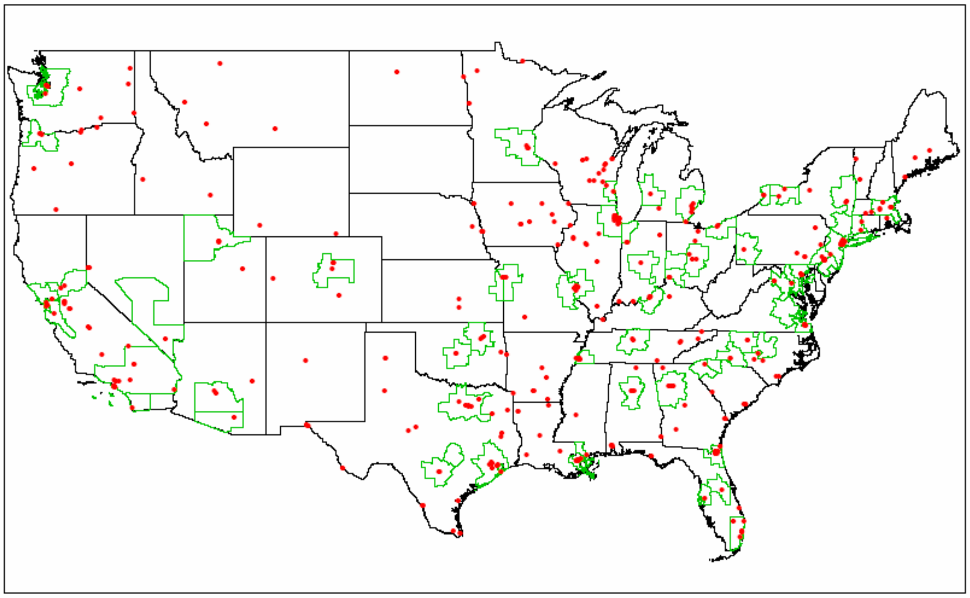


FIGURE 2 Intermodal facilities and the 114 CFS zones.

SOLUTIONS AND IMPROVEMENTS

This paper proposes several data improvements to the CFS products to allow greater functionality and over-all Federal cost savings. Realizing that data needs are great but resources are limited these options are discussed as being relatively low cost or high cost. The low cost options include

1. Release of the CFS network flow data;
2. Linking census and other databases at the confidential level to maximize the greatest leverage for detailed publicly releasable data creation;
3. Strengthening the public, private, and local resources to expand data collection; and
4. Creation of an authority to utilize the raw CFS data for federal analysis with agreements not to release private level data (similar to Confidential Waybill usage).

The higher cost options would include launching an on-the-road survey similar to the Canadian National Roadside Survey or creating a public–private partnership to collect and control the data similar to the Surface Transportation Board.

From both needs and budgetary criterions the primary improvement to the CFS suite of products would be for the Census Bureau to release the mapped commodity flows on the road, rail and waterway networks. Transportation analysis requires having the commodity movements attached to the infrastructure. By focusing on releasing only the geographic zones, the Census is missing a potential opportunity to create useful products from the CFS. Currently the placement of the CFS data onto the road, rail and water networks is done using the summarized public-release results by the FHWA Office of Operations. This duplication of work effort causes the loss of valuable information and a double spending of limited federal dollars.

As part of the Census Bureau's analysis and review process, the CFS shipments are routed on the road network. That flowed-CFS is used to create the ton-miles, commodity shipment distance and estimate the truck drayage to rail intermodal facilities. Currently the road, water and rail networks are not released. This paper proposes that there could be a level of release incorporating the networks. The network release would have mapped flows, similar to the FHWA's Freight FAF maps and geographic information system flows, similar to the Bureau of Transportation Statistics' GeoFreight. Although the FAF currently provides an estimation of the shipments flowed on the road, water and rail networks, it has relied upon inferior summarized CFS shipment data.

Workshop Summary

WORKSHOP SUMMARY

Improving the Use and Accessibility of the 2002 CFS

LISA AULTMAN-HALL
University of Connecticut

SCOTT DRUMM
Port of Portland, Oregon

The primary focus of the 2005 Commodity Flow Survey (CFS) Conference was to discuss the future and to contemplate improvements in both the CFS survey and the resultant data products. It has been suggested that we may not be making the most of what we currently have. The diverse users of the CFS data include planners, economists, engineers and policy makers. This group interacts with various statisticians who design the survey and its products. This diverse set of stakeholders often presents a challenge to CFS-related discussions. Individual perspectives are naturally skewed by experience, training, and uses for the data. By contrast, the diverse participants at this particular session of the 2005 CFS conference were tapped as resources to identify ways in which the pool of CFS data users could be expanded.

The authors hypothesize that there is a critical need to increase both the number of CFS data users and the number of data applications. This could be a key to ensuring long-term local, state, and federal support to continue collecting the data and its future derivatives. While the conference focused primarily on understanding the current data users, this session requested that participants stop for a moment and consider the possible future users. Why don't these potential additional users already use the CFS? Perhaps they do not know about the data, lack an understanding of how to use it, or lack the tools required or the resources to alter it for their needs. If we are to increase the number of users, it is essential that we pursue this marketing exercise and gain this insight.

Although it is an important endeavor to debate the changes needed in CFS in 2007 and beyond, the CFS 2002 is what we currently have to work with. Little is accomplished in discussing how it could have been better. It is a unique survey that includes origins and destinations (O-D) and other information regarding a subset of the nation's commodity shipments. While acknowledging its limitations, including a significantly reduced sample size over 1997, we must address its optimal use.

The window of opportunity for doing so is open now, but it will not be for long. The 2002 data set has just been released. Further, as a result of population numbers from the 2000 Census, newly created metropolitan planning organizations (MPOs) are in need of, but may not be aware of, CFS data. At the same time as we are preparing for CFS 2007, it is the right time to market the data to the private sector. If we are able to understand how they can benefit from the data and then provide the data to them, we are likely to get higher response rates to and better data quality from the 2007 surveys. Finally, there has been substantial of talk about the needs that the Transportation Security Administration may have for this kind of freight data and this conversation needs to be pursued.

The more we can do with 2002, the more improvements we can make beyond 2007. But to do this we must generate more data users and stakeholders. If we can elevate the profile of freight data among elected officials, the private sector, and the general public, the more support will be generated for the CFS program. This support will not only enable us to keep the CFS program going, it will also be influential in our success to improve data collection and dissemination in the future.

This document outlines a comprehensive range of possible endeavors that may increase the use and access to the CFS 2002 data. The suggestions have not been prioritized or evaluated based on their feasibility. Instead, these five categories contain all examples and suggestions that were raised by the session participants and presented by the authors.

Alternative Data Formats

Although the CFS data tabulations have been completed and released for use, reformatting or re-packaging some of these data into a format that could be more readily used by agencies and analysts provides a possible enhancement and incentive. This is different from a re-tabulation that might require the Census Bureau to conduct additional analysis for either confidentiality or statistical validity. For example, a text format O-D matrix, where every row consisted of an O-D pair and each commodity interchange was placed in a column, would save analysts significant time in reformatting and combining the existing tables.

Alternatively, some small re-tabulations might serve the users and potential users. A total freight O-D matrix by three-digit zip code would presumably be releasable and of benefit to many users. If select analysts were allowed access to the micro data, they could develop disaggregate freight generation models (trip rates) which could be provided publicly. Or the Freight Analysis Framework, which is being updated with the CFS 2002 data, can be adjusted to be more valuable to users.

Awareness of the Data

Many potential users may not know that the CFS data exists or how to use it. To this end, a Best Practices Guidebook could be created along with a Data User's Manual. The current website could be augmented to include examples of data fusion as well as "what can I do with CFS" examples.

More long-term awareness of the CFS might be possible through dissemination of ready-made freight-related lecture material for professors to include in transportation planning courses. Homework exercises that use the CFS data, complete with solutions, could be provided.

Training options, including expansion of the National Highway Institute courses might be possible. Materials could make direct use of the CFS data.

Access to the Data

The richness of the CFS 2002 data is not evident in its current public format. This might be achieved if planners could obtain more access to the micro data. This message was clear from the poster presentations at the CFS conference. This access might be accomplished by requesters altering the justification or scope of the research objectives needed for access to the micro data.

If access cannot be made available to all, then some analysts could be given access with the explicit objective to create formats that are more useful for planners, such as regression-based trip generation rates or mode choice models. A freight version of the "Census planning package" could be considered.

Enlarge the Data—Data Fusion

Even though the 2002 data set is particularly small, it might be possible to combine it with older CFS data sets or private data sets which would allow more detailed information, especially geographically, to be available. Many users do conduct data fusion, and perhaps it would make sense for this to be done once and disseminated widely. This would avoid users effectively "re-inventing" fused data sets, which may only vary slightly.

Provide Incentives

Providing incentives was proposed as a way to encourage more creative and extended use of the CFS 2002. While funding research that uses the CFS 2002 is the most obvious and direct method to do this, other options were presented. Funding graduate fellowships would be less expensive but may provide new applications of the data. Awards for use of the data, perhaps by Transportation Research Board (TRB) committees, would draw attention to the data and perhaps provide an incentive. The awards could be stratified by local/MPO use, state, national or international use. TRB committees or others could also submit research problem statements involving the CFS data for consideration for funding through the NCHRP program.

Workshop Summary

Observations on Improving the 2007 CFS

Key Observations from the Conference and Beyond

PAUL BINGHAM
Global Insight, Inc.

OBSERVATIONS

Many possible improvements were identified for the Commodity Flow Survey (CFS) at this conference. These considerations are presented in the context of four factors discussed by participants:

1. Budget and time constraints are important considerations in addressing improvement opportunities.
2. Identification of "low-hanging fruit" improvements and highest payoff improvement investments is important.
3. Desired outputs are defined from intended uses and scope and are a paramount consideration for possible design changes.
4. The shipper focus of the CFS is fundamental within a context of other data sources and intended uses.

Within this context, key points from the conference sessions are documented.

Scope

An expansion of the number of establishments to increase the results that can be published, thus trying to minimize underreporting within scope (e.g., intermodal) is important.

Compatibility

Improvements to the design of the CFS were considered more important than maintaining consistency across years just for the sake of consistency. Assistance can be provide to data users through published analysis, documentation, and bridges between changed dimensions to aid users in coping with any changes.

Shipment Characteristics

Mnay participants noted that the focus of the CFS should be on what is realistically obtainable from shippers in a survey. Shippers should not be expected to provide information that they do not have or are not in a position to know. Also, quality of the basic characteristics is considered more important than obtaining secondary characteristics.

Special Measurement Issues

Improvements to capture growth in intermodal and third parties are important. This could possibly be addressed in the CFS mode questions and in other data efforts outside the CFS that can be integrated or used in context with the CFS. The first leg of import shipments and in-transit shipments are better addressed outside CFS because of its inherent limitations as a U.S. shipper-based survey.

CFS Products

Improvements to the CFS products can be obtained through the addition of access to microdata or special measurement tabs, where possible. The original planning of survey collection and editing could be done to support planned products. Access to the CFS can be improved through value-added efforts that occur outside the Bureau of Transportation Statistics (BTS) and Census.

Survey Methods

Survey methods can be improved by the adoption of an optional web response capability for respondents. This could also improve data edit capabilities and may reduce response burden on shippers while improving response rates. The use of cognitive interviews, pre-canvassing, and pilot testing are tools that could be used to maximize data collection for the available budget.

CONCLUDING STATEMENTS

Federal, state, and local agencies all have a role in the larger community of users of freight data which in turn has a vested interest in the production, application, and use of CFS.

The CFS, by itself, is necessary but not sufficient for a national freight information program. It is one part of the larger national freight data picture. Previous Transportation Research Board Special Reports have detailed this larger need and a framework for improvement, but so far, little action has been taken. Acknowledgment that the CFS is a necessary foundation for decision making by the policy and planning community is important as we look beyond the CFS for other efforts that can fill in the gaps. Significant, difficult work on this, in addition to the effort that is required to prepare and conduct the CFS in 2007, is required. The BTS and Census cannot produce the larger system of freight data by themselves. A higher level of support is needed to gain attention to future freight data programs.

Freight Database for the Future

JOSEPH L. SCHOFER
Northwestern University

TIME TO DEFINE THE NEXT NATIONAL FREIGHT DATABASE

There is ample evidence that transportation researchers and policy makers have developed a love-hate relationship with the Bureau of Transportation Statistics (BTS) and the Census's Commodity Flow Survey (CFS). This survey provides the core database that many use directly and others amplify with additional data sources and analyses to support research and decision-making. At the same time the CFS is limited because of narrow coverage of industries, lack of information on transportation service characteristics, failure to capture the details of intermodal shipments, and most recently, because of its small sample size.

In response to these deficiencies, and in the context of growing demand for freight data in support of policy making, a cottage industry has developed around the CFS to make the original data set more comprehensive, more useful, and more accessible. Some of this work is produced as proprietary products, and other data products come from individual and institutional power users of CFS.

A review of the several surveys conducted by the U.S. Department of Transportation (USDOT) BTS (*1*) characterized many of the deficiencies of the CFS, but recommended

> In view of the widespread use of CFS data for a diversity of applications, BTS should continue to provide data on the flow of goods by mode of transport within the United States. These data should be updated at intervals of no more than 5 years. To ensure that ongoing user needs are met, the CFS should be continued—with some modifications—at least until such time as a viable alternative source of national freight data has been established

The key message in this recommendation is in the last line: "until such time as a viable alternative source of national freight data has been established."

The next version of the CFS is to be implemented in 2007 as a part of the economic census, and many of its characteristics are locked in. Now is the time to think seriously about designing the freight database of the future, so that, by the time of the 2012 Economic Census, we are prepared to collect a commodity flow data set that is truly better and more responsive to user needs.

ROAD TO A BETTER FREIGHT DATABASE

Why Do We Need to Replace the CFS?

The CFS has significant limitations, in industry coverage, sample size (affecting both geographic and industry detail), and information about transportation services. The sampling frame itself—provided by the U.S. Census Bureau—has weaknesses that substantially reduced the responses in 2002 because many sampled firms were either closed or not in scope. Many respondents, all of whom were originating shippers, could not provide detailed and accurate data on transportation modes used.

The future Freight Database (FDB) must respond to critical economic trends including globalization (CFS does not capture imports) and intermodalism—much freight is intermodal, but often the shipper does not know this, at least in detail. Understanding intermodal freight movements is important not only because they are increasingly common, but because

- The private and public infrastructure requirements are substantial and different from historical experience;
- The potential efficiency gains from intermodal shipping are large, and thus structural changes in patterns can be expected to continue;
- Growing intermodalism affects the distribution of flows (and thus capacity requirements and impacts) on the different modes and links; and
- It is so difficult to capture data on the intermodal patterns of shipments.

Link flows and vehicle characteristics, neither of which is included in the CFS, are important factors in decisions about both private and public infrastructure investments. Security concerns have become primary issues, warranting closer tracking of critical and dangerous commodities across transportation networks.

The structure of the industry is also changing, which affects not only how freight is moved, but also who knows how it is moved. Intermodalism is a large part of that change, and the pressure for efficiency has increased specialization, bringing third party logistics firms (3PLs) into the market that take responsibility for freight shipments, determine mode and route patterns, and possess information about freight flows that is often invisible to the shippers who are now the targets of the CFS.

What Do We Need?

It is time to design a new product that avoids the major limitations of the CFS. This will not be an upgraded CFS, but a substantially different survey, perhaps one that gathers data from several different source types—e.g., shippers, carriers, links and terminals. The knowledge gained from the CFS is a starting point, but the CFS itself is probably not.

How Can We Develop the Next FDB?

Developing a new and more responsive FDB will be a complex and difficult task, but the experience of CFS provides a foundation, and the bodies of both interest and knowledge are large and growing (*2*). A clean slate approach driven by user needs and the changing context is likely to produce the most useful and flexible product. Beginning to prepare for a survey beyond 2007 now provides a time frame sufficient to consider substantial changes, new methods, and advancing technology. Importantly, it provides an opportunity to build broad and stable support for the next generation FDB.

Who Should Be Engaged in the Effort?

The constituency for a new FDB is large, and the key resources—the data themselves, the tools, the sampling frame, and money—are held by a variety of entities. The demand for freight data comes from private industry, consultants, researchers, and governments at the metropolitan, state, and federal levels. Each of these organizations has a stake in the content and quality of the next FDB, and each should play a role in planning, organizing, and supporting the collection, preparation, and dissemination of the data. Particularly because of the intense competition for funds, producing the future FDB is necessarily a community effort.

NOT JUST YOUR FATHER'S CFS

Preparing the next generation FDB will not be a simple matter of collecting new data. There is a need to rethink the content (e.g., to include service quality, network, modal trajectory, price, and other data) based on evolving user needs and an understanding of the process that generates both freight and freight vehicle flows.

While it is apparent that the CFS supports a variety of management, marketing, and infrastructure decisions, the fact that there is no major federal funding program for freight infrastructure means that there is no simple and centrally defined set of data requirements in the freight field. Thus it will be necessary to identify and synthesize data needs systematically. For example, supporting capital investment decisions in either private or public sectors requires the ability to predict freight movement demand by corridor or link, as well as the ability to answer if-then questions about alternative investments. Models to do this will need current freight flows, network and service characteristics, as well as economic, scale, and location characteristics of shippers and receivers.

Sample size and sampling methods should also be driven by user needs. The sample size needed to estimate flows by commodity, origin-destination (O-D) by small area, and mode will be considerably larger than the 50,000 target (31,000 actual) collected in 2002. Data collection will likely need to be multidimensional, as no single entity—e.g., the shipper—is likely to know all of the key information about a particular shipment.

The freight generation process is complex and multidimensional; the data collection effort for the future FDB will need to match these characteristics. Advanced freight modeling will require information about shipper (and probably receiver), mode and route service, price and options, and special shipment requirements (size and commodity type, hazardous and sensitive materials, transit time and reliability requirements, etc.). This will almost certainly require collecting data from multiple sources, advanced and creative data fusion, or drilling down into the detail of a sample of shipments, perhaps even to the level of tracking goods through the supply chain.

Information technology and advanced shipment tracking schemes (e.g., radio frequency idenfitication tags) may provide substantial advantages in the future. An increasing number of shippers and carriers rely on these fast-evolving technologies, and there may be opportunities to use their data in assembling the next FDB. Gathering data already in electronic form from selected firms and industries may be a way to boost sample sizes substantially and with great efficiency. Other countries (notably Sweden) are doing this now. Of course there is a need to balance confidentially with data quantity and detail, and thus it is essential for businesses to be closely involved with the development of plans for the next FDB.

The next FDB is far enough into the future to permit radical changes in strategy. Serious consideration should be given to a continuous data collection process that might bring several advantages, including smoothing cost patterns, providing more timely data to track trends and allowing gradual and continuing introduction of new survey methods and technologies.

Finally, closer attention should be paid to making the data more accessible, through easy online access, a simple, web-based analysis engine, and formats that respond to common user needs. This includes providing O-D flows (at least at the metropolitan area level) by commodity type, as well as tonnage and vehicle flows on major national links.

WRITING THE SPECIFICATIONS FOR THE NEXT FDB

Much work has been done to define the requirements for the next FDB, but much more is yet to be accomplished to prepare for post 2007. Given the complexity, the multiplicity of interests and actors, and the competition for resources, it is not too soon to organize this effort. The nexus of this effort must bring together the parties, provide both perspective and continuity, and offer a neutral ground for discussion, debate and consensus.

The Transportation Research Board (TRB) is the logical center for this activity. TRB's long history of engagement in freight planning and thinking about freight data needs provides a foundation for moving to the next FDB. As a meeting ground of information providers and consumers, and public and private entities, TRB is unique and essential to this effort.

TRB should host the development of specifications for the next FDB through its committee structure and conferencing capabilities. Of course it needs resources to accomplish this mission, but their magnitude is not large relative to the needs for the data and the decisions to be supported by the next FDB. There is a strong federal interest in this effort based on government responsibilities for the national economy and infrastructure. There is a clear private interest, as well, and resource support for this effort should be broadly shared.

The consultants and others that make up the cottage industry that creates substantial value added from the CFS have a strong and positive interest in the new database as well. The next FDB will not compete with their current efforts, but will raise the base from which they can create even more important value-added products for the user community. Thus, these value-added providers also need to be a part of the design of the next FDB.

The missions is not only to develop specifications and methods for the next FDB, but to discuss the needs, the plan, and the methods broadly among industry associations and government to build support for implementation as a part of the 2012 economic census. Plans need to be shared with executive and congressional leadership, as well as state, local and industry leaders, to build the coalition necessary to assemble the resources well in advance to deliver a FDB that meets future needs.

CLOSURE

Now is the time to begin to create the next generation freight database. The needs are important and clear; the constituency has emerged and to a substantial degree is working together. The new FDB will be important in planning, management and decision making for industry, federal, state and metropolitan level governments. It will play an important role in supporting the national economy, managing infrastructure, ensuring security, and assuring the ability to respond to disasters. It is an important and worthy investment.

REFERENCES

1. Committee to Review the Bureau of Transportation Statistics' Survey Programs. *Special Report 277: Measuring Personal Travel and Goods Movement.* A Review of the Bureau of Transportation Statistics' Surveys, Committee on National Statistics, Transportation Research Board of the National Academies, Washington, D.C., 2003.
2. Committee on Freight Transportation Data: A Framework for Development. *Special Report 276*: *A Concept for a National Freight Data Program*, Transportation Research Board of the National Academies, Washington, D.C., 2003. http://gulliver.trb.org/news//blurb_detail.asp?id=1730. FHWA's Freight Analysis Framework, http://ops.fhwa.dot.gov/freight/about.htm. Committee on the Future of the FHWA's Freight Analysis Framework, letter report, http://gulliver.trb.org/publications/reports/fafltrfeb2004.pdf, February 9, 2004.

CFS Strategic Remarks

C. MICHAEL WALTON
University of Texas at Austin

Many excellent ideas and issues were offered for enhancing the 2007 version of the Commodity Flow Survey (CFS) which build on the 2002 edition. Several creative and practical ways to leverage the data and information have been discussed. A variety of observations were offered for more immediate short run actions while others are more strategic.

SHORT-TERM TACTICS

First, a need exists to create a strategic business plan for freight data. Such an effort begins with a synopsis of the most recent reports and studies. Perhaps this might be an effort of interest to NCHRP or the recently created Freight Cooperative Research Program. As a minimum, it should contain the following components: vision of the program, possible business models, likely partnerships, technology impact (e.g., intelligent transporation system), impact of global dynamics, trends, SWOT analysis, and others.

Second, leadership and champions need to be identified, developed and energized. For example, a Goods Movement Caucus exists in the U.S. Congress House of Representatives which could provide a forum for more discussion and action on the importance of freight data and the CFS.

Third, promote the idea of an international scan on freight data and freight data surveys. The international scan program is a well established, ongoing, and supported by the FHWA, NCHRP, and others. The action item is for the FHWA and state department of transportation (DOT) staff to submit a request through the appropriate channels promoting such an initiative.

Fourth, establish training programs on the use of the CFS and other freight data. The training program could be tailored after National Highway Institute courses showing best practices, uses and applications, linkages, etc.

Fifth, consider promoting the establishment of an “IDEA” research program for the promotion of the use of freight data similar to other “IDEA” programs (e.g., Long-Term Pavement Performance).

STRATEGIC OR LONGER RUN

First, continue to build an external constituency and champions for freight and freight data needs. As an example, consider the notion of a National Freight Advisory Council at the USDOT Secretary level.

Second, there is a need to energize and leverage the stakeholders of the freight data survey activity. Initiatives could be targeted for specific stakeholders and market segments.

Third, the importance of the utility of accurate, viable freight data is not well recognized. There is a need to build the case and articulate its value as fundamental to national interest.

CLOSURE

The stars are reasonably aligned for promoting all aspects of freight and freight data. Such major initiatives or programs, to mention a few, include: Air 21, SAFETEA-LU, Homeland Security, Global Competitiveness, and perhaps a future SEA-21.

Among the priority of actions is the importance of dynamic leadership and emerging champions for this critical element of the U.S. transport enterprise.

Appendixes

APPENDIX A

List of Abbreviations and Acronyms

3PL	Third-party logistics providers
AADT	Average annual daily traffic
AADTT	Average annual daily truck traffic
AASHTO	American Association of State Highway and Transportation Officials
ABI	Automated Broker Interface
ACE	Automated Commercial Environment
ACS	Automated Commercial System
AERP	Automated Export Reporting Program
ASM	Annual Survey of Manufacturers
BEA	Bureau of Economic Analysis
BTS	Bureau of Transportation Statistics
CAFTA	Central America Free Trade Agreement
CD	Compact Disk
CFS	Commodity Flow Survey
CM	Census of Manufacturing
CMB	constrained matrix balancing
COFC	Container on Flat Car
COT	Census of Transportation
CTS	Commodity Transportation Survey
CVISN	Commercial Vehicle Information Systems and Networks
DOC	Department of Commerce
DOT	Department of Transportation
EIA	Energy Information Administration
EU	European Union
FAA	Federal Aviation Administration
FAF	Freight Analysis Framework
FDB	Freight Database
FHWA	Federal Highway Administration
FMCSA	Federal Motor Carrier Safety Administration
FRA	Federal Railway Administration
GPEA	Government Paperwork Elimination Act
GPS	Global Positioning System
HHI	Herfindahl-Hirschman Index
HOV	High Occupancy Vehicle
HPMS	Highway Performance Monitoring System
IIM	Inoperability Input-Output Model
I-O	Input - output
IPF	iterative proportional fitting
ITDS	International Trade Data System
ITS	Intelligent Transportation System

JIT	Joint Investigative Team
JIT	just in time
LIDAR	light detection and ranging
MA	Metropolitan area
MARAD	Maritime Administration
MPO	metropolitan planning organization
MRT	metric revenue tons
MSAE	Mean Sum of Absolute Errors
MYSTIC	Methodological Framework for Modeling European Passenger and Freight Transport
NA	North America
NAICS	North American Industrial Classification System
NAFTA	North America Free Trade Agreement
n.e.c.	not elsewhere classified
NEMS	National Energy Modeling System
NHI	National Highway Institute
NHTSA	National Highway Transportation Safety Administration
NTARS	National Transportation Analysis Regions
O-D	Origin–destination
ODCM	Origin–Destination Commodity Mode
OM	Origin of Movement
OST	Office of the Secretary of Transportation
PIERS	Port Import Export Reporting Service
PTC	Positive Train Control
RFID	Radio Frequency Identification Devices
RIC	Remote Intelligent Communication
RITA	Research and Innovative Technology Administration
RPC	Regional Purchase Coefficient
SAL	Smart Active Labels
SCTG	Standard Classification of Transported Goods
SIC	Standard Industrial Classification
SITC	Standard International Trade Classification
SSEL	Standard Statistical Establishment List
STB	Surface Transportation Board
STCC	Standard Transportation Commodity Codes
TAZ	Transportation Analysis Zone
TDOT	Tennessee Department of Transportation
TIUS	Truck Inventory and Use Survey
TOFC	Trailer on Flat Car
TRB	Transportation Research Board
TSFD	Transborder Surface Freight Data
VDOT	Virginia Department of Transportation
UAV/MAV	uninhabited autonomous vehicles/micro aerial vehicles
UEM	universal electronic manifest
UN	United Nations
USACE	United States Army Corps of Engineers

USDA	United States Department of Agriculture
VIUS	Vehicle Inventory and Use Survey
WIM	Weigh-in-Motion

APPENDIX B

Conference Sessions and Events

FRIDAY, JULY 8, 2005

What the CFS Is Designed to Do
8:30 a.m.–9:15 a.m.
Arnim Meyburg, Cornell University, presiding

> **Welcome and Conference Objectives**
> *Arnim Meyburg, Cornell University*
> **How CFS Fits in the World of Freight Data**
> *Eric C. Peterson, Deputy Administrator, Research and Innovative Technology Administration (RITA), U.S. Department of Transportation*
> The RITA Deputy Administrator explains how the CFS fits with other data programs to provide a comprehensive picture of freight transportation and why this comprehensive picture is important for decision makers. The evolution of needs for the CFS since the 1960s and the efforts to meet those needs will be summarized; the legislative mandates for the CFS will be highlighted. The presentation concludes with a look at U.S. DOT's expectations for the next CFS and for related freight data programs.

User Applications Overview
9:15 a.m.–10:00 a.m.
Three experienced freight data users will synthesize three groups of CFS user applications. All the applications will be highlighted in the poster session that follows.
Rolf Schmitt, Federal Highway Administration (FHWA), presiding

> **National Transportation Applications**
> *T. Randall Curlee, Oak Ridge National Laboratory*
> **Applying the CFS to State and Local Transportation Issues**
> *Mark Berndt, Wilbur Smith Associates*
> **Uses of the CFS Beyond Transportation**
> *Michael S. Bronzini, George Mason University*

Applications Poster Session
10:30 a.m.–11:15 a.m.
See APPLICATIONS in the main report.

CFS 2007 Design
11:15 a.m.–12:00 noon
The session summarizes the planning for the 2007 CFS.
Frank Southworth, Oak Ridge National Laboratory, presiding

> **Bureau of Transportation Statistics Planning for the 2007 CFS**
> *Mary Hutzler, Bureau of Transportation Statistics (BTS)*
> **Census Bureau CFS 2007 Planning Issues**

Thomas L. Mesenbourg, Bureau of the Census
Summary Remarks and Introduction to the Workshops
Arnim Meyburg, Cornell University

Scope of the Commodity Flow Survey
1:00 p.m.–2:30 p.m.
The scope of the CFS determines the industry sectors to be surveyed from the universe of domestic establishments that ship freight. The 2007 CFS will survey approximately 100,000 establishments from a sample of about 800,000. Participants will consider issues associated with obtaining comprehensive coverage from industry sectors within the sample, ensuring that CFS data are accurate and complete.
Benjamin J. Ritchey, Battelle, facilitator
Catherine T. Lawson, State University of New York, Albany, recorder
Resource Paper
Ron Duych, BTS

CFS Data Comparability Across Years and with Other Data Sources
1:00 p.m.–2:30 p.m.
Many changes in the first three CFSs have affected users' ability to determine trends. In addition, several other data sources have supplemented information provided by the CFS. Participants will discuss potential changes to the 2007 CFS design, the possible impacts that these changes may have on trend analysis, and ways in which the data series can be preserved. Participants also will explore other supplementary freight data sources that provide a more complete picture of commodity flows and will consider issues of data comparability.
Michael Bronzini, George Mason University, facilitator
Jonette Kreideweis, Minnesota Department of Transportation, recorder
Resource Paper
Frank Southworth, Oak Ridge National Laboratory

Shipment Characteristics
3:00 p.m.–4:30 p.m.
Shipment characteristics of great interest to many analysts include shipping cost, time sensitivity of shipment, type of containerization, type of conveyance, piece count, size, empty shipment moves, treatment of multicommodity shipments, and value. Participants will discuss how this information could be used and whether the same or a modified list of characteristics should be collected in the 2007 CFS.
Arnim Meyburg, Cornell University, facilitator
Robert Costello, American Trucking Associations, Inc., recorder
Resource Paper
Bruce Lambert, U.S. Army Corps of Engineers

Special Measurement Issues: Intermodal Shipments and Third Parties
3:00 p.m.–4:30 p.m.
Intermodal shipments and shipments by third parties were difficult to measure from the 2002 CFS. Participants will review the magnitude of the misclassified or missed shipments, definitional problems in distinguishing a shipment by multiple modes versus multiple shipments

by single modes, ways to include third parties without double counting, and ways to measure containerization or other aspects of intermodal movements.

Paul Bingham, Global Insight, Inc., facilitator
Mark Lepofsky, Battelle, recorder
Resource Paper
Rolf Schmitt, FHWA

SATURDAY, JULY 9, 2005

Reports from Breakout Workshops
8:30 a.m.–10:00 a.m.
Arnim Meyburg, Cornell University, presiding

Scope of the Commodity Flow Survey
Benjamin J. Ritchey, Battelle
CFS Data Comparability Across Years and with Other Data Sources
Michael Bronzini, George Mason University
Shipment Characteristics
Arnim Meyburg, Cornell University
Special Measurement Issues: Intermodal Shipments and Third Parties
Paul Bingham, Global Insight, Inc.

Improving CFS Data Products
10:30 a.m.–12:00 noon
The CFS has a diverse customer base, from sophisticated analysts to inexperienced one-time users. Participants will discuss the development of data products and dissemination tools that meet the needs of the principal data users, including approaches for improving the utility and accessibility of CFS data products, customer services, and methods of dissemination for the 2007 survey.

Thomas Zabelsky, Bureau of the Census, facilitator
Ronald Tweedie, Consultant, recorder
Resource Paper
Nathan Erlbaum, New York State Department of Transportation
Jose Holguin-Veras, Rensselaer Polytechnic Institute

Exploring Survey Methods to Enhance CFS Data Quality and Usefulness—
10:30 a.m.–12:00 noon
Design of the 2007 CFS includes consideration of measurement issues, questionnaire design, sampling, and data collection methods—involving new technologies, data processing, and methods documentation. Participants will build on discussions from previous conference sessions and on the currently proposed 2007 CFS design to explore additional methodological changes that would improve and enhance the next CFS. In addition, this session will operate as a "think tank" by stimulating creative thinking about future methodological considerations in the collection of commodity flow data in the CFS.

Joy Sharp, BTS, facilitator
Bruce Dembroski, U.S. Census Bureau, recorder

Resource Paper
Johanna Zmud, NuStats Partners, LP

Reports from Products and Survey Methods Breakout Sessions
1:00 p.m.–1:45 p.m.
Arnim Meyburg, Cornell University, presiding
Improving CFS Data Products
Thomas Zabelsky, Bureau of the Census
Exploring Survey Methods to Enhance CFS Data Quality and Usefulness
Joy Sharp, BTS

Improving Use and Accessibility of 2002 CFS
1:45 p.m.–2:45 p.m.
Pat Hu, Oak Ridge National Laboratory, presiding
Two freight data experts synthesize suggestions from earlier sessions on improving the use and accessibility of 2002 CFS data and add their observations.
Observation Presentations
Scott Drumm, Port of Portland
Lisa Aultman-Hall, University of Connecticut

Panel Discussion: Key Observations on Improving 2007 CFS
3:15 p.m.–4:45 p.m.
Arnim Meyburg, Cornell University, presiding
This panel reviews the conference discussions and offers observations on the use of the 2002 CFS products and on potential improvements.
Panelists:
Paul Bingham, Global Insight, Inc.
Joseph L. Schofer, Northwestern University
C. Michael Walton, University of Texas

APPENDIX C

Workshop Discussion Questions

Scope of the Commodity Flow Survey

- What gaps in the scope and industry coverage of the CFS are the most necessary to address in the 2007 CFS?
- Are there certain CFS commodity groups where industry coverage is not comprehensive in the North American Industry Classification System (NAICS) coding scheme?
- How can the lack of comprehensive coverage in certain CFS commodity groups covered by the CFS be identified and addressed?
- When does it become feasible to accept less than complete industry coverage in a CFS Standard Classification of Transported Goods (SCTG) commodity group?
- Have changes in supply chain logistics affected the desirable scope of the CFS, especially in the role of third-party logistics (3PL) providers and auxiliaries?
- Are there any CFS commodity groups or industries that could be dropped or modified from the 2007 CFS?
- Are there any industries that have been traditionally excluded from the scope of the CFS that should be considered for inclusion in 2007?

CFS Data Comparability Across Years and with Other Data Sources

- What are the biggest CFS gaps that need to be filled in with other data sources?
- What other data sources can we use to fill these gaps?
  - How can we use these sources to fill in missing data cells, and add spatial and/or commodity detail?
  - How compatible are these datasets across commodities, regions, and modes of transportation?
  - What data modeling techniques are available for merging data sets?
  - Can we use these techniques to combine CFS data from the 1993, 1997, and 2002 surveys (to fill in missing data cells)?
- How consistent are the CFS results over the 1993, 1997, and 2002 surveys? (Every version of the CFS has changed—in geography, industry and commodity coding, modal detail, etc.)
  - How reliable is a CFS trend analysis?
- How important is this?
- What error level is acceptable?
  - Would a continuously sampled CFS help?
- Can a redesigned CFS capture more information?

Shipment Characteristics

- What shipment characteristics are most important to policy makers, planners, or power users (ranging from the uninformed to the informed)?
  - Are these characteristics linked to other economic databases to assess economic relationships?
  - What characteristics are reported most often in various documents (trade journals and publications, economic research, regulatory analysis, and planning documents)? Can the report be constructed to capture these linkages?
  - Are there terms relevant for use in the industries actually being surveyed?
  - Should these be included in the website and on the distribution CD?
- Generating results: can the information be presented in a meaningful manner to diverse audiences?
  - How is the average mile per shipment calibrated by mode and destination? Can this be linked to other research areas?
  - Which of the reported shipment characteristics are the most useful? Which are the least useful?
  - Does the reporting format present shipping characteristics in a relevant manner?
- Data inputs and coding: how are these calculated and coded? What are the review checks and balances?
  - What other shipment characteristics are reported, by mode, for other federal databases? What are the coverages or linkages between such sets?
  - Should the survey actually report shipments and not gross totals (the over forty rule)? Some assumptions are made when the form is completed and when it is coded.
  - Any statistical checks on Item G—the monthly value of shipments?
- Are respondents to the CFS likely to be able to answer questions about the missing shipment characteristics?

Special Measurement Issues: Intermodal Shipments and Third Parties

- How serious are the deficiencies in data covering third-parties (brokers, warehouses, and others) who arrange shipments in the CFS?
  - How important are these deficiencies?
  - How can the actions of these third-parties be captured?
- How do the representation of third-parties in the NAICS and the nature of shipping documents reduce or exacerbate the problems of double counting individual shipments in the CFS?
- What is the appropriate definition of an intermodal shipment? How serious are the deficiencies in data covering intermodal movement sin the CFS?
- How should containerized shipments be addressed? What do we need to know about containerized shipments?
- Are there efficient ways to improve the coverage of these areas in CFS for 2007?
- Is a change in scope or a completely separate survey needed?
- Can we get at intermodal shipments and third-parties by changing the way in which the questions are asked more effectively than by changing the CFS scope?

Improving CFS Data Products

- What are the most important changes that would improve the current CFS data products?
- How would you prioritize geographic detail, commodity detail, and industry detail in future CFS data products? What is the minimum acceptable level of detail of these for commodity flows?
- What sorts of "value added" or interpretive analyses, if any, would you like to see accompany the release of the CFS data?
- Users have urged the Census Bureau to produce a public micro data file. The law requires confidentiality of the data collected. To produce such a product, the Census Bureau would have to strip some data from the individual records to maintain confidentiality. To do so, they would use top code, additional noise, or a combination of some or all of these. Is there a preference? Do alternatives exist?

Exploring Survey Methods to Enhance CFS Data Quality and Usefulness

- Is the current CFS questionnaire able to respond to industry changes and accurately capture the information of interest?
- Are there pre-existing issues with the design of the questionnaire that may result in poor quality of the information collected or unduly burden on the respondent?
- How might other anticipated changes (e.g., industry coverage) influence the design of the questionnaire?
- Are multiple data collection instruments needed to assist establishments with providing accurate shipment information?
- Are there other modes of data collection that should be explored? Web-based questionnaires have been effectively used in other establishment surveys and, in many cases, help to alleviate the burden placed on respondents. Is this a feasible option for the CFS?
- Are the accompanying respondent instructions adequate and comprehensive? How might they be improved?
- Are there changes that should be implemented in the response process to ensure questionnaire and survey materials reach the appropriate respondent(s) in the establishment?
- Although a mandatory survey, the CFS also suffers from lack of responses. What additional measures could be implemented to improve the overall response rate and reduce potential bias?
- The CFS is conducted every 5 years as part of the economic census. Should it be conducted more frequently? Should it move towards an annual or continuous collection? Should establishments continue to report quarterly throughout the year? Less? More? Or differentially according to industry category (or other characteristic)?

APPENDIX D

Conference Attendees

Felix Ammah-Tagoe
Senior Research Consultant
MacroSys@USDOT

Bob Armstrong
Transportation Industry Analyst
U.S. Department of Transportation/Volpe Center

Lisa Aultman-Hall
Associate Professor
University of Connecticut

Tim Baker
Mobility Unit Manager
Colorado Department of Transportation

Mark Berndt
Senior Freight Analyst
Wilbur Smith Associates

Paul Bingham
Principal
Global Insight, Inc.

Loretta Bitner
Transportation Specialist
Federal Motor Carrier Safety Administration

Jock Black
Branch Chief
U.S. Census Bureau

Thomas Bolle
Director, Intergovernmental Affairs
DOT/RITA/BTS

Rob Bostrom
Kentucky Trans. Cabinet
rob.bostrom@ky.gov

Ruth Bramblett
Source Data Coordinator
Bureau of Economic Analysis

James Brogan
Associate
Cambridge Systematics, Inc.

Michael Bronzini
Dewberry Chair Professor
George Mason University

Brian Canepa
Senior Program Associate
Transportation Research Board

Patrick Canning
Senior Economist
U.S. Department of Agriculture

Harold Cerveny
Senior Associate
Tioga Group

Promod Chandhok
Survey Statistician
Bureau of Transportation Statistics

Richard Chard
Economist
Bureau of Economic Analysis

Shih-Miao Chin
Oak Ridge National Laboratory

Ed Christopher
Planner
FHWA Resource Center

Paul Ciannavei
Principal
Global Insight

Michael Cohen
Supervisory Survey Statistician
Bureau of Transportation

Robert Copp
Division Chief
California Department of Transportation

Bob Costello
Chief Economist and Vice President
American Trucking Associations, Inc.

Kenneth Crowther
Graduate Research Assistant
University of Virginia Center for Risk
Management of Engineering Systems

T. Randall Curlee
Distinguished Scientist
Oak Ridge National Laboratory

Scot Dahl
U.S. Census Bureau

Bruce Dembroski
U.S. Census Bureau

Scott Dennis
Economist
Bureau of Transportation Statistics

Ruth Detlefsen
Assistant Division Chief
U.S. Census Bureau

Virginia Dick
ATRI
American Transportation Research Institute

Adhi Dipo
Transportation Research Analyst
MacroSys@USDOT

Scott Drumm
Manager, Research and Market Information
Port of Portland

Ronald Duych
Senior Transportation Specialist
U.S. DOT RITA/BTS

Nathan Erlbaum
Associate Transportation Analyst
NYS

Jerry Everett
Research Director
University of Tennessee Transportation Center

David Fitzpatrick
Associate
Booz Allen Hamilton

Juan Flores
Freight Coordinator
AASHTO

Susie Forde
Chief, Data Management Section
Wisconsin Department of Transportation

John Fowler
U.S. Census Bureau

Glenda Fuller
Traffic Survey Manager
Idaho Transportation Department

Stephen Fuller
Texas A&M University

James Golden
Manager, Transportation Statistics Office
Florida Department of Transportation

Peter Gordon
Professor
University of Southern California

James Gosnell
Deputy Executive Director
Southern California Association of Governments

Lance Grenzeback
Senior Vice President
Cambridge Systematics, Inc.

Robin Grier
Assistant Division Administrator
Virgina Deptartment of Transportation

Yacov Haimes
Professor of Systems & Information Engr.
Center for Risk Management of
Engineering Systems

James Hall
Assistant Professor
University of Illinois @ Springfield

Kathleen Hancock
Associate Director
Center for Geospatial Information Technology
Virginia Tech

Jose Holguin-Veras
Associate Professor
Rensselaer Polytechnic Institute

Mary Hutzler
Associate Director
Office of Statistical Programs
Bureau of Transportation Statistics

Barbara Ivanov
Director of Freight
Washington State Department of Transportation

Derek Jaeger
Research Analyst
Port of Portland

Deborah Johnson
Transportation Specialist
RITA/BTS

Jonette Kreideweis
Director
Minnesota Department of Transportation

Bruce Lambert
Economist
U.S. Army Corps of Engineers

Catherine Lawson
New York University at Albany

Mark Lepofsky
Senior Research Scientist
Battelle

Steve Lewis
USDOT/RITA
steve.lewis@dot.gov

John Madden
Civil Engineer 2
New York State Department of Transportation

Robert Mannell
Statewide Planning Manager
Virginia Deptartment of Transportation

Michael Margreta
Survey Statistician
Bureau of Transportation Statistics

Caroline Marshall
Principal Planner
Atlanta Regional Commission

Wassel Mashagbeh
Senior Economist
Maritime Administration
U.S. Department of Transportation

Lee Maynus
Director, Office of Transportation Statistics
New York State Thruway Authority

James McQuirt
Administrator, Office of Technical Services
Ohio Department of Transportation

Warren Merrell
Manager. Systems Planning Office
Florida Department of Transportation

Thomas Mesenbourg
Assistant Director for Economic Programs
U.S. Census Bureau

Arnim Meyburg
Professor
Cornell University

Bruce Millar
Meeting Coordinator
Transportation Research Board

Mabel Ocasio
Head of Transport Statistics
U.S. Census Bureau

John Palm
Senior Project Manager
Reynolds Smith and Hills

Thomas Palmerlee
Transportation Data Specialist
Transportation Research Board

Qisheng Pan
Associate Professor
Texas Southern University

Ki Young Park
Argonne National Laboratory

Eric Peterson
Deputy Administrator
RITA/USDOT

Pierce Pierce
Statistician
Battelle

Michele Provencher
Senior Statistician
Transport Canada

Lisa Randal
Planning Team Leader
Federal Highway Administration

Reed Reed
Principal
Cons. Svcs. Int'l

Benjamin Ritchey
Vice President & General Manager
Battelle Memorial Institute (Battelle)

Charles Rombro
Team Lead Economist
FMCSA

Rolf Schmitt
Freight Policy Team Leader
Federal Highway Administration

Joseph Schofer
Professor
Northwestern University

Mikael` Schöllin
Statistics Sweden

Joy Sharp
Statistician
Bureau of Transportation Statistics

Irwin Silberman
Transportation Specialist
RITA/BTS/USDOT

Caesar Singh
Transportation Specialist
RITA/USDOT

Howard Slavin
President
Caliper Corporation

Frank Southworth
Senior R&D Staff
Oak Ridge National Laboratory

Michael Sprung
International Transportation Specialist
RITA/BTS/USDOT

Jack Stickel
Transportation Planner
Alaska Deptartment of Transportation

Tianjia Tang
Transportation Specialist
Federal Highway Administration

Ronald Tweedie
Consultant
Tweedie Construction Co.

Ava Vollbrecht
Director, Market Research
Norfolk Southern

Mark Wallace
Division Chief, SSSD
U.S. Census Bureau

C. Michael Walton
Ernest H. Cockrell Centennial Chair in Engineering
University of Texas at Austin

Sharon Weekley
Industrial Engineer
U.S. Army Corps of Engineers

Joyce Wenger
Senior Associate
Booz Allen Hamilton

Lars Werke
Project manager
Statistics Sweden

Todd Westhuis
Traffic Monitoring Section Supervisor
New York State Department of Transportation

Karen White
Economist
FHWA

Rhonda Young
Assistant Professor
University of Wyoming

Thomas Zabelsky
Assistant Division Chief
U.S. Census Bureau

Johanna Zmud
President
NuStats Partners, LP

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. William A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both the Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. William A. Wulf are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is a division of the National Research Council, which serves the National Academy of Sciences and the National Academy of Engineering. The Board's mission is to promote innovation and progress in transportation through research. In an objective and interdisciplinary setting, the Board facilitates the sharing of information on transportation practice and policy by researchers and practitioners; stimulates research and offers research management services that promote technical excellence; provides expert advice on transportation policy and programs; and disseminates research results broadly and encourages their implementation. The Board's varied activities annually engage more than 5,000 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. **www.TRB.org**

www.national-academies.org

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