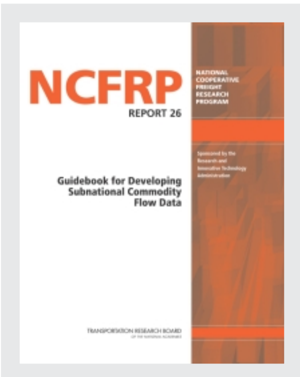


Guidebook for Developing Subnational Commodity Flow Data

DETAILS

155 pages | 8.5 x 11 | PAPERBACK
ISBN 978-0-309-28353-3 | DOI 10.17226/22523

AUTHORS

Cambridge Systematics, Inc.

NATIONAL COOPERATIVE FREIGHT RESEARCH PROGRAM

NCFRP REPORT 26

Guidebook for Developing Subnational Commodity Flow Data

Cambridge Systematics, Inc.
Atlanta, GA

WITH

Ken Casavant
Pullman, WA

Anne Goodchild
Seattle, WA

Eric Jessup
Freiburg, Germany

Catherine T. Lawson
Albany, NY

Subscriber Categories
Data and Information Technology • Freight Transportation • Planning and Forecasting

Research sponsored by the Research and Innovative Technology Administration

TRANSPORTATION RESEARCH BOARD

WASHINGTON, D.C.
2013
www.TRB.org

NATIONAL COOPERATIVE FREIGHT RESEARCH PROGRAM

America's freight transportation system makes critical contributions to the nation's economy, security, and quality of life. The freight transportation system in the United States is a complex, decentralized, and dynamic network of private and public entities, involving all modes of transportation—trucking, rail, waterways, air, and pipelines. In recent years, the demand for freight transportation service has been increasing fueled by growth in international trade; however, bottlenecks or congestion points in the system are exposing the inadequacies of current infrastructure and operations to meet the growing demand for freight. Strategic operational and investment decisions by governments at all levels will be necessary to maintain freight system performance, and will in turn require sound technical guidance based on research.

The National Cooperative Freight Research Program (NCFRP) is a cooperative research program sponsored by the Research and Innovative Technology Administration (RITA) under Grant No. DTOS59-06-G-00039 and administered by the Transportation Research Board (TRB). The program was authorized in 2005 with the passage of the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU). On September 6, 2006, a contract to begin work was executed between RITA and The National Academies. The NCFRP will carry out applied research on problems facing the freight industry that are not being adequately addressed by existing research programs.

Program guidance is provided by an Oversight Committee comprised of a representative cross section of freight stakeholders appointed by the National Research Council of The National Academies. The NCFRP Oversight Committee meets annually to formulate the research program by identifying the highest priority projects and defining funding levels and expected products. Research problem statements recommending research needs for consideration by the Oversight Committee are solicited annually, but may be submitted to TRB at any time. Each selected project is assigned to a panel, appointed by TRB, which provides technical guidance and counsel throughout the life of the project. Heavy emphasis is placed on including members representing the intended users of the research products.

The NCFRP will produce a series of research reports and other products such as guidebooks for practitioners. Primary emphasis will be placed on disseminating NCFRP results to the intended end-users of the research: freight shippers and carriers, service providers, suppliers, and public officials.

NCFRP REPORT 26

Project NCFRP-20
ISSN 1947-5659
ISBN 978-0-309-28353-3
Library of Congress Control Number 2013948813

Published reports of the

NATIONAL COOPERATIVE FREIGHT RESEARCH PROGRAM

are available from:

Transportation Research Board
Business Office
500 Fifth Street, NW
Washington, DC 20001

and can be ordered through the Internet at:

http://www.national-academies.org/trb/bookstore

Printed in the United States of America

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. C. D. Mote, Jr., 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 Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. C. D. Mote, Jr., are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 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

COOPERATIVE RESEARCH PROGRAMS

CRP STAFF FOR NCFRP REPORT 26

Christopher W. Jenks, *Director, Cooperative Research Programs*
Crawford F. Jencks, *Deputy Director, Cooperative Research Programs*
William C. Rogers, *Senior Program Officer*
Charlotte Thomas, *Senior Program Assistant*
Eileen P. Delaney, *Director of Publications*
Ellen M. Chafee, *Editor*

NCFRP PROJECT 20 PANEL
Freight Research Projects

John W. Fuller, *University of Iowa, Iowa City, IA* (Chair)
Monica M.H. Blaney, *Transport Canada, Ottawa, ON*
Kuo-Ann Chiao, *New York Metropolitan Council, New York, NY*
Barbara A. Ivanov, *Washington State DOT, Olympia, WA*
Diane Jacobs, *California DOT, Los Angeles, CA*
Marygrace Parker, *I-95 Corridor Coalition, Greenwich, NY*
Rakesh Shalia, *FedEx Services, Memphis, TN*
Irwin Silberman, *Roswell, GA*
Ronald J. Duych, *RITA Liaison*
Michael Sprung, *RITA Liaison*
Thomas Zabelsky, *Other Liaison*
Thomas Palmerlee, *TRB Liaison*

FOREWORD

By William C. Rogers
Staff Officer
Transportation Research Board

NCFRP Report 26: Guidebook for Developing Subnational Commodity Flow Data (Guidebook) provides state DOTs and other subnational agencies with a comprehensive discussion of how to obtain and compile commodity flow data useful for their analysis. The *Guidebook* contains descriptions of existing public and private commodity flow data; standard procedures for compiling local, regional, state, and corridor databases from these commodity flow data sources; new and effective procedures and methodologies for conducting subnational commodity flow surveys and studies; and methods for using commodity flow data in local, regional, state, and corridor practice.

Commodity flow data are critical to conducting transportation planning at state, regional, and local jurisdictional levels and in corridors (collectively called subnational levels). Commodity flow data are used to understand which industries generate the most demand on the transportation system. These data also provide a key link between economic trade relationships and freight demand and are used in modal diversion studies. These data are also a key input to multimodal trade corridor studies and air quality assessments. Currently, there are a number of useful commodity flow data sources at the national level that are of limited application to subnational planning because they lack the appropriate geographic detail for flow origins and destinations. National level commodity flow data sources cannot easily be used to identify available data sets for subnational use. State DOTs and other subnational agencies need a variety of tools to help them tailor existing commodity flow data for their specific needs and to develop additional sources of data.

Under NCFRP Project 20, Cambridge Systematics was asked to (1) describe and review examples of recent or current efforts at the subnational level to compile and use commodity flow information for transportation planning and analysis; (2) describe national data sets and their use and limitations for application at subnational levels; (3) develop procedures and methodologies for conducting commodity flow surveys at subnational levels; (4) demonstrate the application of the procedures and methodologies by applying them in a test case to address such issues as modal diversion, air quality, and/or public freight investment prioritization; and (5) prepare a guidebook that illustrates how the data should be collected and used in models of decision making as well as providing guidance for compiling commodity flow data sets appropriate for subnational analysis.

In addition to the *Guidebook,* two subtask reports from NCFRP Project 20 are available at www.trb.org/Main/Blurbs/169330.aspx: "Review of Subnational Commodity Flow Data Development Efforts and National Freight-Related Data Sets" and "Demonstration of Application of Establishment Survey."

CONTENTS

1 Chapter 1.0 Overview of the *Guidebook* and Key Issues
1 1.1 Introduction
3 1.2 Background on Commodity Flow Data
8 1.3 Geographic Issues Related to Subnational Commodity Flow Data
10 1.4 Commodity Issues Related to Subnational Commodity Flow Data

16 Chapter 2.0 Collecting Subnational Commodity Flow Data Using Establishment Surveys
16 2.1 Introduction
17 2.2 Step-by-Step Process for Conducting Establishment Surveys
58 2.3 Next Steps

59 Chapter 3.0 Collecting Subnational Commodity Flow Data Using Roadside Truck Intercept Surveys
59 3.1 Introduction
60 3.2 Step-by-Step Process for Conducting Roadside Truck Intercept Surveys
91 3.3 Next Steps

92 Chapter 4.0 Developing Subnational Commodity Flow Data Using Supplemental Sources of Local Economic Activity
92 4.1 Introduction
93 4.2 Step-by-Step Process for Collecting Locally Available Data
102 4.3 Example of Developing a Diesel Fuel Local Commodity Flow Database
107 4.4 Detailed Example of Developing a Potato Commodity Flow Database
115 4.5 Next Steps

116 Chapter 5.0 Developing Subnational Commodity Flow Data Using Disaggregation
116 5.1 Introduction
117 5.2 Step-by-Step Process for Disaggregating Commodity Flow Data
135 5.3 Next Steps

136 Chapter 6.0 Playbook
136 6.1 Introduction to the Playbook
136 6.2 Getting Started—What Should Be Known About Problems and Resources
138 6.3 Considerations in Answering Sample Freight Questions
141 6.4 Self-Diagnosis of Available Freight Resources
143 6.5 Examples of Preparing a Game Plan to Answer Freight Questions
151 6.6 Sample Game Plan Worksheet
152 6.7 Playbook Recap and Conclusions on Developing Subnational Commodity Flow Data

153 References
154 List of Acronyms

Note: Many of the photographs, figures, and tables in this report have been converted from color to grayscale for printing. The electronic version of the report (posted on the Web at www.trb.org) retains the color versions.

CHAPTER 1.0

Overview of the *Guidebook* and Key Issues

1.1 Introduction

Information about commodity flows is one of the most critical data needs for freight planning. Commodity flows describe the quantity of commodities that are shipped between origins and destinations and the transportation modes that are used. In other words, commodity flow data describe what moves, where it moves from and to, how much of it moves, and the manner by which it moves. This information is critical to freight planning for the following reasons:

- Commodity flow data provide a direct link between the inputs and outputs of an economy and the freight flows that the economy gives rise to. Commodity flow data indicate what is produced and consumed in an economy and can be used to generate estimates of demand for freight transportation. Commodity flow information also can be used to understand which industries generate demand for freight and therefore benefit from freight investments. This is a critical input for decisions that prioritize freight investments.
- Commodity flow information—the types of commodities, the amount being shipped, and the distances commodities are being shipped—is a critical determinant of mode choice. Therefore, commodity flow data are critical in modal diversion studies. In order to optimize multimodal transportation systems, state departments of transportation (DOTs), metropolitan planning organizations (MPOs), and federal transportation agencies are often interested in examining the costs and benefits of competing modal investments. Commodity flow information is critical to determining how much freight is divertible from one mode to another based on transportation agency investment decisions.
- Commodity flow data are often used as inputs to state and multistate models that forecast transportation demand. Commodity flow data are critical in multistate corridor studies because they provide the best representation of freight transportation demand at this level of geography. Commodity flow data also are increasingly being used to estimate air quality impacts of the transportation inputs to various industrial activities.

Purpose of *NCFRP Report 26: Guidebook for Developing Subnational Commodity Flow Data*

The purpose of *NCFRP Report 26: Guidebook for Developing Subnational Commodity Flow Data* (*Guidebook*) is to provide state DOT and MPO freight transportation planners with an easy-to-use manual on how to develop subnational commodity flow databases to meet freight transportation planning needs in their region. The *Guidebook* describes the limitations of using existing national commodity flow databases for local planning. The *Guidebook* also describes methods to develop primary commodity flow data using local data collection along with how to augment local data collection efforts with information from published data sets and commodity flow disaggregation techniques. The *Guidebook* provides a detailed, step-by-step description of each method. It also describes the limitations of each method and how these limitations can be mitigated.

Uses of the *Guidebook*

There are several ways that the *Guidebook* can be used. A *Guidebook* user may want to learn how to apply one of the methods of developing subnational commodity flow data described in the *Guidebook.* The *Guidebook* provides detailed instructions on how to use the method, what types of problems it might best be applied to, and its limitations. Chapters 2.0, 3.0, 4.0, and 5.0 describe methods of developing subnational commodity flow data.

A *Guidebook* user may want a comprehensive approach that uses the different methods in combination to produce a subnational commodity flow database, taking into consideration the relative strengths and weaknesses of each of the methods. The *Guidebook* provides instructions for a procedure that combines the methods. Also, users may not have sufficient resources to build a comprehensive commodity flow database solely from primary research. In this case, combining primary data collection methods with methods that use existing public sources may be necessary.

A *Guidebook* user may want information on the strengths and weaknesses of specific freight data sources. This can be found in a subtask report associated with the development of the *Guidebook* titled "Review of Subnational Commodity Flow Development Efforts and National Freight-Related Data Sets," which includes descriptions of existing databases and references on what is included in these databases and is available at www.trb.org/Main/Blurbs/169330.aspx.

For primary data collection, the *Guidebook* provides information about sampling frames and their cost and availability as well as illustrative survey instruments that have been demonstrated during the course of this project.

Structure of the *Guidebook*

The *Guidebook* is structured to emphasize collecting new freight flow data for developing subnational commodity flow data. It is complemented by data analysis efforts such as using locally available freight data and disaggregating national commodity flow data. This structure is designed to allow transportation agencies to choose the approach that best fits with their freight planning needs and available resources. The specific *Guidebook* sections are as follows:

- **Chapter 1.0—**Overview of the *Guidebook* and Key Issues
- **Chapter 2.0—**Collecting Subnational Commodity Flow Data Using Establishment Surveys
- **Chapter 3.0—**Collecting Subnational Commodity Flow Data Using Roadside Intercept Surveys
- **Chapter 4.0—**Developing Subnational Commodity Flow Data Using Supplemental Sources of Local Economic Activity
- **Chapter 5.0—**Developing Subnational Commodity Flow Data Using Disaggregation
- **Chapter 6.0—**Playbook

Following this introduction, the *Guidebook* is divided into two main sections. Chapters 2.0 through 5.0 comprise the "Methods" section. They are designed to introduce the user to specific methods that can be used to develop subnational commodity flow data. Each chapter describes a particular method and is structured to provide step-by-step procedures for implementing that particular method. Examples are provided from actual practice, and there are worksheets that can be used by practitioners to design an application of the method for their own planning studies.

Chapter 6.0 is the "Playbook" section. Borrowing an analogy from football, the Playbook section is designed to help practitioners design a "Game Plan" consisting of mixing and matching different methods from Chapters 1.0 through 5.0 (the Methods section) into their own "Playbook" for attacking a particular freight planning problem. The Playbook section also includes examples of actual planning problems and guides the user through the process of selecting and combining methods. Many users will start in the Playbook section after reading this introduction. They will figure out what they need to know and start designing their game plan; the Playbook will guide them back to the Methods section to do the detailed planning of the surveys, data mining, or data disaggregation that they may need to solve their problems. The Playbook section is designed to work together with the Methods section and guide the reader back and forth between the two sections.

Other users, those who want to get a complete picture of the whole topic, can start at the beginning of the *Guidebook* and just read from cover to cover.

1.2 Background on Commodity Flow Data

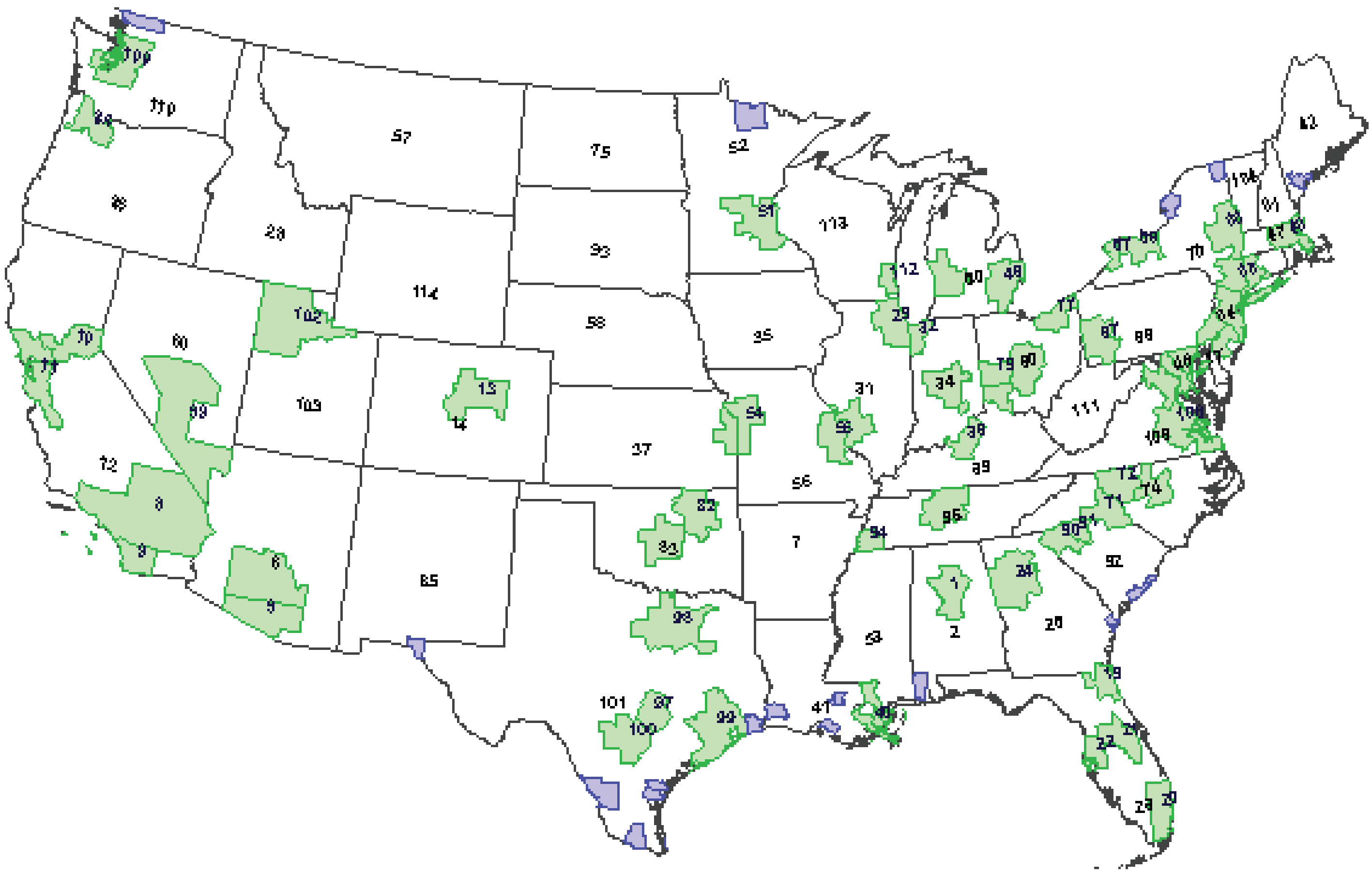
The most commonly used commodity flow database is FHWA's Freight Analysis Framework (FAF) database. In 2011, FHWA released the third version of this database. FAF pivots off of the Bureau of Transportation Statistics (BTS) Commodity Flow Survey (CFS), and then integrates data from a variety of sources to create estimates for tonnage and value by origin, destination, commodity, and mode for a base year (currently 2007 to be consistent with the CFS) and forecasts, currently through 2040.

The FAF database is national in scope and the sampling and data collection embodied within it (such as the CFS) are designed to produce statistically valid data at the national level and for very limited subnational geographies. FAF provides data for each state and for 89 regions around the country. Figure 1.1 shows the FAF regions. Table 1.1 shows a snippet of FAF data for a handful of agricultural commodities originating in Alabama and destined for any location. Table 1.2 shows truck flows from the Norfolk, Virginia, metropolitan region to the Chicago, Illinois, metropolitan region.

While there is a wealth of information in the FAF database, states and MPOs need subnational databases that provide more disaggregate geographies for the origins and destinations of flows. Subnational commodity flow databases might involve county-level detail, city-level detail, or even traffic analysis zone (TAZ)-level detail. These databases can be used in several types of applications, including the following:

- Trade flow analysis
- Development of truck trip tables for travel demand model
- Feasibility analysis for new modal services
- Estimating the economic impact of freight activity
- Regional or statewide freight plans
- Corridor studies of freight-intensive roadways
- Modal diversion studies

Trade flow analyses include studies of the trading partners for states or substate regions (such as counties or MPOs). These analyses are useful for understanding the relationship between geographic entities such as the relationship between two different states or between an MPO region and an adjacent, external region. Trade flow analyses also provide information on the relative size of internal, internal-external, and through freight trips for a region. Trade flow analyses can be done for total freight flows, specific modes, and specific commodities or commodity groups. They are useful for understanding how a region's economies relate to the economies in other regions. These analyses also provide a sense of where freight is trying to move to and from for a region. Trade flow data can be depicted graphically as desire lines of freight moving between



Source: FHWA FAF.

Figure 1.1. FHWA FAF regions.

regions. Figure 1.2 shows a map depicting trade flow between MPO regions in Georgia using a preliminary output from the state's travel demand model.

The development of a truck trip table is one of the initial steps in creating a truck component for a travel demand model. These tables provide estimates of the number of trucks moving between TAZs in the model area, including external regions. In recent years, a number of states and MPOs have developed truck trip tables from commodity flow data by gathering data on the average payload (or loaded weight) for trucks carrying different commodities and adjusted for empty movements. If the total tonnage of a commodity moving by truck between an origin and destination is known, this can be divided by an average payload for trucks carrying that commodity to determine the number of truck trips.

Generally, the last step in the travel demand model process is route choice, in which the truck trip table is assigned to the road network. While routed commodity flow data can be useful to calibrate and validate travel demand models or to determine which industries use different corridors, routed flows are not required since the model assigns trucks based on predetermined criteria. It should be noted that it is theoretically possible to develop vehicle trip tables for other modes using commodity flow data, but the routing logic for these nontrucking modes may be more complex and is usually not incorporated into state or MPO models.

Feasibility analyses for new modal services include studies on the demand for new or altered service for a specific mode, for example, a study to determine the need for a new rail intermodal yard.

Table 1.1. Example of FHWA FAF data.

			2007	
Origin	Commodity	Mode	Total Tons (Thousands)	Total Dollars (Millions)
Alabama	Live animals/fish	Truck	1,569.38	1,691.28
Alabama	Live animals/fish	Rail	0.01	0.05
Alabama	Live animals/fish	Air (include truck-air)	0.38	14.51
Alabama	Live animals/fish	Multiple modes and mail	0.01	0.03
Alabama	Cereal grains	Truck	3,070.71	642.71
Alabama	Cereal grains	Rail	0.36	0.08
Alabama	Cereal grains	Air (include truck-air)	0.02	0.04
Alabama	Cereal grains	Multiple modes and mail	8.65	1.57
Alabama	Other agricultural products	Truck	3,678.31	1,457.25
Alabama	Other agricultural products	Rail	583.76	293.73
Alabama	Other agricultural products	Air (include truck-air)	0.63	6.15
Alabama	Other agricultural products	Multiple modes and mail	216.07	149.31
Alabama	Other agricultural products	Other and unknown	187.27	27.83
Alabama	Animal feed	Truck	3,997.60	1,169.09
Alabama	Animal feed	Rail	2.54	9.60
Alabama	Animal feed	Air (include truck-air)	0.06	1.83
Alabama	Animal feed	Multiple modes and mail	70.13	40.76
Alabama	Animal feed	Other and unknown	334.94	20.38
Alabama	Meat/seafood	Truck	2,069.60	4,705.65
Alabama	Meat/seafood	Rail	0.02	0.02
Alabama	Meat/seafood	Air (include truck-air)	3.58	4.01
Alabama	Meat/seafood	Multiple modes and mail	4.49	12.48
Alabama	Meat/seafood	Other and unknown	68.77	128.17

Source: FHWA FAF[3] Database, 2007.

In the case of intermodal rail, the key factors for determination of the feasibility of a new service include the types of commodities that have a trip end near the yard location and the distance that the goods are being carried. For this type of study, the specific route being used by current traffic is not relevant, therefore routing information is not needed. However, a commodity flow database could be used to estimate the commodities that are demanded and supplied at a particular location, in addition to the distances that those goods travel to reach their corresponding trip ends.

Estimating the economic impact of freight activity also is something that can be accomplished without using routing information. This type of study is generally used to determine the size of the freight-related components of a region's economy relative to the regional economy as a whole. In this case, commodity flow databases can be used to estimate the size of the freight-related components of the economy. These databases also can be used to estimate how freight-related economic activity is distributed among various commodities and industries within a region.

Regional or statewide freight plans can be developed in several different fashions. However, in order to understand freight activity in a comprehensive regional fashion, it is critical to understand how many freight vehicles there are at different locations on the freight network, the travel patterns of these vehicles, and the freight network that is used. Typically, routed flow information is not provided with commodity specificity, but information on truck routing on the highway network is a critical component of developing a statewide or regional freight plan.

Corridor studies on freight-intensive roadways are the most direct example of freight planning efforts in which national- or state-level commodity flow information by itself is insufficient.

Table 1.2. Sample truck flow data from FHWA FAF database.

			2007		
DMS[a]_ORIG	DMS_DEST	SCTG2	DMS_ MODE	Total Tons (Thousands)	Total Dollars (Millions)
Norfolk VA-NC MSA[b] (VA Part)	Chicago IL-IN-WI CSA[c] (IL Part)	Live animals/fish	Truck	0.04	0.13
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Cereal grains	Truck	0.01	0.01
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Other agricultural products	Truck	7.89	17.33
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Animal feed	Truck	0.44	0.20
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Meat/seafood	Truck	8.17	20.96
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Milled grain products	Truck	2.06	3.39
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Other foodstuffs	Truck	16.62	88.40
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Alcoholic beverages	Truck	11.46	26.70
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Tobacco products	Truck	0.02	0.19
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Building stone	Truck	2.55	0.06
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Natural sands	Truck	24.70	1.18
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Gravel	Truck	0.01	0.00
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Nonmetallic minerals	Truck	15.05	0.07
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Coal	Truck	0.00	0.00
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Gasoline	Truck	0.04	0.02
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Fuel oils	Truck	0.28	0.50
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Coal-n.e.c.	Truck	0.07	0.01
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Basic chemicals	Truck	6.25	21.05
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Pharmaceuticals	Truck	0.00	0.00
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Fertilizers	Truck	29.65	4.69
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Chemical products	Truck	1.98	7.93
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Plastics/rubber	Truck	23.05	36.75
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Logs	Truck	0.03	0.01
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Wood products	Truck	9.90	10.12
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Newsprint/paper	Truck	9.16	5.12
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Paper articles	Truck	0.50	1.46

Table 1.2. (Continued).

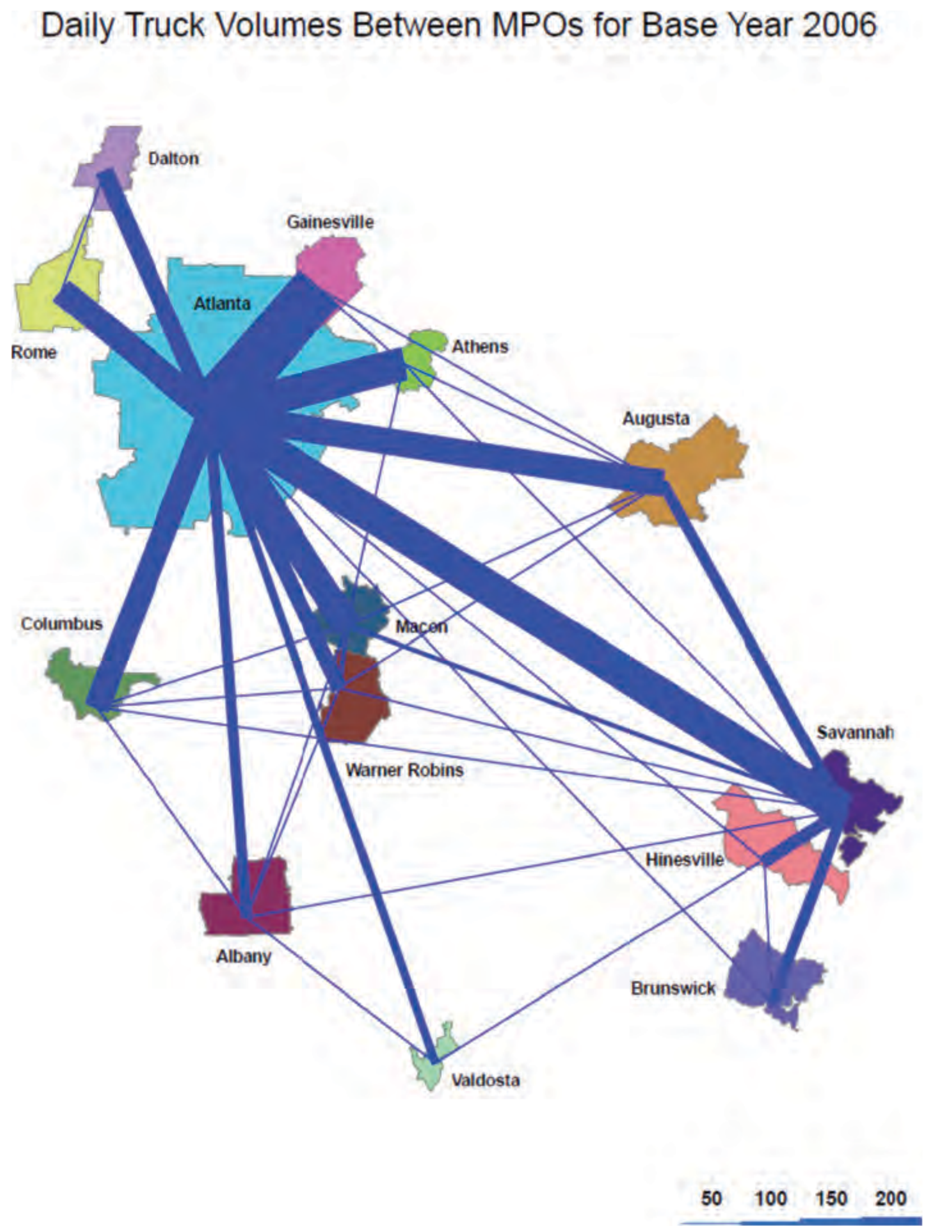
DMS_ORIG	DMS_DEST	SCTG2	DMS_ MODE	2007 Total Tons (Thousands)	2007 Total Dollars (Millions)
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Printed products	Truck	2.28	5.61
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Textiles/leather	Truck	0.00	0.00
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Nonmetal min. products	Truck	36.40	16.28
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Base metals	Truck	13.23	24.93
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Articles-base metal	Truck	14.31	25.46
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Machinery	Truck	43.87	361.79
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Electronics	Truck	10.29	76.05
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Motorized vehicles	Truck	10.60	67.33
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Transport equipment	Truck	0.08	5.36
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Precision instruments	Truck	0.51	12.21
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Furniture	Truck	0.50	0.54
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Misc. mfg. products	Truck	7.67	46.93
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Waste/scrap	Truck	0.68	0.09
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Mixed freight	Truck	2.48	1.64
Norfolk VA-NC MSA (VA Part)	Chicago IL-IN-WI CSA (IL Part)	Unknown	Truck	0.00	0.00

[a]DMS = Domestic
[b]MSA = Metropolitan Statistical Area
[c]CSA = Consolidated Statistical Area

Source: FHWA FAF, 2007.

Commodity flow information needs to be complemented or supplemented by specific routed flow information. This can come in many forms. To estimate truck flows using a standard travel demand model, commodity flow information must be disaggregated to the TAZ level and assigned to the network taking into account the corresponding routing of passenger cars and other vehicles. Truck roadside intercept surveys also are a useful tool for understanding corridor-level truck flows. They can be used to augment or validate disaggregated commodity flow data. A comprehensive roadside survey program can be used as a stand-alone tool to develop commodity flow information for a corridor. Similarly, establishment survey data can be used to better understand commodity flow information at the corridor level.

Modal diversion studies also require information about routed flows to fully understand the impact of freight moving on one mode relative to another. Typically, these studies are designed to determine the effect of diverting freight from highway to rail, and the studies attempt to quantify the impact on congestion, safety, road maintenance, and the environment. This requires an



Source: GDOT Statewide Travel Demand Model as of October 20, 2010.

Figure 1.2. Georgia MPO trade flow map.

understanding of the current freight flow routing that is likely to divert from highway to rail. Similar to corridor studies, travel demand models are one useful source of these types of estimates. These models typically rely upon disaggregated commodity flow databases, and they can be augmented by roadside surveys or establishment surveys.

1.3 Geographic Issues Related to Subnational Commodity Flow Data

As stated earlier, one of the primary issues related to subnational commodity flow data is that available freight flow data are not provided with the geographic detail needed to conduct several different types of common state and MPO freight planning activities. Figure 1.3 shows the differ-

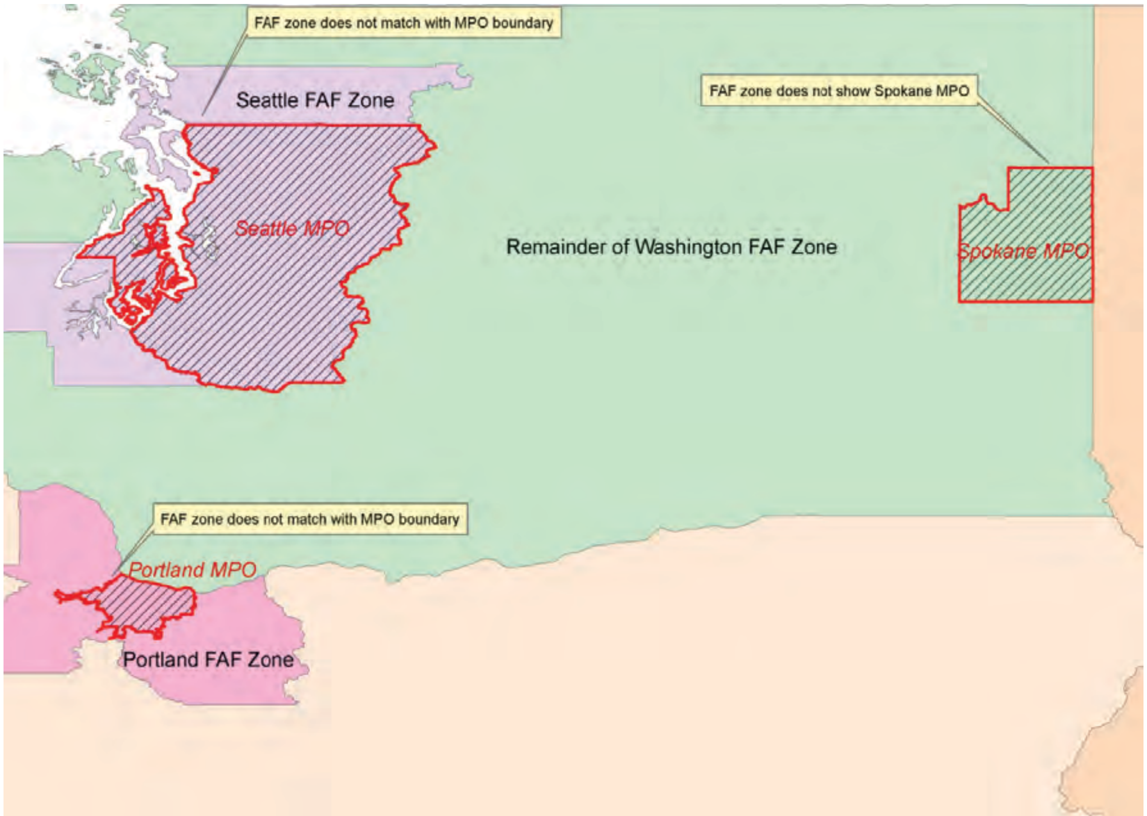


Figure 1.3. FAF zones and MPO regions in the Pacific Northwest.

ence between FAF zones and MPO regions in the Pacific Northwest. For the Seattle and Portland metropolitan regions, the MPO boundaries and the urban boundaries that are defined in FAF do not match. Also, county-level detail within these two metropolitan regions is too disaggregated to be covered by FAF.

The Spokane MPO falls into an FAF zone characterized as "remainder of Washington." It includes over 75 percent of the state and by itself would not be able to provide meaningful information on Spokane's regional freight flows.

As noted previously, national commodity flow databases are developed from samples that do not provide statistically valid data at the subnational geographies that are needed for most state and MPO freight analysis. One approach that has been used extensively to overcome this obstacle is to disaggregate national databases to the subnational level by using economic data as an indicator of the level of commodity production or consumption in the subnational zones. This approach is described in more detail in Chapter 4.0 of the *Guidebook*. Many researchers have noted that in some cases there are weak relationships between commodity production/consumption and the economic indicator data that are often used in the disaggregation procedures. For example, industry-specific employment is often used as an indicator variable for consumption or production within an industry. However, it has been found that for certain commodities there is not a strong correlation between the quantity of goods produced in an industry and the number of people employed in the industry. Chapter 4.0 of the *Guidebook* describes a number of disaggregation techniques, identifies some of the weaknesses in the relationships between commodity production/consumption and specific economic indicator data, and discusses how best to use disaggregation techniques in combination with other approaches.

Primary data collection is the most straightforward approach for developing subnational commodity flow data. Chapter 2.0 of the *Guidebook* examines data collection methods for establishment surveys. Chapter 3.0 of the *Guidebook* examines data collection methods for roadside truck intercept surveys. The use of global positioning system (GPS) data is an emerging field of tracking freight flows. The following is an overview of these data sources:

- **Establishment Surveys—**The establishment survey procedures described in this *Guidebook* include sampling and recruiting strategies, survey instrument design, sample sizes for different levels of geography, and sampling frames for identifying potential recruits. One of the issues associated with using establishment surveys at the subnational level is the need to survey for both outbound and inbound shipments. National establishment surveys, such as the CFS, only need to survey for outbound shipments. Establishment surveys also must weigh the tradeoffs associated with trying to get high levels of geographic detail on the origins and destinations of shipments relative to achieving high survey response rates.
- **Roadside Intercept Surveys—**These data are extremely useful for identifying commodity flows through a specific geographic location. The primary disadvantage is that the locations at which data can be collected are severely limited and likely to be insufficient to establish comprehensive geographic coverage. Roadside surveys also are not useful for collecting data on commodity flows within a metro area. Additionally, these surveys can only be applied to highway flows, so they are not truly multimodal. Finally, roadside intercept surveys provide information on truck origins and destinations, but not the true trip pattern of the cargo being carried.
- **GPS Databases—**It is now possible to purchase GPS data directly from vendors of truck fleet tracking systems. These databases are very useful for understanding truck operational characteristics. The problem in using these for commodity flow studies is that there is no information about the commodities carried. Commodity information can sometimes be inferred from the land uses at trip ends, but detailed land use data are not available in most locations. Additionally, it is difficult to control the sampling procedures used for GPS databases, so the statistical validity of the data is questionable. Also, similar to roadside intercept surveys, the data are only available for the truck mode.

There are a number of specialized databases, many in the public domain, that provide data on commodity production that can be used to enhance or develop subnational commodity flow data. These are described in the *Guidebook.* Typically, these databases do not provide geographic detail that is more disaggregated than county-level detail. There also are establishment-level databases that provide employment and value of shipment data at the level of the individual firm that can be used to enhance subnational commodity flow databases with much more geographic detail. These establishment-level databases also are described in the *Guidebook.* Economic input-output model databases can be used in conjunction with specialized databases to develop information about consumption of commodities. These economic input-output databases are typically used in conjunction with data on the industries that consume these commodities. Chapter 5.0 of the *Guidebook* describes this approach to developing subnational commodity flow data. The biggest limitation of these specialized databases is that few include information about the origins or destinations of flows. Therefore, the greatest use of specialized databases tends to be in establishing control totals for inbound and outbound flows at subnational geographies of interest.

1.4 Commodity Issues Related to Subnational Commodity Flow Data

There are several commodity classification issues that arise in developing subnational commodity flow databases. The largest establishment survey in the United States is the BTS CFS.

However, it only covers certain commodities. The industries that are covered by the CFS fall into the following North American Industry Classification System (NAICS) sectors:

- Mining
- Manufacturing
- Wholesale trade
- Select retail trade industries—specifically electronic shopping, mail-order houses, fuel dealers, and publishers
- Auxiliary establishments (i.e., warehouses and managing offices) of in-scope multi-establishment companies. An advance survey of approximately 40,000 auxiliary establishments was conducted to identify auxiliary establishments with shipping activity

The CFS does not include establishments classified in the following sectors:

- Farms
- Forestry
- Fishing
- Utilities
- Construction
- Government-owned entities (except government-owned liquor stores)
- Transportation
- Most retail and services industries
- Foreign-based business importing to the United States. However, in theory, domestic portions of imported shipments can be captured in the CFS once arriving at a U.S.-based establishment (assuming it is an eligible shipping establishment included in the CFS)

Level of Commodity Detail and Variation in Commodity Characteristics

The level of commodity detail that is desired in a primary data collection effort will also have a significant impact on the size of the sample required and the complexity of the survey procedures. The tradeoff between level of detail in the final database and cost/complexity of the data collection effort is described in Chapter 2.0 of the *Guidebook.* If disaggregation of national commodity flow databases is utilized, then the strength of the relationship between the economic indicator variable and the estimated tonnage generated will be directly impacted by the level of commodity aggregation that is used.

For commodities that are included in the CFS, data accuracy varies across commodity, mode, and origin-destination patterns. This is in part due to the varying sample sizes used for different industries. FAF is derived from CFS and therefore inherits its accuracy issues for in-scope commodities. For out-of-scope commodities, FAF weaves together a variety of sources; this generates new accuracy issues that must also be considered. Note that for the last three CFSs, the Standard Classification of Transported Goods (SCTG) coding scheme was utilized. SCTG codes were developed for the 1997 CFS and are based on the Harmonized System (HS) codes.

The level of commodity detail is important for freight planning applications that require analysis of the value of goods moving in particular corridors or amongst regions. The level of commodity detail can also have a significant impact on the ability to accurately estimate vehicle flows from commodity flow data because these estimates are typically calculated on an average ton per vehicle basis. However, commodity detail can vary significantly even within two-digit commodities. Table 1.3 shows the ton-value ratio of the textile industry (SCTG 30) and compares it to the ratios for its three-digit subcomponents. At the two-digit level, the ton-value ratio is 99. However, for the textile industry's three-digit subcomponents, the ton-value ratio ranges from 46 for footwear to 209 for textiles not elsewhere classified. Therefore, converting vehicles

Table 1.3. Ton-value ratios for textiles and textile subcomponents, 2007.

SCTG Code	Commodity Description	Value (Millions Dollars)	Tons	Ton/ Value Ratio
30	Textiles, leather, and articles of textiles and leather	473,610	46,728,000	99
301	Textile fibers, yarns, and broadwoven or knitted fabrics, except coated or treated	56,837	10,954,000	193
302	Textile clothing and accessories, and headgear (except safety)	262,867	12,452,000	47
303	Textiles and textile article, not elsewhere classified	98,757	20,627,000	209
304	Footwear	37,234	1,706,000	46
305	Leather and articles, luggage of related materials, and dressed furskins and articles	17,916	990,000	55

Source: 2007 BTS Commodity Flow Survey.

to tons at the two-digit level for textiles can lead to inaccurate results if the actual commodities being moved are actually concentrated into one of the three-digit categories.

Variation of commodity characteristics at more disaggregated levels of commodity detail as compared to a more aggregate classification scheme also is important for expanding survey data collected from new sources at the local level. If similar expansion factors are used for commodities that are the same at the two-digit level, but vastly different at the three-digit commodity level (for example, with respect to commodity shipment weights, ton-value ratios, or commodity production/consumption per employee), then inaccurate expanded survey data may be generated. Additionally, accurate commodity classification data are important when disaggregating commodity flow data to the local level since often the local data are provided in units of dollar value and these must be translated into tons.

Commodity Classification Systems

Commodity flow databases need to use a classification system, and there are a variety of choices, including the following:

- **Standard Classification of Transported Goods (SCTG).** This system was developed by the U.S. Department of Transportation, U.S. Bureau of the Census, Statistics Canada, and Transport Canada to replace the Standard Transportation Commodity Code (STCC) for the CFS and to integrate the U.S. coding system with the Canadian coding system. SCTG tends to be directly tied to industries that create and ship goods. SCTG identifies major commodities carried by each mode of transportation and each significant intermodal combination, and it can be easily linked to the classifications used for international trade. This system has been used in the last two versions of CFS and FAF.
- **Standard Transportation Commodity Code (STCC).** This system was developed initially in the 1960s by the Association of American Railroads for analyses involving the railroad industry. Railroad waybill data, a comprehensive rail commodity flow database available to states, is still published using the STCC coding system. The first two versions of FAF and the CFS also used this system.
- **Harmonized System.** This system was developed by the World Customs Organization as a customs tariff and statistical tool. It also is used by governments, international organizations, and the private sector for setting trade policies, monitoring price and quota controls, and compiling national accounts.

- **The U.S. Census Vehicle Inventory and Use Survey (VIUS) System—**VIUS was part of the U.S. Economic Census, but it has been discontinued. It was used extensively by freight analysts to develop payload factors for converting commodity tonnage into vehicle trips (see previous discussion of creating truck trip tables from commodity flow data). Some analysts continue to use this older data for this purpose. This product developed its own classification system roughly equivalent to SCTG, but aggregated to better match the collected data.
- **Waterborne Commerce Statistical Center (WCSC) Codes.** These codes were developed by the U.S. Army Corps of Engineers. These codes have been standardized to reflect the hierarchical structure of the Standard International Trade Classification Codes, but are focused on commodities that are most likely to utilize water transport. These codes are closely tied to HS codes.
- **U.S. Department of Agriculture (USDA) Crop Report Codes.** These are detailed codes with classification schemes that differentiate specific crops (e.g., winter wheat versus spring wheat versus durum wheat). USDA Crop Report codes can be easily rolled into other commodity classification codes.
- **North American Industry Classification System (NAICS)—**NAICS is a collaborative effort by Mexico's Instituto Nacional de Estadística, Geografía Informática (INEGI), Statistics Canada, and the U.S. Office of Management and Budget. The system is designed to be compatible with the United Nations Statistical Office's International Standard Industrial Classification System (ISIC), and it was designed to replace the Standard Industry Code (SIC) codes which were previously utilized. Updated NAICS versions are released every 5 years. As noted in its definition, NAICS is not actually a commodity classification system but rather is an industrial classification system. However, some analysts have used it to classify commodities based on the idea that the industry code can be used to represent the primary product produced by each industry. This makes the linkage between industry data and commodity data more transparent for some users. The research team for the *Guidebook* discourages use of the NAICS system for the classification of commodities.

Other freight flow data sources such as the Port Import-Export Reporting Series (PIERS) and the Census Foreign Trade Data Bureau also have their own classification systems. These classification systems are generally based off the more commonly used codes, but are tailored to specific uses. New commodity classifications are encountered (for example, specialized classifications developed by trade associations to represent their products) when incorporating local data sources. Similarly, when conducting surveys of the private sector, researchers may find that each industry (and sometimes each company) has a unique method of classifying the commodities they move.

When conducting local surveys to collect subnational commodity flow data, researchers may find it most useful to collect data and present the final results in SCTG format so that the data can be compared with CFS and FAF data. CFS and FAF data may also be used in state and regional planning to provide control totals; thus, having consistency in the commodity classification scheme is important. However, there is the potential to use other commodity classification codes for specific data collection efforts. These efforts can include surveys focused on the railroads or surveys in which the survey sample of companies already has been predefined into a different system that does not easily translate to SCTG.

If different sources of data are going to be used to develop a subnational commodity flow database, they may incorporate various different classification systems. In this case, a methodology will have to be developed to bridge from one classification system to the other. Developing these bridge tables must be tailored to the commodity detail that is desired in the final database and the commodity detail that is available from existing sources. Inevitably, there is some loss of accuracy in utilizing a bridge table. For example, in bridging from STCC 24 (Lumber or Wood Products) to SCTG, there are two potential commodity codes that can be utilized: SCTG 25 (Logs and other Wood in the Rough) and SCTG 26 (Wood Products). While a bridge table can be improved by using commodity detail with more specificity (e.g., three- and four-digit commodity

Table 1.4. Commodities produced by the food manufacturing industry.

Industry	$ Millions (2002)	Percent Total
Food Products	447,676	99%
Basic Chemicals	2,233	0.5%
Beverage Products	1,183	0.3%
Pharmaceuticals and Medicines	555	0.1%
Soaps, Cleaning Compounds, and Toiletries	191	0.04%
Other Chemical Products	38	0.01%
Plastics and Rubber Products	27	0.01%
Yarn, Fabrics, and Other Textile Mill Products	22	0.005%
Converted Paper Products	11	0.002%

Source: BEA Use Table after Redefinitions, 2002.

codes), the usefulness of more specificity will be limited by the amount of detail available in the existing data sources.

Relationship Between Commodities and Industries

Conducting local establishment surveys typically includes collecting information on both outbound and inbound freight flows for individual companies. To relate outbound and inbound flows, the relationship between industries and commodities needs to be well understood. Generally, industries produce commodities that are generally consistent with their industrial classification—textile mill companies produce textile products and food product companies produce food products. Table 1.4 shows that essentially all commodities (99 percent) produced by the food manufacturing industry are food manufacturing products, and Table 1.5 shows that 80 percent of commodities produced by the textile mill industry are textile mill products.

However, it is common for industries to consume products within the industry as well as products from outside the industry. For example, the food manufacturing industry consumes not only food products (34 percent), but also animal products (31 percent) and crop products

Table 1.5. Commodities produced by the textile mill industry.

Industry	$ Millions (2002)	Percent Total
Yarn, Fabrics, and Other Textile Mill Products	35,724	80%
Resins, Rubber, and Artificial Fibers	4,774	11%
Nonapparel Textile Products	2,441	5%
Basic Chemicals	526	1%
Plastics and Rubber Products	492	1%
Apparel	318	1%
Medical Equipment and Supplies	89	0.2%
Printed Products	63	0.1%
Paints, Coatings, and Adhesives	61	0.1%
Soaps, Cleaning Compounds, and Toiletries	59	0.1%
Other Miscellaneous Manufactured Products	48	0.1%
Converted Paper Products	39	0.1%
Furniture and Related Products	35	0.1%
Scrap, Used, and Secondhand Goods	14	0.03%

Source: BEA Use Table after Redefinitions, 2002.

Table 1.6. Commodities consumed by the food manufacturing industry.

Commodities	$ Millions (2002)	Percent Total
Food Products	83,159	34%
Animal Products	74,883	31%
Crop Products	37,559	15%
Converted Paper Products	16,445	7%
Plastics and Rubber Products	8,445	3%
Boilers, Tanks, and Shipping Containers	3,974	2%
Fish and Other Nonfarm Animals	3,835	2%
Basic Chemicals	1,276	1%
Other Commodities	7,775	6%

Source: BEA Use Table after Redefinitions, 2002 (excluding service sectors).

(15 percent). It also consumes products that are in completely different industry sectors, such as converted paper products, plastics, and rubber products as well as many others. This contrasts significantly with the commodities it produces, which essentially all fall under the food products category, as seen above. In addition, the textile mill industry consumes 38 percent of textile products, 35 percent of resins and fibers, as well as more than 20 percent of products from other industries, including crop products, basic chemicals and so on. Tables 1.6 and 1.7 show the distribution of goods consumed by the food manufacturing and textile mill industries. A comparison of Tables 1.4 and 1.5 with Tables 1.6 and 1.7 shows how the goods consumed by the food manufacturing and textile mill industries differ from the industries themselves according to make and use tables produced by the U.S. Department of Commerce Bureau of Economic Analysis (BEA).

Therefore, in collecting subnational commodity flow data for specific industries, it is important to understand who the producers and consumers of a specific commodity are. For example, to develop a subnational commodity flow database of the food manufacturing industry, it would not only be necessary to survey companies in food manufacturing, it would also be necessary to survey companies in industries that produce animal products, crop products, and paper products on the inbound side. At the metropolitan and state level, it is likely that several of the producers of a key commodity will be located outside of the jurisdiction of the transportation planning agency. Therefore, it would likely be necessary to survey companies that consume the commodity and are located within the agency's jurisdiction.

Table 1.7. Commodities consumed by the textile mill industry.

Industry	$ Millions (2002)	Percent Total
Yarn, Fabrics, and Other Textile Mill Products	8,020	38%
Resins, Rubber, and Artificial Fibers	7,416	35%
Crop Products	1,725	8%
Basic Chemicals	914	4%
Soaps, Cleaning Compounds, and Toiletries	616	3%
Semiconductors and Electronic Components	508	2%
Other Fabricated Metal Products	360	2%
Apparel	317	2%
Converted Paper Products	259	1%
Animal Products	246	1%
Nonmetallic Mineral Products	226	1%
Plastics and Rubber Products	178	1%
Paints, Coatings, and Adhesives	164	1%

Source: BEA Use Table after Redefinitions, 2002 (excluding service sectors).

CHAPTER 2.0

Collecting Subnational Commodity Flow Data Using Establishment Surveys

2.1 Introduction

This section provides an examination of how to develop subnational commodity flow data using establishment surveys. Establishment surveys are conducted by interviewing representatives of specific businesses and gathering information about freight flows in and out of specific physical establishments. The *Guidebook* identifies the following 10 general steps that need to be addressed in administering an establishment survey data collection program:

- Step 1—Geographic boundary of concern
- Step 2—Industry/commodity classification scheme
- Step 3—Universe of companies to survey
- Step 4—Determining sample size
- Step 5—Establishing data elements
- Step 6—Survey questionnaire
- Step 7—Conducting the survey
- Step 8—Database assembly
- Step 9—Data expansion
- Step 10—Data accuracy and validation

Many of these steps are interrelated, but the *Guidebook* discussion of each step is ordered as shown in the list above. The description of each step is structured to focus on the following four key elements described in the Playbook (Chapter 6.0):

- **Key Considerations—**A brief description of the main issues encountered and tradeoffs that will need to be made for the step.
- **Implementation Process—**A detailed description of how to implement the step.
- **Example—**An example of how this step has been implemented in other studies. Many of the examples in this chapter are taken from demonstration establishment surveys conducted in Seattle and Spokane as part of NCFRP Project 20.
- **User's Guide Worksheet Punch List—**Simple bulleted instructions that *Guidebook* users can check off to ensure that they have implemented each of the major steps involved in developing an establishment survey.

Each of these four elements is designed to focus on different aspects of conducting an establishment survey and to reflect the types of activities that might be undertaken by a state or local transportation agency. For transportation agencies that are considering hiring a contractor to develop an establishment survey, reading the "Key Considerations" section of each step will likely provide enough information for the generation of a request for proposals (RFP) on the topic. Transportation agencies that want to understand the details of how to conduct an establishment survey should focus on the "Implementation Process" sections in addition to the "Example" section. The "Example" section also will provide specific references to efforts in other regions that can be compared with what already has been done in the agency's region and with responses to RFPs submitted to the agency. After transportation agencies have a sufficient background in all of the aspects related to developing an establishment survey, the "User's Guide Worksheet Punch List" sections can be used to walk the agency through all of the specific steps that need to be done to implement the survey. This section can also be used to compare to a response to RFPs submitted to the agency to determine the completeness of the submittals.

2.2 Step-by-Step Process for Conducting Establishment Surveys

This section provides a comprehensive examination of the steps involved in developing subnational freight data using establishment surveys. These examples are designed to provide a detailed description of all of the necessary steps, including addressing relevant implementation issues, that a typical regional planning office, local agency, or state department of transportation may experience when considering and implementing future establishment survey efforts.

As part of the research conducted to develop the *Guidebook,* two pilot-scale demonstrations were conducted of actual establishment surveys. The surveys were conducted in the Seattle metropolitan region and in Spokane (both in Washington state). These demonstrations also tested a number of aspects of the methods described in order to advance the state of practice for subnational commodity flow surveys. Several of the examples provided in this chapter are based on the pilot-scale demonstration surveys (also referred to as the demonstration survey in this chapter).

Step 1—Geographic Boundary of Concern

Key Considerations

Typically, the study area of an establishment survey is consistent with the jurisdiction of the agency conducting the survey. So a state DOT is obtaining information for an entire state and an MPO is obtaining information in their metropolitan region. However, the definition of subregions of concern within the study area and the definition of external regions from/to which commodities may be shipped to/from the region have ramifications throughout the survey process. For example, the definition of these regions becomes one dimension of the sampling matrix. The more geographic zones there are from a sampling perspective, the larger the number of cells that will need to be filled, and the more surveys that will need to be collected.

However, if the defined geographic subregions are very large, it decreases the ability to relate the collected survey data to the local freight network if any of the applications of the data involve routing the flows on the local freight network. It will be especially difficult to determine the network elements used to access other subregions within the geographic area of concern. Similarly, size considerations need to be used in the definition of external regions.

An additional consideration will be features that are unique to the local area of concern. For example, there may be specific subregions that are important either because they are known to be freight hubs or because they include key freight facilities of concern. Ideally, subregions will be defined based on boundaries that create unique land use or population characteristics within a region.

Implementation Process

The first step in this process is to define the overall boundary of concern for the establishment survey. Typically, this will coincide with the jurisdiction of the agency conducting the survey. Based on the process used for conducting establishment surveys, it will be easier for an MPO region to define its boundaries at the county level to streamline the data expansion and validation processes. However, if county-level boundaries are not feasible or more geographic focus is required for a particular commodity study, there are supplemental estimation methods that can be incorporated at later steps to allow for subcounty estimates.

The next step is to identify the number of subregions within the boundary of concern. Some of the subregions to consider include TAZ, city, county, BEA zone, metropolitan statistical area, or BTS CFS zone. The selection of zones should be based on the types of planning activities that are likely to occur in the region along with the boundaries of other data sets that the transportation agency utilizes for other activities. If the purpose of the establishment survey is to feed into a MPO-level travel demand model, then it is likely that the smallest feasible geographic unit will be utilized. For example, there are a substantial number of county-level employment and sales data, number of establishments data, and commodity production data that are published nationally or compiled at the MPO level, but subcounty data are more difficult to find. As these county-level data sets may be used for establishing control totals for databases, inputs to estimate missing data, and expansion factors for the survey itself, collecting data at the county level is a very attractive option.

Developing a freight flow database at very refined geographic levels (such as zip code or TAZ) is sometimes desired. One of the biggest issues with developing these databases is the difficulty in getting responses to survey questions at this level of detail. Often this level of detail in an establishment survey is considered proprietary by the survey respondent or too time consuming to realistically include in the survey effort.

If the purpose of the establishment survey is to better understand the flows of a specific commodity, then geographic zones should be determined based on previous information regarding production locations of the commodity. In other situations, the subregions will be defined based on regions that are known to be freight hubs within larger geographic areas.

The final step is to define the boundaries for the external regions. Options for defining external regions are numerous. A few of the common approaches include the following:

- A single external region that tracks interregional flows relative to intraregional flows.
- External regions defined in terms of the direction(s) in which freight leaves and arrives in the region—north, south, east, or west. This approach can be complicated because it implies knowledge of the routing patterns of the flows.
- Regions that are mapped to corridors that are thought to capture large portions of external traffic. For example, external regions can be developed based on likely destinations for truck traffic as it leaves the metro area.
- Regions that are mapped to specific clusters of economic activity at the metropolitan, state, or multistate level. This can include regions such as BEA zones, individual state, Northeastern United States, the Midwest, the West Coast, etc.

Examples

As noted above, there are several options for defining geographic regions for an establishment survey. For the demonstration survey conducted in Seattle and Spokane, the counties that most closely aligned with the MPO boundaries were chosen as the regions of interest. The study area for intraregional flows was defined as consistent with the MPO boundaries, and the internal zones were defined as counties within the MPO boundary. For the Seattle region, this included King, Pierce, and Snohomish counties. For the Spokane region, Spokane County was the sole internal geographic unit. For both surveys, the external region was defined as a single region (i.e., all flows to/from outside Seattle or Spokane were lumped in a single external geographic zone). This was in large part due to the limited sample that would be taken for the demonstration survey.

In 2005, the Georgia DOT conducted establishment surveys of warehouse and distribution facilities in the Port of Savannah subregion. For these surveys, the study area for intraregional flows was defined as the area bordered by I-95, I-16, and the Savannah intercoastal waterway. This study area was then divided into internal zones based on pods or clusters of container activity. The external regions were defined as north, south, west, and east. These zones were not consistent with any existing set of zonal boundaries, but were created especially for the study. This way of defining external regions required respondents to the survey to provide street addresses for where shipments were coming from or going to. Then, the Georgia DOT, using a geographic information system (GIS), placed the flows in one of the four external regions. The north and south trucks were assumed to leave the region using I-95. The west trucks were assumed to leave the region on I-16. The trucks going to the northeast were going to the Savannah port terminals. Figure 2.1 shows a general schematic of truck trip directions and how they were assigned to external regions from what was termed the "West Pod" in the Savannah port subregion. This method was successful in estimating the number of trucks on each of the subregion major freight highway corridors.

Zones also can be set up at the zip code level. Figure 2.2 shows the zip code zones for Dane County in Wisconsin. Zones set up at the zip-code level are convenient for measuring freight activity, because this level provides submetro-level detail without the need for populating the hundreds of zones that are often used for passenger vehicle zone systems. Figure 2.3 shows a zone system for the Seattle metropolitan region. This is similar to the zone system that is discussed in the detailed example discussed later in this chapter.

A hypothetical example of how to lay out external regions would have to be consistent with the BTS CFS. This would allow a comparison of results between the two establishment survey efforts. For example, if the state of Tennessee were to conduct an establishment survey, it could define the state's counties as internal regions. For external regions, it could use boundaries of aggregated FAF regions, such as those shown in Figure 2.4. Assuming that the four zones within Tennessee are the internal zones and the five zones outside of Tennessee are external zones, the number of origin-destination combinations would be calculated as 4 multiplied by 5, or 20.

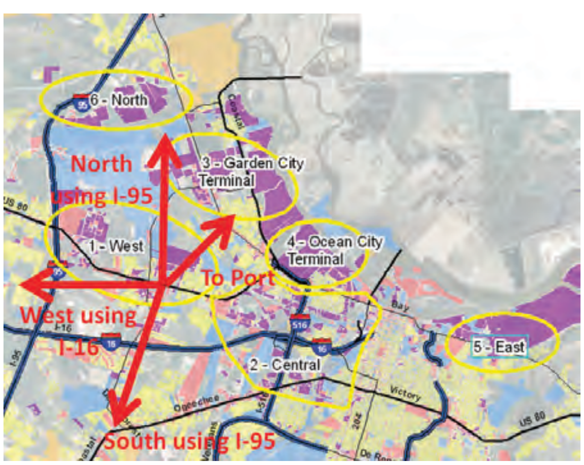


Figure 2.1. Matching directions to external regions in Savannah.

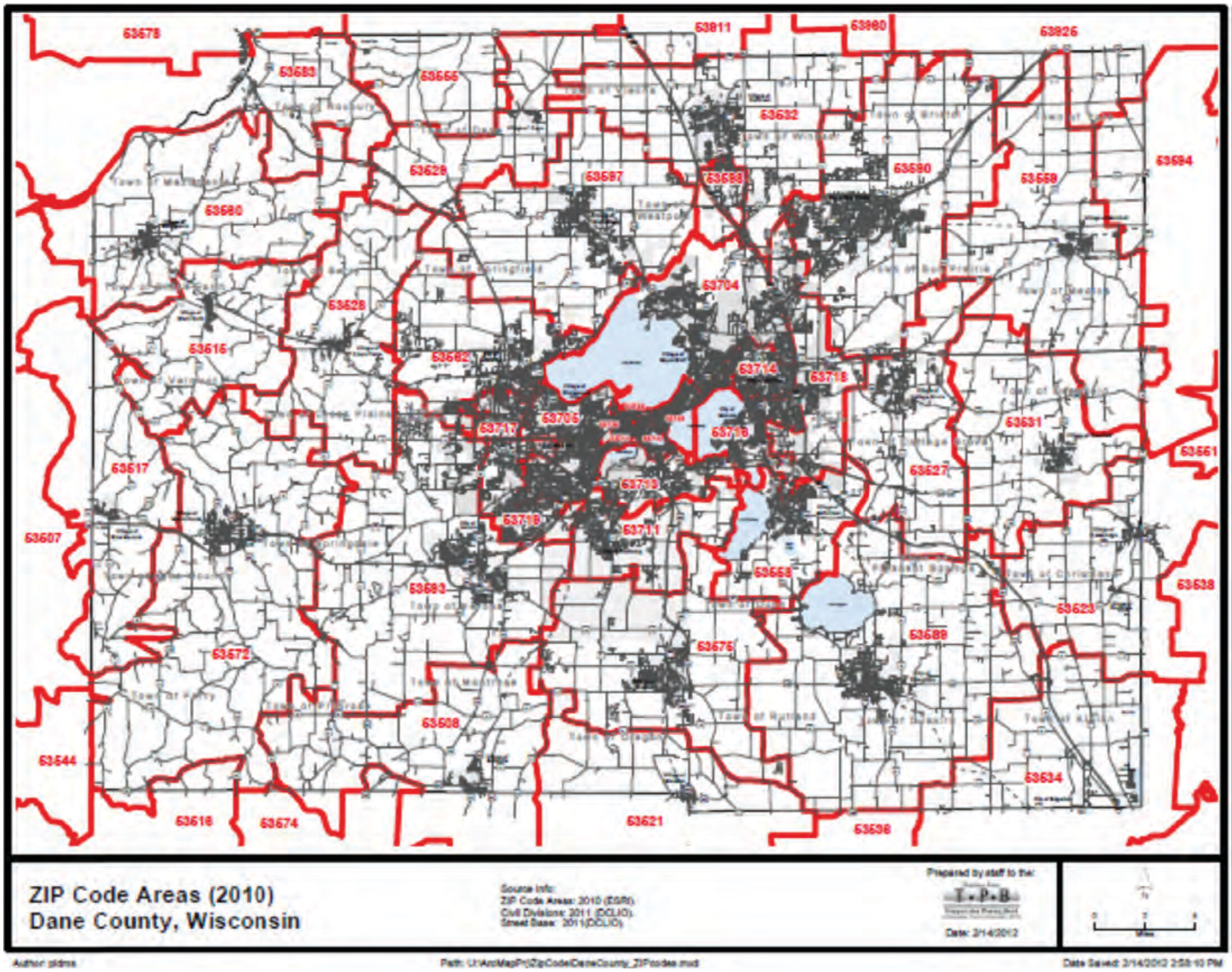


Figure 2.2. Zip-code-level map for Dane County (Wisconsin).

Another hypothetical example could be considered for an MPO region that was developing a travel demand model. This region could collect establishment data internal to its region at the zip code level and then make external regions consistent with the corridors that are commonly used for external trips. Using one of the disaggregation methods discussed in Chapter 4.0, the zip-code-level data could be later disaggregated to the TAZ level needed to feed into the travel demand model.

User's Guide Worksheet Punch List

- Determine your study area for the establishment survey.
- Identify the boundaries for internal regions for the study area.
- Identify the boundaries for the external regions for the study area.
- Calculate the number of unique origin-destination pairs in your region. This is the number of internal regions multiplied by the number of external regions.

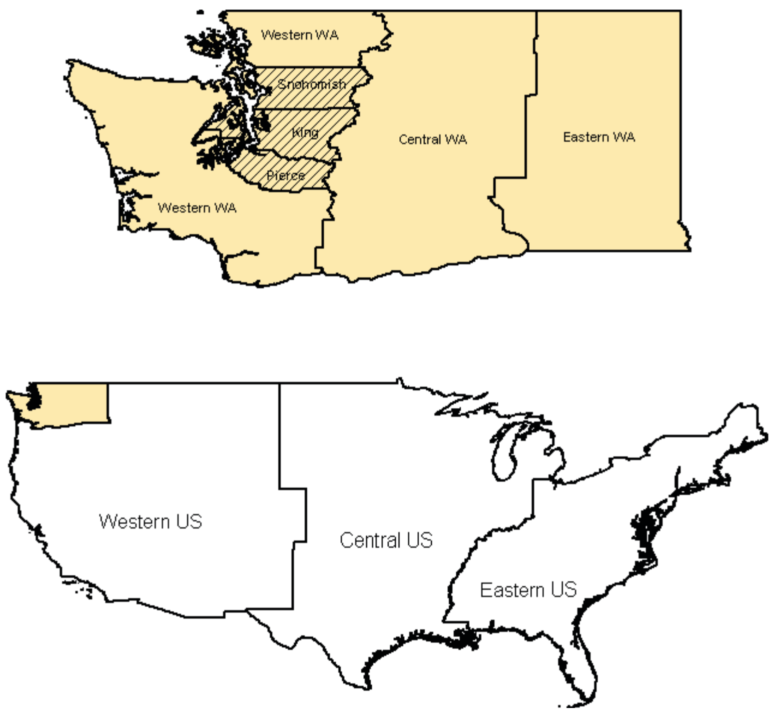
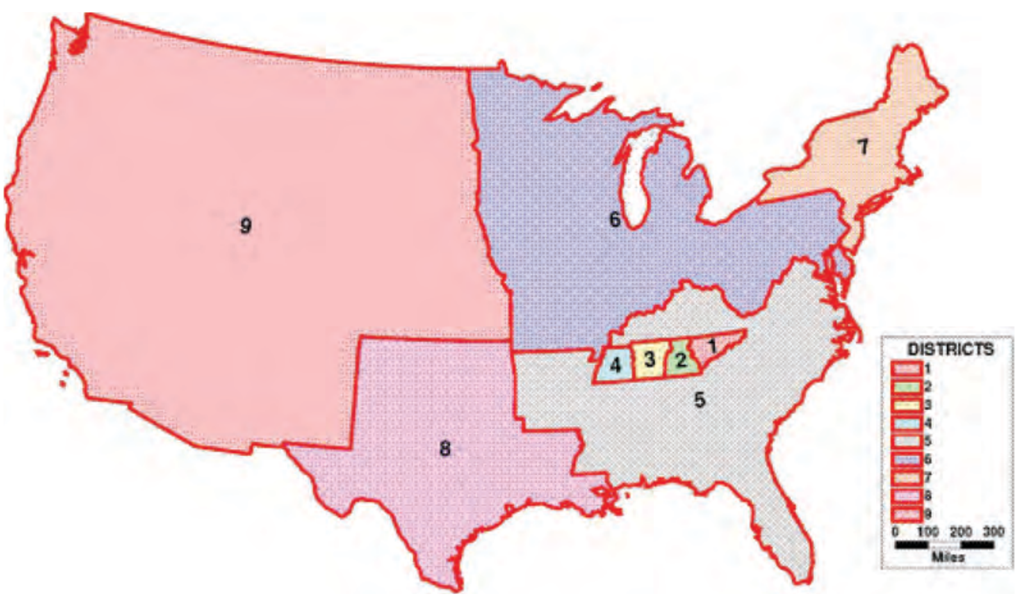


Figure 2.3. Zone system for Seattle metropolitan region establishment survey.



Source: (Cambridge Systematics, Inc. 2007)

Figure 2.4. Tennessee example of aggregated FAF regions used to define external regions.

Step 2—Industry/Commodity Classification Scheme

Key Considerations

The topic of commodity classification schemes and their relationship to industry classification was introduced in Chapter 1.0. For an establishment survey, both an industry classification system and a commodity classification system need to be adopted. The industry classification system will be used to determine which companies to survey and to obtain information about these companies/industries when expanding the survey data (see Step 9). Commodity classification is based on which commodities are of interest, the specific classification system (which will be selected on the basis of whether the data need to be combined with or compared to another data set that uses a particular classification scheme), and how recognizable the commodity classifications are to potential survey respondents. For example, it has already been noted that it may be useful to classify commodities using the SCTG system that has been adopted by both the CFS and FAF so that it will be possible to use these two databases for control totals at an aggregate level of geographic detail (e.g., FAF data on commodity flows may be available for an entire MPO region and the survey is attempting to collect data for counties within the region) or to fill gaps in the survey data. To determine how well the SCTG codes will work during data collection, it might be useful to conduct a pilot survey or outreach interviews with representatives of the industries to be surveyed asking them to identify the products/commodities they ship and receive and then see how easily these "self-classifications" can be converted into the commodity classification system that is chosen.

It also will be important to understand the relationship between commodity classifications and industry classifications. As mentioned in Chapter 1.0, an input-output model's "make-use" table can be very helpful in making this connection because it shows what industries make and use each commodity. An example of a "make-use" table is shown in Table 2.8. If an establishment survey is going to focus on only a few very important commodities in the region, it will be important to know which industries produce and consume those products so that the proper companies are surveyed. It is useful if commodities are classified in a manner that makes it easy to see which industries produce which commodities. Additionally, it will be necessary to consider how to classify companies that are in multiple industries and/or produce multiple types of goods.

The next consideration is the level of detail to include in the industry/commodity classification scheme. There is a tradeoff between the amount of commodity detail that is included in the survey responses and the sample size needed to accomplish the survey design and therefore the resources needed to implement the survey. It also is possible that there are particular industries or commodities of concern to a local area. These can include industries that employ a large number of people in the local area, industries that produce a high dollar value of goods in the local area, or commodities whose production is known to be especially sensitive to the performance of the transportation system (for example, those that are involved in "just-in-time" supply chains). Ideally, these specialized industries or commodities will be considered as a unique classification to maximize the potential of the establishment survey to estimate its freight flows.

> Another consideration is the industry or commodity classification of the database that will be used to identify companies to survey. To the extent that the chosen classification scheme is consistent with the scheme used to categorize companies in this database, it will be easier to select companies that are consistent with the desired sample for each industry or commodity.

Implementation Process

The first step is to review some of the industry and commodity codes that are typically used for establishment surveys. Many freight flow databases, including CFS and FAF, are based on the SCTG at the two-digit level. Other commonly used commodity codes include the STCC and the HS Code. Commonly used industry codes include the North NAICS and the SIC. SCTG codes and STCCs are listed and described in Tables 2.1 and 2.2, respectively.

The next step is to determine whether the survey will attempt to cover all industries and commodities in the study area or only a subset of industries or commodities. If the survey is industry/commodity specific, then special care should be taken to ensure that the industries and/or commodities are well defined. Additionally, both the industries that produce and consume a commodity should be considered for incorporation into the survey. For example, the apparel industry generates several inbound commodity flows due to the commodities that it consumes in its manufacturing process. The commodities that it consumes include yarn, textiles, apparel, and even printed matter. Table 2.3 shows a range of commodities consumed by select industries based on input-output data developed by BEA. Following the determination of what industries and commodities will be included in the survey, a classification scheme can be chosen. If the survey covers all industries and commodities, then consistency with national freight flow data sets, business listing databases, and other local freight data sets is a more important consideration.

Note that the selection of an industry/classification scheme will be revisited following consideration of sampling issues in Step 4.

Example

Only a small number of industries were included in the demonstration survey of both the Seattle and Spokane regions. The two primary criteria for selecting industries were (1) the importance of an industry to each region's economy and (2) the need to demonstrate commodity shipment patterns that are expected to be important to regional economies but that may not be well captured in the CFS or FAF.

To identify the largest industries in each region, the U.S. Census County Business Patterns (CBP) database was used to determine number of employees, annual payroll, and number of establishments by industry sector. For the Seattle region, the transportation equipment manufacturing industry was found to be the largest manufacturing sector in terms of both employment and economic output measured in dollars. The selection of the second industry to survey in the Seattle region was targeted towards an industry with heavy reliance on the Port of Seattle. This decision was based on the exclusion of imported freight flows in the CFS survey process. Data were examined related to containerized import commodities at the Port of Seattle. Additionally, FAF data were reviewed for the Puget Sound region, and the determination was made that apparel is a major commodity moving through the port and also a major consumer retail commodity for the region. Due to the existing commodity classification coding options for

Table 2.1. Two-digit SCTG codes and descriptions.

SCTG Code	Description
01	Live Animals and Fish
02	Cereal Grains (including seed)
03	Other Agricultural Products, except for Animal Feed
04	Animal Feed and Products of Animal Origin, n.e.c.
05	Meat, Fish, and Seafood, and Their Preparations
06	Milled Grain Products and Preparations, and Bakery Products
07	Other Prepared Foodstuffs, and Fats and Oils
08	Alcoholic Beverages
09	Tobacco Products
10	Monumental or Building Stone
11	Natural Sands
12	Gravel and Crushed Stone
13	Nonmetallic Minerals, n.e.c.
14	Metallic Ores and Concentrates
15	Coal
16	Crude Petroleum Oil
17	Gasoline and Aviation Turbine Fuel
18	Fuel Oils
19	Coal and Petroleum Products, n.e.c.
20	Basic Chemicals
21	Pharmaceutical Products
22	Fertilizers
23	Chemical Products and Preparations, n.e.c.
24	Plastics and Rubber
25	Logs and Other Wood in the Rough
26	Wood Products
27	Pulp, Newsprint, Paper, and Paperboard
28	Paper or Paperboard Articles
29	Printed Products
30	Textiles, Leather, and Articles of Textiles or Leather
31	Nonmetallic Mineral Products
32	Base Metal in Primary or Semifinished Forms and in Finished Basic Shapes
33	Articles of Base Metal
34	Machinery
35	Electronic and Other Electrical Equipment and Components, and Office Equipment
36	Motorized and Other Vehicles (including parts)
37	Transportation Equipment, n.e.c.
38	Precision Instruments and Apparatus
39	Furniture, Mattresses and Mattress Supports, Lamps, Lighting Fittings, and Illuminated Signs
40	Miscellaneous Manufactured Products
41	Waste and Scrap
43	Mixed Freight

n.e.c. = not elsewhere classified

Table 2.2. Two-digit STCCs and descriptions.

STCC	Description
01	Farm Products
08	Forest Products
09	Fresh Fish or Other Marine Products
10	Metallic Ores
11	Coal
13	Crude Petroleum, Natural Gas or Gasoline
14	Nonmetallic Minerals; except Fuels
19	Ordnance or Accessories
20	Food or Kindred Products
21	Tobacco Products; except Insecticides – see Major Industry Group 28
22	Textile Mill Products
23	Apparel, or Other Finished Textile Products or Knit Apparel
24	Lumber or Wood Products; except Furniture – see Major Industry Group 25
25	Furniture or Fixtures
26	Pulp, Paper or Allied Products
27	Printed Matter
28	Chemicals or Allied Products
29	Petroleum or Coal Products
30	Rubber or Miscellaneous Plastics Products
31	Leather or Leather Products
32	Clay, Concrete, Glass or Stone Products
33	Primary Metal Products, including Galvanized; except Coating or other Allied Processing
34	Fabricated Metal Products; except Ordnance – see Major Industry Groups 19, 35, 36 or 37
35	Machinery; except Electrical – see Major Industry Group 36
36	Electrical Machinery, Equipment or Supplies
37	Transportation Equipment
04	Business Services Division
38	Instruments, Photographic Goods, Optical Goods, Watches or Clocks
39	Miscellaneous Products of Manufacturing
40	Waste or Scrap Materials Not Identified by Producing Industry
41	Miscellaneous Freight Shipments
42	Containers, Carriers or Devices, Shipping, Returned Empty
43	Mail, Express or Other Contract Traffic
44	Freight Forwarder Traffic
45	Shipper Association or Similar Traffic
46	Miscellaneous Mixed Shipments
47	Small Packaged Freight Shipments
48	Hazardous Wastes
49	Hazardous Materials
50	Bulk Commodity Shipments in Boxcars

Table 2.3. Use table for select industries.

Distribution of Select Commodities Consumed by Industries	Select Industries			
	Crop Prod	Food Mfg	Apparel Mfg	Wood Product Mfg
Crop products	8,063.3	37,533.7	0.0	92.0
Animal products	368.6	74,690.5	0.0	0.0
Forestry and logging products	0.0	239.3	5.4	16,468.2
Fish and other nonfarm animals	0.0	3,835.3	0.0	0.0
Support activities for agriculture and forestry	10,760.9	0.0	0.0	0.0
Oil and gas extraction	0.0	5.1	0.0	5.1
Coal mining	0.0	279.2	1.6	2.0
Metal ores mining	0.0	0.0	0.0	0.0
Nonmetallic mineral mining and quarrying	506.9	0.0	0.0	0.0
Mining support services	0.0	0.0	0.0	0.0
Food products	0.0	79,681.4	0.0	9.5
Beverage products	0.0	101.4	0.0	0.0
Tobacco products	0.0	0.0	0.0	0.0
Yarn, fabrics, and other textile mill products	28.5	0.0	9,064.3	371.7
Nonapparel textile products	218.2	283.2	526.0	229.9
Apparel	0.0	0.0	2,292.0	0.0
Leather and allied products	0.0	19.4	155.0	8.9
Wood products	440.1	72.9	0.0	18,274.7
Pulp, paper, and paperboard	0.0	555.9	20.7	113.0
Converted paper products	300.7	16,353.3	69.3	267.7
Printed products	23.4	248.5	245.1	26.2
Petroleum and coal products	4,476.4	722.9	24.1	238.0
Basic chemicals	892.9	1,275.3	16.9	228.6
Resins, rubber, and artificial fibers	0.0	250.7	108.6	454.1
Agricultural chemicals	7,897.3	0.0	0.0	0.0
Pharmaceuticals and medicines	0.0	908.1	0.0	13.5
Paints, coatings, and adhesives	5.9	25.0	0.0	409.5
Soaps, cleaning compounds, and toiletries	4.1	391.0	0.0	86.7
Other chemical products	40.5	1,075.4	7.5	80.4
Plastics and rubber products	737.9	8,393.9	36.1	781.6
Nonmetallic mineral products	14.0	1,097.8	0.0	966.3
Warehousing and storage	548.5	626.7	101.8	185.8

apparel, apparel manufacturing was the industry that was surveyed in the demonstration survey. But much of the apparel that is moved into and out of the Seattle region (and all of the apparel imported through the Port of Seattle) is actually flowing to businesses other than apparel manufacturing (such as apparel wholesalers and retailers). So it is important to identify these additional industries for sampling inbound commodity flows. The identification of these industries (that purchase apparel) involves the “make-use” table.

For the Spokane region, the largest manufacturing industry in the region in terms of employment and number of establishments was the fabricated metal manufacturing industry (NAICS 332). Another large industry was the food manufacturing industry (NAICS 311). This industry was selected because Spokane is the largest metro area in Eastern Washington, and food manufacturing in an urban area surrounded by rural agricultural regions may exhibit some unique commodity flow patterns. Additionally, the differences in produced and consumed commodities highlight one of the key differences between national and regional establishment surveys.

The final industries selected for the demonstration survey were the following:

- **Transportation Equipment Manufacturing (Seattle).** The major commodity produced by this industry is easily classified in the SCTG system as Transportation Equipment (SCTG code 37).
- **Apparel Manufacturing (Seattle).** Examining the commodity classification coding options for apparel, either the STCC classification (STCC 23, Apparel) or SCTG 304 seem to be appropriate.
- **Food Manufacturing (Spokane).** The STCC classification system provides a simpler bridge of commodity classifications for the outbound shipments of this industry (STCC 20, Food, and Kindred Products).
- **Fabricated Metal Products Manufacturing (Spokane).** The primary commodities shipped outbound by this industry can be classified as SCTG code 33, Articles of Base Metals.

User's Guide Worksheet Punch List

- Review the most commonly used industry and commodity codes.
- Review any existing state and local freight flow databases. Note the classification schemes that are used.
- Determine if the local establishment survey will develop a final database based on industry or commodity codes.
- Review available economic data for the study area. This should include both employment and output data by industry. The U.S. Economic Census is a commonly used database for this type of data. State and regional economic development agencies often have industry-specific economic data as well.
- Determine if the local establishment survey will focus on a subset of industries and commodities or a full set.
- Determine which classification scheme to utilize for the establishment survey. Utilize two-digit SCTG codes as a default unless the responses to the bullets listed above indicate otherwise.

Step 3—Universe of Companies to Survey

Key Considerations

Identifying the universe of companies to survey can involve use of preexisting business databases, information from local trade associations, and local knowledge of major companies in the region. Business databases are proprietary lists of companies that can be purchased for developing comprehensive lists of businesses. However, these databases are compiled through a variety of different sources with businesses categorized in different fashions. Therefore, it is important to review these databases with other local sources to ensure that at least the major known companies in key industries are included. Trade associations, along with staff at economic development agencies or chambers of commerce, may also be able to assist in this review.

State and federal labor agencies typically maintain a list of firms participating in unemployment insurance programs, which may be a good source of business data. Licensing agencies may also maintain lists of companies in certain industries. However, this information may not be made available to the agency conducting the survey due to confidentiality restrictions.

It also is important to capture as much information as possible about each company in the list of potential companies to survey. This will assist in stratification, as there is typically a desire to oversample companies that are larger. Larger companies are typically identified through either the size of the facility, number of employees, sales at the location, or local knowledge.

Implementation Process

The survey effort will need an accurate list of firms in the region. There are a number of sources for the necessary data elements, including names, addresses, contact person, and phone number, with some sources able to provide additional administrative information, such as on-site activities (e.g., NAICS code designations), number of employees, and value of shipments. One source may be from another local agency. For example, economic development agencies assemble lists of firms, and labor agencies have a list of firms participating in unemployment insurance programs (e.g., "Quarterly Census of Employment and Wages" data) or have assembled a list for a particular project. Trade associations will have a list of their members, and licensing agencies have lists of firms as well. If agencies have sufficient interest in improving and maintaining the integrity of their databases, it may be possible to use their data and update any out-of-date information using a memorandum of understanding (MOU) or data-sharing agreement.

If no local sources are available or no local partners are willing to share their lists, there are vendors who specialize in lists of firms for any metropolitan area. Contracts regarding the sharing of these data contain some restrictions; however, with special documentation and confidentiality agreements, it is possible for a state DOT to purchase the list of all firms in its state and share these data with MPOs and researchers. Vendors charge varying prices for the fields, including specific contact information and geocoding information. Data elements should include company name, address, industry sector, number of employees, value of shipments, latitude/longitude of physical address, and contact person(s) phone number(s)/e-mail(s).

The CBP database, produced annually by the U.S. Bureau of the Census, is extracted from the Business Register, which has the most complete, current, and consistent data for business establishments, but the CBP database only publishes aggregate statistics about businesses in each U.S. county and does not provide names of individual establishments. InfoUSA and Dun & Bradstreet data are obtained from less systematic information-gathering processes. Sources include business licenses, trade associations, phone book directories, and other proprietary sources that are identified by these companies. In addition, the categorization of businesses may be somewhat different between these two sources and relative to the CBP data. If the survey budget allows, it is recommended to purchase data from both of these companies to obtain as comprehensive a list of companies as possible. This will provide as exhaustive a list of potential companies to survey as possible. This is especially important given the relatively low response rate to establishment surveys. If a database like InfoUSA or Dun & Bradstreet is used, it will be useful to compare information such as the number of establishments and the size distribution of establishments as reported in CBP with similar aggregate statistics from the commercial databases to detect any systematic biases in the commercial databases (for example, under-representation of small companies). If these types of biases are detected, it will be useful to work with trade associations and the local business community to identify firms that are under-represented to include in the sampling frame.

Using geocoded firm locations in a GIS environment makes it possible to build a sampling frame with geographic specificity rather than choosing a random sample of firms. The information developed from the surveying effort is intended to be descriptive statistics rather than inferential statistics. The collection of descriptive statistics is preferred since it is more easily translatable into a quantifiable commodity flow database. The descriptive statistics need to be coupled with geographic targeting to ensure that specific inputs necessary for transportation planning are available. The data will be representative of the freight activities and freight community perspectives on problems located in specific areas.

Example

For the demonstration surveys in Seattle and Spokane, a list of firms was developed in each of the four industries using purchased establishment data from both Dun & Bradstreet and InfoUSA. For the survey purposes, the following data were received:

- Contact name
- Contact phone number
- Company web site
- Estimated revenue of the company
- Size of establishment by square feet
- Size of establishment by number of employees

All of these data items are useful in collecting establishment survey data. These items also are useful for expanding collected data to represent the full population across an entire industry. However, it should be noted that the quantitative data provided in the commercial databases are typically provided in ranges, so estimates of these data were ultimately developed. For example, rather than reporting the actual employment at an establishment, a range of employees is reported. The U.S. Census CBP data also were used to compare to the purchased databases. Table 2.4 shows the number of firms identified in each industry and geographic region. As shown in Table 2.4, there was wide variability in the number of establishments that were found in each industry by the separate databases. This is likely the result of the different methodologies that are used to identify and define establishments by each of the sources and the frequency and procedures for updating and purging the database of inactive companies. For the vendor-provided data, it also is important to understand how multiple branches, subsidiaries, etc., are captured in the data and how the industry sector is determined (for example, if a company reports itself in multiple industries, how this is reported in the database).

Table 2.4. Number of establishments identified by industry and region.

		Number of Establishments		
Region	**Industry**	County Business Patterns (CBP)	Info USA	Dun & Bradstreet
Seattle	Transportation Equipment Manufacturing	229	263	536
	Apparel Manufacturing	59	16	272
Spokane	Food Manufacturing	50	72	75
	Fabricated Metal Product Manufacturing	112	28	125

User's Guide Worksheet Punch List

- Contact a state's labor association, licensing authorities, and economic development agencies to determine the availability of business establishment databases in the study area. Capture as much information about individual companies as allowable through these sources.
- Extract establishment count data from U.S. Census Bureau CBP.
- Contact trade associations and chambers of commerce to solicit support for the overall establishment survey and to obtain any available list of companies by industry in the study area.
- Determine whether the list obtained thus far is sufficient for the survey by comparing the CBP count of companies to those identified through the sources listed in the above bullets.
- If needed, purchase a business list database from one of the proprietary sources. Include information on each business to assist in determining the relative size of the business and industry for each business.
- Develop a comprehensive list by combining information from all sources.
- Review the comprehensive list with local specialists at economic development agencies and local chambers of commerce to ensure that there are no major oversights in the list.

Step 4—Determining Sample Size

Key Considerations

There is always a question as to how large the sample size should be. Even in the case of the national CFS, sample size is significantly impacted by budget constraints. In designing a sampling strategy, it is preferred to establish statistical criteria for the data in advance of the survey. There is a tradeoff between confidence level achieved, the dimensions of the establishment survey, and the number of samples needed to survey. Additionally, data collection costs tend to increase proportionally relative to the number of samples desired. The dimensions of the survey include the geographic zones included in the survey, the commodities included, and the modes included. A survey that intends to cover 40 commodities, 20 geographic zones (a 20 x 20 origin-destination matrix), and 5 modes would have a sampling matrix with 8,000 cells. While not all cells will have meaningful data, a sample with 10 observations in each cell would require data on 80,000 shipments. It is easy to see how sample size can grow as the dimensions of a sampling matrix increase. Following the data collection effort, it may be necessary to aggregate cells from a desired comprehensive matrix to one that achieves the statistical criteria desired for the survey and can still be accomplished within the available budget.

From a pragmatic perspective, sample size considerations are often impacted by constraints. Oftentimes, the budget for a survey effort will be prescribed prior to the survey development process. In these instances, it will be important to understand the statistical confidence levels achievable with survey design elements that are more likely to be under the surveyor's control such as the number of geographic zones and the number of commodities surveyed.

Implementation Process

Sample size is probably the most important determinant of precision for the information collected from the drawn sample, and it is jointly determined by (1) the distribution of a variable in the study population, which is reflected by the variable's mean and standard deviation, and (2) the desired degree of precision and the statistical confidence level with which the analysis needs to be conducted.

For establishment surveys, the number of surveys required to generate sufficient accuracy at a regional level will be far less than the number that would be required at a subregional level such as zip codes. Similarly, the number of surveys needed is impacted by the number of external regions that are desired for the survey process. Therefore, defining the geographic level of concern for both the internal region and external regions is important prior to commencing an establishment survey effort. Generally, internal and external regions should be defined with only enough detail to match the freight planning activities that are being considered by the transportation agency. Additionally, it should be considered that future freight planning efforts that require more detailed data can be accompanied by smaller data collection efforts designed to validate whether processes such as disaggregation can be used as a surrogate for collecting new establishment surveys. Disaggregation is discussed in greater detail in Chapter 4.0.

Sample size also is an important determinant of costs in most data collection efforts. Given the budget constraints of a study, it is important to recognize the tradeoffs among the selected sampling method, the desired levels of precision and statistical confidence, and the corresponding sample sizes.

Given the distribution of the variable values in the population, there are two ways to approach the analysis of sample size questions. The analyst could determine (1) the sample size required to achieve a desired level of precision and statistical confidence (statistical confidence is the probability that an estimated value falls within a specific range) for selected variables of interest or (2) the degree of precision and the confidence level that would be expected for each variable of interest under a range of sample sizes.

The approach of identifying the necessary sample size assumes that a reliable estimate of the variance of key variables is available. This estimate can be checked at the conclusion of the survey to confirm its accuracy. With the variance known, Equations 1 through 3 below can be used to arrive at the sample size estimate.

Equation 1 is the standard normal expression for a sample mean (the mean is the average or expected value of a sample) stating that with 95 percent confidence, the mean will lie within two standard deviations. (The standard deviation is the measure of the variation or dispersion that exists from the mean. A low standard deviation indicates that data points are located close to the mean, while a high standard deviation indicates that data points are located relatively far from the mean):

$$\left(\bar{x}-\frac{2\sigma}{\sqrt{n}},\ \bar{x}+\frac{2\sigma}{\sqrt{n}}\right) \qquad \text{(Eq. 1)}$$

Equation 2 is derived from the first equation:

$$\frac{4\sigma}{\sqrt{n}}=W \qquad \text{(Eq. 2)}$$

Equation 3 is achieved by solving for the sample size, n:

$$n=16\sigma^2/W^2 \qquad \text{(Eq. 3)}$$

W is the width in units of the confidence interval. So the wider the confidence interval, the lower the sample size necessary to maintain the 95 percent confidence level. The mean is represented by $\bar{x}$, and the standard deviation is σ.

It is up to the survey designer to decide the tradeoff between expected variance and width of the confidence interval that is acceptable.

An alternative approach would be to develop the sample size based upon some acceptable threshold of error. In this case, Equations 4 through 6, reflecting the probability that $\hat{p}$ lies within two standard deviations of the mean, would be used:

$$\left(\hat{p} - 2\sqrt{0.25/n}, \hat{p} + 2\sqrt{0.25/n}\right) \qquad \text{(Eq. 4)}$$

$$4\sqrt{\frac{0.25}{n}} = W \qquad \text{(Eq. 5)}$$

$$n = \frac{4}{W^2} = \frac{1}{B^2} \qquad \text{(Eq. 6)}$$

Again, W refers to the width in units of the confidence interval. B refers to the allowable error. Thus, to allow a 10 percent error (90 percent confidence) in a normal population, sample size would have to be at least 100. Likewise, in order to achieve 99 percent confidence (allowing only 1 percent error), sample size would have to be 10,000.

Example

To illustrate these geographical considerations and others related to industry classification, sample size, data attributes, questionnaire design, and data extrapolation, consider a specific example. The example involves obtaining subnational commodity flow data for the Spokane region at the county level. The general structure of the type of information desired is represented by an origin-destination matrix shown in Table 2.5.

Table 2.5 assumes only 4 geographical units (two origins, two destinations), 10 industry categories, and 5 modes of transportation. In this simple example, the number of cells to populate with freight flow information is 200, which is calculated as

Number of cells = the number of origins (2) ∗ number of destinations (2)
∗ number of industries (10) ∗ number of modes (5).

Unfortunately, not many organizations are interested in only three geographic units. But it is evident that modification of the scale of data collection may occur from any of the three factors of the matrix (geographical units, industries, and modes). It may be that within certain parts of the county or city, only a small number of industries exist, and therefore including a large number of industrial categories is not necessary. Likewise, the county of Spokane has no water freight transportation facilities and possibly no pipelines, thus reducing the magnitude of the origin-destination matrix considerably by lowering the number of modes from five to three.

These tradeoffs are illustrated in Table 2.6, showing the outcomes of changing each of these three variables. Example 4 in Table 2.6 offers a relatively manageable matrix to populate, with 25 origin-destination combinations (5 origins and 5 destinations), 10 industries and three modes of transportation, resulting in 750 cells.

Number of cells = the number of origins (5) ∗ number of destinations (5)
∗ number of industries (10) ∗ number of modes (3).

Table 2.5. Sample origin-destination freight flow matrix.

Origin	Industry	Mode	Volume of Freight	Destination
Origin A and B	Ag			Destination B and A
		Truck		
		Rail		
		Water		
		Air		
		Pipeline		
	Mining			
		Truck		
		Rail		
		Water		
		Air		
		Pipeline		
	Manufacturing			
		Truck		
		Rail		
		Water		
		Air		
		Pipeline		
	XX			
		Truck		
		Rail		
		Water		
		Air		
		Pipeline		
	XX			
		Truck		
		Rail		
		Water		
		Air		
		Pipeline		
	Ten Industries			
		Truck		
		Rail		
		Water		
		Air		
		Pipeline		

Table 2.6. Different combinations of origin-destination matrix scales.

Example	Geographic Units (origins/destinations)	Industries	Modes	Number of Cells to Populate
Example 1	2	10	5	100
Example 2	10	3	3	90
Example 3	4	5	5	100
Example 4	25	10	3	750
Example 5	100	10	5	5,000

It also is evident from Table 2.6 that attempting to obtain detailed information on all three factors can produce a matrix of cells that becomes a massive data collection effort, as would be the case if you had 100 geographical units and industry information at the four-digit NAICS level (110) on five modes of transportation producing a 55,000 cell matrix.

In this hypothetical example, the county of Spokane is interested in commodity flows into and out of the county and has allocated resources to conduct an establishment survey. If the variance associated with each two-digit estimate of the number of establishments (see Table 2.7) is known, the necessary sample size for each industry category within Spokane County can be identified. This calculation is based upon the assumption of a normal distribution and a population that is independent and identically distributed, where the sample size is calculated as:

$$\text{Sample size} = 16 * (\text{variance})/(\text{Confidence Interval Width})^2.$$

Table 2.7 shows that depending on the confidence interval width (+/– 1, 2, 3 units), the sample size changes, increasing in size as confidence interval becomes smaller. In this hypothetical example, only the number of establishments at the different two-digit NAICS level is considered, but it is possible to use additional information regarding the size of each industry (employment, payroll, or output) to develop a stratified sampling design.

The type of geographical unit associated with origin-destination pairs of shipments will not impact the sample size within each industry for the county if it is assumed that industries within the county have similar origin-destination patterns and shipment volumes relative to the indicator variables within the county. Since the sample is designed to capture enough information to be statistically valid (95 percent confidence interval and assuming normal population properties) for the establishments, wherever they physically exist, increasing or decreasing the geographic scale of the commodity flow activity will only impact the number

Table 2.7. Calculating sample size for Spokane County industries.

NAICS Code and Code Description		Employees	Payroll	Total Establishments	Variance	Sample Size Assuming Different Confidence Interval Widths		
						1	2	3
---	Total for all sectors	177,847	$6,492,586	12,515	375	6,007	3,004	2,002
11	Agriculture, Forestry, Fishing and Hunting	105	$5,807	23	1	11	6	4
21	Mining, Quarrying, and Oil and Gas Extraction			28	1	13	7	4
22	Utilities			15	0	7	4	2
23	Construction	10,999	$547,263	1,535	46	737	368	246
31	Manufacturing	14,361	$616,393	548	16	263	132	88
42	Wholesale Trade	10,545	$478,700	735	22	353	176	118
44	Retail Trade	25,492	$667,276	1,646	49	790	395	263
48	Transportation and Warehousing	5,100	$181,864	285	9	137	68	46
51	Information	4,203	$208,232	217	7	104	52	35
52	Finance and Insurance	11,083	$565,205	901	27	432	216	144
53	Real Estate and Rental and Leasing	3,525	$101,663	659	20	316	158	105
54	Professional, Scientific, and Technical Services	8,782	$418,127	1,282	38	615	308	205
55	Management of Companies and Enterprises	2,357	$162,382	89	3	43	21	14
56	Administrative and Support and Waste Management and Remediation Services	9,482	$232,658	625	19	300	150	100
61	Educational Services	6,397	$149,144	140	4	67	34	22
62	Health Care and Social Assistance	35,391	$1,507,822	1,439	43	691	345	230
71	Arts, Entertainment, and Recreation	4,168	$81,750	161	5	77	39	26
72	Accommodation and Food Services	16,472	$245,715	1,026	31	492	246	164
81	Other Services (except Public Administration)	7,657	$173,849	1,142	34	548	274	183
99	Industries not classified	17	$302	19	1	9	5	3

of questions to be asked during the establishment survey or allocated after information has been obtained from the survey and processed. The establishments themselves represent the beginning and ending of the freight shipment activity and if the survey captured enough data to be statistically valid relative to the number of establishments within each industry category, then likewise information regarding the origin/destination of flows has been statistically represented.

However, if it is believed that similar industries within the county may have vastly different shipment patterns in terms of either volumes relative to an indicator variable or origin-destination patterns, then the number of sample sizes will increase proportionally relative to the number of geographic units that are developed at the subcounty level. Therefore, assumptions about commodity flow patterns within a county are critical to the determination of sample size for a region.

For the demonstration surveys conducted in Seattle and Spokane, there were two additional considerations that impacted the survey samples:

- **Random versus Nonrandom Samples.** Using the establishment data from sampling frames, the research team determined the largest companies in each of the industry categories. These establishments were included in the sample to maximize the usefulness of the responses received by ensuring that those industries that represent a disproportionate share of total commodity flows are not excluded from the sample. This is considered a nonrandom sample. The remaining samples were selected randomly. As a rule of thumb, it is recommended that any company that represents 10 percent or more of an industry within the study area should be included in the survey as a nonrandom sample. For comparison purposes, in the CFS, approximately 40 percent of the sample is nonrandom.
- **Precanvassing.** This refers to conducting an advance survey of selected companies to finalize the survey questionnaire and provide information on the most effective survey processes. These surveys also can be used to collect field data, but that was not done for this particular demonstration survey effort.

User's Guide Worksheet Punch List

- Determine the number of cells to populate for the desired survey by multiplying the origins, destinations, commodities, modes, and any other relevant variables together.
- Estimate the mean and variance for each of the variables using existing sources.
- Determine the desired confidence interval for each variable.
- Calculate the number of samples needed for each industry based on Equations 1 through 6 provided above.
- If the number of samples for each industry seems higher than reasonable, aggregate cells until the sample size becomes manageable.
- Determine the need for nonrandom samples based on the distribution of company size within each industry. This will need to be added to the survey sample.
- Review the User's Guide Worksheet Punch List from Steps 1 through 3 to determine whether the results of the sampling size calculations impact these previous survey design elements.

Step 5—Establishing Data Elements

Key Considerations

Data elements can be grouped into categories that are generally consistent across establishment surveys such as information about the establishment, origin-destination information for goods moving in and out of the establishment, and commodity information. However, there are several different ways of requesting these data, and these result in major differences in how the surveys are administered and the type of data that are collected. One of the key differences across establishment surveys is whether information is collected about specific shipments (such as the last 20 shipments) or whether aggregate information is collected across a specific period of time (such as the most recent year). Specific shipment information tends to be more difficult to obtain from companies participating in a survey. A request for aggregate information is likely to result in a higher response rate, but is also likely to be estimated by the individual company representative participating in the survey.

Another key consideration is the collection of inbound survey data. Typically, companies have less information about inbound shipments. However, as the research team has noted, because inbound flows for a region represent a large fraction of the flows, it is important to collect data on inbound shipments for a local commodity flow survey (this is not done in the national CFS). Collecting data on inbound flows is more important at the local level than at the national level because only a relatively small fraction of total commodity flows in and out of the 50 states comes from outside of the United States (although that fraction has been increasing with globalization). At the local level, a substantial portion of commodity flows will be coming from outside of the study area of interest.

There also is the need to consider how many different types of commodities to collect data on at each facility. Some surveys focus on the primary commodity at the location, while others may collect information on all commodities. Obviously, if a survey can be limited to only the primary commodities, it will take less time for the respondent to provide the data, and thereby response rates may increase. However, it will be difficult to determine how much has been missed. Other specific elements that need to be considered are the units of collected data (e.g., tons, containers, value, shipping units). The level of detail requested on origin and destination data should be matched to the geographic boundaries of the internal and external regions for the study. However, more detailed geographic information should be obtained from the survey if it can be done without reducing responsiveness or significantly increasing the duration of the survey.

Implementation Process

To identify data elements for the survey, it is useful to divide the data collection process into four components: (1) background information about the facility, (2) freight flow information on outbound shipments, (3) freight flow information on inbound shipments, and (4) open-ended questions.

Background information on the facility is needed to confirm that the company has the operating characteristics that are expected. This information is also helpful in ensuring that there is appropriate representation of different types of firms in the stratified sample. The data elements potentially included in background information include address, revenue or sales information, type of business, and size of facility. Information on the size of the facility is particularly important as it is often used to stratify samples, expand collected data, and determine relationships between company size and shipment volumes. The size of the facility should be asked about using multiple methods to provide greater flexibility to the later statistical processes. The size can be estimated based on the number of employees, square footage of the facility, or sales/revenue volumes. If it is anticipated that a significant proportion of interviewers will be hesitant to provide specific information, then an alternative process would be to have the interviewer accept an answer in predetermined range categories.

Freight flow information on outbound shipments is the core of the survey process. The type of shipment data that will be requested from the participating companies should be determined first. The CFS asks about a fixed number of shipments over a 1-week period during four seasons in the year. The number of shipments requested is based on the size of the company, but is between 20 and 40 for each company surveyed. If this shipment information can be successfully replicated in subnational surveys, then the steps of data expansion, validation, and analysis in the local survey would potentially benefit significantly. However, given the effort required on the part of a company being surveyed, the lack of a mandate for companies to participate in these surveys, and the limited resources available for conducting surveys at the local level, it may be difficult to collect detailed shipment data at the local level.

An alternative to the CFS approach would be to ask for data from only 1 week in the most recent season and an estimate of the degree to which shipments fluctuate in other seasons. A closely related alternative is to request estimates of annual volumes and percentages of shipments in each of the four seasons in order to understand seasonal variations in the data. This alternative tends to be the simplest to implement for both the surveyor and the respondent. One challenge with this approach is that origin-destination patterns may also vary during the year and providing separate origin-destination information for each season may prove too burdensome for survey respondents. If the goal of the survey is to understand peak volumes, then it may be necessary to ask directly about volumes and timing during the peak season and then pivot other volume and origin-destination information off of that response.

Another implementation issue to be resolved is whether or not commodities will be preclassified by the survey team or whether survey respondents will define commodities on their own. Preclassifying commodities tends to standardize respondents' answers. It also increases the likelihood that respondents remember to include all of the major inputs and outputs into their facility. However, preclassification can make the survey longer as there will be a need to explain the different commodity categories and a dialogue may be needed for respondents to match their inputs and outputs to the predetermined commodities.

If preclassifying commodities is selected, then the specific inputs for each industry will need to be identified in advance of the survey so that they can be incorporated into the survey questionnaire. This is done with the assistance of input-output tables for each industry. Input-output tables describe the industries that serve as customers and suppliers for other industries. Table 2.8 shows a BEA input-output table created in 2002. More recent data can be purchased through proprietary economic databases. For each industry, it is typically preferable for the survey to cover 90 percent of the likely inputs and outputs for each facility, which may include between one and four commodities, depending on the industry.

Open-ended questions also can be asked as part of an establishment survey. These can be tailored to current hot-button issues in the region or to issues such as the desire to participate in the long-range transportation planning process. The benefit of asking these types of questions is that they provide respondents with an opportunity to bring their concerns to public officials through the survey process, and this may make them feel that the survey will benefit them. However, because establishment surveys tend to be quite lengthy, it may be desirable to not include

Table 2.8. BEA input-output table, 2002 (select commodities and industries, dollars in millions).

Commodity Code	For the distribution of commodities consumed by an industry, read the column for that industry. For the distribution of industries consuming a commodity, read the row for that commodity.	Crop Production	Animal Production	Forestry and Logging	Fishing, Hunting and Trapping	Support Activities for Agriculture and Forestry	Oil and Gas Extraction
	Commodity Code	1110	1120	1130	1140	1150	2110
1110	Crop products	8,063.0	14,996.4	5.4	6.7	239.7	0.0
1120	Animal products	369.0	19,635.6	70.0	0.0	366.0	0.0
1130	Forestry and logging products	0.0	0.0	12,923.6	0.0	11.8	0.0
1140	Fish and other nonfarm animals	0.0	0.0	0.0	0.0	0.0	0.0
1150	Support activities for agriculture and forestry	10,760.0	1,748.4	2,811.2	20.1	0.0	0.0
2110	Oil and gas extraction	0.0	0.0	0.0	0.0	0.0	1,989.9
2121	Coal mining	0.0	118.4	0.4	0.0	0.0	125.4
2123	Nonmetallic mineral mining and quarrying	507.0	76.0	0.0	0.0	1.5	0.0
2130	Mining support services	0.0	0.0	0.0	0.0	0.0	1,569.9
3110	Food products	0.0	13,817.9	43.5	22.7	169.3	0.0
3121	Beverage products	0.0	34.8	0.0	0.0	0.0	0.0
3122	Tobacco products	0.0	0.0	0.0	0.0	0.0	0.0
3130	Yarn, fabrics, and other textile mill products	28.5	0.0	0.0	0.0	53.6	0.0
3140	Nonapparel textile products	218.2	21.5	0.0	32.0	240.1	0.0
3150	Apparel	0.0	0.0	0.0	0.0	0.0	0.0
3160	Leather and allied products	0.0	50.2	0.0	0.0	0.0	0.0

Source: BEA.

open-ended questions in the same effort. Furthermore, if information is collected about issues of concern to freight industries, the agency collecting the information must be prepared to act on this information and ensure that the businesses that provided the information are kept abreast of the actions being taken to address their concerns. Note that the confidentiality needs of survey participants also are a key consideration for selecting data elements. Typically, state and local agencies, similar to federal agencies, are able to ensure the confidentiality of data provided by individual respondents, even in the case of legal proceedings.

Example

The demonstration surveys conducted in Seattle and Spokane collected data on several data elements from the four industries surveyed. The survey data elements included the following:

- Identifying information for the establishment
- Square footage of the facility
- Number of employees at the facility
- Inbound annual tons
- Inbound annual value
- Inbound seasonal information
- City, state, zip code, and port of entry information for the largest four commodities arriving at the establishment

- Outbound annual tons and value
- Outbound seasonal information
- City, state, zip code and port of exit information for the largest four commodities leaving the establishment.

The response rate for several data elements in this survey was calculated and is shown in Table 2.9. These response rates can be used to estimate response rates for similar questions in other survey efforts.

User's Guide Worksheet Punch List

- Determine the background information to include in the survey and how the size of establishment will be captured.
- Determine the shipment information to request in the survey.
- Determine whether the survey will focus on outbound goods, inbound goods, or both.
- Determine whether or not to preclassify commodities for the survey.
- Determine the level of geographic specificity at which to survey. Refer to the geographic zones developed in Step 1 and the response rates at different levels of geography shown in Table 2.9.
- List all data elements on which data will be collected.

Step 6—Survey Questionnaire

Key Considerations

It is most effective to design a survey questionnaire that moves from the easiest to hardest questions. With this design, if the respondent decides to stop participating in the middle of the survey, at least some data can be collected. Partial data can still be used in many cases. For this reason, respondents are typically asked to provide general information about the facility first. From there, the survey goes on to request information regarding outbound shipments and inbound shipments, and it concludes with open-ended questions.

Within the portions of the survey addressing outbound and inbound shipments, origin-destination information is generally the most complex. Therefore, it is typical to ask for commodity information first, then information about modal usage, and then origin-destination information for each commodity. Within the origin-destination questions, it is generally easiest to ask for state information, then city information (if easily available), and then about subcity geographic units, if needed.

The questionnaire should be designed so that data are captured in a format that best serves the purpose of the survey. However, this aim must be balanced with the need to capture information in a form that is comfortable for survey respondents. Additionally, there is a tradeoff between the length of the survey and response rates.

For industry-specific surveys, it is generally beneficial to have an industry expert review a draft of the survey to confirm that the questions are reasonable given the typical structure and sourcing patterns for each industry.

Table 2.9. Response rates for individual questions in demonstration surveys.

	Metropolitan Region and Industry				
Item	Spokane Food	Spokane Fabric. Metal	Seattle Apparel	Seattle Trans. Equip.	Total
Background Information Response Rates					
Total Establishments Contacted	41	34	30	42	147
Percent Establishments Responding	25%	30%	33%	24%	27%
Number of Establishments Responding	10	10	10	10	40
Provided Response to Revenue Information	60%	90%	30%	70%	63%
Provided Response to Background Information	100%	100%	100%	100%	100%
Provided Response to Size of Facility	100%	100%	90%	90%	95%
Inbound Shipment Response Rates					
Provided Response to Inbound Annual Tons or Other Shipping Units	100%	100%	100%	100%	100%
Provided Response to Inbound Value of Shipments	0%	0%	0%	0%	0%
Provided Response to Inbound Seasonal Question	30%	20%	70%	90%	53%
Percent of Respondents that Provided Inbound Seasonal Information	30%	10%	70%	30%	35%
Provided Response to Inbound Distribution of Comm. 1	100%	100%	70%	80%	88%
Provided Response to Inbound Distribution of Comm. 2	0%	30%	50%	0%	20%
Provided Response to Inbound Distribution of Comm. 3	30%	10%	20%	30%	23%
Provided Response to Inbound Distribution of Comm. 4	0%	N/A	N/A	0%	0%
Provided Response to Comm. 1 Origin Attribute					
City	40%	90%	70%	60%	65%
State	100%	100%	80%	80%	90%
Zip Code	0%	0%	0%	0%	0%
Country	100%	100%	70%	80%	88%
Port of Entry	0%	0%	60%	0%	15%
Mode	100%	100%	80%	80%	90%
Provided Response to Comm. 2 Origin Attribute					
City	50%	60%	10%	60%	45%
State	80%	70%	10%	60%	55%
Zip Code	0%	0%	0%	0%	0%
Country	80%	80%	10%	70%	60%

Port of Entry	0%	0%	10%	0%	3%
Mode	80%	80%	10%	60%	58%
Provided Response to Comm. 3 Origin Attribute					
City	10%	20%	0%	30%	15%
State	20%	50%	0%	40%	28%
Zip Code	0%	0%	0%	0%	0%
Country	30%	50%	0%	20%	25%
Port of Entry	0%	0%	0%	0%	0%
Mode	30%	50%	0%	40%	30%
Provided Response to Comm. 4 Origin Attribute					
City	30%	N/A	N/A	10%	20%
State	30%	N/A	N/A	10%	20%
Zip Code	0%	N/A	N/A	0%	0%
Country	30%	N/A	N/A	30%	30%
Port of Entry	0%	N/A	N/A	0%	0%
Mode	30%	N/A	N/A	30%	30%
Outbound Shipment Response Rates					
Provided Response to Outbound Annual Tons or Other Shipping Units	100%	100%	100%	100%	100%
Provided Response to Outbound Value of Shipments	0%	0%	0%	0%	0%
Provided Response to Outbound Seasonal Question	0%	0%	0%	0%	0%
Percent of Respondents that Provided Outbound Seasonal Information	N/A	N/A	N/A	N/A	N/A
Provided Response to Outbound Distribution of Comm. 1	100%	80%	90%	20%	73%
Provided Response to Outbound Distribution of Comm. 2	0%	40%	30%	70%	35%
Provided Response to Outbound Distribution of Comm. 3	10%	30%	0%	40%	20%
Provided Response to Outbound Distribution of Comm. 4	N/A	N/A	N/A	N/A	N/A
Provided Response to Comm. 1 Destination Attribute					
City	50%	50%	30%	0%	33%
State	90%	90%	70%	20%	68%
Zip Code	0%	0%	0%	0%	0%
Country	80%	100%	100%	20%	75%
Port of Entry	10%	0%	0%	0%	3%
Mode	90%	100%	100%	20%	78%

(continued on next page)

Table 2.9. (Continued).

Item	Metropolitan Region and Industry: Spokane Food	Spokane Fabric. Metal	Seattle Apparel	Seattle Trans. Equip.	Total
Provided Response to Comm. 2 Destination Attribute					
City	40%	60%	10%	0%	28%
State	60%	80%	60%	20%	55%
Zip Code	0%	0%	0%	0%	0%
Country	70%	80%	70%	20%	60%
Port of Entry	20%	20%	0%	0%	10%
Mode	80%	90%	70%	20%	65%
Provided Response to Comm. 3 Destination Attribute					
City	0%	10%	N/A	0%	3%
State	40%	40%	N/A	10%	30%
Zip Code	0%	0%	N/A	0%	0%
Country	60%	50%	N/A	10%	40%
Port of Entry	0%	10%	N/A	0%	3%
Mode	60%	50%	N/A	0%	37%

Implementation Process

As mentioned in the previous step, the more that state and DOT establishment survey questionnaires can be made similar to the CFS instrument, the more the data collected in the two efforts can be used in a complementary fashion. However, there are significant differences between this national survey and surveys that would be done at the state or metropolitan level. These differences mark the key decision points for developing the questionnaire.

An alternative to using the CFS questionnaire is to consider a survey structure that describes a standard set of 16 questions, based on the work of "Approach to Collecting Local Freight Information" (Thompson et al. 2010). This questionnaire was developed specifically for state and local transportation agencies in the hope of creating a standardized format that also would allow for comparison across subnational establishment surveys.

This survey questionnaire begins by confirming general information about the facility, including the following:

- Company name
- Address of the actual site to be visited
- Street number
- Street
- City
- State
- Zip
- Name of person(s) to be interviewed

Table 2.10 contains the description of the 16 standardized questions (based on "Approach to Collecting Local Freight Information"[Thompson et al. 2010]). The questions must be asked and recorded with a level of rigor that allows for more probing at any point in the interview, but all of the questions need to be asked and recorded to complete the local commodity flow database.

We recommend reviewing the survey questionnaire in Table 2.10 to observe the similarities and differences that are possible. The survey questionnaire developed for each region should then be customized based on local freight planning needs, available resources, and current understanding of potential survey respondents' willingness to answer key questions.

It is recommended that experts within the industries being surveyed review and comment on the data elements that are being considered for the survey as well as on the entire survey questionnaire. These experts can be located through trade associations, chambers of commerce, or previous local knowledge of representative companies in each industry.

An alternative to using known experts is to conduct precanvassing of select companies within targeted industries to discuss the data elements and the survey questionnaire that will be used. Feedback from the precanvassing process can be used to revise the survey. This is described in greater detail in Step 7 (Conducting the Survey).

Example

For the demonstration establishment surveys conducted in Seattle and Spokane, the survey questionnaire had four main sections:

1. **Background information about the company and surveyor—**filled out in advance, but confirmed via the survey process.
2. **Size of the facility—**measured in terms of square footage and number of employees.
3. **Outbound shipments—**tonnage and value, timeframe, seasonality, modes, and origins and destinations.
4. **Inbound shipments—**tonnage and value, timeframe, seasonality, modes and origins, and destination.

Table 2.10. Questionnaire elements.

Question Number	Topic	Interviewer Instructions/Explanation
Q1	Business Description	Keywords used by interviewee to capture primary business activity(ies) – to be converted into single industry sector designation for commodities using the 43 categories available in the SCTG. Note other activities that generate freight.
Q2	Number of Employees	Current number of full- and part-time employees at the site (e.g., head count). If response is only available for multiple locations in the region, note this aggregation and make sure other data elements also are aggregates.
Q3	Shipments by Mode	Capture how the company receives and ships most of its goods—clearly indicating if the goods are being received (inbound) or being shipped (outbound).
Q4	Deliveries Received by Mode WEEKLY	Average number of deliveries received weekly. If interviewee can provide only monthly or annual numbers, convert these figures to weekly data in postprocessing procedures.
Q5	Shipments Generated by Mode WEEKLY	Average number of shipments generated weekly. If interviewee can only provide monthly or annual numbers, convert these figures to weekly data in postprocessing procedures.
Q6	Origins of Inbound Shipments	Capture major origins where shipments come directly to the site. For example, if interviewee knows origin is California, but last leg is from Dallas, record information as "California through Dallas." Continue to probe until it is possible to determine the origins by percentages of their total activities (to sum to 100 percent). Probe for the direction for "within an MPO" or "outside the MPO" (use a compass icon to illustrate directions) and indicate any other location information (e.g., ports by name).
Q7	Destinations for Outbound Shipments	Capture major destinations for shipments from the company. For example, if interviewee knows destination is California, but last leg is through Dallas, record information as "California through Dallas." Continue to probe until it is possible to determine the destinations by percentages of their total activities (to sum to 100 percent). Probe for the direction for "within an MPO" or "outside the MPO" (use a compass icon to illustrate directions) and indicate any other location information (e.g., ports by name).
Q8	Size of Shipment	Check for each mode used for inbound and outbound activities and whether most of these shipments are "less than full load" or "full load." For containers, indicate size of container (e.g., 20′ or 40′).
Q9	Weight of Shipment	NORMAL weight of a full shipment (not including vehicle weight) inbound and/or outbound by all modes. Verify rail spur if rail is mentioned and connection to port if barge is mentioned. Otherwise, query for shipments actually arriving/leaving by truck. Containers assumed as truck trip, but indicated separately as container on truck. Indicate weight in tons or pounds; however, only enter pounds in database (e.g., pounds x 2,000 = tons).
Q10	Size of Facility	Size in square feet under roof. Indicate outdoor space used as separate information.
Q11	Expansion Plans	Note any comments indicating plans for expansion within the next 5 years. Note the anticipated year and amount of increase in size, converting to square footage if percentage is provided.

Table 2.10. (Continued).

Question Number	Topic	Interviewer Instructions/Explanation
Q12	Value of Goods	Total value in dollars of goods received/shipped for most recent year. If unavailable, indicate annual sales amount as proxy for shipment value.
Q13	Annual Volume of Shipments — Actual	Note year and the total ANNUAL number of shipments inbound/outbound for most recent year. (Check by multiplying Q4 and Q5 by 52 and comparing weekly answers to annual).
Q14	Annual Volume of Shipments — Forecasted	Note the year and the ANNUAL total number of shipments the company expects 5 years in the future (or sooner).
Q15	Problems at the Location	Note any location or site-specific problems described by the interviewee. Probe for details, including impacts. When noting comments, be sure not to leave the impression that problems will be fixed, only that their descriptions will be forwarded to the MPO and used for transportation planning infrastructure/operations.
Q16	Problems in the Area	Make note of any route or significant problems in the area. Probe to get enough information to clarify any issues mentioned. When noting comments, be sure not to leave the impression that problems will be fixed, only that their descriptions will be forwarded to the MPO and used for transportation planning infrastructure/operations.

Source: Thompson et al. 2010

The survey questionnaires and the survey design also allowed for the respondent to provide information at different levels of detail. For example, if the respondent was only willing to provide county detail for origins and destinations, this was collected, but if they were willing to provide city or zip code detail, this was collected. Both inbound and outbound freight flows were covered.

The primary outbound and inbound commodities that were expected to be received or shipped by each industry were preclassified, and information about these commodities was requested. Respondents also were given the opportunity to identify other commodities that they shipped or received. While the identification of outbound commodities was generally fairly straightforward in the cases that were tested, these industries may purchase supplies that represent a broad range of different commodities. For the design of the survey, an input-output table (otherwise known as the "make-use" table of an input-output model) was used to identify the principal commodities purchased by each of the industries that were surveyed, and this information was used in the survey design for outbound shipments. Thus, the questionnaires were customized for each industry at least in the naming of commodities in the inbound and outbound shipments section of the questionnaire. The final survey questionnaires used are located in a subtask report associated with the development of the *Guidebook* titled "Demonstration of Application of Establishment Survey," which is available at www.trb.org/Main/Blurbs/169330.aspx.

Precanvassing was conducted to finalize the survey questionnaire for the demonstration surveys. The precanvassing was done in in-person interviews by a team member from Washington State University. A total of 11 firms were included in the precanvassing (see Table 2.11). Five of the firms were located in the Seattle region and six were located in the Spokane region. Three of the firms in Seattle were in the transportation equipment industry, and two of the Seattle firms

Table 2.11. Number of companies included in precanvassing by industry.

Region	Industry	Number of Companies Interviewed In Person
Seattle	Transportation Equipment Manufacturing	3
Seattle	Apparel Manufacturing	2
Spokane	Food Manufacturing	4
Spokane	Fabricated Metal Product Manufacturing	2
Total		**11**

were in the apparel industry. Four of the Spokane firms were in the food manufacturing industry, and two were in the fabricated metal product manufacturing industry. These firms were identified based on previous research team contacts.

The precanvassing process led to several key conclusions regarding the full demonstration survey. All 11 interviewees were interested in the goal of the study and felt it was a worthwhile endeavor. This seemed to be borne out by response rates in the final survey. This indicates that if the survey is relatively short and focused, it may not be necessary to expend significant time and resources convincing interviewees that surveys are important or that survey data can be used for transportation planning.

Interviewees in the precanvassing process thought that commodity information would be available from the demonstration survey respondents. Some of the interviewees thought that the companies should provide commodity information rather than having it identified beforehand. The interviewees were of the opinion that the identification of subregional flows could be accomplished using the questionnaire. However, they also believed that survey respondents would be providing varying levels of detail. The interviewees doubted that revenue or value data would be provided. The interviewees felt that providing shipment value information would be equivalent to providing proprietary information about the company such as rate schedules and profit margins.

Nine of the 11 interviewees were somewhat concerned that too much detail was being requested. They believed that survey respondents would not have detailed data available and that only general responses would be received. This was seen as an issue particularly for the origin and destination data for both inbound and outbound shipments. Giving respondents the ability to respond for "typical" patterns and allowing them to adjust the level of geographic detail at which they were willing to respond were approaches recommended by the precanvassed interviewees.

The interviewees consistently suggested that the questionnaire script should be rephrased to request "estimates of data," "typical years," or "last year" shipments or attributes. Such an approach was expected to generate more complete responses, even if these would need to be considered expert estimates rather than precise data amounts. Overall, the interviewees suggested that the conversation with the phone interviewer be "dynamic" rather than scripted. This would allow for a dialogue to unfold that would generate on the origin-destination patterns of shipments and the percent allocation by location and temporal factors.

These results led to some changes in the questionnaire. Specifically, different or fewer categories were offered in the revised questionnaire. Additionally, space was changed on the physical questionnaire to allow the interviewer to capture information that was being offered. Also, information that was known prior to the survey (e.g., name and location of firm) was moved to an earlier part of the instrument. The final survey questionnaires used are located in a task report associated with the development of the *Guidebook* titled "Demonstration of Application of Establishment Survey," which is available at www.trb.org/Main/Blurbs/169330.aspx.

The determination of whether or not precanvassing is needed in a local survey effort depends on two factors: (1) the degree of familiarity with the industry being surveyed and (2) the types of questions that are being included in the establishment survey.

If transportation agencies are comfortable with their familiarity with an industry and they are asking standard questions, then precanvassing is generally not needed. However, if "out-of-the-box" questions are being considered, then a precanvassing survey would be recommended. Additionally, if a survey is being conducted on an industry that the transportation agency is not familiar with, then precanvassing also would be recommended.

Precanvassing surveys such as the surveys conducted in this process are most effective when they are implemented using companies that already are familiar with the local MPOs or DOTs. These companies are most likely to take the time to assist the transportation agency, and they are more likely to provide thoughtful, complete answers than companies that are unknown prior to the survey effort.

Alternatively, companies also can be identified by using state or local chambers of commerce, industry associations, and establishment information provided from companies such as InfoUSA or Dun & Bradstreet. This implies that conducting an establishment survey should be one component of a broader private-sector freight stakeholder effort that DOTs or MPOs operate. This ongoing outreach activity will improve precanvassing efforts, provide a sounding board to confirm reasonableness of full survey efforts, and provide guidance on freight planning decisions that are made based on collected data.

User's Guide Worksheet Punch List

- Develop a full survey questionnaire for the region using the data elements identified in Step 5.
- Categorize these questions into general information, outbound flows, inbound flows, open-ended questions, and other relevant categories for the questionnaire.
- Sort the categories of questions from easiest to hardest to respond to, based on general knowledge of the interviewees.
- Develop a draft questionnaire based on fine-tuning the sorted list for any needed practical considerations.
- Provide the draft questionnaire to industry experts to confirm the reasonableness of the data elements and the survey process. Use a formal precanvassing effort, if needed.
- Edit and finalize the survey based on feedback from industry experts.
- Document potential changes in the region's freight stakeholder outreach based on the experience of developing the survey.

Step 7—Conducting the Survey

Key Considerations

With the survey questionnaire finalized and potential companies to interview determined, the next step is conducting the survey. One of the key considerations in this process is the considerable effort involved in identifying and confirming a time to interview a specific individual. Another key consideration is that surveyors should have some background in freight transportation to successfully implement the survey. This familiarity assists in the dynamic conversation that is needed to extract information that will support the statistical analysis of survey data and generalized freight planning efforts. Surveyors should briefly research each company prior to the survey to assist in this effort.

Implementation Process

The first step is to identify the correct individual at each company to survey. This can sometimes be determined from a vendor list or recommendations from trade associations. Often, they can provide an initial point of contact. Large firms may have a logistics specialist, while smaller firms may have plant managers with sufficient knowledge. Surveyors should maintain a log of each of the companies contacted and individuals spoken to, with summaries of each conversation and next steps. All of these items should be date and time stamped, so that follow-up can be done in an appropriate amount of time. It is not uncommon for establishment surveys to have response rates of 10 to 25 percent, so many potential surveys will turn out to be dead ends.

Campaigns to encourage participation carried out with local trade associations and the Chamber of Commerce could help improve response rates. If the state or MPO has a freight advisory council, then they also can be used to spread the word about the survey and encourage participation by target companies. Another option would be taking a freight neighborhood approach. This kind of approach would make it possible to let freight "neighbors" know who else has already participated. In this approach, those who have completed the process are asked if they would be willing to encourage neighbors to participate by allowing their participation to be part of the initial contact information. This strategy may work best with a mix of industry types to reduce any concerns with competition among similar firms.

At least two hard copies of the questionnaire should be available, one for the interviewer and one for the survey participant (provided only if requested, however, as it could be a distraction). At the beginning of the interview, the interviewer should reconfirm the time needed to complete the data collection and remind the participant that all the information will be held in confidence. Business cards should be offered to persons met during the interviewer's visit. It is not advisable for an interviewer to leave partially completed forms and expect firm employees to fill in and return information. Due to the flexibility required to capture information from establishment surveys, they are typically done with pen and paper. However, even if the survey is conducted electronically, hard copies of the questionnaire should still be made available to the survey participant.

The interviewing process is intended to build trust and confidence, but should not result in unrealistic expectations of how the information collected will be used. This can especially be an issue with any open-ended survey questions that ask participants to list problems and issues. The interviewer needs to confirm that this information will be submitted as input to the long-range transportation planning process for future consideration.

The interviewer should thank the person(s) participating in the interview and encourage them to call at any time if they want to offer additional information. As soon as possible, review the data collected to ensure that the information was correctly written/typed and the appropriate units were recorded (e.g., weekly or annual totals). As a final outreach, staff at the transportation agency should either mail a note of thanks for participation or send an e-mail within a week of the interview.

Example

Graduate students at Washington State University were trained to administer the demonstration survey in Seattle and Spokane. Surveyors briefly summarized the purpose of the survey, the transportation agency that was implementing the survey, what the data would be used for, and the privacy controls on the data that would be collected. Additionally, surveyors were physically located such that respondents could call them back through a general number at the university to confirm that the survey was being conducted by the party stated. When surveys are conducted by outside consulting firms, then the sponsoring transportation agency contact and contact information should be readily provided to the respondent.

The surveyors were encouraged to check the web sites of the selected firms before making their initial call. Interviewer knowledge of the firm characteristics information, prior to the phone call, was expected to make the phone survey more effective and efficient in terms of increasing response rates and receiving more precise responses.

The contact information provided in the two sampling frames (Dun & Bradstreet and InfoUSA) for establishments was found to be very accurate. Only 2 out of roughly 120 contacts had incorrect phone numbers. However, there were some cases in which area codes were not provided and had to be researched online. It was generally found that each company's receptionist was able to identify the correct person to respond to the survey, even though it often took multiple attempts to speak to this individual. Tracking down and receiving approval from this individual is the key determinant in the response rate for the overall survey.

The introductory conversations generally began with the following:

> Hello, my name is _____________ and I'm a student at Washington State University, in Pullman, WA. We are working with the U.S. Department of Transportation to survey businesses in the Pacific NW in regards to their use of the transportation infrastructure. In particular, we are interested in your company's annual shipments into and out of facilities and the locations where shipments originate and final destinations. Is there someone within your business that I can talk to in regards to this study?

Initially, no time length was mentioned in the introduction; however, once the survey began it became clear that without some stated bounds on the time required to complete the survey, respondents were not willing to participate. Thus, the students began stating that the survey would take 5 minutes or less.

Approximately 15 percent of survey respondents ended up taking more than 5 minutes, but respondents were relieved if the 5-minute time limit was maintained. Given that the proprietary establishment lists also contained information on company web sites, the students were asked to spend a few minutes browsing the company web site prior to calling to familiarize themselves with the company and the freight activities that this type of business was likely to conduct. This additional preparation increased survey labor time, but it also improved the students' ability to understand who they were calling and target their questions toward collecting the desired information.

In all cases, during the survey, respondents were asked if they shipped or received any of the preclassified commodities, but the respondents were also given the option to name commodities they ship and receive using their own terminology.

In this demonstration survey, there was a trend of smaller companies being more responsive to the survey process both in terms of overall responsiveness and specific responses to individual questions. Many of the larger firms stated that they were too busy to respond to the survey or that they needed approval from staff at corporate headquarters that was not colocated within the establishment. The bureaucracy of larger organizations caused them to have a lower response rate than the smaller firms. At smaller firms, individuals felt more empowered to provide this information to the surveyor.

The smaller companies also seemed more confident in the responses that they did provide. This is probably related to it being easier for a single person to understand shipping practice in full at a smaller firm than it is at a larger firm. At larger firms, shipments are likely to be managed by a team of people with expert knowledge of only the shipment types under their purview.

The lower response rate of larger firms underscores the need to incorporate an establishment survey process as part of a larger freight stakeholder outreach effort. The larger companies will likely need more time to approve survey participation, and this approval is more likely to occur if they have worked with the transportation agency extensively in previous efforts and if their participation is likely to impact actual decisions that are made by the agency. Having senior-level executives serve on an ongoing freight advisory council can accelerate this approval process and allow for transmittal of more detailed shipment information.

It is difficult to obtain detailed origin/destination information from larger companies. Many of the larger firm respondents indicated they ship or receive from "all over the United States," and attempting to ask for their top three was problematic. Smaller firms had origin-destination information more readily available and seemed to be more confident in the accuracy of the information that they were providing.

The students' affiliation with the local university seemed to make respondents much more willing to cooperate, since many expressed some association themselves (either a graduate or had children attending there). A private phone survey company may have had lower participation and response rates.

When companies refused to provide a value of shipments, the interviewers were able to adjust and ask for company revenue information. Initially, companies were unwilling to provide this, but by offering a broad range of revenue categories (0 to 5 million dollars, 5 to 10 million dollars, 10 to 20 million dollars . . .), it was possible to obtain revenue information from approximately 63 percent of respondents.

It can be challenging to provide categories of inbound or outbound shipments for some industries. For the Seattle transportation equipment industry, the outputs were simply too many to mention or categorize. These ultimately were labeled "aerospace parts," even though they may have been electronic assemblies, mechanical components, commercial, military, etc.

As mentioned above, collecting origin-destination information at the state level is the most easily accessible, at the city level it is somewhat less accessible, and zip code level information is nearly impossible to capture using this survey approach.

Outbound data were generally more readily available than inbound data. This confirmed that requests for outbound data should be made first in establishment surveys. However, the availability of data depends in large part on the structure of the industry and the complexity of individual companies' supply chains. Therefore, as part of the precanvassing effort, companies should be asked which portion of the supply chain is easier for the industry to provide.

User's Guide Worksheet Punch List

- Begin conducting the survey. Track the responses of all companies contacted, including failed surveys.
- For successfully scheduled surveys, allow for surveyors to make additional interesting notes that can be reviewed in the near term and long term.
- Following the successful completion of 10 percent of the survey in terms of either desired sample or completed responses, review the collected data to confirm that it includes the information most important to support the region's freight planning efforts. This includes the need to expand the data to represent the full region.
- Complete survey.

Step 8—Database Assembly

Key Considerations

It should be expected that a significant amount of effort will be needed to process the information that is collected in a single database. The survey can be designed to capture information in a way that is easiest for the surveyor to process. However, many of the respondents will provide information in a fashion that is inconsistent with the survey structure. This information will require postsurvey calculations by the surveyor in order to fit into the survey structure. Expansion variables should be determined prior to survey implementation and postprocessing to ensure that collected data are sufficient to support the data expansion process.

Implementation Process

This implementation consists of entering the collected data into a database format. Ideally, the data entry is done by the staff that conducted the survey to better interpret the results. The field names and data relationships need to follow a formal system to allow for statistical analysis of results in later steps. However, there should be flexibility regarding some of the entries, including comment fields that allow the database to retain nuances captured by the surveyor that do not easily fit into a quantitative format. There will likely be several instances of needing to translate units into a consistent format. For example, respondents may respond in terms of tons, containers, or pallets, and this will need to be converted into a consistent format across the entire survey. Additionally, the commodity information is likely to be captured in varying levels of detail, particularly if the survey participant is given the option to self-describe the goods that are shipped to and from the location. The survey database should preserve the original response of each participant, but it should also convert the response into the classification scheme that will be used for future analysis. It may be possible to develop a "seed" file in the database that could be made available for easy data entry and report production. An example of a similar tool is the Land-Based Classification Standards (LBCS). (See www.planning.org/lbcs/implementation.)

There will need to be a quality control check on the data to confirm that the entries are readable and reasonable. This quality control check should include a review of origin-destination information for consistency with logical shipping patterns. Additionally, there should be a check of the shipping units. Problems that can occur in terms of volumes include recording a response in tons that should be in pounds and developing the right conversion factor for containers to a weight-based unit. In select instances, follow-up with survey respondents may be necessary to clarify entries. Alternatively, a small portion of the data may need to be removed from the larger data set if it fails the quality control check.

Example

In the Spokane and Seattle demonstration surveys, responses were recorded using pencil and paper. The collected data were input into a Microsoft Excel spreadsheet. A review of the data identified that there were no issues with reasonableness or consistency. The commodity information was captured at several different levels. The raw responses were captured in the spreadsheet. Additionally, each response was classified into the two-digit SCTG code consistent with the CFS commodity classification.

A snippet of the raw spreadsheet database is shown in Table 2.12. This table shows the data elements that were asked about for general information on the company and inbound shipments. Additional information included in the spreadsheet is the modal percentage for each commodity and the origin information for Commodities 2 to 4. All of this data was also collected for outbound shipments. The full data entry spreadsheet can be found in a subtask report associated with the development of the *Guidebook* titled "Demonstration of Application of Establishment Survey," which is available at www.trb.org/Main/Blurbs/169330.aspx.

User's Guide Worksheet Punch List

- Input raw data into an electronic format.
- Check raw data responses for reasonableness.
- Refine data by making any necessary conversions such that all entries can be analyzed at the same level of commodity and geographic detail.

Table 2.12. Raw survey database for Seattle and Spokane Demonstration Survey (select fields).

	Background Information									Inbound Shipments																
Company	Company Name	Address	City	State	Zip	Name	Date of Survey	Method	Number of Employees	Annual Tons	Other Shipping Units	Value	Seasonal Peaks (Y/N)?	Timing of Peaks	Comm. 1	Comm. 1 %	Comm. 2	Comm. 2 %	Comm. 3	Comm. 3 %	Comm. 4	Comm. 4 %	C1 – City	C1 – State	C1 – Zip	Etc…
Co. 1																										
Co. 2																										
Etc…																										

Step 9—Data Expansion

Key Considerations

The key to data expansion is identifying the appropriate indicator variable and estimating the correct control totals. Typically, employment is used as the expansion (or indicator) variable because employment data tend to be relatively easy to collect through surveys. Additionally, employment data are available or can be estimated for various regions or subregions of concern within the study area.

Implementation Process and Example

Since only a sample of establishments is surveyed, it is necessary to develop statistical weights to expand the sample data to reflect the characteristics of the entire population of establishments for each industry. The key step in this process is to determine an appropriate expansion (or indicator) variable. The most straightforward expansion variable is based on information already contained in the establishment database from which the survey sample was drawn. Sample expansion variables can be based on number of employees, amount of output, or square footage size of the establishment. Employment is the most commonly used expansion variable due to its ease of collection during a survey and availability for various geographies and industries across the country. Using data on total employment by industry from sources such as CBP to establish the control total or population estimate, it is possible to estimate the fraction of total employment in an industry the sample represents. The fraction of total employment captured in the establishments surveyed represents the sampling fraction. By taking the reciprocal of the sampling fraction and multiplying this factor by whatever variable is being estimated (for example, total tons shipped), it is possible to get an estimate of the totals for the region.

Employment at individual companies can be inquired about via the survey process. It can also be cross-referenced with information included in establishment databases. In establishment databases, the information is provided in ranges; the midpoint for each establishment range should be used to estimate the total employment in the industry and the total employment of the companies included in the survey.

Expansion factors can be developed by dividing the indicator variable amount in the control total to the level of the indicator variable in each of the geographic regions in the survey study area. This needs to be done for each industry that is included in the survey.

For example, if the total employment in the metropolitan Seattle transportation equipment industry was found to be 10,000 and the employment of the companies surveyed in the transportation equipment industry was 1,000, then all of the data collected for this industry would be multiplied by 10 to develop estimates of freight flow patterns for transportation equipment.

It also is possible to use external employment data to conduct an expansion. For example, if the Census Bureau's Economic Census estimates that there are 20,000 employees in the transportation industry in Seattle, then it may be deemed appropriate to expand the data by multiplying by 20 rather than 10. However, it is generally preferable to utilize an expansion variable that was obtained from the original survey sample data to avoid circumstances where different data estimation processes can generate significantly different estimates. One method for refining the employee expansion process during the survey is to ask survey respondents how their level of productivity (in terms of revenue or tonnage per employee) compares to other companies in their field and in the survey region. Additionally, survey respondents can be asked what they think the average level of productivity is in their industry. This will enable the development of specific local employment estimates from local sources.

Another option for survey expansion variables would be the information provided in freight flow databases such as the BTS CFS. The CFS provides shipment values at the metropolitan level across two-digit commodities and several freight modes. In the previous example, the commodity-specific tonnage totals for the Seattle region can be considered the control total. The tonnage collected through the survey can be expanded based on the proportion of tonnage surveyed relative to the CFS total. Therefore, if the CFS estimates 40 million tons of transportation equipment shipped outbound and the survey estimates 1 million tons of transportation equipment shipped outbound, then an expansion factor of 40 can be applied to the surveyed data to estimate freight flow patterns of transportation equipment in the Seattle region.

It should be noted that this process of using indicator variables is limited by three factors:

- The statistical correlation between the indicator variable and the amount of freight used.
- The detail available for indicator variables at the subregional level (e.g., if apparel data is available at the zip code level).
- The accuracy of the indicator variable at the subregional level (e.g., some databases have estimates of these variables that are developed from other estimation processes).

Because of these limitations, the expanded data should be checked to confirm that the expansion process produced results that can be used for planning purposes. The CFS can be considered one source to validate the reasonableness of the results. The freight flows in each of the CFS regions can be compared to the expanded data. Members of trade associations for related industries also can be used to check the results for reasonableness.

User's Guide Worksheet Punch List

- Determine the indicator variable for the survey.
- Determine the control total for the indicator variable for each geographic zone and each industry within the study area.
- Determine the expansion factors for the study area by dividing the value of the indicator variable by the control total for each geographic zone.
- Confirm reasonableness of expanded data.

Step 10—Data Validation and Accuracy

Key Considerations

Data validation should be done by comparing the expanded survey data to other freight data sources in the region. Options include the CFS, local trade association data, and economic output data. In select instances, truck count data or roadside intercept survey data may be used to confirm local truck volumes and generalized commodity distributions estimated through the establishment survey process. Discrepancies between the expanded survey and these sources indicate that additional surveys may be needed to fully validate the local survey process.

Data accuracy is measurable through detailed statistical analysis of several of the variables collected in the survey process. There is a tradeoff between the number of samples collected in each geographic zone for each industry and the accuracy of the freight flow estimates.

Implementation Process

For establishment surveys using simple random sampling, all potential companies have an equal probability of being selected to be part of the sample. This sampling method creates an element of error due to the very randomness of the way in which the sample is chosen. The selected sample may result in sample data that are not necessarily representative of the whole population of interest, and this, in turn, results in inaccuracy in the estimates developed through the establishment survey process.

Equations 7 and 8 show how to assess the accuracy (or precision) (D) of a given sample size and confidence level using two steps. First, the standard error (SE) of the mean ($\bar{x}$) is calculated as a function of sample size, population, and standard deviation. The relative or absolute accuracy is then calculated as a function of the standard error, and the desired confidence level is reflected in the value of the z-statistic.

Standard Error: $$SE(\bar{x}) = \sqrt{\frac{\sigma^2}{n} * \frac{(N-n)}{N}} \qquad \text{(Eq. 7)}$$

where σ^2 = population variance,
n = sample population, and
N = total population.

Precision: $$D = SE(\bar{x}) * z \qquad \text{(Eq. 8)}$$

Similarly, Equations 7 and 9 show how to calculate in two steps the confidence level that corresponds to a given sample size and the desired degree of variable accuracy. First, the standard error of the mean is calculated as a function of sample size, population, and standard deviation. The confidence level (as reflected in the value of the z-statistic) is then calculated as a function of the relative or absolute precision and the standard error. The confidence level associated with various z-statistics is shown in Table 2.13.

Standard Error: $$SE(\bar{x}) = \sqrt{\frac{\sigma^2}{n} * \frac{(N-n)}{N}} \qquad \text{(Eq. 7)}$$

z-statistic: $$z = \frac{D}{SE(\bar{x})} \qquad \text{(Eq. 9)}$$

Table 2.13. Correlation of *z*-statistic and confidence levels for a given sample size.

z-statistic	Confidence Level
0.67	50%
0.84	60%
1.04	70%
1.28	80%
1.64	90%
1.96	95%
2.58	99%
3.29	99.9%

For full establishment survey efforts, it is possible to statistically compare the data collected from companies in different strata to determine whether shipping characteristics are related to other variables. For example, the relationship between employment and output for small and large companies in Seattle's apparel industry could be compared. Similarly, this relationship could be tested in Spokane's food manufacturing industry in the eastern side of town and the western side of town. These types of comparisons can also provide clues as to the types of future survey efforts that are most critical for a region.

Another type of test that can be done identifies whether different industries have similar truck trip generation characteristics. Industries with similar truck trip generation ratios may be candidates for consolidation. Conversely, within a single industry, if truck trip generation is found to have a large standard deviation, then further examination may be needed to determine whether there are other factors that need to be incorporated into the development of a truck trip generation function.

As sample size increases, the margin of error for collected data decreases and the confidence interval for collected data narrows. The percent margin of error at the 95 percent confidence level can be calculated at various sample sizes. This relationship can be plotted to illustrate how margin of error decreases as the sample size increases. The generalized relationship between sample size and margin of error for variables that are normally distributed is shown in Figure 2.5. At the 95 percent confidence level, there will be a 14 percent margin of error when 50 samples are taken. When 250 samples are taken, the margin of error in the estimate decreases to 6 percent. It is important to note that the sample size represents the number of completed survey responses for the specific variable not for the overall survey. Therefore, different variables that are estimated through the survey will have different levels of accuracy.

Example

An example of calculating accuracy and sample sizes is useful in illustrating the theory described in the previous section. This *Guidebook* cannot use the demonstration survey results for this example due to privacy guarantees given to respondents during the survey process. Therefore, a hypothetical example of an establishment survey of the paper products industry in the Seattle region will be used. For purposes of this analysis, Seattle will be divided into two regions: West Seattle and East Seattle. The data collected in this hypothetical example are shown in Table 2.14.

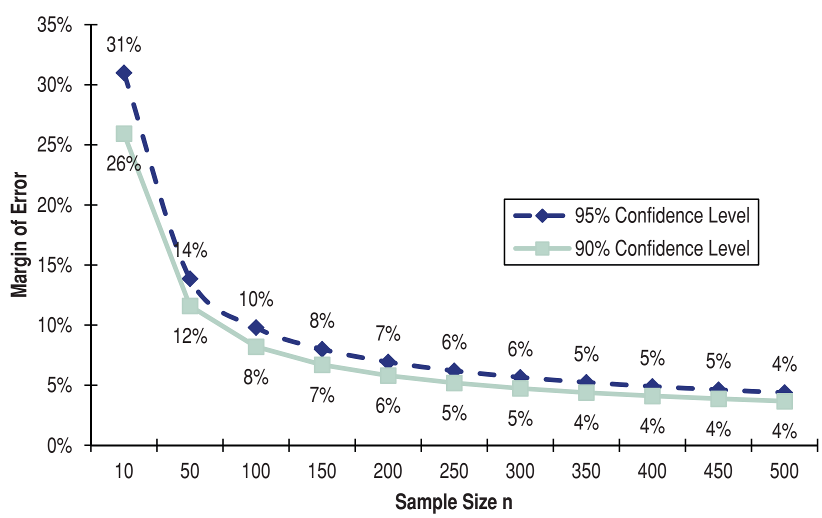


Figure 2.5. Relationship between sample size and margin of error.

Table 2.14. Data from hypothetical survey of the paper products industry in Seattle.

Survey ID	Tons Shipped to West Seattle	Tons Shipped to East Seattle	Total Tons Shipped
West Seattle Company 1	10	100	1,000
West Seattle Company 2	20	150	1,500
West Seattle Company 3	30	200	2,000
West Seattle Company 4	40	250	2,500
West Seattle Company 5	50	300	3,000
East Seattle Company 1	25	700	5,000
East Seattle Company 2	35	750	5,500
East Seattle Company 3	45	800	6,000
East Seattle Company 4	55	850	7,000
East Seattle Company 5	65	900	8,000

To determine the accuracy of the estimate of the total tons of paper products shipped from the Seattle region, the first step is to calculate the mean and the standard deviation of this value for the data collected, using the data in the far right column of Table 2.14. The mean is estimated using the following formula:

$$\text{Mean} = (1{,}000 + 1{,}500 + 2{,}000 + \ldots + 7{,}000 + 8{,}000)/10 = 4{,}150 \text{ tons}$$

$$\begin{aligned}\text{Standard Deviation} &= (1{,}000 - 4{,}150)^2 + (1{,}500 - 4{,}150)^2 + \ldots \\ &\quad + (7{,}000 - 4{,}150)^2 + (8{,}000 - 4{,}150)^2 \\ &= 2{,}461 \text{ tons}\end{aligned}$$

The next step is to calculate the standard error using the following formula:

$$\text{Standard Error} = 2{,}461/(\text{square root}(10)) = 778$$

Using a confidence level of 90 percent gives us a z-statistic of 1.64. This is used to calculate the precision of the estimate using the following formula:

$$\text{Precision} = D = 778 * 1.64 = 1{,}275.9$$

Therefore, the average number of tons produced by these 10 firms is 4,150, while the precision of the estimate of this average is +/–1,275.9 tons. This precision can be applied to the total tons estimated to be produced from this sample as well. In this case, the total tonnage estimated by the sample is 41,500 tons and the precision of this estimate with a 90 percent confidence interval is +/–1,275.9 tons. The precision value can be applied to the fully expanded data as well.

To determine the accuracy of the estimate of the total tons of paper products shipped from the Western Seattle region to the Eastern Seattle region in this hypothetic example involves the same formulas, but different values within the formulas. The mean is estimated using the following formula:

$$\text{Mean} = (100 + 150 + 200 + 250 + 300)/5 = 200 \text{ tons}$$

$$\begin{aligned}\text{Standard Deviation} &= (100 - 200)^2 + (150 - 200)^2 + (200 - 200)^2 + (250 - 200)^2 + (300 - 200)^2 \\ &= 79.1 \text{ tons}\end{aligned}$$

The next step is to calculate the standard error using the following formula:

$$\text{Standard Error} = 79.1/(\text{square root}(5)) = 35.3$$

Using a confidence level of 90 percent gives us a z-statistic of 1.64. This is used to calculate the precision of the estimate using the following formula:

$$\text{Precesion} = D = 35.3 * 1.64 = 58.0$$

Therefore, the total tons estimated to be produced from this sample is 1,000 and the precision of this estimate with a 90 percent confidence interval is +/–58.0 tons. The precision value also can be applied to the fully expanded data.

The formulas described in this section also can be used on the collected data to determine the sample size needed to achieve a specific confidence level for specific types of origin-destination combinations. Using this process, future surveys can target specific data elements within surveys for which confidence levels and precision are desired to be increased.

User's Guide Worksheet Punch List

- For each industry, estimate the mean, standard deviation, and standard error of tons for the companies included in the survey.
- Calculate the precision of the estimate of the tons produced based on the z-statistics and the desired confidence level.

2.3 Next Steps

The information presented in this chapter was designed to serve multiple purposes depending on where transportation agencies are in terms of considering development of an establishment survey. To read a description of collecting subnational commodity flow data using roadside surveys, proceed to Chapter 3.0. To identify the best next steps for your specific effort, refer to Chapter 6.0, "the Playbook."

CHAPTER 3.0

Collecting Subnational Commodity Flow Data Using Roadside Truck Intercept Surveys

3.1 Introduction

This section provides a comprehensive examination of methods for developing subnational commodity flow data using roadside truck intercept surveys. Roadside truck intercept surveys are conducted by interviewing truck drivers while they are in the process of operating their vehicle. Data is gathered from each truck driver regarding recent stops, future stops, goods being carried, and other specialized information related to the trip. These surveys typically occur at the side of a roadway, but also can occur at the entrance or exit of a major freight facility such as a port terminal gate or an intermodal rail yard.

Unlike establishment surveys, which were described in Chapter 2.0, roadside intercept surveys cannot be used to develop comprehensive commodity flow surveys or studies because roadside surveys focus exclusively on truck movements and miss shipments by any other mode. If the shipments that are intercepted are actually part of an intermodal shipment (for example, intercept surveys conducted at the gates of an intermodal terminal) they will not capture the true origin and destination of the shipment. Because the respondents to the survey are carriers instead of shippers or receivers of goods, the truck drivers may have only limited information about the characteristics of the shipment. This is especially true if the carriers are carrying mixed shipments (as would be the case with a less-than-truckload carrier).

Even with these disadvantages, there are some advantages to conducting roadside surveys. The biggest advantage is that these surveys provide more accurate data about corridor-level flows. As noted in the introduction to the *Guidebook,* there are a number of applications that benefit from this corridor-level detail. Roadside survey data also can be used in combination with other sources of commodity flow data to validate the results of efforts to model the routing of commodity flows.

The *Guidebook* identifies the following nine general steps involved in administering a roadside truck survey data collection program:

- Step 1—Site Selection
- Step 2—Questionnaire Design
- Step 3—Selecting Survey Dates and Times
- Step 4—Sampling Issues
- Step 5—Interviewer Training
- Step 6—Site Preparation (including traffic control)
- Step 7—Utilizing Uniformed Officers and Vehicles
- Step 8—Data Quality Control Procedures
- Step 9—Preparing Data for Usage in Commodity Flow Development.

Many of these steps are interrelated, but the discussion of each step is ordered as shown in the above bulleted list. The description of each step is structured to focus on the following four key elements described in the Playbook (Chapter 6.0):

1. **Key Considerations—**A brief description of the main issues encountered and tradeoffs that will need to be made for the step.
2. **Implementation Process—**A detailed description of how to implement the step.
3. **Example—**An example of how this step has been implemented in other studies. Many of the examples in this chapter are taken from two roadside intercept surveys conducted in Washington, the Eastern Washington Intermodal Transportation Study (EWITS) and the Strategic Freight Transportation Analysis (SFTA) study.
4. **User's Guide Worksheet Punch List—**Simple bulleted instructions that *Guidebook* users can check off to ensure that they have implemented each of the major steps involved in developing an establishment survey.

Each of these four elements is designed to focus on different aspects of conducting a roadside truck intercept survey and to reflect the types of activities that might be undertaken by a state or local transportation agency. For transportation agencies that are considering hiring a contractor to conduct a roadside survey, reading the "Key Considerations" section of each step will likely provide enough information for the generation of a request for proposals (RFP) on the topic. Transportation agencies that want to understand the details of how to conduct a roadside survey should begin by focusing on the "Implementation Process" sections, and then move on to the "Example" section. The "Example" section will describe previous specific efforts that have been undertaken in other regions. These examples can provide important lessons to improve existing efforts in an area or help to evaluate contractor responses to RFPs.

After transportation agencies have a sufficient background in all of the aspects related to developing a roadside truck intercept survey, the "User's Guide Worksheet Punch List" sections can be used to walk the agency through all of the specific steps that need to be completed to implement the survey. This section also can be used to help evaluate RFP responses.

3.2 Step-by-Step Process for Conducting Roadside Truck Intercept Surveys

Step 1—Site Selection

Key Considerations

The choice of locations for conducting roadside surveys is primarily driven by the practical constraints of feasible data collection locations. The ideal locations for roadside truck surveys are truck weight and inspection stations, because they have the space to accommodate parking for large trucks and they have equipment that can safely intercept trucks from the traffic stream. However, the locations of these facilities in most states is such that they tend to over-represent intercity trucks when compared to the universe of commodity flow patterns that might be encountered in urban areas. Additionally, using these facilities also requires cooperation from the facility operators, which can be challenging to secure if there is any sense that the surveys could interfere with the inspection and enforcement activities that are the primary function of these facilities. It also is worth noting that the operating hours of weigh stations may be limited, and this can make these locations less than ideal for conducting robust roadside surveys. Finally, surveys at weigh stations need to include trucks that are cleared to bypass the weigh stations in advance using a system such as NCPass.

Implementation Process

Because the major constraint in identifying roadside surveys is finding feasible locations, the first step in this process is to identify all of the feasible locations in the transportation agency's area of concern. This identification process will likely include truck weigh stations and truck inspection stations, but it also can include truck pull-out areas, public rest areas, and large parking lots located immediately adjacent to the corridor of interest. Potential sites will need to be screened to ensure that there is sufficient physical space to conduct the surveys. This screening can often be done using a satellite mapping service, but in some cases site visits may be needed to ensure that there are sufficient truck and automobile parking spaces, roadways with sufficient turning radii and acceleration/deceleration distances to get trucks safely in and out of the survey site.

Another requirement of the site is that it provides surveyors with the capability to communicate with truck drivers while they are on the road and request that they participate in the survey. This can be done through variable message signs, static arrow signs, and, in some low volume/speed cases, a flagman or flagwoman directing trucks to the survey location. Most weigh stations have this technology built in as they often need to visually inspect vehicles. Similarly, gate surveys have natural stopping points for trucks that enable surveys to be conducted easily. At other types of locations, such as rest areas and truck pull-out locations, a variable message sign that can be quickly changed at the discretion of the survey site manager will be needed to communicate with trucks regarding their participation in the survey.

After potential site locations have been screened, they should be mapped and compared to the origin-destination patterns for the types of trucks for which information is desired. This will help ensure coverage of key corridors at critical points and ensure that key origin-destination pairs are captured. One potential method of identifying key truck corridors is to locate the top truck count locations in the region and ensure that there are survey locations that capture trucks at or near these locations. An origin/destination check involves identifying already known, concentrated generators of freight in the region and determining whether the roadside survey locations will capture trucks moving between these locations. This determination tends to be more of a judgment call that does not have specific quantitative rules to follow.

Truck count data can also be helpful for site selection. Ideally, this count is collected with sufficient detail to indicate time of day and day of week variability. Having this information can ensure that the highest volume locations are incorporated into the study. Additionally, it ensures that the specific data collection time periods match with the most important days and times for truck activity on the corridor.

The next step is to contact the facility operator to obtain approval to conduct roadside truck surveys. For weigh stations and rest areas, the facility operator is often a department within the state DOT, but in some cases, state safety departments or highway patrol agencies operate these facilities. International land border crossings are operated by U.S. Customs and Border Protection and often require security issues to be addressed prior to approval. Marine ports can be operated by state or local agencies or authorities, but many terminals are privately operated, with the public entity acting as the "landlord." In these cases, it is important to obtain approval from both the terminal owner and operator. Intermodal rail terminals are privately operated by the railroads. The approval process in some instances may simply involve permission from the day-to-day operator at the facility. In other instances, approval may be required from higher-ranking officials at an organizations' headquarters.

Example

In the early 1990s, the state of Washington commissioned the Eastern Washington Intermodal Transportation Study (EWITS) to better understand freight flows in the eastern portion of the state. This was followed by the Strategic Freight Transportation Analysis (SFTA) project with the goal of better targeting investments in freight infrastructure. To support both of these efforts, a series of roadside truck origin-destination surveys were conducted throughout the state.

To identify potential survey site locations for the EWITS and SFTA, preliminary data analysis was conducted using the state's traffic count and vehicle classification data in order to identify the primary freight concentrations using the state highways. These data were collected at weigh-in-motion stations and permanent traffic-recorder stations across the state. These data indicated that the heaviest truck volumes were on the long-haul Interstate corridors, in particular, I-5 and I-90. Secondarily, the major state highways had large volumes, particularly highways connected to the state's agricultural regions. This included U.S. 395, U.S. 97, and U.S. 12. The state also was interested in understanding the travel patterns of trucks traveling between Washington and Canada. Therefore, survey locations at or near to border crossings were included as well.

Figure 3.1 shows the freight economic corridor map for the Washington DOT (WSDOT). The color of each highway segment is based on tonnage estimates: the orange-colored highways carry 10 million tons or more per year and the green-colored highways carry between 4 and 10 million tons per year. Alternatively, truck count data can be used to identify high truck volume locations along the corridor. For the EWITS and SFTA surveys, the feasible roadside truck survey locations were found to match up well with the high truck volume locations that were indicated by truck count data. Note that for cases where potential survey sites are not located near count sites, supplemental classification count data may need to be collected to expand the raw roadside truck origin-destination survey data.

Another check on the data was to determine whether the survey locations captured all of the key origin/destination patterns. Washington's largest metropolitan region, Seattle, also is the state's largest freight generator. Intercity truck trips from Seattle were well covered via the stations at I-5 in Everett, I-5 at Kelso South and I-90 in Cle Elum. Spokane is the second largest metropolitan region in Washington, and it likely also generates a relatively large fraction of freight. Spokane was well covered by the survey locations on I-90 in Tokio and at the border with Idaho. It also was covered to the north by a survey location at Deer Park on U.S. 395. However, there were no surveys to the south of the Spokane region, so truck trips between Spokane and points dead south of Spokane were not covered. This would include truck trips from the Spokane region to Boise, Idaho, and Salt Lake City, Utah. It is important to note these deficiencies at the point of site selection, so that a decision can be made regarding whether enough of the key origin-destination pairs are being captured to justify implementation of the full survey effort. For the roadside surveys in Washington, it was determined that the lack of coverage of southward, Spokane-generated truck trips was not problematic enough to discontinue the study.

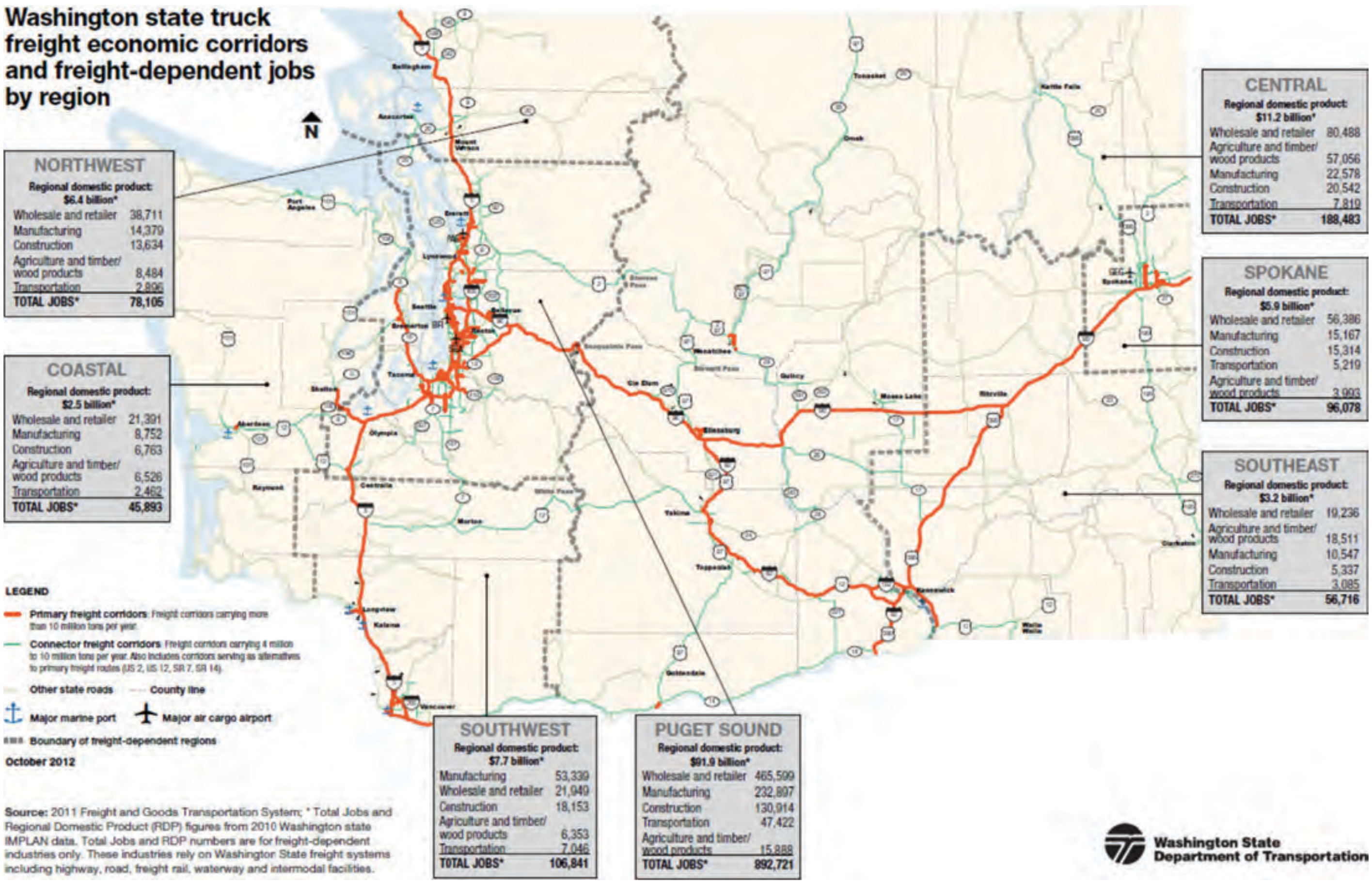


Figure 3.1. Washington freight economic corridor map.

As noted earlier, it is oftentimes possible to conduct roadside truck surveys at rest areas, truck fuel stops, and designated truck parking areas near metropolitan areas. One of the disadvantages of using truck stops and truck parking areas is that trucks are not required by law to enter these areas, whereas they are required to enter weigh stations. For the Washington roadside truck surveys, it was determined that truck stops and rest areas would not be used due to the additional work that would be needed to develop representative samples at those locations and the availability of more suitable locations. The survey locations were restricted to weigh stations and ports of entry. The final locations are shown in Figure 3.2.

To gain permission to use weigh stations, ports of entry, and border crossings to conduct roadside surveys, it was first necessary to make an official request and obtain approval from the respective agencies overseeing each site. In Washington, this began with scheduling a meeting with senior Washington state DOT officials, Washington State Patrol officials and U.S. and Canadian customs officials to explain the purpose of the survey, describe the benefit of data obtained from the study, and provide assurance that the survey would not interfere with the typical inspection activities of the state patrol and customs officials. It is not a given that all state organizations would approve of similar surveys.

Coordinating information was then sent to all regional offices explaining the survey process and asking staff members to comply. In addition, information on operating hours for each location was obtained so that surveys could be scheduled on days when weigh stations already would be staffed.

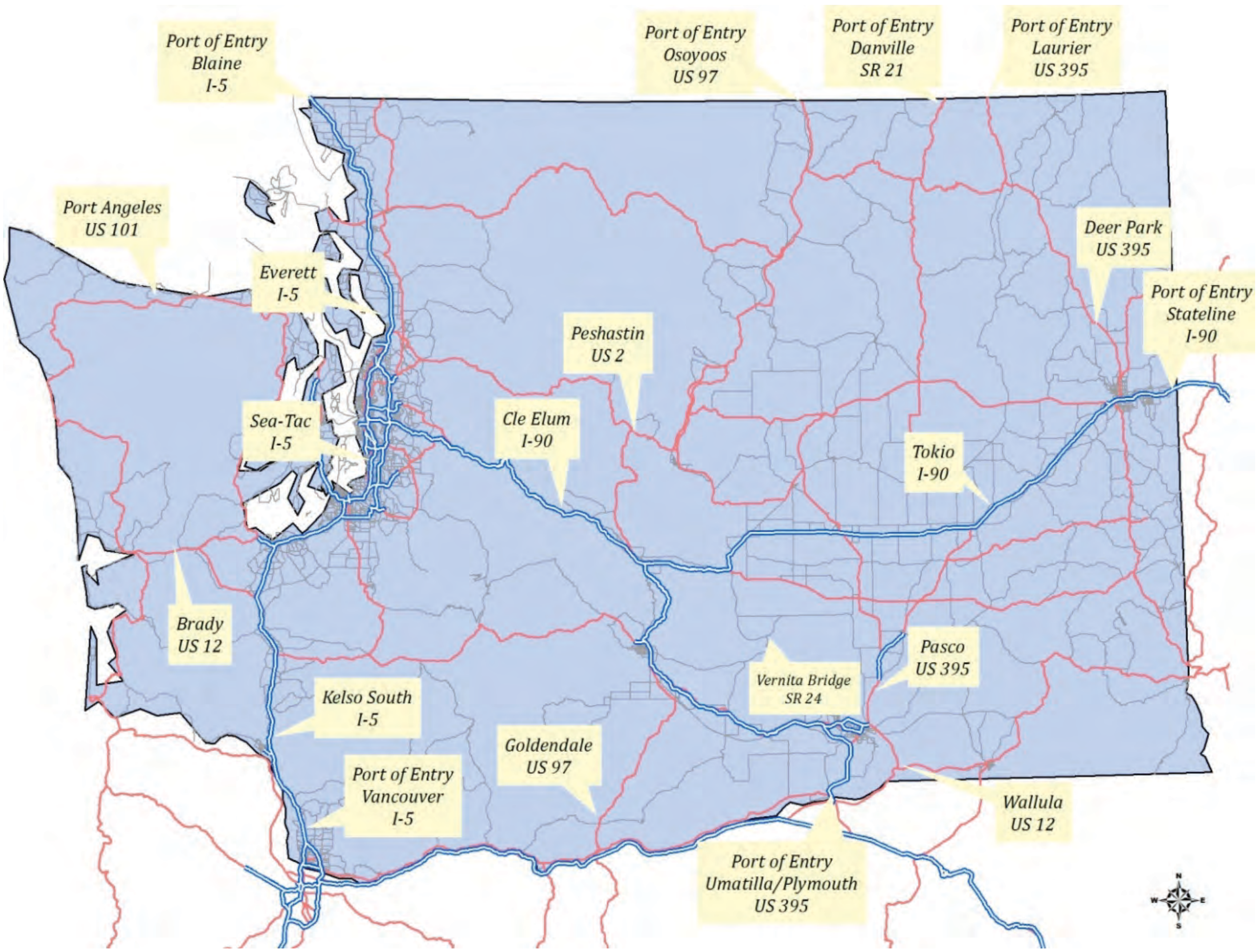


Figure 3.2. Roadside survey sites for the SFTA origin-destination roadside survey in Washington.

User's Guide Worksheet Punch List

Conduct the following activities for the area of interest:

- Locate all truck weigh stations, rest areas, truck pull-out areas, and points of entry through contact with operating agencies or Internet research.
- Locate major port terminals and rail intermodal yards.
- Confirm the feasibility of all of these locations for conducting roadside surveys using a satellite mapping service, site visits, and/or contacting facility operator.
- Determine truck corridors of interest using truck volume information and other relevant information specific to the region.
- Identify corridors (or segments along corridors) of interest that are not well covered by the existing locations.
- Identify concentrations of freight generation that represent truck trip origin/destination pairs in the region. Determine which of these origin/destination pairs are not well covered by the potential locations listed in the bullets above.

- Locate major truck stops and large parking lots nearby locations along corridors of interest and in between major freight generation locations that could potentially be used to fill in gaps for roadside surveys where other types of locations are not readily available.
- Contact facility operators to confirm their willingness to participate.

Step 2—Questionnaire Design

Key Considerations

Designing roadside truck survey questionnaires requires a clear definition of the purpose of the data collection effort and also involves anticipating truck driver response to questions and a knowledge of the surveyors' familiarity with trucking operations. Additionally, there are tradeoffs that involve the number of questions asked relative to the response rate and the number of questions asked relative to the number of surveys that will be conducted over a fixed period of time. To the extent that a smooth flow of questions can be preserved, it is best to ask the most important questions first, followed by questions of less importance. It is ideal that someone with intimate knowledge of the trucking industry be involved in the survey design process, including potentially having the questionnaire reviewed by the state motor trucking association.

Implementation Process

The general structure of a roadside truck survey questionnaire is to start with questions that the surveyor can answer by visually observing a truck as it approaches the survey location. These questions are followed by requests for information that must be collected from the driver about the current trip, and are followed by requests for general information about the trucking company or broad questions about services provided. The survey usually concludes with open-ended questions of interest to the transportation agency. Items that can be noted as the truck is approaching include the following:

- Time of day
- Vehicle and trailer configuration
- Number of axles
- Hazardous material placard (if any)
- Name of trucking company
- Identifying information on trailer (e.g., DOT registration number)

This information is typically easy to identify. There are a few cases where some or all of the hazardous material placard information may not be legible from a distance.

Information that can be collected from the driver about the current trip includes the following:

- Location of last stop and next stop of the truck
- Location of last pickup/delivery and next pickup/delivery
- Ultimate origin and destination of cargo being transported (can be read from a manifest or bill of lading)

- Facility type (home, intermodal facility, warehouse, port, etc.) at last stop, next stop, last pickup/delivery, next pickup/delivery, ultimate origin and destination
- Commodity being carried
- Weight of cargo
- Empty, partially loaded, or fully loaded
- Identification of routes used or to be used on this trip

An important distinction is between a location where a truck stops, a location where a truck picks up or delivers goods, and the locations where the cargo begins or terminates the shipment. A location where a truck stops can include diesel filling station, hotel, home base, or restaurant. The location of last pickup or delivery is the last point at which the cargo inside the truck changed and may have occurred a few days prior to the survey or can occur a few days following the survey. The ultimate origin and destination of the goods may be different than the last pickup or delivery, particularly for less-than-truckload trucks or intermodal goods. It is important to note that the driver may or may not know the ultimate origin and destination of the cargo, because these moves may be outside of the responsibility of the truck driver.

Each of these pieces of information is important for different reasons. Typically, information on last stop and next stop is the most important for truck modeling. Commodity flow models tend to benefit most from information on ultimate cargo origin and destination, because that provides useful information on producers and consumers of goods, which is the basis of commodity flow modeling. However, it is common for truck drivers to not know whether the ultimate cargo origin and destination is different from the last pickup/delivery and next pickup/delivery. Therefore, the pickup/delivery information tends to be the best information available from roadside truck surveys regarding the movement of the cargo.

Another key element of truck roadside surveys is the determination about the geographic specificity requested for origin and destination information. The most commonly geographic information requested through these surveys is city and state. However, there are several circumstances that require more refined trip information at one or both of the truck trip ends being surveyed. In these circumstances, street address information can be requested. However, truck drivers are sometimes not willing to provide such specific information about their customers. Another option is to ask for zip code information at trip ends. Truck drivers may or may not have this information readily available, and they may not want to spend the time to look up the information on the bill of lading or other documents related to their shipments. Other options include getting information on freeway entrance or exit points, which can serve as a proxy for final origins or destinations. Alternatively, the survey instrument can include a map with prespecified zones that are at the subcity level, and the truck driver can pick from one of the zones on the map. In cases where multiple stop information is needed, the survey questionnaire can be modified to reflect the need to collect this information. For these surveys, it is recommended that an electronic data collection device be utilized, because hard copy surveys would need to be very lengthy to allow for collecting information on several stops.

It also is important to determine the amount of commodity detail that is needed from the roadside truck origin/destination survey. Truck drivers tend to know in general the goods that are being carried, but, on occasion, additional probing is needed to specify this to the level that is needed for the survey. Additionally, there may be instances in which commodities are predefined into categories for the truck driver to choose from. Just as truck drivers may not know the specifics of the goods being carried, truck drivers may not have information regarding cargo weight or whether or not the truck is partially or fully loaded. Inquiries regarding routes selected on the trip are typically best conducted by preselecting no more than five routes of interest and having the truck driver identify the ones that will be used. When detailed information is needed, it is typically best to provide the truck driver with a map and let the driver sketch the route directly onto the map.

General information about the trucking company or broad questions about the trucking services offered include the following:

- Location of truck's home base
- Type of trucking firm (truckload, less-than-truckload, private)
- Number of vehicles in fleet
- Number of times that the truck driver passes the survey location (or alternative location) in a day, week, or year
- Number of years of experience of the truck driver

Open-ended questions can include questions on infrastructure, operations, and/or on policy that span a wide range of topics. Survey design decisions about these kinds of questions require balancing the information needs by the transportation agency with the knowledge of the truck driver and the time needed to conduct the survey. Typically, these questions should be carefully phrased so that they do not imply that any immediate action is going to occur. Examples of these open-ended questions are the following:

- What improvements should be made to this corridor and where should they be made?
- Are there any safety issues with this corridor in regard to truck operations?
- Would you be willing to pay a toll to use this corridor if additional capacity were added?
- In your opinion, in the last 5 years has congestion on this corridor gotten worse, better, or stayed roughly the same?

There are several interesting pieces of information regarding the cargo that the truck driver is not likely to have or would be unwilling to answer, such as the value of the cargo, the date that the product was produced, or the amount paid by the customer to move the goods.

After deciding which information is sought through the surveys, the next step is the physical construction of the questionnaire. In recent years, it has been common for many of these surveys to be conducted electronically using handheld devices. The advantage of an electronic method is that it can automatically capture information in a database format that allows for easier processing at a later time. The advantage of the traditional paper survey format is that it is easier to review the previous questions answered by the truck driver to ensure consistency throughout the survey. The paper survey format also makes it easier for the survey manager to monitor the responsiveness of truck drivers and make field adjustments, if necessary, to improve the survey process. If a traditional paper survey is conducted, then the paper survey should be no more than two pages long to allow for an easy data collection process. Printed questionnaire cards have been used in several roadside surveys as a method to slightly increase response rates. The cards can be filled out by drivers at their convenience and then be mailed back to the survey team.

Additionally, the mechanics of the field survey need to be considered, especially if the surveyor will be referencing a map, trying to hold paper and a clipboard, and writing or typing questions all at the same time. Upon completion of the survey questionnaire, a practice survey should be conducted in the office with one surveyor pretending to be a truck driver. This will assist in fine-tuning the survey. It also will provide an estimate of the length of time it takes to complete the questionnaire, which assists in the overall survey development process.

Example

The SFTA roadside truck surveys asked questions in the following categories:

- Time of day
- Vehicle and trailer configuration (e.g. dry van, flatbed, bulk, tanker)
- Number of axles
- Hazardous material placard (code)
- Name of trucking company and home base

- Empty weight of the vehicle/trailer
- Loaded weight (current payload weight)
- Commodity on-board
- Cargo origin (city, state)
- Facility type for cargo origin
- Cargo destination (city, state)
- Facility type for cargo destination
- Identification of highways used on this route (two maps, state level, and urban map for Puget Sound)
- Number of times in past 7 days this route has been traveled.

Data on the first five items listed above (time of day, vehicle configuration, number of axles, hazardous material placard, and trucking company home base) were observed visually by the surveyor while the truck was approaching the survey location. The first question addressed to the truck driver was whether they were loaded or empty, and the survey continued through the questionnaire from there (see Figure 3.3).

For vehicles that were making many stops/deliveries, the SFTA survey questionnaire was modified slightly to include recording each stop along the route at the city and state level within Washington or province level for Canada. This modification was found to allow for questions that were easy for truck drivers to answer, but the responses were likely to lack adequate detail for evaluating freight activity within large cities such as Seattle.

User's Guide Worksheet Punch List

Conduct the following activities for the roadside truck survey:

- Identify the most important pieces of information to be obtained from the survey.
- Specify the survey questions that will solicit this information.
- Identify additional questions that are likely to provide information that will assist in freight planning efforts over the next 10 years. These questions should be limited to those that truck drivers are likely to answer in a time-efficient manner, thereby focusing the resources of the survey effort on the most important questions to be answered.
- Develop a survey questionnaire based on the responses to the above bullets.
- Based on the length of the survey questions and available technology for the survey, determine whether it will be conducted using electronic or paper survey instruments.
- Conduct a practice run of the questionnaire to estimate the length of time it takes to complete the survey questionnaire as this has implications for future steps in the roadside truck survey process.

Step 3—Selecting Interview Dates and Times

Key Considerations

The primary considerations for selecting interview dates and times are any restrictions inherent in the survey locations, likely sampling issues, the resources available to conduct the survey effort, and/or the need to collect data during different periods within typical truck cyclical behavior (e.g., time of day, day of week, and season in year).

For Office Use Only
Survey ___________
QCA ___________
Input ☐

Washington State Department of Transportation and
Washington State University
Truck Traffic Survey, Spring 1994

Please Remember - your Club is depending on YOU for the Quality Control Award!

✓ *Write neatly!* ✓ *Do not abbreviate!* ✓ *Complete all required questions!*

Thank You!

CONFIDENTIAL

1) Station Location: Tokio Westbound

2) Initials of Interviewer: ___________

3) Interview shift:

1. Day Shift	2. Evening Shift	3. Night Shift
7:00 a.m. - 3:00 p.m.	3:00 p.m. - 11:00 p.m.	11:00 p.m. - 7:00 a.m.

4) Time of interview: AM PM

5) Is this truck a part of the "official sample"? 1) ☐ Yes 2) ☐ No

6) Truck Configuration *[Check only one truck configuration]* ***[see Quality Control Notes for definitions]***	7) Trailer Style *[If appropriate, check more than one trailer style]* ***[see Quality Control Notes for definitions]***
1. ☐ Straight truck	1. ☐ Van (without temperature control)
2. ☐ Truck and trailer	2. ☐ Van with temperature control
3. ☐ Tractor only	3. ☐ Flatbed
4. ☐ Tractor and trailer	4. ☐ Car carrier
5. ☐ Tractor with two trailers	5. ☐ Hopper or belly dump
6. ☐ Other (specify) ___________	6. ☐ Stake and rack
	7. ☐ Concrete mixer
	8. ☐ Tanker
	9. ☐ Float or low boy
	10. ☐ Dump
	11. ☐ Container
	12. ☐ Wood Chip
	13. ☐ Animal Carrier
	14. ☐ Logging
	15. ☐ Belt
	16. ☐ Other (specify) ___________

8) Total number of axles on the ground: ___________

9) Is a hazardous material placard displayed? 1) ☐ Yes ID#. 2) ☐ No

Figure 3.3. Roadside survey questionnaire used in Washington state (EWITS).

(continued on next page)

[Please ask the following questions] **CONFIDENTIAL**

10) Trucking company name: ____________________
Very Important! Do Not Abbreviate! Be exact!

Trucking company home base: City: ____________ State/Province: ____________

12) What is the unloaded weight of this vehicle? ________ lbs.

13) Is this vehicle carrying cargo or is it empty? ☐ carrying cargo *[Ask Q14-21]* ☐ empty *[Ask Q22-27]*

14) What is the major commodity on board: ____________________
DO NOT ABBREVIATE! BE SPECIFIC!

15) How much does the cargo you are carrying today weigh? ________ lbs.

Complete only the one column that applies to this trip. No round-trip information, please!

Trucks CARRYING cargo:

Where did you pick-up this cargo?

16) City: ____________

17) State/Province: ____________

18) Facility: *[see Quality Control Notes]*

1) ☐ trucking yard
2) ☐ railroad yard
3) ☐ river or ocean port
4) ☐ airport
5) ☐ factory, processing plant, or sawmill
6) ☐ warehouse/distribution center or post office
7) ☐ farm or forest
8) ☐ retail store or gas station
9) ☐ job or construction site
10) ☐ other ____________

What is the destination of your cargo?

19) City: ____________

20) State/Province: ____________

21) Facility: *[see Quality Control Notes]*

1) ☐ trucking yard
2) ☐ railroad yard
3) ☐ river or ocean port
4) ☐ airport
5) ☐ factory, processing plant, or sawmill
6) ☐ warehouse/distribution center or post office
7) ☐ farm or forest
8) ☐ retail store or gas station
9) ☐ job or construction site
10) ☐ other ____________

Trucks WITHOUT cargo:

Where did this trip without cargo begin?

22) City: ____________

23) State/Province: ____________

24) Facility: *[see Quality Control Notes]*

1) ☐ trucking yard
2) ☐ railroad yard
3) ☐ river or ocean port
4) ☐ airport
5) ☐ factory, processing plant, or sawmill
6) ☐ warehouse/distribution center or post office
7) ☐ farm or forest
8) ☐ retail store or gas station
9) ☐ job or construction site
10) ☐ other ____________

Where will your trip without cargo end?

25) City: ____________

26) State/Province: ____________

27) Facility: *[see Quality Control Notes]*

1) ☐ trucking yard
2) ☐ railroad yard
3) ☐ river or ocean port
4) ☐ airport
5) ☐ factory, processing plant, or sawmill
6) ☐ warehouse/distribution center or post office
7) ☐ farm or forest
8) ☐ retail store or gas station
9) ☐ job or construction site
10) ☐ other ____________

28) What Washington highways were used to travel between the two locations identified above?

____________________ *(Remember, accurately highlight attached map!)*
Write out the highways used to get between the two locations identified above.

29) ***Including this trip***, how many times has this truck traveled the above route in the past 7 days?
________ Times ☐ Don't know

Figure 3.3. (Continued).

Implementation Process

In practice, the selection of interview dates and times is typically constrained by the resources available for the survey effort, restrictions inherent at survey locations, and the number of locations that will be surveyed. Therefore, it is important to first understand whether there are limitations at any of the survey locations of interest. At weigh stations, dates and times can be limited by the operating hours of enforcement activities. Often times, weigh stations are only open during daytime hours, and they may be closed on weekends as well. If evening and nighttime surveys or weekend surveys are desired, that often requires restricting surveys to the few dates when the weigh stations are operating during these hours or requesting special permission from the weigh station operator.

Rest areas and truck pull-out locations may become infeasible during nighttime hours as reduced visibility creates safety concerns for truck drivers, surveyors, or other users of the rest areas. Similarly, truck stops also may be limited in terms of their hours of operation. Additionally, truck stops have safety and security considerations that must be addressed when conducting nighttime surveys.

Many roadside truck survey efforts also are constrained by the resources available. If a roadside survey is funded to the level of $500,000 for field data collection, and there are 40 locations that are to be included in the survey, that leaves $12,500 to collect data at each location. If each location is staffed with three surveyors at a cost of $40 per hour per surveyor, then that allows for a little over 100 hours of survey activity at each location. The selection of appropriate survey dates and times is then made using these 100 hours per location as a general guidepost.

Once the number of hours available at each location has been estimated, it is important to consider the cyclical nature of the trucks that are to be surveyed at different locations in order to determine at which points of the cycle the survey is desired. For instance, volumes of some agricultural commodities peak in late summer, but this can be highly variable depending on the crop. International intermodal container volumes tend to peak in the early fall in advance of the Christmas shopping season. Other types of goods have other types of volume peaks throughout the year. To guard against the potential of seasonal bias in the collected data, it is ideal to conduct one survey in each of the four seasons. Based on resource availability, this may be done only at locations where seasonal variability is considered to be most the important, such as agricultural locations.

There also is the potential for variation by day of week and time of day in collected survey data. Typically, nighttime and weekend truck movements include a higher proportion of long-haul trucks relative to daytime, weekday truck movements. Therefore, a fully representative sample of data would include data collected during day and night and during the week and on weekends. However, this may again be restricted by resource availability. Additionally, if the purpose of the data collection effort is primarily to support the development of a travel demand model, then the most important data to collect may be data on typical weekday activities. Another alternative is to collect survey data during the nighttime and weekends at select locations and then use the off-peak data at these select locations to factor the data at locations where more limited data were collected. Similarly, if focus is on trucks most likely to be impacted by congested traffic conditions, then it may be important to collect data during the daytime.

Example

The dates and times for data collection in Washington were selected to ensure the collection of information on the seasonality of freight shipments and to ensure that fluctuations in the large and diverse agricultural economy across the state could be tracked in the collected data. To accomplish these aims, it was decided to survey each site during each of the four seasons. Each survey was conducted on a Wednesday, in order to capture midweek flows that were expected to be "normal" freight flows and avoid unusual traffic peaks or valleys at the beginning or end of the week.

The survey schedule also was designed to minimize double counting of trucks on a single day. Sites surveyed on a given day were selected to be somewhat mutually exclusive. The hours of operation of each weigh station also influenced when data could be collected. In Washington, most weigh stations on heavy freight corridors (I-5 and I-90) are open 24 hours during days of operation. Other stations were only open for short portions of the day. At these locations, data collected from nearby vehicle classification counters were used to estimate truck activity for the full 24-hour period. During the implementation of the survey, there were multiple unforeseeable events that resulted in rescheduling of data collection at a particular site. The need to reschedule could be caused by highway accidents, last-minute state patrol officer reallocation, and issues with assembling the survey team.

User's Guide Worksheet Punch List

- Estimate the approximate budget available per survey by dividing the anticipated resources for data collection by the number of locations as identified in Step 1. Note that the resources available for data collection will be only part of the total resources for the survey because some resources must cover the costs associated with survey setup, travel costs, data analysis, and report writing.
- For each survey location, determine whether the effort will be to have the survey capture seasonal variation.
- For each survey location, determine whether the effort will be to have the survey capture time of day or day of week variation.
- Estimate the number of surveys that can be conducted at each location by dividing the total number of hours per location by the estimated number of surveyors at the location and the estimated hourly cost per surveyor.
- Adjust the available budget at each location to optimize the data collection at each location relative to available survey hours while maintaining a constant budget for total data collection activities.
- Optional Task: Redo the above four bullets assuming that all weigh stations in the sampling plan are open only during the weekday between the hours of 8:00 a.m. and 6:00 p.m.

Note that this step in the worksheet will need to be repeated following completion of Step 4 on sampling issues.

Step 4—Sampling Issues

Key Considerations

Sampling is perhaps the most complex quantitative issue related to setting up a roadside truck survey program. The main issue with sampling is determining the necessary number of surveys and the necessary timing of those surveys that will ensure statistical confidence in the survey results. This determination depends on several factors, including the key survey variables of concern, presurvey assumptions about variability in truck behavior, and feasible hours of surveying at each location.

In practice, the overall sample size tends to be determined by resource availability rather than statistical analysis. However, the statistical analysis can still be used to assist in determining which survey locations should feature additional samples and whether additional surveys are needed for particular times of the day, days of the week, or seasons in the year.

Implementation Process

Steps 1 and 3 have covered identifying a set of potential survey sites along with the dates and times of potential surveys at each site. This step addresses how many trucks to survey at each site, given the fiscal constraints of the roadside intercept survey process, and the types of freight questions that are usually answered through truck roadside intercept surveys.

From a statistical perspective, to determine the appropriate number of samples it is important to isolate a few key variables collected from the survey questionnaire that are deemed the most critical. One example of a variable that is often considered important for origin-destination surveys is commodity. To determine the sample size needed to estimate for a particular commodity, the first step is to generate an estimate of the commodity percentage and variability through the location. This estimate can be developed using nearby or state-level commodity distribution percentages such as those available in the FAF database. If resources allow, pilot surveys can be conducted to develop estimates of the mean and variation for key variables. Alternatively, these estimates can be confirmed for accuracy at the conclusion of the roadside survey effort. The next step is to develop a confidence level for the estimate. The confidence level typically used is 95 percent. Then, the approach used in Step 4 of Chapter 2.0 can be applied to this scenario. As stated in Step 4 of Chapter 2.0, there are three key equations used to estimate sample size:

- Equation 1 is the standard normal expression for sample mean stating that with 95 percent confidence, the mean will lie within two standard deviations.
- Equation 2 is derived from the first equation.
- Equation 3 is achieved by solving for the sample size, *n*.

W is the width in units of the confidence interval. So the wider the confidence interval, the lower the sample size necessary in order to maintain the 95 percent confidence level. The mean is represented by $\bar{x}$ and the standard deviation is σ.

$$\left(\bar{x} - \frac{2\sigma}{\sqrt{n}}, \bar{x} + \frac{2\sigma}{\sqrt{n}}\right) \quad \text{(Eq. 1)}$$

$$\frac{4\sigma}{\sqrt{n}} = W \quad \text{(Eq. 2)}$$

$$n = 16\sigma^2/W^2 \quad \text{(Eq. 3)}$$

As in Step 4 of Chapter 2.0, an alternative approach would be to develop the sample size based upon some acceptable threshold of error, in which case, Equations 4 though 6 would be used reflecting the probability that $\hat{p}$ lies within two standard deviations of the mean. Again, *W* refers to the width in units of the confidence interval and *B* refers to the allowable error. Thus, for a 10 percent error (90 percent confidence) in a normal population, a sample of at least 100 would be needed. Likewise, in order to achieve 99 percent confidence (allowing only 1 percent error), a sample of 10,000 would be needed.

$$\left(\hat{p} - 2\sqrt{0.25/n},\ \hat{p} + 2\sqrt{0.25/n}\right) \quad \text{(Eq. 4)}$$

$$4\sqrt{\frac{0.25}{n}} = W \quad \text{(Eq. 5)}$$

$$n = \frac{4}{W^2} = \frac{1}{B^2} \quad \text{(Eq. 6)}$$

With the sample size determined for key variables, the next step is to determine the survey periods to consider as part of the analysis. Survey periods are generally determined based on observed fluctuations of truck count data at a particular location, but could vary based on a number of different factors that are known at a location. For each period, the truck activity is assumed to be different. Therefore, it is important to collect the correct number of samples from each time period. Dividing the time period by the time it takes to conduct the survey, then multiplying it by the number of surveyors at each location will enable a determination of whether the sample size can be collected over the course of a day or whether a multiday period will be needed. Ideally, the number of surveyors can be altered during the design phase to ensure that the optimal number of surveyors is present at each location to meet the sample size requirements at each location. However, the number of surveyors at a location may be limited by the physical constraints of the survey site and labor availability of the survey team.

It is very likely that origin-destination patterns and potentially commodity distributions change over the course of the day. Nighttime truck operations tend to include a higher proportion of long-haul truck trips. The difference between daytime and nighttime operations may be more pronounced at very urban locations, where daytime trucks are often used for the purposes of urban distribution and service facilities that are open during the daytime hours. Roadside intercept surveys are often used to estimate through truck trips relative to internal truck trips at a particular point. To most accurately answer this question, collecting nighttime surveys is preferable. For this reason, it is ideal to collect 24-hour data, at least for a handful of locations, to enable an estimation of the variability between daytime and nighttime operations. This variability can be established based purely on time of day or also can be developed separately for single commodities or commodity groups.

If it is believed that there is a significant seasonal variability in origin-destination patterns or commodity distribution in the state, then the survey process should be repeated during each of the four seasons at a minimum. One indication of the potential of seasonal variability at a location is the occurrence of wide fluctuations in truck volumes during different seasons. The potential for seasonal variability also can be obtained by talking to experts in the industries that produce the commodities that are identified in initial intercept surveys. An alternative to conducting full survey efforts during each season to capture seasonal variability is to conduct seasonal surveys at a handful of locations. From the data collected at these locations, seasonal variability factors can be developed for each commodity or for commodity groups. This is similar to the approach for capturing seasonal variability using a limited set of survey locations.

As mentioned previously, in practice most roadside intercept locations are constrained by budgets. Therefore, the vast majority of roadside intercept surveys are conducted such that each location is surveyed over the course of a single day. Sometimes the survey period is roughly during the midday peak hours of the truck operations. Surveys with larger budgets can capture trucks over a 24-hour period at each location. The goal of most surveys is to have some data available in as many locations as possible given the limits of the budget.

Example

One of the specific purposes of the roadside intercept surveys conducted in the EWITS and SFTA studies was to create a truck origin-destination matrix for the state of Washington. To achieve this, survey team staffing was one of the critical factors examined in the survey design process.

Generally, survey team staffing throughout the day was determined based on time-of-day freight traffic distributions from WSDOT traffic count data, knowledge concerning each survey site, and physical parking limitations, which limit the number of trucks that can be interviewed at any one time. Increasing the number of days that the survey is conducted doesn’t impact the sample size because the population of traffic likewise increases for the additional days surveyed.

If the maximum number of questionnaires collected at a given site is 20 percent of the total freight traffic, it is most likely to be the same on the additional days that the survey is conducted (assuming the traffic conditions are similar).

Sampling on additional days certainly enriches the data and provides increased confidence in the proportion of freight traffic that is obtainable. Given the dynamic and variable nature of freight traffic conditions, the proportion of freight traffic that may be sampled varies throughout the day and across different data collection sites. Realizing these constraints, the stated goal of the Washington roadside surveys was to obtain as many completed survey questionnaires as physically and safely possible during the dedicated data collection periods. Once the samples were collected, a weighting scheme would be developed to expand the sample averages so as to represent characteristics of the total traffic population.

In almost all cases, available parking at each weigh station limited the number of vehicles that could be surveyed at a given time, so regardless of whether survey staff were available, only vehicles that arrived when parking was available were surveyed. At the lower volume survey sites, available parking was less of a problem, and between 60 and 80 percent of freight truck traffic was interviewed (at a few very-low-volume sites along the Canadian border, almost 100 percent of trucks were interviewed). At the very-high-volume survey sites (I-5 and I-90), a range of 5 to 20 percent of the freight truck traffic was captured in the survey sample. The variation in sample size is practically unavoidable given the fluctuations in traffic conditions and available parking areas, but represents the best compromise given the prevailing surveying conditions and the goal to obtain as much information as possible.

As was previously mentioned, the state of Washington, similar to many states, has a program where freight shippers and truckers may register their vehicle with the state electronic bypass program, Commercial Vehicle Information Systems and Networks (CVISN), and install vehicle transponders that allow shippers to register trip and shipment details online and bypass weigh stations that have CVISN capability. Currently, there are 12 of these weigh stations in Washington, but during the EWITS origin-destination survey, this technology was not available, and during the SFTA origin-destination survey, the technology had only just become available and only two weigh stations possessed CVISN capability. However, the data obtained from the CVISN program is available (via request to the CVISN data office), and this information for each survey site can be used in determining sample characteristics and the population of trucks that used the program during the survey time period. In addition, many of the questionnaire attributes are available from the CVISN data and can be incorporated into the data validation/verification process at relevant sites.

User's Guide Worksheet Punch List

- Identify a key variable collected through the roadside survey effort (e.g., payload, origin-destination pairs, commodity).
- Estimate the mean and variance of the key variable for truck trips through a survey location using FAF data or a pilot survey.
- Determine the desired confidence level for the sample size. Ninety-five percent can be used as a default value.
- Estimate the sample size needed to estimate the variable using Equations 1 through 3 provided in the implementation section of this step.
- Repeat this process for as many variables and locations that are of interest to the survey effort.

Step 5—Interview Team Recruiting and Training

Key Considerations

The structure of the interview team will impact both the quality of the data and the cost of collecting the data. Using an interview team with significant experience in conducting roadside truck origin/destination surveys tends to provide the best data quality. Additionally, the time and resources needed to train experienced data collectors should be minimal. However, the hourly rate for experienced data collectors is likely to be higher than the hourly rate for other potential data collectors. In contrast, using general temporary labor will decrease the hourly cost, but will increase the need for upfront training and potentially result in somewhat lower data quality. Regardless of the experience level of the team involved, training will be needed to ensure safety during data collection operations and effectiveness in asking truck drivers questions.

Implementation Process

The data collection field team can include transportation professionals, data collection specialists, or temporary employees. Typically, the most cost-effective strategy is to use temporary employees to collect the bulk of the survey data and have a transportation professional manage the operation at each survey location to provide quality control of the data collected. In addition to being cost-effective, using temporary employees to collect data also has the advantage of providing the most bandwidth of all of the options. With a small number of transportation professionals, it is possible to conduct dozens of surveys simultaneously or within very short periods of time. Site managers also can collect data, but may have their collection interrupted by other activities. In the case of conducting multiple surveys simultaneously, the site manager may be responsible for several locations and may need to travel between locations to monitor operations.

Temporary employees will need to be tested for the minimum levels of knowledge required to conduct a survey, including geography, basic math, reading, and writing. This testing can be provided by the temp agency from which employees are recruited. Using temporary employees will require the most upfront training of staff. It also requires a higher level of monitoring of the data collection process. Temporary employees will need instruction on the purpose of the survey effort, a detailed description of each question that is included in the questionnaire, and detailed safety training for each type of survey location. An option related to using temporary employees would be the use of volunteer data collectors. These can potentially come from local charitable organizations.

An alternative structure would be to staff the team entirely with transportation professionals. This option would incur a higher cost relative to using temporary employees, but it will save time. It also has the potential to streamline the entire process, because team members who collect the data in the field also can do data entry in the office.

Data collection specialists are firms that specialize in collecting data across a number of different topics such as market research and polling. Using one of these firms tends to cost less than using transportation professionals, but cost more than using temporary employees. While the employees of firms that specialize in data collection are trained in the data collection aspect of conducting a survey, they will still need detailed training for each question in the questionnaire and safety training on the sites. These employees will typically require much less monitoring than temporary employees.

Regardless of the structure of the team, each data collection location should have a site supervisor that manages the coordination needed with facility operators, monitors the quality of the collected data, ensures that safety guidance is being adhered to at each location, and collects completed survey forms or data from electronic devices. This site supervisor should be an experienced transportation professional that has participated in previous roadside truck survey efforts.

For surveys that are done at a handful of locations, it is typically most effective to use transportation professionals to save time and money on training and ensure data quality over the survey effort. Transportation professionals or data collection specialists also are recommended if the survey is asking complex questions of truck drivers or questions that will require a certain amount of probing or pivoting from one answer to the following question. This would include asking for detailed origin/destination data such as assigning each trip end to a prescribed set of zones. For surveys that occur at several dozen locations, temporary employees are likely to be the most effective. These employees can be trained once and then be used repeatedly over a short period of time. If temporary employees are used, then the questionnaire should be made relatively straightforward with a finite set of responses that are expected to be captured through the survey effort. If the transportation agency elects to hire contractors to perform the data collection services, then proposals can be accepted from firms offering alternate team structures, and the agency can select the team that provides the best overall services in terms of cost and data quality.

Staff training should include a description of the purpose of the survey, including what is driving the need to conduct surveys and what the data will ultimately be used for. Then, each question in the survey should be reviewed with the field staff. For temporary staff, this will need to be a detailed question-by-question review of the survey questionnaire. For each question, staff should be trained in how to phrase the question, how to record the answer, and what types of responses are likely. For staff not familiar with freight terminology, basic descriptions of key items (such as rail intermodal yard) will be needed. For electronic surveys, staff will also need to be trained in using the actual devices. As with paper surveys, the actual survey instrument should be available and reviewed by the staff being trained.

There also will need to be special alerts for responses to questions that will require follow-up or additional probing. This is the case most commonly with responses to origin and destination questions where either not enough specificity is given or landmarks rather than addresses or city/state combinations are provided. If special maps will be used in the survey, then these maps should be developed and presented during the training session. If there are predefined categories for any of the questions, then these need to be reviewed at the training session as well. As part of the training session, a dry run of a survey should be conducted so that all surveyors can get a sense of timing and protocol and have a chance to raise any other logistical questions.

The training session also should include training on site safety. This aspect of the training will include a typical site layout for each of the different types of survey facilities and will include a diagram showing how the site will operate, including where trucks are diverted from the traffic stream, how they approach the survey location, where the surveyors will be stationed, how trucks will pull out from the survey location, and the approximate time interval between surveys. The training will need to cover how surveyors should approach a vehicle and where they should position themselves as a truck approaches and pulls away. Surveyors should also be trained in how to handle truck drivers that decline to participate in the survey or decline to answer specific questions.

The final portion of training will address how to store and transfer the data for future analysis. This training will be needed whether the data are collected using paper and pen or electronic devices.

Example

The SFTA roadside surveys were conducted using state service clubs (Lions Clubs and Kiwanis Clubs). Recruitment of these teams began by contacting club presidents near the selected data collection sites so that survey teams would not have long travel times to and from survey sites.

- 9:00 a.m. – 9:15 a.m. Introductions
- 9:15 a.m. – 9:30 a.m. Overview of project and survey process
- 9:30 a.m. – 10:00 a.m. General site setup and safety requirements
- 10:00 a.m. – 10:15 a.m. Items to bring to survey locations
- 10:15 a.m. – 11:00 a.m. Question-by-question instructions including typical responses
- 11:00 a.m.– 11:30 a.m. Mock surveys
- 11:30 a.m. – 11:45 a.m. Wrap-up and Q&A

Figure 3.4. Hypothetical agenda for roadside truck survey training session.

Using service club members was ideal in several respects. First, they proved to be a relatively large and geographically diverse short-term labor force, which matched the needs of the survey. Use of these service club members allowed for 15 to 18 surveyors at each data collection site for the 24 hours of data collection, equating to 5 to 6 surveyors per 8-hour time period. Because the service club members lived near the survey locations, they also had extensive local knowledge of the roads and cities being accessed by the truck population. In lieu of individual salaries, contributions were made to the service clubs.

Once interview teams were identified, each team received classroom training, typically lasting about 2 to 3 hours, which included the following agenda (see Figure 3.4):

- Explanation of the overall project goals of the study
- Description of each question on the questionnaire, including typical and atypical answers to each
- Safety requirements
- Site setup/operation
- Items to bring to the survey location such as comfortable clothes/shoes, hats, sunscreen, rain gear, etc.
- A question/answer session.

Each survey member was provided with a hard copy of a safety manual that outlined key safety considerations such as the need to always wear safety/reflective vest. Other safety instructions provided in the safety manual were never approach a moving truck, wait for it to stop instead; do not enter or climb onto the truck; never allow traffic congestion at survey sites; and always be mindful of traffic flow and other moving vehicles.

Training of each team continued during the actual implementation of roadside surveys, as each survey site had a site manager that oversaw all operations and monitored completed surveys in order to quickly identify problems in conducting the surveys. As teams became more experienced, the role of the site manager became less important, except for survey members who may not have participated on earlier roadside surveys.

User's Guide Worksheet Punch List

- Determine the ideal mix of transportation professionals, data collection specialists, and temporary staff desired for each survey location.
- Develop a training session agenda for field data collectors.
- Develop all needed training materials for the training session. This should include a description of how to phrase each question, the types of responses that are likely to be heard, how to record the response, and typical responses that require follow-up by the surveyor.

Step 6—Site Preparation and Traffic Control

Key Considerations

Site preparation and traffic control needs vary for each survey site. Coordination with the facility operator will be needed to ensure adherence to site-specific safety guidelines. Additionally, a transportation professional should determine whether there is a need for a traffic control plan on site. Additional preparations will be needed if off-site equipment such as variable message signs is used.

Implementation Process

Weigh stations are considered to be the ideal roadside truck survey location because they are designed specifically for managing truck operations. This includes the ability to intercept any truck from the traffic stream and inspect it for long periods of time if needed. Weigh stations require the least amount of site preparation and in some instances will not require a formal traffic control plan. The survey manager should meet with the on-site operator of the weigh station to determine what the specific survey operations will be. This includes the process for ensuring that the trucks selected for the survey are random, deciding who will select the trucks, and settling on how communication will occur between the survey site manager and the weigh station operator. From this meeting, the survey site manager should, at a minimum, develop a sketch drawing detailing the location of the survey, the number of truck parking spaces available for the survey, and the path of the trucks as they enter and leave the survey location.

Locations that are not weigh stations will require the development of a traffic control plan. This plan should be developed with oversight of the facility operator and take into account the sample size requirements that were identified in Step 4. The plan should also be reviewed with the transportation agency that is responsible for the roadway where trucks are being intercepted to ensure that the plan does not create truck queues or speed deceleration events within the mainline traffic area. If the site is not capable of handling the number of trucks expected, then Step 4 may need to be revisited to ensure that sufficient data can be collected and to determine whether different survey dates and times are required.

At locations that are not weigh stations, it is likely that variable message signs will be needed to intercept trucks. The survey site manager will spend much of their time at these sites managing the selection of trucks from the traffic stream, including changing the information on the variable message sign as needed. These locations also will require standard signs to direct trucks to the survey location and assist them in navigating away from the survey site. The survey site location also should be isolated from vehicular activity not associated with the survey, so there is no comingling of traffic at the site. It is likely that a series of traffic cones, signs, tape, and other markings will be needed to ensure a safe data collection site.

The site preparation and traffic control plans may need to be altered during non-daylight hours depending on the availability and quality of lighting and any different traffic patterns that occur during this time period.

Example

The site-specific preparation for conducting roadside surveys at weigh stations and ports of entry generally followed physical layouts for weigh stations of two sizes (see Figure 3.5). The smaller sites on lower volume highways often only allowed enough space to park two to four vehicles, as is depicted in Figure 3.5. A large reflective sign indicating that there would

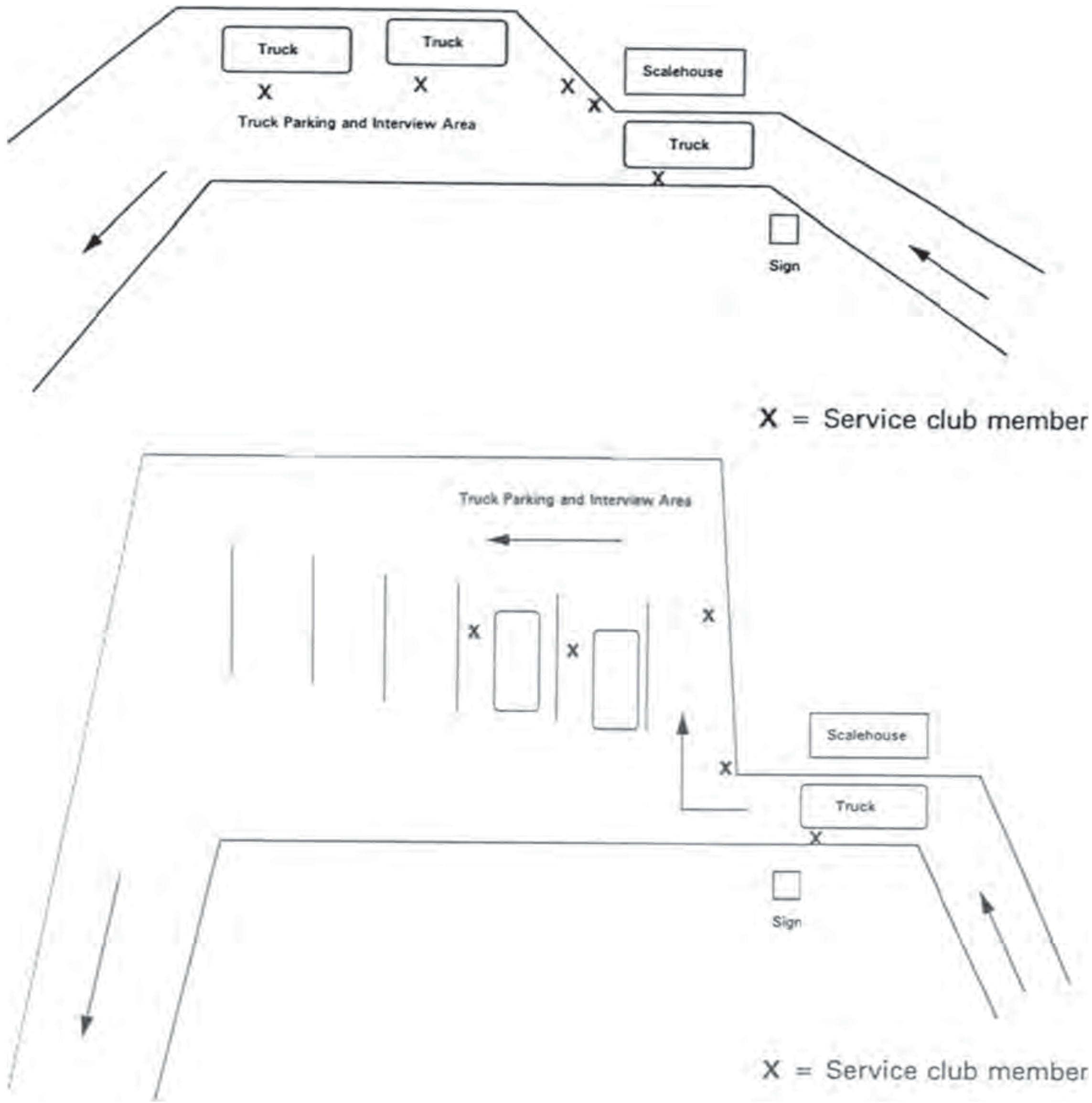


Figure 3.5. Typical traffic control plan for Washington roadside truck surveys.

be a survey team ahead was first placed at the entry of each weigh station to alert approaching drivers of the survey team's presence. Then, depending upon the individual officer(s) operating the weigh station and how busy they happened to be with their normal enforcement and inspection activities, one of two approaches was used to request that drivers participate in the survey.

One approach was to place a survey team member on the weigh scales or just beyond the scales; once the officer indicated the truck was "OK" to continue, the team member would approach the vehicle and ask if the driver would be interested in completing a short survey. If the driver agreed, the survey team member would then ask the driver to pull ahead and park, and one of the other survey team members would immediately conduct the survey. Once the survey was complete, the team member would thank the driver and the driver would continue on. If the driver indicated no interest in participating in the survey, then the team member would ask the next truck driver.

The person who is first interacting with the driver has the added responsibility of maintaining traffic flow, asking drivers if there is an available place to park, and finding an available staff person to conduct the survey. It is important not to interfere with the normal operations of state patrol personnel. The larger sites generally operated in the same manner, but could often have enough space to park 6 to 10 vehicles or more.

Another approach for indicating which vehicles to survey would be for the state patrol officer and one of the survey team members to work together inside the weigh station. The survey team member inside the weigh station would monitor the activities of the other team members as well as parking availability to determine when there was availability for another truck to be surveyed. When parking and a survey team member were available, the team member inside the weigh station requested that the state patrol officer stop a particular truck for a survey. The officer would turn the switch of the weigh station light to indicate that the designated truck should pull into the survey site location.

In some cases, the vehicle that was being asked to park was also being stopped for inspection or weight violations, in which case the officer would tell the team member inside the weigh station to communicate to the survey team member conducting the survey that once it was completed, the driver should bring their log book inside the station. Generally, the second approach worked much more efficiently and safely.

There were periods during survey sampling when the weigh station would temporarily shut down or stop accepting trucks for inspection. This would occur when there was only one state patrol officer at a specific site and it was necessary to pursue a truck driver who had avoided the scales. There also were periods during which the highway was closed due to accidents nearby or issues such as brush fires. In each case, the survey team would note the exact time of the closure, and loop detector data was used to determine the number of freight vehicles that passed during this time.

User's Guide Worksheet Punch List

- Contact facility operators at each location to determine whether there is a need for a traffic control plan.
- Develop a traffic control plan for one weigh station and one non-weigh station location, if applicable.
- Compare traffic control plans with needs identified in the sample size step (Step 4) and the survey dates and times step (Step 3).
- Submit these plans for review by facility operators and relevant transportation agencies.

Step 7—Using Uniformed Law Enforcement Officers and Their Vehicles

Key Considerations

Trucks traveling on the Interstate system are accustomed to stopping at enforcement locations. The use of enforcement officers at data collection sites is therefore useful from both a response rate perspective and a site safety perspective.

Implementation Process

It is recommended that uniformed law enforcement officers and/or their vehicles are used at every roadside survey site location. This will happen automatically if weigh stations are used as survey site locations because uniformed law enforcement officers are always stationed at these facilities. Truck drivers tend to feel more comfortable participating in surveys if they are aware that law enforcement is involved. Additionally, law enforcement officers are trained to develop and administer safe vehicle control plans for all types of sites.

At non-weigh-station locations, visible uniformed law enforcement officers are even more useful than they are at weigh stations. Most truck drivers are aware that mobile inspection stations may divert them out of the traffic stream on virtually any Interstate or state highway. This familiarity will extend to visible uniformed officers at non-weigh-station locations for the purpose of administering roadside truck surveys. The uniformed officers also should review the traffic control plans at all sites to confirm their agreement and where they will be stationed throughout the survey effort. There is typically an additional cost to using these officers that will need to be incorporated into final budgets. Another option is to have an unmanned patrol car stationed adjacent to the survey site to indicate that the survey has the approval of local law enforcement, even if an officer is not available to monitor the site during the survey effort.

Example

For roadside truck intercept surveys conducted at weigh stations in California, several unique site locations were used, including truck stops, rest areas, truck pull-out locations, and large parking lots located adjacent to Interstates and state highways. At each of these locations, the survey team found it extremely helpful to locate law enforcement vehicles just upstream of where truck drivers were requested to pull off of the road. Truck drivers are trained to anticipate truck inspections happening not just at weigh stations, but at other locations as well. The presence of the law enforcement vehicle put truck drivers on notice that they might be requested to exit the traffic stream. Using this method generated a high survey response rate for trucks at nonstandard locations throughout the state and enabled the identification of sites based on data needs rather than on weigh station locations.

User's Guide Worksheet Punch List

- For all survey locations that are not weigh stations, determine whether or not use of uniformed law enforcement officers will occur.
- Contact the relevant law enforcement agency and record the protocol in terms of the timing, cost, and role of uniformed enforcement officers.

Step 8—Quality Control Procedures

Key Considerations

Quality control procedures need to be considered as early as the questionnaire design phase and be carried through the analysis of data collected in the field. There are key intervention points when collected data should be reviewed to ensure that data quality is preserved throughout the process.

Implementation Process

Quality control procedures need to be interspersed throughout the survey effort. One of the initial steps in quality control is for the survey site manager to review the survey process for each surveyor while they are in the field. This will ensure that questions are being asked in a similar format for each truck driver and that there are no issues with the survey questionnaire. At the end of the first day, the survey site manager should meet with other senior staff on the survey team and review the mechanics of the survey process. Issues that tend to arise at this stage could include the following:

- Identification of questions that are complex to ask or answer in the questionnaire
- Adjustments in how one or more surveyors ask questions during the survey process
- Improvements in the ease of use of the data collection device, including paper surveys, electronic data collection software, or supporting materials such as maps
- Confirmation of the time needed to complete the survey
- Clarification of confusing questions that frequently are omitted or answered incorrectly by truck drivers

When the field data collection is complete, the raw data should be analyzed to identify missing data items and incorrect data items. Missing data items are typically the result of truck drivers electing not to answer a question, truck drivers not knowing an answer to a question, or a field data entry error. Due to the data quality check that should have happened early in the survey process, missing data items should be minimal at the end of the survey. Data items with unexpectedly high missing elements (perhaps more than 20 percent) indicate that there are questions that should be reconsidered for future roadside truck survey efforts. In some cases, missing survey items can be filled in following the survey. For example, if the city was recorded, but state information was not recorded, then it can usually be inferred and corrected in the database.

Incorrect data items can sometimes be identified through responses that are inconsistent with other known information about the survey. For example, a truck driver may state an origin/destination pair that is inconsistent with the driver's direction of travel. Alternatively, the truck driver may state a payload amount that is inconsistent with legal limits or inconsistent with its noted truck configuration or practical loading considerations. In other instances, incorrect data items can be the result of data entry errors such as recording the number of axles as being different than what is implied in the truck configuration. Some incorrect data items can be easily corrected such as backwards origin/destination pairs. Other incorrect data items should be completely deleted such as a payload weight that is not feasible. It is possible to develop programs that check the data automatically for accuracy. It is rare for an entire truck record to require deletion, but there may be multiple missing items for a single truck record.

The raw data collected through the survey process should be preserved for future reference. This includes the paper copy of survey responses and the entry of the raw data from paper into spreadsheet form. For electronic surveys, the raw electronic data should be preserved for future reference as well. Data edits should be done in a separate file.

Example

Efforts to increase data quality and accuracy began with survey team training and continued through the survey itself and on to the inputting of the data into computer databases. Having a site manager at each survey site significantly reduced the likelihood that someone conducting the surveys was doing so incorrectly or completing the survey form incorrectly or illegibly. Not every survey questionnaire was double checked by the site manager, but the site manager did double check many questionnaires and was especially diligent when new team members began or when there was a shift change. Once problems were addressed and corrected, oversight became less rigorous for each survey shift and team.

The completed questionnaires were then thoroughly reviewed prior to being input into a relational database. During this process, much of the missing or incorrect data was corrected. In most cases, miscommunication between the driver and the survey team member caused incorrect information to be recorded on the questionnaire. But by thoroughly reviewing each questionnaire, the research team could evaluate the data to ensure that the information was logically and rationally consistent and correct any problems. Some of the common errors on completed questionnaires included the following:

- Incorrect weight (payload and empty weight considerably above legal limit)
- Incorrect state or province associated with shipment origin/destination
- Missing commodity or some other missing attribute
- Number of axles didn't agree with truck configuration specified
- Hazardous material code incorrect (didn't match real code)

In almost all cases, corrections to the data were developed from reviewing other information on the survey and using deductive, logical reasoning or by reviewing other survey questionnaires that came from the same carrier, same origin-destination, or some other common attribute. Most of the cargo weight errors involved the driver providing the gross vehicle weight or maximum legal weight of the vehicle. It was fairly easy to deduce the cargo (payload) weight once the research team had the vehicle weight and the gross weight. If the vehicle weight was not known, but the exact truck combination was known, then the research team could verify the empty vehicle weight by using known empty weights (provided by state patrol officers) of each configuration. Most of the incorrect or missing information from the origin state could be ascertained from the highlighted route the drivers provided, and the specific type of hazardous material being shipped could often be obtained by the name of the shipper or the business.

User's Guide Worksheet Punch List

- Document the protocol for site survey managers to check field data collection for each survey location.
- Make sure to include data checking and editing in the survey schedule and budget.

Step 9—Preparing Data for Usage in Commodity Flow Database Development

Key Considerations

The key element in the data preparation process is data expansion. There are several data expansion methods that can be used, but hourly truck count data are typically the best alternative for roadside truck intercept surveys.

Implementation Process

This step takes the data sample collected through the survey effort and expands it to estimate the full set of truck activity at each location. Truck count data are used to expand the sample, because the truck count data can be used to determine the percentage of trucks that are captured at the location throughout the day.

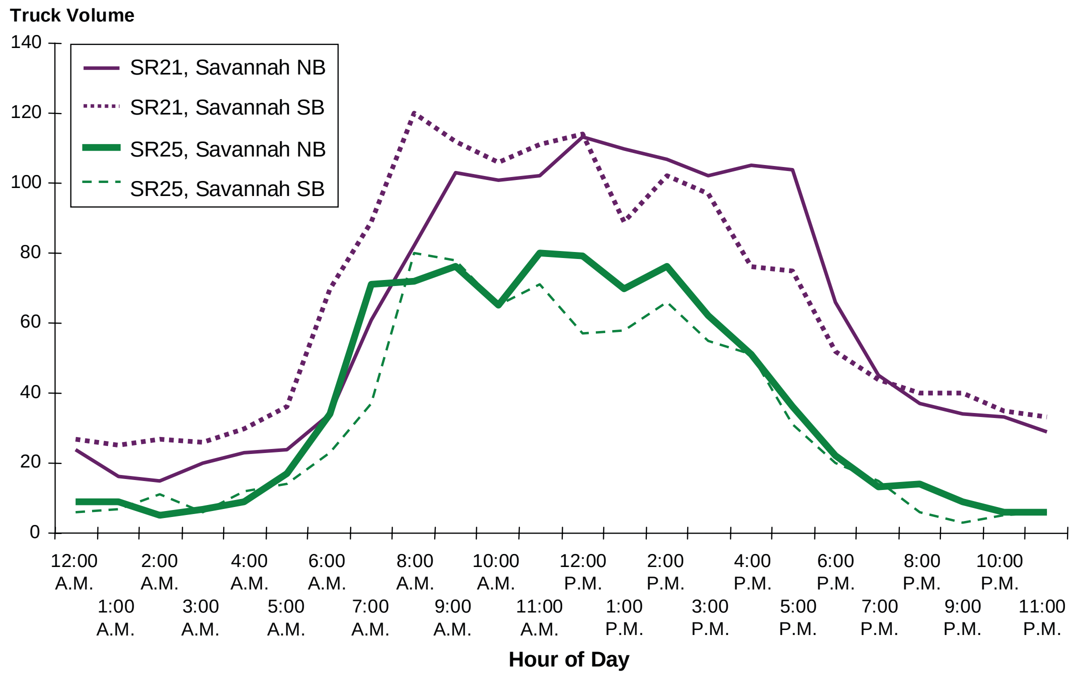


Figure 3.6. Hourly truck count data near the Port of Savannah.

The first step is to divide the day into time periods that are believed to be consistent with fluctuations in truck activity. An hourly truck count curve at a location is a good indicator of these fluctuations. Figure 3.6 shows an example of hourly truck counts near the Port of Savannah. Based on this curve, there appear to be four time periods of activity for trucks—a morning buildup in truck activity, a midday peak, an evening drop off in truck activity, and a nighttime low in truck activity. These fluctuations may vary at each site, but for simplicity it is reasonable to use the same time periods at each location. For each time period, the sampling factors are developed by calculating the percentage of vehicles that were captured by the survey, then multiplying that by each of the variables that are being expanded.

As a simple theoretical example, consider a scenario with two time periods, daytime and nighttime. Fifty percent of the survey data are collected in each time period, while 80 percent of the truck traffic occurs in the daytime and 20 percent of the truck traffic occurs at night. The variable to be expanded is average truck trip length through the survey location. In the daytime, the survey data indicate an average truck trip length of 100 miles, and at nighttime the survey data indicate an average truck trip length of 200 miles. To calculate the average truck trip length at the location during a typical 24-hour period, the calculation would be the following:

$$\text{Average truck trip length} = (80\% * 100) + (20\% * 200) = 120 \text{ miles}$$

Note that the average truck trip estimate does not depend on the number of samples collected during each period. Rather, it is dependent on the expansion factors that are developed using the periodic truck count data. The next section provides three detailed examples of how to expand surveyed data.

Examples of Data Expansion

There are three examples of data expansion that are presented in this section. First, is the expansion of collected data to a 24-hour time period. This example is presented using hypothetical truck count data and origin-destination data. The second example is how to expand data collected in different seasons to develop annual averages. This example is provided using an example from the EWITS survey using payload data. This example also shows how to develop systemwide averages of data collected at numerous sites. The third example describes uses of commodity flow data collected using a roadside intercept survey.

Hourly truck count data are typically used to expand data collected at a single site. For this hypothetical example, it is assumed that the hourly data are as shown in Table 3.1. Examination of the hourly count data indicate that four time periods of analysis make sense:

1. Morning period—6 a.m. to 9 a.m.
2. Midday peak—9 a.m. to 5 p.m.
3. Evening period—5 p.m. to 9 p.m.
4. Late night period—9 p.m. to 6 a.m.

If hourly truck count patterns are drastically different across locations, then it may be necessary to develop different truck count time periods for different survey locations.

Based on the time periods defined above, the origin-destination patterns collected in the survey can be aggregated into time periods for the data expansion analysis. An aggregation of these data is shown in the column labeled "Truck Count" in Table 3.2. The expansion factors shown in the third column of Table 3.2 represent the percentage of total daily truck trips that are in each time period. The next step is to calculate the percentage of through truck trips during each time period. This is shown in the fourth column of Table 3.2. In this example, through truck trips represent truck trips that travel through the state of Washington without conducting pickup- or delivery-type activities.

The percentage of truck trips can be multiplied by the expansion factor and summed to arrive at the expanded estimate for the percentage of through truck trips at the location. The percent-

Table 3.1. Hypothetical truck count data.

Hour Beginning	Truck Count
12:00 a.m.	250
1:00 a.m.	200
2:00 a.m.	200
3:00 a.m.	250
4:00 a.m.	250
5:00 a.m.	300
6:00 a.m.	400
7:00 a.m.	500
8:00 a.m.	600
9:00 a.m.	600
10:00 a.m.	700
11:00 a.m.	800
12:00 p.m.	850
1:00 p.m.	800
2:00 p.m.	850
3:00 p.m.	800
4:00 p.m.	850
5:00 p.m.	650
6:00 p.m.	550
7:00 p.m.	450
8:00 p.m.	400
9:00 p.m.	250
10:00 p.m.	250
11:00 p.m.	250
Total	**12,000**

Table 3.2. Data expansion calculation to estimate through truck percentage.

Time Period	Truck Count	Expansion Factor	Percent of Through Truck Trips Survey During Each Period	"Expanded" Percent of Through Truck Trips
Morning Period	1,500	13%	30%	4%
Midday Peak	6,250	52%	25%	13%
Evening Period	2,050	17%	45%	8%
Late Night Period	2,200	18%	80%	15%
Totals	**12,000**	**100%**		**40%**

age of through truck trips for each time period is shown in the fifth column of Table 3.2. The total expanded estimate of through truck trips is 40 percent. Any variable can be substituted for through truck trips and a similar data expansion process can be conducted. Due to the vast number of variables included in origin-destination surveys, this operation is typically carried out in a spreadsheet format.

For surveys that are conducted across multiple days, there is also the need to develop estimates of truck characteristics and activity across each of the time periods. In the EWITS survey, each location was surveyed four times—once for each season. Given that the sample was collected for each season, an origin-destination matrix for each season could be developed or one aggregated matrix could be developed in which each season's sample was aggregated into a total using a weighted average of each season's sample. The choice of matrix type largely depends upon the kinds of questions to be addressed and whether seasonal separation is necessary. Table 3.3 shows the calculation of average payload for the EWITS survey across different seasons and locations. A similar method could be used to develop an origin-destination matrix by expanding the data across each of the potential origin-destination pairs in the study area (including external regions).

Once the data from the Washington surveys was compiled into a relational database, it was used in a variety of applications. One use was to develop a matrix of shipments (or tonnage of freight) between all origin and destination pairs in the database. Table 3.3 illustrates how this can be achieved with roadside data.

The volume of tons shipped between any pair is calculated by taking all observations across all data collection sites with that origin and destination, applying the site-specific expansion factor that accounts for the proportion of freight traffic captured at each respective site, and multiplying by the average payload weight for that site to arrive at the average daily total across all sites. This type of matrix may be developed for each season or averaged (seasonal weighted average) across all seasons, depending on how the matrix will be used. This process is shown in Table 3.4.

In addition, for each origin-destination cell, it is possible to identify specifically which commodities comprised those shipments. In most cases, at the commodity-level detail there will not be very many observations per commodity type for each origin-destination pair. Ideally, the goal should be to have at least 30 observations per origin-destination cell in order to overcome small sample size issues, but often this is not possible because of limitations on the proportion of traffic that may be sampled. But even when the variance may be wide for cells where the freight shipments are based on two or three observations, it is informative to know that some shipments were captured with some information regarding commodity.

In addition to developing a freight origin-destination matrix of shipments, one may wish to identify the relationship in terms of freight traffic between any one point on the transportation network and all other areas, similar to stream flow analysis (up and down stream flows). Given

Table 3.3. EWITS seasonal expansion of sample data.

Survey Site	Season	Expansion Weight	Relevant Observations	Avg Payload (tons)	Total Expanded Number of Trucks	Average Daily Total Payload
A (Cle Elum)	Spring	6.5	235	8	1,528	12,220
	Summer	5.8	257	10	1,491	14,906
	Fall	5.9	246	14	1,451	20,320
	Winter	6.2	237	10	1,469	14,694
Site Total			975		5,939	62,140
B	Spring	1.2	25	18	30	540
	Summer	1.5	36	17	54	918
	Fall	1.3	18	13	23	304
	Winter	1.7	47	15	80	1,199
Site Total			126		187	2,961
C	Spring	3.5	64	12	224	2,688
	Summer	4.8	59	13	283	3,682
	Fall	6.1	48	14	293	4,099
	Winter	3.3	62	15	205	3,069
Site Total			233		1,005	13,538
D	Spring	2.70	123	16	332	5,314
	Summer	3.00	142	14	426	5,964
	Fall	2.90	151	17	438	7,444
	Winter	4.40	117	10	515	5,148
Site Total			533		1,711	23,870
Total All Sites (Straight Average)	Spring		447		528	5,190
	Summer		494		563	6,367
	Fall		463		551	8,042
	Winter		463		567	6,027
Total All Sites			1,867		2,209	25,626
Total All Sites (Weighted Average)	Spring		447		928	8,302
	Summer		494		936	9,976
	Fall		463		945	13,661
	Winter		463		918	9,355
Total All Sites			1,867		3,727	41,294

that each route in the database has been geocoded to represent all highway arcs comprising each observation, it becomes easy to provide this powerful analytical capability that develops a connection between all data attributes collected in the roadside questionnaire and the state geography or highway network. As an example, Figure 3.7 depicts the entire database of information (all routes) collected from the SFTA roadside survey in 2003/2004, across all sites. This graphic, when compared with the actual data from permanent traffic recorders, mirrors actual freight traffic intensity, indicating that in aggregate across the entire region the statewide coverage was adequate. In Figure 3.8, all observations that are common to Weigh-In-Motion site P21 (selected at random) may be depicted across the entire network, illustrating the geographical reach (in this case both origins and destinations) and intensity of flows that pass through this one particular

Table 3.4. Matrix of freight tons between origins and destinations.

			Internal									
	Origins / Destinations	Aberdeen, WA	Addy, WA	Adna, WA	Alderdale, WA	Algona, WA	Amanda Park, WA	Anacortes, WA	.	.	.	.
	Aberdeen, WA	1,359	1,288	1,325	1,069	1,302	678	1,047	971	895	1,185	551
Internal	Addy, WA	539	-	837	1,354	690	905	-	1,181	460	156	737
	Adna, WA	-	1,156	-	1,439	1,278	162	23	266	-	881	299
	Alderdale, WA	724	1,264	1,024	-	573	1,092	775	1,341	940	1,327	687
	Algona, WA	797	847	911	1,123	-	268	52	-	1,326	116	574
	Amanda Park, WA	700	313	1,493	1,212	253	-	1,411	984	491	1,380	1,477
	Anacortes, WA	448	119	-	1,199	720	856	1,183	1,174	174	84	404
	.	-	700	232	637	705	645	825	262	857	-	1,330
	.	441	606	1,388	862	248	549	1,400	500	1,051	-	604
	.	1,441	1,042	65	637	1,166	672	-	1,268	696	1,097	1,345
	.	290	275	1,087	-	323	1,085	1,159	445	265	670	1,307
External	Abbotsford, BC	977	241	-	1,210	667	719	515	1,303	1,205	651	957
	Abilene, TX	1,037	883	170	118	1,081	1,109	1,251	917	149	641	148
	Acton, Ontario	-	1,420	552	-	1,241	930	1,483	583	712	943	459
	.	1,141	1,433	751	1,029	349	1,459	73	1,143	-	152	1,188
	.	263	492	846	99	438	836	1,246	797	265	508	1,319
	.	-	979	945	672	930	752	1,465	1,036	1,023	930	34
	.	1,327	199	391	656	447	932	1,023	845	571	988	1,181

Origins / Destinations	Aberdeen, WA
Anacortes, WA	448

Survey Site	Expansion Weight	Observations	Avg Payload (tons)	Average Daily Total
A	20.34	4	8	651
B	6.3	5	10	315
C	5.8	2	14	162
D	4.2	13	10	546
E	18.2	5	11	1,001
F	1.6	1	8	13
				448

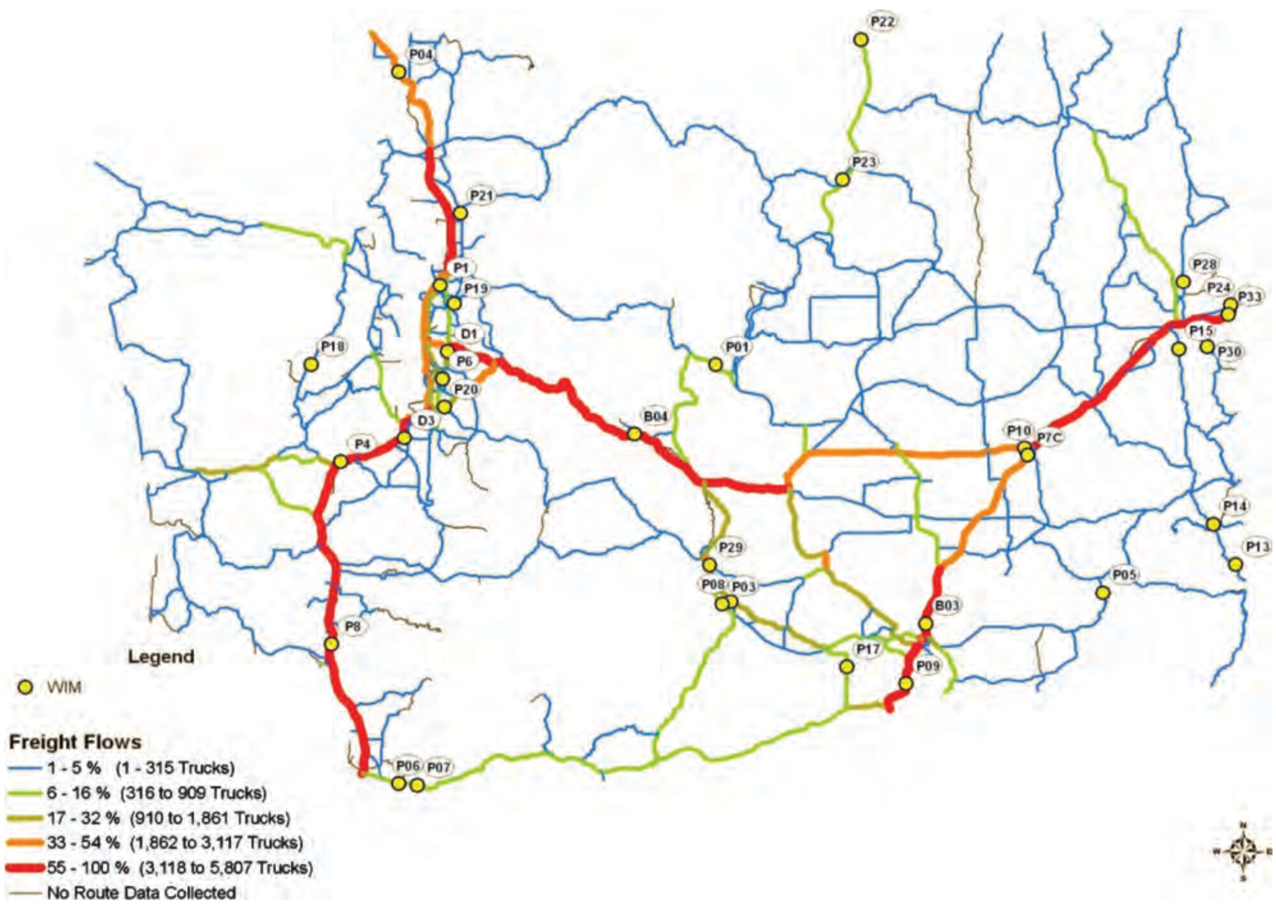


Figure 3.7. Intensity of freight routes collected from statewide origin-destination survey.

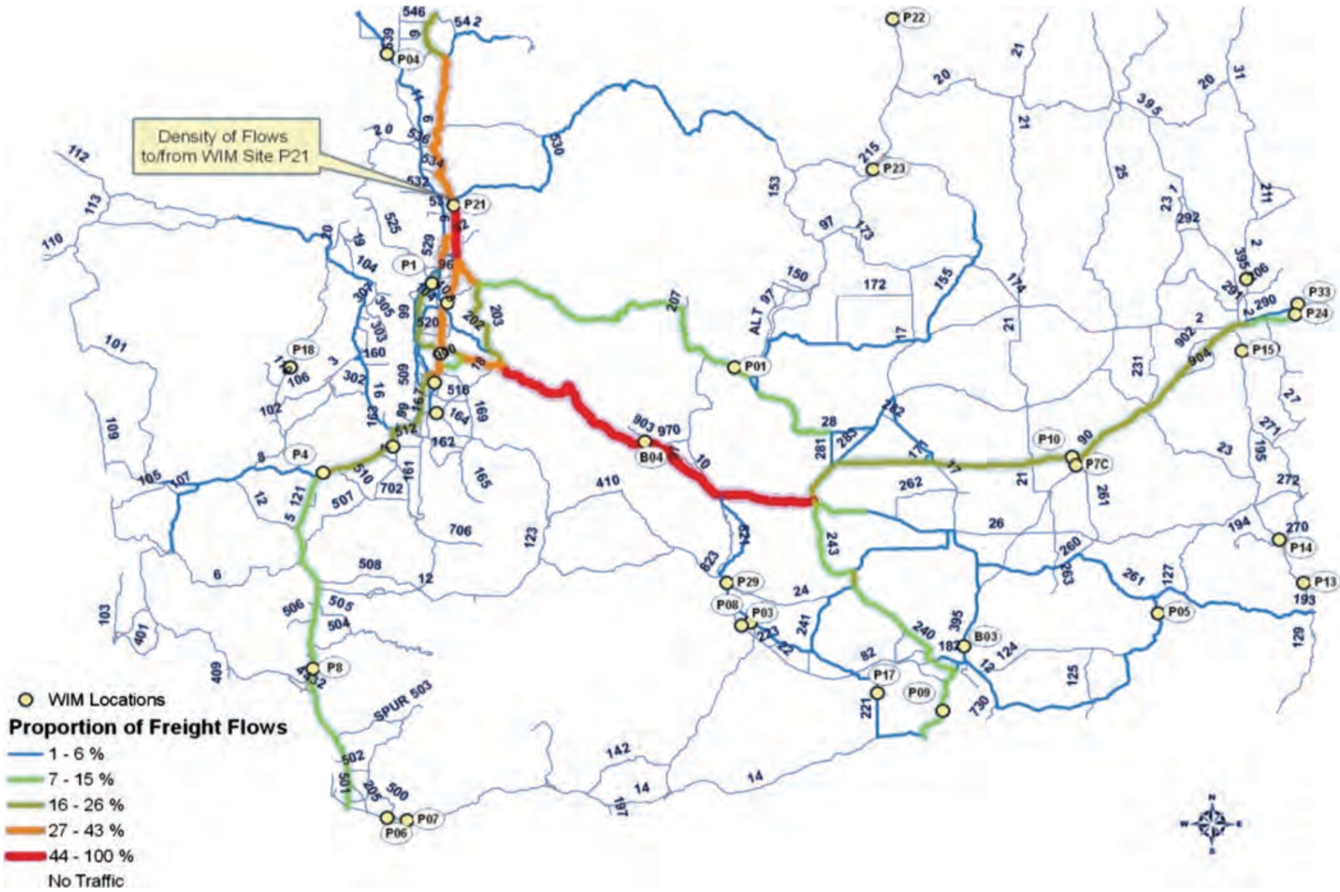


Figure 3.8. Relationship between freight traffic at one location on the network and all other areas.

area on the network. This same type of analysis also may be performed for any attribute, including (but certainly not limited to) the following:

- All freight flows common to a specific origin or destination
- All freight flows of a given commodity or value
- All freight flows of a specific vehicle type or hazardous material type
- All freight flows by a specific time of day

These analyses could even become more advanced geographical analyses that include combinations of many attributes (depict all statewide flows of a specific commodity type, using a straight truck configuration, which passed through King County). In addition, this type of analysis could be used to identify/estimate the economic linkages between cities or counties or regions. The possibility of different combinations is limitless, but further illustrates the value of route data that is perhaps best captured in roadside surveys.

User's Guide Worksheet Punch List

- Identify the most important variables to expand in the survey.
- Obtain hourly truck count data for the region, ideally as close to potential survey locations as possible. At a minimum, the hourly truck count data should be located on one of the same roadways as the roadside truck intercept survey.

- Develop expansion factors at each survey location based on the observed fluctuation in the hourly truck count data.
- Expand data to estimate truck activity for a 24-hour period at each survey location.
- Analyze and map the distribution and origin-destination pairs for key truck trip variables at individual locations and for the roadside survey program as a whole.

3.3 Next Steps

This chapter provides a detailed description of several elements involved in developing a roadside truck intercept survey. Nine basic steps have been identified that can be followed to fully implement a survey. Each step has been divided into the following four elements:

1. **Key Considerations.** Refer to this section to identify key concepts to keep in mind when initially considering conducting a roadside truck survey.
2. **Implementation Process.** Refer to this section to get information on the step-by-step process for conducting each step.
3. **Example.** Refer to this section to see how other agencies have conducted this step in separate roadside truck intercept survey programs.
4. **User's Guide Worksheet Punch List.** Refer to this section when you are ready to begin to work through several of the specific items that are needed to develop a roadside truck intercept survey.

Refer the Playbook section (Chapter 6.0) to identify the next portion of the *Guidebook* that will be most relevant to where your transportation agency is in the data collection process.

CHAPTER 4.0

Developing Subnational Commodity Flow Data Using Supplemental Sources of Local Economic Activity

4.1 Introduction

This section provides an examination of how to develop subnational commodity flow data from supplemental sources of data on local economic and goods movement activity. These data on economic activity can come from local governmental sources. These data can also come from state and local trade associations. Often these supplemental data need to be combined with other pieces of information to convert them into a standard commodity flow template, which includes origins, destinations, commodity, mode, and volumes.

The *Guidebook* identifies the following three steps that need to be addressed in administering a commodity flow data disaggregation technique:

- Step 1—Determine industries and commodities of special interest
- Step 2—Assemble data on local economic and goods movement activity
- Step 3—Estimate missing data

Some of these steps are interrelated, but the *Guidebook* discussion of each step is ordered as shown in the above bulleted list. The description of each step is structured to focus on the following four key elements:

1. **Key Considerations—**A brief description of the main issues encountered and tradeoffs that will need to be made for the step.
2. **Implementation Process—**A detailed description of how to implement the step.
3. **Example—**An example of how this step has been implemented in other studies. Note that this chapter includes brief examples of each of the steps and then provides two detailed examples of estimating local commodity flows for potatoes and diesel fuel at the end of the chapter.
4. **User's Guide Worksheet Punch List—**Simple bulleted instructions that *Guidebook* users can check off to ensure that they have implemented each of the major steps involved in conducting a commodity flow data disaggregation.

4.2 Step-by-Step Process for Collecting Locally Available Data

Step 1—Determine Industries and Commodities of Special Interest

Key Considerations

There are many potential reasons to focus supplemental data development efforts on specific industries or commodities. The industry or commodity may be one of special interest to the region of concern such as a major employer or a strategic industry from an economic development perspective. Alternatively, a transportation agency may have a preexisting commodity flow database and there is a desire to verify or adjust a specific commodity to ensure it matches with local activity. It is important to define these special industries or commodities as narrowly as possible to focus the supplemental data collection effort.

Implementation Process

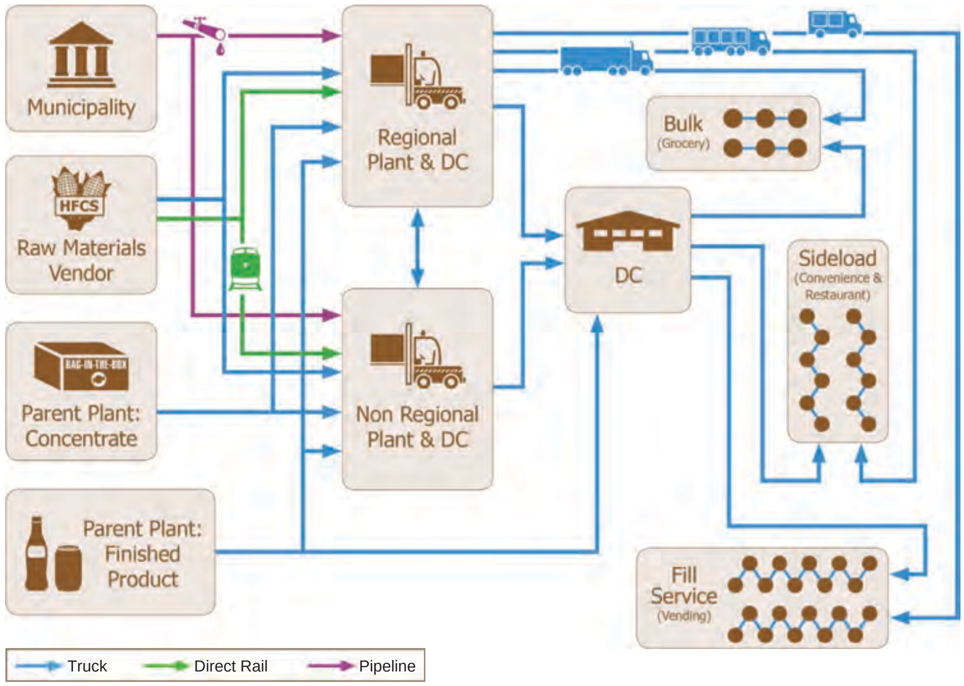
This process begins by identifying the industry or commodity of special interest. It is important to specify whether the interest is in an industry or a commodity. In situations where a transportation agency is interested in the impacts of a decision on its regional economy, it is typically an industry of interest. When the goal is to improve a preexisting commodity flow database, it is typically a specific commodity or group of commodities that are of special interest.

Then, it is important to understand the relevant supply chain for the industry or commodity. For industry-specific supply chains, it will be important to understand what commodities are produced by the industry and what commodities are used to supply the industry. Similarly, the upstream industries that produce the supplies and the downstream industries that consume the products will be important to understand. For commodity-specific supply chains, it will be important to understand the set of industries that produce the commodity along with the set of industries that consume the product.

Next, it is important to develop an understanding of the supply chain to determine the types of facilities that the goods typically move through. Are the commodities extracted from specific types of locations (such as coal or nonmetallic minerals)? Are the commodities stored in warehouses and distribution centers? Are the commodities purchased by the final consumer at retail locations or used as industrial inputs for a future processing activity?

The final portion in developing the supply chain is to understand the types of modes and vehicles that are used for each leg of the supply chain. It is common for several legs of the supply chain to be serviced by multiple types of modes and vehicles with specific companies tailoring their individual supply chain to their specific set of suppliers and customers. A graphic of a supply chain for soft drink beverages from *NCFRP Report 14: Guidebook for Understanding Urban Goods Movement* is shown in Figure 4.1 (Rhodes et al. 2012).

Many specific supply chains already have been mapped through previous research and can be found through Internet research on specific industries and trade associations. Economic input-output data are a more quantitative source of supply chain information. Input-output data describe the commodities consumed and services purchased by each industry along with the commodities produced by each industry. Input-output data also can be used to quantify the percentage of a commodity that is purchased by different industries. A comprehensive reference describing input-output data and models is *Input-Output Analysis: Foundations and Extensions,* most recently revised in 2009 (Miller and Blair 2009).



Source: Rhodes et al. 2012, Exhibit 3-1, p. 18.

Figure 4.1. Supply chain for soft drink beverages.

A publicly available source of input-output data is the U.S. Department of Commerce Bureau of Economic Analysis (BEA). These data are comprehensive in terms of their coverage of industries and commodities. However, the most recently available data from BEA are from 1997 and 2002. Therefore, the BEA data are more helpful in identifying the relationships between industries and commodities than they are in quantifying the current amounts produced and consumed. More recent sources of these data are proprietary, but can be obtained through economic modeling and analysis companies such as IMPLAN, Regional Economic Models, Inc. (REMI), PECAS, and TREDIS.

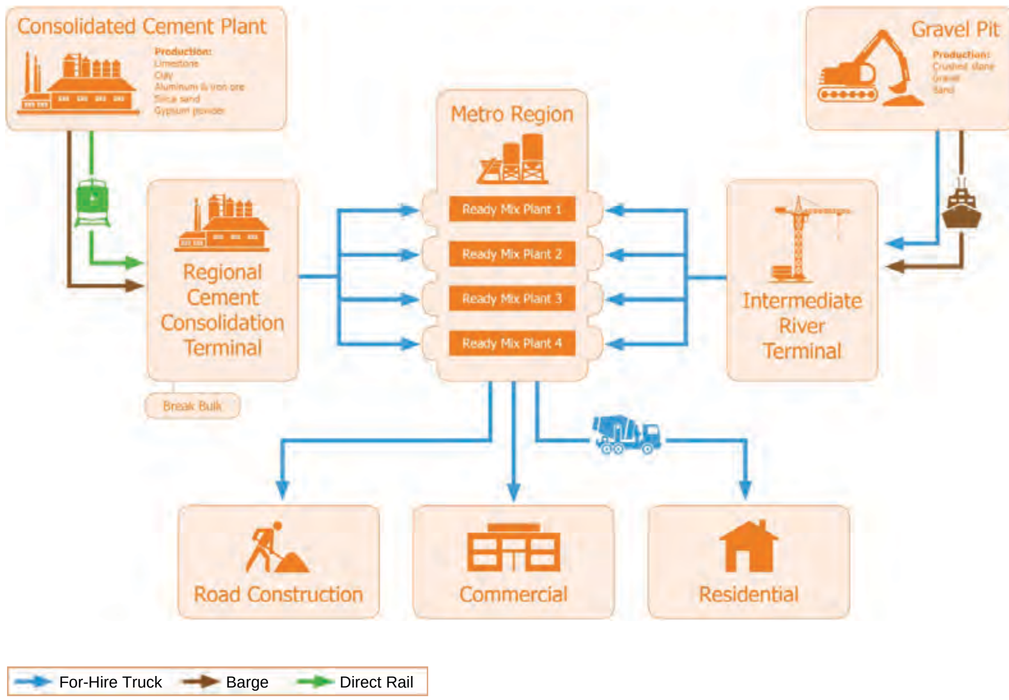
The BTS Transportation Satellite Account (TSA) data provide information on the amount of transportation purchased by industries. These data can be used as a starting point to understand the modes that are used to ship goods between suppliers and consumers of commodities.

Examples

Figures 4.1 through 4.3 provide supply chain examples for soft drink beverages, construction materials, and gasoline and petroleum fuels, respectively.

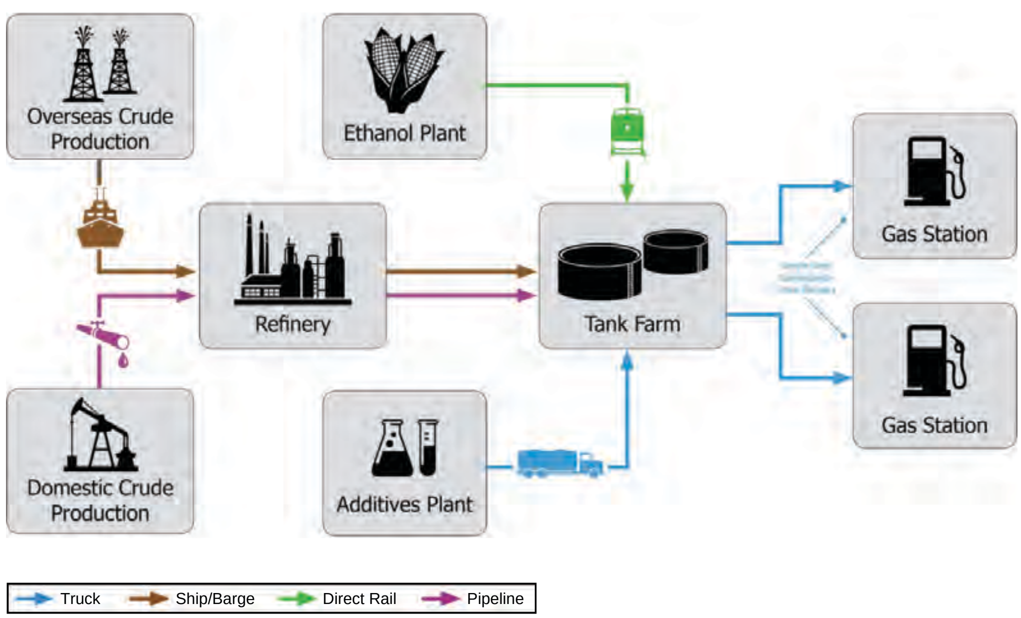
User's Guide Worksheet Punch List

- Determine industry or commodity of interest.
- Determine relevant upstream and downstream industries and commodities.
- Identify relevant freight facilities used.
- Determine modes and vehicles used to move goods between relevant freight facilities.
- Develop supply chain schematic for industry or commodity of interest.



Source: Rhodes et al. 2012, Exhibit 3-4, p. 22.

Figure 4.2. Supply chain for construction materials.



Source: Rhodes et al. 2012, Exhibit 3-2, p. 19.

Figure 4.3. Supply chain for gasoline and petroleum fuels.

Step 2—Assemble Data on Local Economic and Goods Movement Activity

Key Considerations

Sources of industry and commodity-specific information can range from regularly maintained databases to snapshots of freight activity from ad hoc data collection processes to anecdotal information from trusted industry experts. Each component of the supply chain schematic should be considered as a potential source of information for the industry or commodity of specific interest.

Implementation Process

The supply chain schematic developed in Step 1 can be used to identify potential supplemental sources of local economic and commodity flow data for the industry or commodity of interest. For each element in the supply chain, the following stakeholders should be considered for potential outreach to obtain supplemental data:

- **Government agencies** that regulate or monitor the industry or commodity. These agencies exist at the federal, state and local level, and they often maintain publicly available data on commodity movements on their web sites. Specific data requests should also be attempted with these agencies as they often make available more data to other public-sector entities. Typical types of data available include sales, production location, and permit data. Some data elements may be suppressed at more refined levels of geography.
- **Trade organizations** that represent private-sector producers in the relevant industry or commodity. Trade organizations also often track the amount of goods produced and consumed along with information on the importance of the industry for the broader economy. Trade organizations can include state mining associations, county farm cooperatives, or statewide industrial groups such as the Texas Petrochemical Association. Transportation agencies should be prepared to sign a nondisclosure agreement with trade organizations to obtain the most detailed data. Trade organizations also can be used to confirm the accuracy of the supply chain schematic and identify specific names and locations of shippers, receivers, and carriers used for the commodity. They may also provide names of specific industry experts that have access to quantitative and anecdotal information that can be useful for understanding specific industries and commodities.
- Local **research institutions** may be operated by academic, government, or private-sector organizations. They may maintain regularly updated data on the industry or commodity of interest. They may also have produced specialized reports that may be useful for understanding local commodity movements. Like trade organizations, they can validate the supply chain schematic, provide specific information about the location, size, and importance of key elements within the supply chain, and provide the names of individuals who may have useful information.
- There may be **major companies** that dominate an industry or commodity in a local region. If no specific contact has been previously identified for relevant major companies in your local region, then a logistics manager or government relations officer at the company should be contacted. By providing information on their own activities, they will provide insight on how the major local commodity flow movements occur. It will be important to ask them how big they are relative to the entire local market.
- There may also be **major carriers** that dominate an industry or commodity in a local region. These may include trucking firms, railroads, air cargo carriers, or shipping lines. For example, while there are several coal companies that have operations in West Virginia, only two railroads move the majority of the coal out of the state. The railroads could be contacted to obtain

information on the nature of these commodity flows. Similarly, FedEx and UPS dominate parcel delivery in most local areas. Therefore, they should be considered for potential outreach to understand how these types of goods are moved.

- **Facility operators** such as marine ports and airports also have information on commodity movements through their facilities. Typically, volume and commodity information can be made publicly available. Trade flows to other facility operators may be available at aggregated levels. Facility operators may also have collected specialized surveys providing them with information on ultimate origins and destinations at customer locations. These data may also be made available in some sort of aggregate form.
- Key individuals or **industry experts** also are useful for developing local commodity flows. They can verify the accuracy of the supply chain along with the currency of information obtained from other sources. In some instances, estimates provided by these experts may be the best data available for certain elements of the supply chain. Additionally, these individuals often have well-informed opinions about key trends that are shaping their industry in the near future.

From each of the sources mentioned above, the key types of commodity flow data to capture provide information on locations of origins, locations of destinations, types and amounts of commodities produced at each location, and the modes and vehicles used to transport the goods. Additionally, information on shipment sizes, vehicle volumes, and intermediate handling locations should be requested.

Typically, information on volumes produced is the most readily available. This information may be available at a substate level. Information on modes and vehicles used to transport goods may be available and is very likely to be available at least at an anecdotal level.

Table 4.1 provides a list of potential sources of supplemental local commodity flow data along with some specific examples of data sources from across the country. This table shows examples for select commodities, but similar organizations and associations can be found for the full range of commodities. The Examples section of this step provides snippets of some of these data sources. Often the data provided from these sources will provide a transportation agency with some, but not all, of the information needed to develop the specific commodity flow data of interest. In these cases, information from other sources may be used to fill in the missing data and transform the supplemental data into a full commodity flow database.

In some instances, the supplemental data needed may be heavily connected to a specific freight mode. This can occur because the industry or commodity of interest has a component of its supply chain that relies heavily on a single mode. Alternatively, this situation can arise if the transportation decision being considered has impacts that primarily affect a specific mode. Two of the key sources of mode-specific data that should be considered as potential supplemental sources on local goods movement activity are the Surface Transportation Board's (STB's) rail Carload Waybill Sample and U.S. Army Corps of Engineers Waterborne Commerce Statistics Center.

The rail Carload Waybill Sample provides detailed geographic data about loading and unloading points and interline locations along with information on tonnage and value. The data available far exceed the detail and accuracy of CFS and FAF data on rail activity. Rail Carload Waybill data include both public use files and confidential files. The public use files provide information at the BEA level of geography. The confidential files are available at the facility level such that rail volumes by commodity and configuration type (e.g., carload, intermodal, and bulk) are available for each rail line for the Class I railroads and for each major railyard in the United States. The confidential files are provided to a state agency upon request to the STB. Information must always be displayed in a manner that protects the confidentiality of the railroads and their shippers. Rail Carload Waybill data already are in a commodity flow format and therefore do not require the addition of missing data to be combined with other commodity flow data.

Table 4.1. Potential sources and specific examples of supplemental local data sources by commodity.

Commodity Type (Based on SCTG)	Potential Local Sources of Commodity Flow Data	Source of Examples of Supplemental Data
Cereal Grains Other Agricultural Products Animal Feed Other Foodstuffs	State agricultural department, U.S. Department of Agriculture, state fishery agency	Data from Montana Department of Agriculture, North Carolina Farm Bureau, Idaho Pork Producers Association
Natural Sands Gravel Nonmetallic Minerals Nonmetallic Mineral Products Base Metals	Local or state mining trade association, local or state base metal trade association	Data from Georgia Mining Association, Colorado Stone, Sand, and Gravel Association, Aluminum Association of Florida
Coal	Local or state mining trade association, U.S. Energy Information Association, State department of ecology	West Virginia Coal Association, California Municipal Utilities Association, University of Tennessee Department of Ecology
Crude Petroleum Gasoline Fuel Oils	U.S. Energy Information Association, state department of ecology, state petroleum producers association	U.S. EIA Petroleum Consumption Data for Washington, Louisiana Oil and Gas Association, Montana Petroleum Association
Basic Chemicals Fertilizers Chemical Products Plastics/Rubber	Chemical manufacturers association, state and local trade associations	American Chemistry Council, Ohio Rubber Group, Texas Fertilizer Association
Logs Wood Products	State trade associations, state forestry agency, wood products trade group	Southeastern Wood Producers Association, North Carolina Association of Professional Loggers, Pellet Fuels Institute
Waste/Scrap	State environmental agency, local recycling cooperative, state waste haulers association	New Jersey Department of Environmental Protection, Pennsylvania Independent Waste Haulers Association, Florida Recyclers Association

The Carload Waybill data are the source of the rail data that are used in the FHWA FAF database. Therefore, state DOTs and MPOs will not need to access the rail Carload Waybill Sample unless the more geographically refined data are useful. Obtaining the more detailed Waybill data is preferable to disaggregating FAF data due to the loss of accuracy that can occur in the disaggregation process. The Carload Waybill data do not include any forecasts, but these can be developed by associating rail activity with other economic activity for which forecasts already exist. State DOTs and other state agencies have the right to obtain the confidential files. Local agencies, such as councils of local governments and MPOs, will need to ask states to retrieve the Carload Waybill data. However, the STB may deny these pass-through requests and desire a direct request from local planning agencies, depending on the purpose and use of the data.

The U.S. Army Corps of Engineers Waterborne Commerce Statistics Center data provide tonnage, 20-foot-equivalent units (TEU), and value information by commodity for each port in the United States. However, there are no data available on trading partners or on inland origins or destinations that move through the port. All data are publicly available and can be obtained at the following link: http://www.ndc.iwr.usace.army.mil/wcsc/wcsc.htm.

Example

This section provides example snippets of data from the sources listed in Table 4.1. Table 4.2 shows sample data from the Montana Department of Agriculture. These data include specific crop produced, the production year, the county of production and the amount produced. For this snippet of data, it will first be necessary to convert the production unit into the desired unit for the commodity flow database. Typically, this unit will be tons. The conversion can be done using standard conversion charts for each commodity. Table 4.2 shows the county of production for the crop, which serves as the origin of the goods, but there is no information on destinations. To find information on destinations, it is useful to consult the supply chain schematic developed in Step 1 and/or to contact trade organizations that specialize in the crop being produced, major companies producing the crop, truck and rail companies that move the crop, local research institutions that study crop development, and consultants to the crop industry.

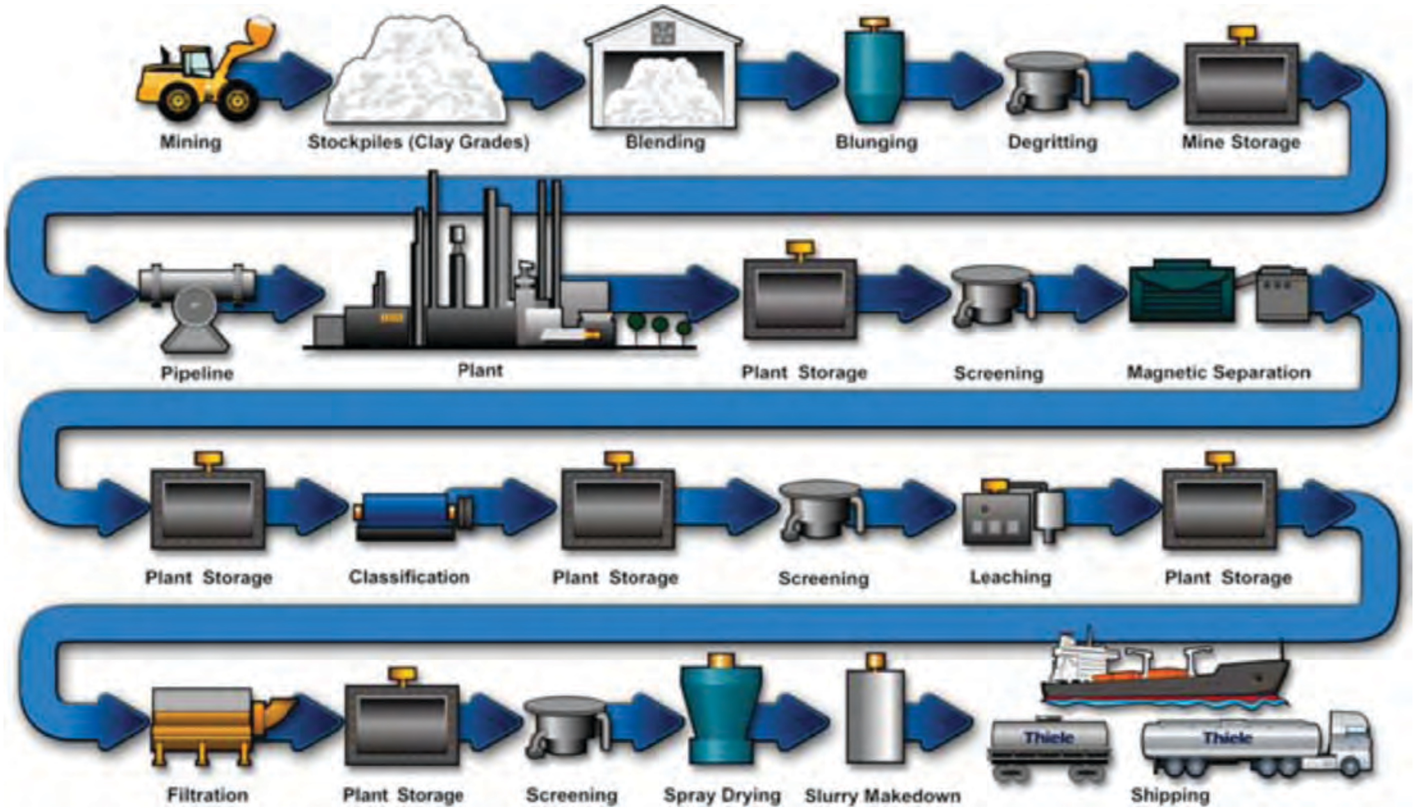
Figure 4.4 shows an example of a supply chain schematic that was provided through outreach to the Georgia Mining Association. The Georgia Mining Association is a trade organization representing private-sector mining interests that operate primarily in the center portion of the state and specialize in extracting kaolin. The schematic was developed by one of the association's member companies, but is used by the association to provide an overview of how kaolin is developed. The schematic can be used to verify portions of a broader supply chain that includes the locations of facilities across the state along with other key pieces of information.

Table 4.3 shows the production and consumption of energy across very specific categories for the state of Washington. These data can be used in conjunction with missing facility and origin-destination information to develop a commodity flow database across several fuel types. The example presented in the next section describes this process in greater detail.

Table 4.2. Sample data from the Montana Department of Agriculture.

Commodity	Year	State	County	Production	Prod. Unit
Barley; all	2001	MT	Flathead	502,000	Bushel
Barley; all	2001	MT	Lake	163,000	Bushel
Barley; all	2001	MT	Powell	82,000	Bushel
Barley; all	2001	MT	Ravalli	35,000	Bushel
Barley; all	2001	MT	Sanders	8,000	Bushel
Barley; all	2001	MT	Blaine	609,000	Bushel
Barley; all	2001	MT	Chouteau	515,000	Bushel
Barley; all	2001	MT	Glacier	2,137,000	Bushel
Barley; all	2001	MT	Hill	314,000	Bushel
Beans, All Dry Edible	2001	MT	Treasure	12,700	Hundredweight
Beans, All Dry Edible	2001	MT	Yellowstone	18,900	Hundredweight
Beans, All Dry Edible	2001	MT	Custer	20,000	Hundredweight
Beans, All Dry Edible	2001	MT	Prairie	45,200	Hundredweight
Beans, All Dry Edible	2001	MT	Rosebud	12,300	Hundredweight
Canola	2001	MT	Daniels	10,090,000	Pounds
Canola	2001	MT	Richland	2,827,000	Pounds
Canola	2001	MT	Roosevelt	3,347,000	Pounds
Canola	2001	MT	Sheridan	14,760,000	Pounds

Source: Montana Department of Agriculture.



Source: Thiele Kaolin Company.

Figure 4.4. Sample supply chain data acquired through the Georgia Mining Association.

Table 4.3. U.S. Energy Information Association consumption data for Washington.

Energy Data	Units	Period
By Type		
Total Energy	2,037 trillion Btu	2010
Total Petroleum	138.7 million barrels	2010
Motor Gasoline	64.1 million barrels	2010
Distillate Fuel	25.3 million barrels	2010
Liquefied Petroleum Gases	4.2 million barrels	2010
Jet Fuel	19.3 million barrels	2010
Natural Gas	285,865 million cu ft	2010
Coal	72.7 million short tons	2010
By End-Use Sector		
Residential	478,794 billion Btu	2010
Commercial	380,074 billion Btu	2010
Industrial	564,920 billion Btu	2010
Transportation	612,728 billion Btu	2010
For Electricity Generation		
Petroleum	2,000 barrels	Week of 7/1/2012
Natural Gas	3,108 million cu ft	Week of 7/1/2012
Coal	1,000 short tons	Week of 7/1/2012

User's Guide Worksheet Punch List

- Identify government agencies that regulate or track all industries and commodities identified in the supply chain schematic. Make sure to research federal, state, and local agencies.
- Identify trade organizations for all industries and commodities identified in the supply chain schematic.
- Identify academic institutions and research organizations that have conducted research on the industries and commodities identified in the supply chain schematic.
- Assemble publicly available information from the three types of sources listed above. Contact each type of source to determine if they have or are aware of other commodity flow information for your local region.
- Contact specific individuals at major shippers, receivers, and carriers for your industry or commodity of interest. Confirm information collected from the types of organizations listed above and determine whether there is additional information on local commodity flows.

Step 3—Estimate Missing Data

Key Considerations

The key elements in a commodity flow database are origin, destination, commodity, mode, and volume. Most supplemental data sources will be missing one or more of these data elements. Methods for estimating the missing data include making inferences from quantitative sources and anecdotal data and collecting new data. Creativity and resourcefulness are often needed to fill in these data gaps.

Implementation Process

There are several types of information that may be missing from supplemental data sets. In many instances, the volume of a specific commodity will be known at the desired geographic level, but the destinations of the commodity are not known. Alternatively, the origin and destination information may be known, but not at the desired level of geographic specificity.

A first step to filling in missing data is to reach out again to the freight stakeholders identified through the development of the supply chain schematic to determine whether there are data sources that have been overlooked and to determine, from people close to the data, a reasonable range for the values of the missing data.

Another option is to consult alternative data sources, such as FHWA FAF data, to determine proxies for the data that are missing. For example, in a case where the agricultural flows from a county to external locations are desired, this information can be approximated by consulting FAF data on flows to external locations at the regional or state level. Similarly, local employment data and land use data can be used to approximate data on specific origins and destinations of commodity flows that are missing from supplemental data sets.

The most resource-intensive option to consider is filling in missing data by collecting new local commodity flow data. This option can include conducting an establishment survey on a specific industry, as described in Chapter 2. It also can include collecting roadside truck origin-destination surveys at the gates to an intermodal marine terminal or a private-sector truck terminal. Roadside truck surveys are described in Chapter 3. The survey instrument used in these kinds of data collection efforts can be much shorter than a typical survey instrument due to the very specific type of information that will be requested; this allows an increase in the number of attempted surveys and, hopefully, leads to an increase in the survey response rate.

It may also be necessary to expand supplemental data to estimate missing data. For example, if supplemental commodity flow information were received regarding four of six paper mills in a local region, then data expansion could be used to approximate the commodity flows for the entire region by multiplying the supplemental data by 150 (6 divided by 4).

Finally, another method that might be needed is proportioning supplemental data that were developed using a geographic scheme that does not exactly match the commodity flow database desired. The simplest way to correct for this is to adjust the supplemental data based purely on the sizes of the geographic regions that are being considered. Alternatively, proxy variables, such as employment by a specific industry, can be used to proportion supplemental data to the desired geographic boundaries.

Example

Sections 4.3 and 4.4 provide detailed examples of how to generate local commodity flow databases for diesel fuel and potatoes, respectively. Examples of filling in missing data along with other aspects of using supplemental commodity flow data are detailed in these sections.

User's Guide Worksheet Punch List

- Identify missing data type(s) amongst the origin, destination, commodity, mode, and volume characteristics.
- Reach out to stakeholders identified in the supply chain to determine whether missing data or expert estimates can be identified.
- Identify broader commodity flow databases (such as FAF) that can be used to approximate missing data.
- Identify local activity variables (such as employment) that can be used to approximate missing data.
- If options for approximating missing data are not sufficient, collect new data to fill in missing data gaps.

4.3 Example of Developing a Diesel Fuel Local Commodity Flow Database

Step 1—Determine Industries and Commodities of Special Interest

In this example, the commodity of interest is diesel fuel. The researchers were interested in the commodity flow within Washington state at a substate level of geography. The commodity flow information desired included the types of commodities, the amount being shipped by each mode, and the distances being shipped by each mode. Specifically, the interest was in the number of truck trips on a typical day between a specific set of origins and destinations within the state of Washington.

In this step, qualitative information was gathered regarding diesel distribution in Washington state, including the following reports:

- WSDOT's *Washington Transportation Plan Update Freight Movement* (September 2008). This report contains an overview of the delivery and supply system for petroleum-based fuel in Washington state. The report summarizes the flow of refined products from the five active refineries in the state to end users at fueling stations, Seattle-Tacoma International Airport (Sea-Tac), the maritime industry, and private homes for home heating.
- *Review of Pipeline Utility Corridor Capacity and Distribution for Petroleum Fuels, Natural Gas and Biofuels in Southwest Washington* (ICF International 2007). In addition to addressing pricing and supply and demand patterns, the report describes the infrastructure of the diesel distribution system within Washington state. The report identifies the Portland (OR)/Vancouver (WA) area as being the hub of the distribution network in the Northwest, as it receives and distributes product via pipeline, marine vessel, and tanker truck.
- *Washington State Freight and Goods Transportation System (FGTS) 2007 Update* (February 2008). This document provides an understanding of the roadway network within the state as it applies to the movement of freight.

Fuel distributors and industry representatives were also interviewed to obtain preliminary information and begin to structure a general supply chain for the industry. It was determined that the major oil companies (BP, Shell, ConocoPhillips, U.S. Oil, and Chevron—defined by ownership of their own crude oil reserves) refine crude oil into diesel and transport their product by way of pipeline, marine vessel, or tanker truck. (Tesoro produces diesel as well but was not considered a major oil company because it purchases crude oil from the major oil companies rather than sourcing it itself. Chevron does not operate a refinery in Washington state, but it imports diesel into the state via the Chevron Pipeline from Utah.) The pipelines are each owned and operated by one of the major oil companies (Olympic Pipeline—BP; Yellowstone Pipeline—ConocoPhillips; and Chevron Pipeline—Chevron). Independent operators in the marine industry contract with the major oil companies to move diesel by barge or tanker ship, and the tanker trucks are operated by independent companies known as marketers.

Diesel that is transported by a pipeline or marine vessel is offloaded at one of the 27 terminal locations in the state. The terminals owned and operated by a major oil company are known as proprietary terminals, and those owned and operated by an independent terminal operator are known as common terminals. For instance, BP and Shell have terminals in Seattle on Harbor Island, and ConocoPhillips has terminals in Renton and Tacoma. Common terminals are owned by NuStar in Tacoma and by Kinder Morgan on Harbor Island. From the terminals, a marketer purchases diesel and makes delivery to a cardlock location, other fueling stations, directly to fleets of vehicles, or as specified by its customers. In addition, the major oil companies can contract directly with the marketers to transport diesel directly from a refinery. The diesel distribution supply chain has two additional features of note:

1. Marketers (these are the companies that move diesel from refineries or terminals to retail sales locations) purchase the fuel, rather than simply carry the fuel.
2. Given the volume of diesel consumed and its liquid form, it is distributed by marine vessel or pipeline whenever possible. Truck transportation is used only for last-mile distribution.

Following these initial conversations, a model of the diesel supply chain actors was developed. This model is shown in Figure 4.5.

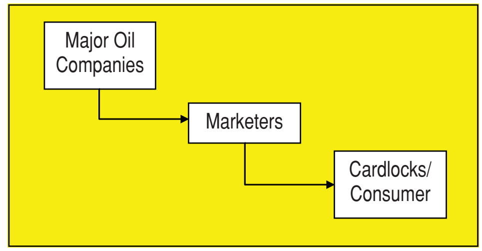


Figure 4.5. General supply chain for diesel fuel.

Step 2—Assemble Data on Local Economic and Goods Movement Activity

The research team identified the following sources of information:

- Commercial Fueling Network (CFN) and Pacific Pride. The company web sites for these companies list the locations of cardlock facilities. These are the primary distribution locations for diesel trucks (www.cfnnet.com/ and www.pacificpride.com/).
- The Washington State Department of Ecology (ECY) regulates active underground storage tanks (USTs). The agency provided a database containing 2005 data for tank volume, fuel type, geographic coordinates, and physical addresses for 10,869 USTs, of which 2,378 were classified as holding diesel fuel. Although current information for each tank could be queried by a variety of selection options at the following ECY Web portal (https://fortress.wa.gov/ecy/tcpweb reporting/reports.aspx), the information was provided only in summary. The database acquired directly from ECY personnel was in a Microsoft Excel spreadsheet containing all locations in list format. The ECY personnel verified that the data provided in the 2005 database were substantively comparable to the current year database and that the current year database was not available in a format that could be as easily extracted as the 2005 database.
- The U.S. Environmental Protection Agency (EPA) regulates above-ground storage tanks. A database containing all current year above-ground storage tanks was acquired upon request. Although there was less certainty as to the specific formulation of the database in comparison to the USTs database, the above-ground storage tank database contained tank volumes, fuel types, and physical addresses for 67 identified tanks containing diesel.
- The Washington State Department of Revenue provided a list of active terminal locations in Washington state upon request (where fuel is distributed to trucks from refineries, barges, or pipelines). There were 27 terminal locations, including the 5 refineries.

The 27 terminal racks in the state define the origins of interest. These are shown as green squares in Figure 4.6. The locations of USTs and above-ground storage tanks (ASTs) were compared to the complete list of cardlock locations provided by the CFN and the Pacific Pride diesel fueling network. Actual cardlock locations were obtained from each company's web site and matched with the locations of USTs and ASTs. In total, 376 of 433 (86.8 percent) cardlock locations were matched with actual diesel tank locations. This data set provided a reliable group of known diesel distribution destination locations with tank volume data. These cardlock locations (racks) are the destinations of interest and are shown as pink circles in Figure 4.6.

Data regarding fixed infrastructure and capacity, while somewhat time consuming to track down, were readily available in comparison to data on movements. Data on movements were inherently more difficult to obtain, as they vary quickly over time and space.

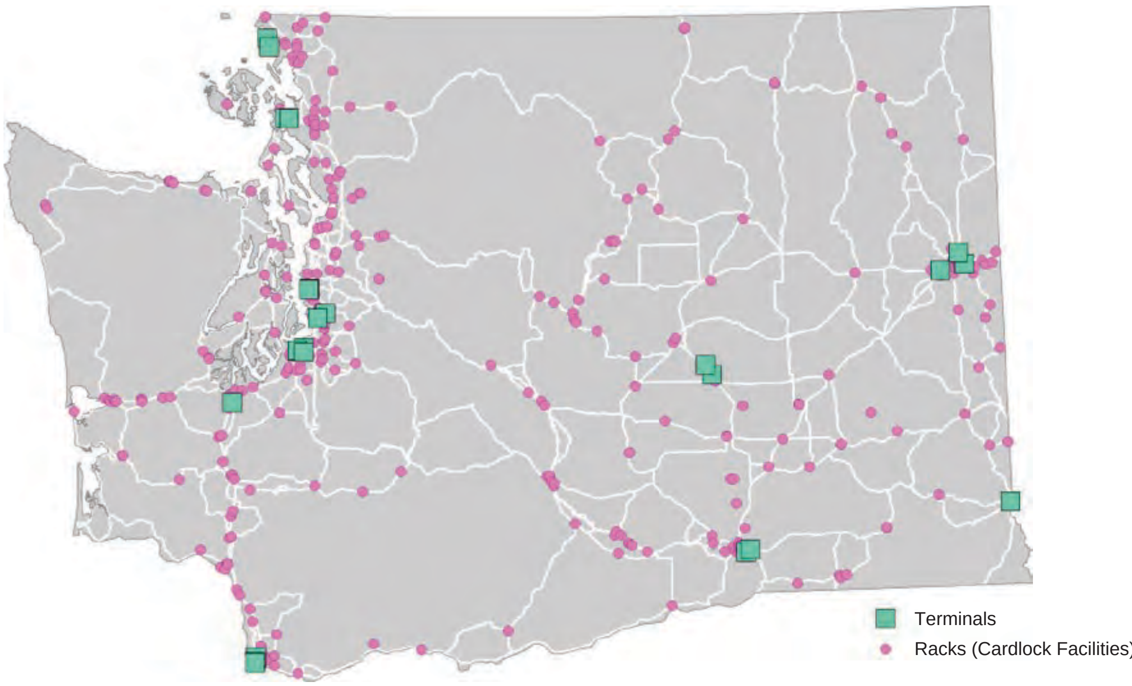


Figure 4.6. Location of diesel terminals and racks (cardlock facilities) in Washington.

The SFTA at Washington State University conducted a roadside origin-destination survey of freight trucks in 2003 and 2004. The database is categorized by UN placard number (1202 and 1203 for petroleum products), payload weight, and origin-destination (city, state) for the surveyed truck trips. Although this database contains empirical trip counts with pertinent associated data, the trip counts for petroleum products are quite low, the type of petroleum product being moved is undefined, and expanding the trips to incorporate the entire state or a section of the state was determined to be insufficient due to uncertainties in how the survey was conducted.

The FAF is an origin-destination database published by FHWA based on data collected during the 2002 Commodity Flow Survey. *Report No. S6—Petroleum Products National Totals* (n.d.) (http://ops.fhwa.dot.gov/freight/freight_analysis/faf/faf2_tech_document.htm) provides national aggregated movement totals for petroleum products by pipeline, water, highway, and rail. The information provided in the FAF was not found to be useful because the report does not have specific data at the substate level for the state of Washington. Additionally, there are discrepancies between the total fuel-related commodities estimated through FAF relative to other, more reliable data sources.

Both the Washington State Utilities and Transportation Commission and the Washington State Office of Financial Management were contacted, but they did not possess any pertinent information.

The Washington Oil Marketers Association provided a significant amount of information describing the Washington state diesel distribution network. It made note of the relationship between the marketers and the major oil companies and verified the general diesel supply chain network. It also provided names and contacts for diesel distribution marketers.

Eight marketers were contacted with questions about the distribution of diesel by their companies. Three companies responded and provided valuable information that connected many of the missing pieces of the Washington state diesel distribution network not previously uncovered by the research.

The FMCSA was contacted to obtain information about the reporting requirements of the diesel distributors and marketers. According to federal regulations, a carrier is required to have a HAZMAT placard issued by the FMCSA and a Certificate of Registration issued by the PHMSA to conduct diesel delivery operations (these last for several years). The carrier is only required to report an individual movement in the case of a hazardous material spill.

The information collected from these disparate sources resulted in a large amount of information on origins, destinations, commodities, and modes. However, volume information was still missing. The third step describes how this remaining information was estimated.

Step 3—Estimate Missing Data

No commodity flow data were retrieved from the sources listed above to estimate the diesel flow volume between terminals and cardlock facilities (or racks). The research team used two sources to fill in these missing data. First, the team identified the closest terminal (shortest travel time on the network) to each cardlock facility. It was assumed that the closest terminal was the one that was used as the origin of diesel fuel destined for each cardlock facility. This assumption is reasonable particularly if it can be assumed that the price for diesel was relatively equivalent at each terminal. With origins and destinations matched, the team defined the preferred service area for each terminal. These terminal service areas are shown in Figure 4.7.

Information was also missing on the number and routing of truck trips between terminals and cardlock facilities. The number of truck trips could not be estimated because the volume information was missing. The routing information, while not included in commodity flow databases, was also not available from any existing sources. To fill in this missing information, the research team used nearby vehicle count data to estimate diesel flows for the state. For this commodity flow application, it was determined that the number of origin-destination pairs using each link in the road network would be a sufficient proxy for volumes. However, average annual daily traffic (AADT) could be used to estimate the amount of diesel consumed at each cardlock. AADT at the nearest count location possible on roadways proximal to the cardlock facility was used to

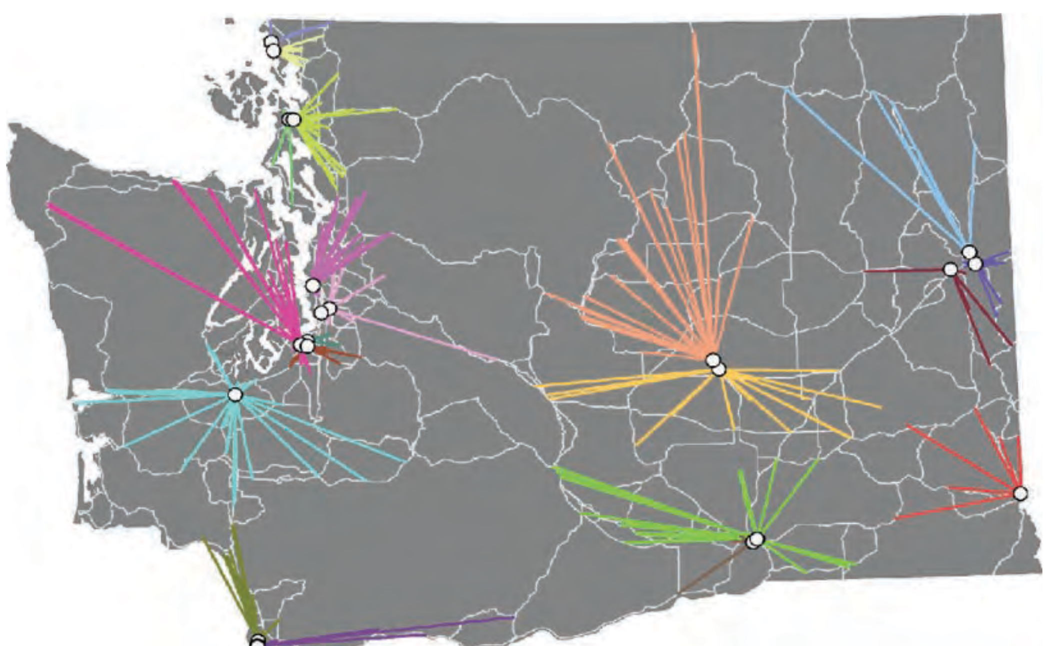


Figure 4.7. Terminal service areas in Washington.

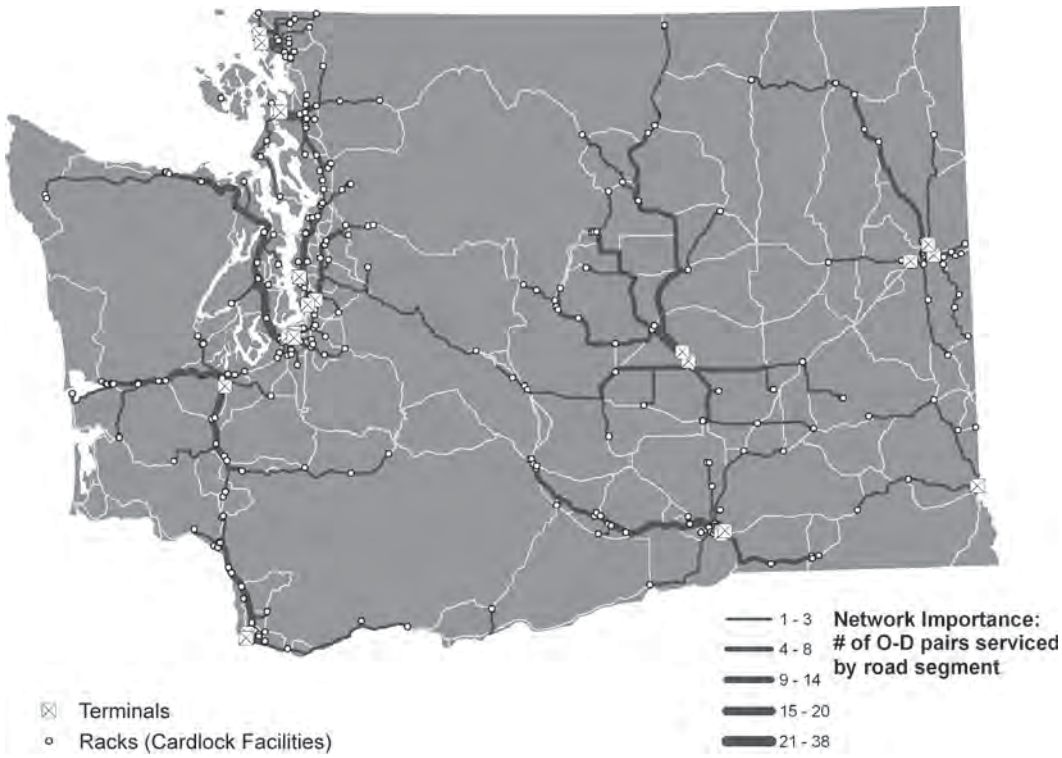


Figure 4.8. Diesel network flow map.

distribute known volume of diesel produced by terminals. Each cardlock received a portion of the total diesel dispensed equal to its AADT over the sum or all AADT. This assumes that cardlocks on roadways with more traffic distribute more fuel. AADT is available for all state highways in Washington through WSDOT.

The final result of this analysis is the estimation of truck flows between terminals and cardlock facilities. Figure 4.8 shows the number of origin-destination pairs that use each link in the network, where origins are terminals and destinations are cardlock facilities. Some links service almost 40 origin-destination pairs.

4.4 Detailed Example of Developing a Potato Commodity Flow Database

Step 1—Determine Industries and Commodities of Special Interest

In this example, the commodity of interest is potatoes and potato products (frozen potatoes, dehydrated potatoes, and potato chips). An estimate was desired for the commodity flow within Washington state and between specific origins and destinations within the state that reflect the locations of production, processing, export and consumption of potatoes and potato products.

Commodity flow information includes the types of commodities, the amount being shipped by each mode, and the distances being shipped by each mode. The goal of this commodity flow database development was the number of truck trips, for each potato product, on a typical day, between a specific set of substate origins and destinations within the state of Washington.

To develop a sense of the supply chain for potatoes, a meeting was arranged between the research team and the Washington State Potato Commission. The Washington State Potato Commission defines itself on its web site as a trade organization that has as one of its primary goals to "enhance trade opportunities, to advance environmentally sound production and cultural practices through research, and to represent the growers' interests in areas and issues relating to public and industry education, trade barriers, irrigation, transportation and crop protection" (http://www.potatoes.com/our-commission/mission/). As such, the commission staff have good knowledge of who produces and consumes the commodity. In addition, the staff compile data from a number of data sources to understand industry trends and conditions. Given this extensive background, the Washington State Potato Commission was considered to be a prime source of data.

Following the initial conversation with the Washington State Potato Commission, a schematic of the potato supply chain was developed. This is shown in Figure 4.9.

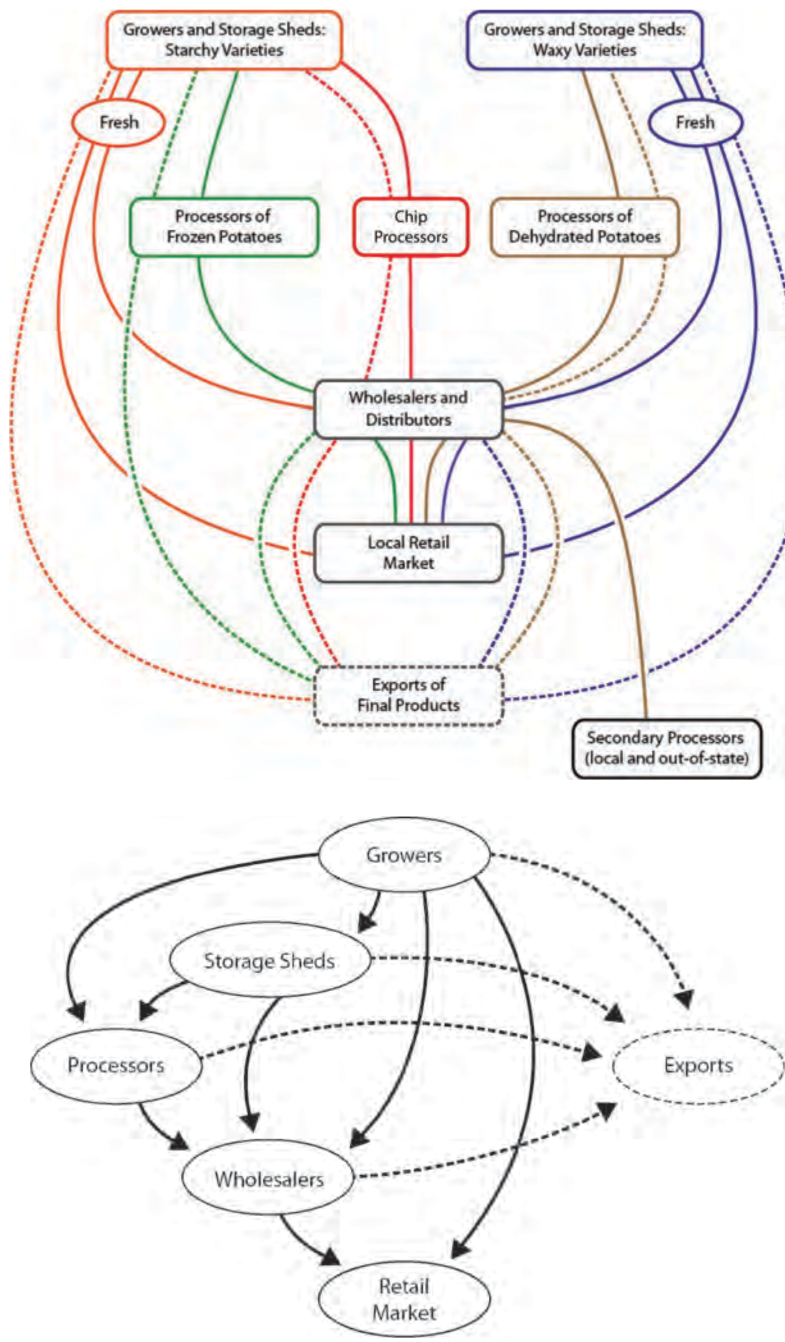


Figure 4.9. Supply chain schematic for potatoes.

Step 2—Assemble Data on Local Economic and Goods Movement Activity

Several available data sources were identified by Washington State Potato Commission staff and members of the research team (prior knowledge). These include the following:

- U.S. Department of Agriculture (USDA) field production data. These data are publicly available via the USDA's Economic Research Service. U.S. and state data on potato acreage, production, value, and use are provided. The USDA's Economic Research Service also includes U.S. data on prices, price spreads, and consumer and producer price indexes; stocks of fresh and processed potatoes; and trade in potatoes and potato products. At the national and state level, no data suppression issues are present. The research team believes this data to be of high quality.
- U.S. Census Bureau Foreign Trade Statistics, Total U.S. Exports (Origin of Movement) via Washington state. These data are available from the U.S. Census Bureau at the state level for the top 25 commodities and are also available by country of export. These data are readily accessible via the web and thought to be of high quality, with no suppression issues.
- SFTA Statewide Origin and Destination Truck Surveys completed by Washington State University in 2003 and 2007. SFTA data can be requested through the SFTA web site (http://www.sfta.wsu.edu/data_inquiries/data.htm). The research team uses statistical methods to estimate total truck flows from the sample. As the actual flows are not known, the statistical quality of the data cannot be judged.
- Washington State Department of Agriculture surveys conducted between 1999 and 2007. Potato acreage and production information for the state can be obtained from the Washington State Department of Agriculture. The township-range-section (TRS) level acreage data used in this study are a compilation of the results from a continuum of surveys conducted by the Department of Agriculture between 1999 and 2007. The potato production figures for each county were allocated to the TRS level by using yield and acreage information calculated for each county (this work was completed by the Washington State Department of Agriculture). These data are available by request and are thought to be of high quality. There were no suppression issues with this set of data.
- 2007 Washington State Potato Commission Survey. The Washington State Potato Commission conducted a survey of its members in 2007. This information is available upon request. It included destination and route. This information is thought to be of high quality. Requests can be made to the Washington State Potato Commission, who reserves the right to share or not share the information and to determine in what form information is provided.
- Washington State Potato Commission member data. The commission asks its members to submit monthly information on shipments of potatoes by variety, product, and destination. This information is thought to be of high quality. Requests for this information can be made to the Washington State Potato Commission who reserves the right to share or not share the information and to determine in what form information is provided.
- Industry experts and representatives. Although not explicitly based on data provided to the research team, more generalized industry knowledge was used to estimate some parameters.
- United States Census Bureau Population and Housing Unit Estimates. The Census provides city and town population estimates. This information is available on the U.S. Census Bureau web site and is thought to be of high quality. For the size of city considered in this example, there are no suppression issues.

As mentioned, the research team's interest was in estimating daily truck trips on the state's road network. The Washington State Department of Agriculture had estimated Potato Production by Township and Range for 2006 as shown in Figure 4.10. These data indicate that potatoes are produced in three regions of the state: the Skagit Valley, the Lower Basin, and the Upper Basin. These three regions are considered origins of fresh potatoes. Centroids of the region are identified as the origins of truck trips. Potato production volumes for each region are estimated

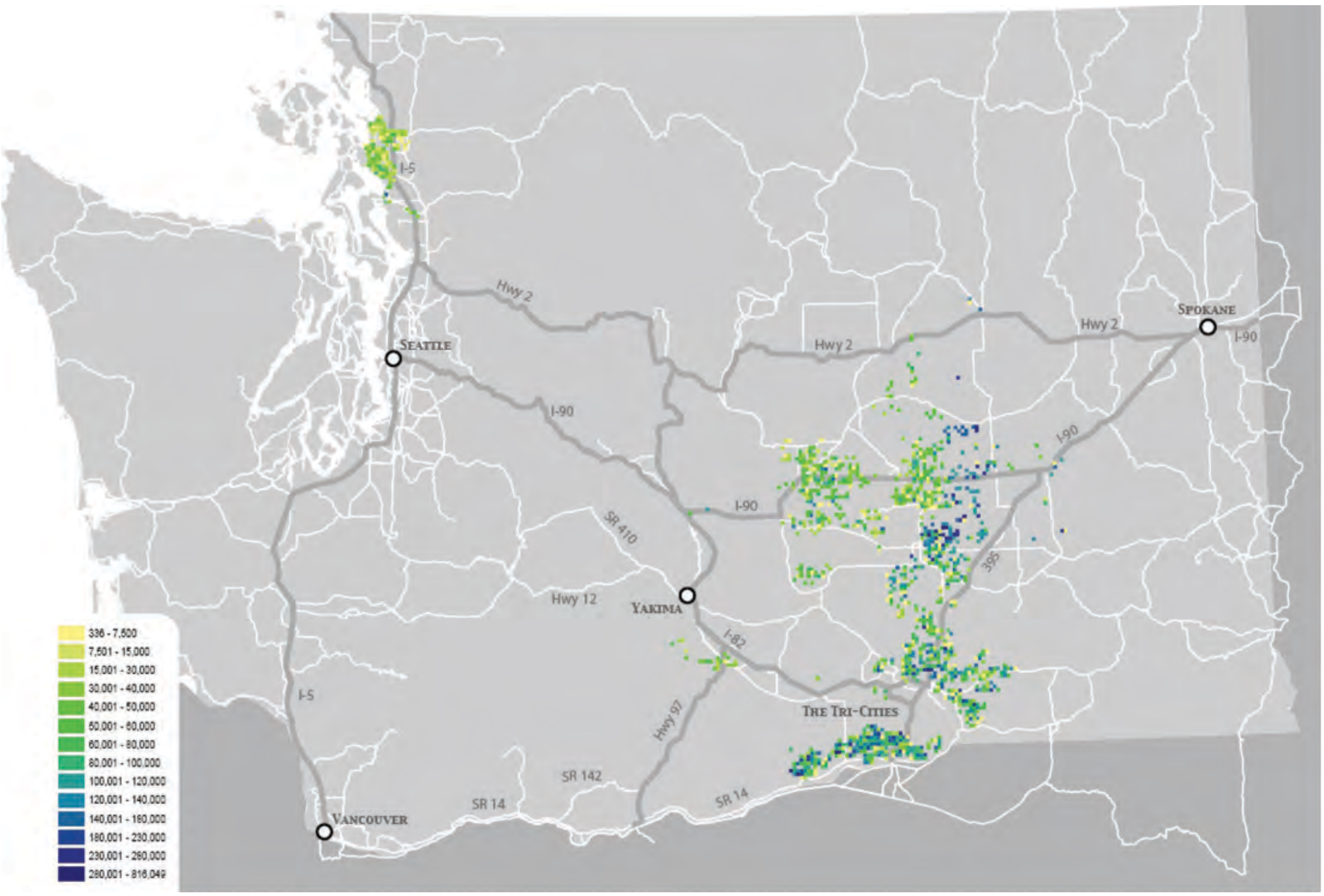


Figure 4.10. Washington potato production, 2006 (hundredweight by township and range).

by summing the production in the region, less the loss rate (6 percent). The loss rate was provided by the Washington State Potato Commission.

The Washington State Potato Commission also estimates the percentage of each potato product produced in each region. The Skagit Valley is estimated to produce 100 percent fresh potatoes, the Upper and Lower Basins are estimated to produce 14 percent fresh potatoes, 73 percent frozen potatoes, 11 percent dehydrated potatoes, and 2 percent potato chips.

Processing

The Washington State Potato Commission provided information on the location of all potato processing facilities in Washington (as shown in Figure 4.11) and the ratio of truckloads of fresh potatoes to truckloads of processed potatoes. The ratio for fresh potatoes to fresh potatoes is of course 1:1; the ratio of fresh potatoes to frozen potatoes is 2:1; the ratio of fresh potatoes to potato chips is 4:1; and the ratio of fresh potatoes to dehydrated potatoes is 6:1. Thus, for every truckload of dehydrated potatoes leaving a processing facility, six truckloads of fresh potatoes are required. Note, however, that a truckload of fresh potatoes is assumed to be equivalent to 22.22 tons, and a truckload of frozen potatoes is assumed to be equivalent to 20 tons.

No information was available regarding the location of dehydration processing for potatoes grown in the Skagit Valley. Dehydration could take place at processing facilities in either the Upper or Lower Basins. On the basis of interviews with industry experts of the Washington State

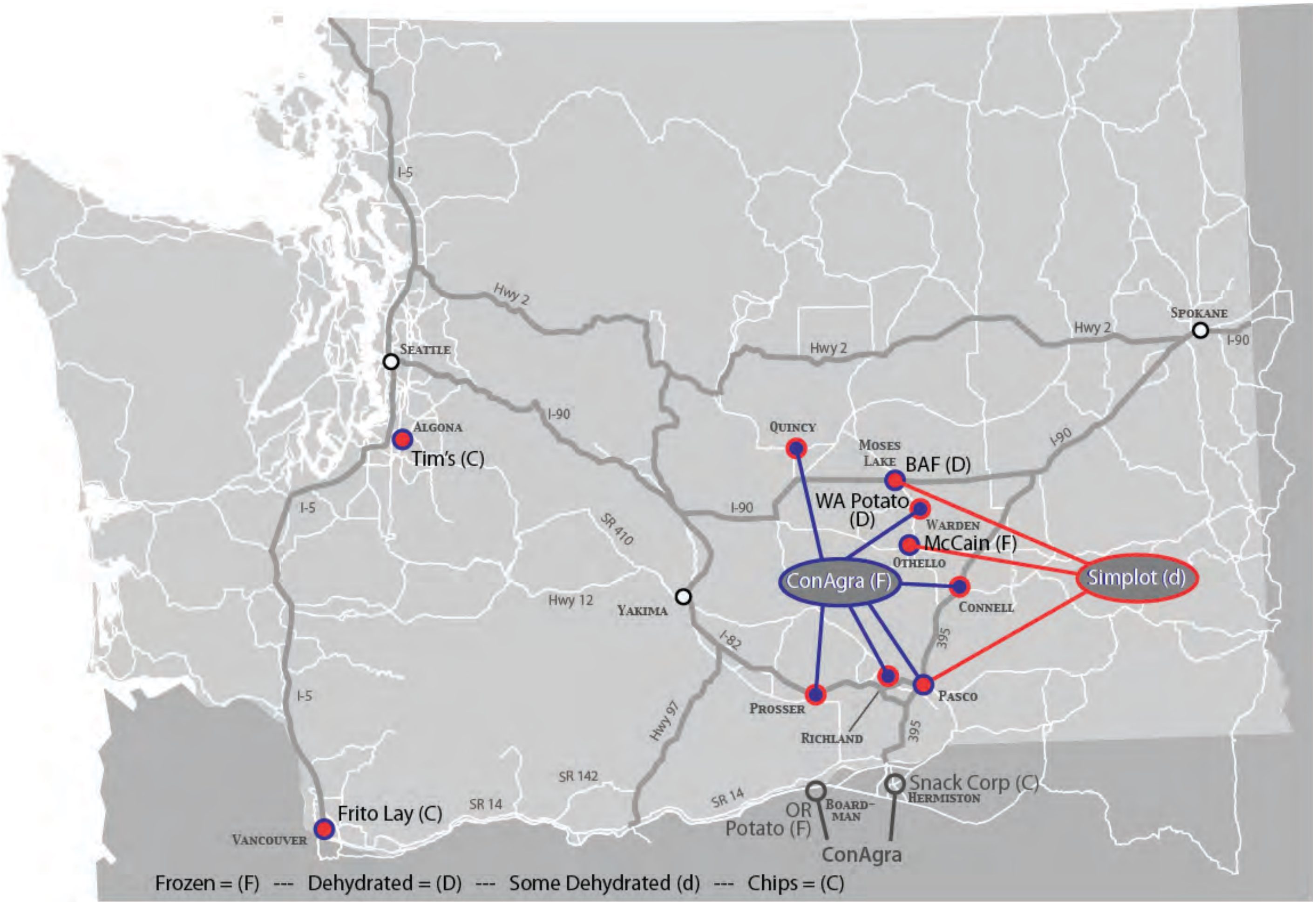


Figure 4.11. Location of Washington potato processors.

Potato Commission, it was estimated that 25 percent of these potatoes would be dehydrated in the Upper Basin, and 75 percent would be dehydrated in the Lower Basin. Potatoes grown in the Upper Basin were assumed to be processed in the Upper Basin, and potatoes grown in the Lower Basin were assumed to be processed in the Lower Basin.

Mode

It was assumed that 25 percent of the frozen potatoes processed in the state leave the state via rail. This assumption was supported by conversations with an industry expert at one of Washington's largest frozen potato producers, ConAgra. It also was assumed that 11 percent of fresh, dehydrated, and chipped potatoes are shipped out of the state on rail. Typically, these potatoes are destined for regions east of the Mississippi River.

Potato Consumption and Export

Potatoes grown and processed in Washington are either consumed in Washington or exported. The percentage of potatoes to major destination by region (estimated by the 2007 Washington Potato Commission Survey) is shown in Table 4.4.

Potatoes destined to international locations are exported via the Port of Seattle. According to the 2007 Washington Potato Commission Survey, these shipments use the routes shown in Table 4.5.

Table 4.4. Percentage of potatoes to major destinations by region.

Major Destinations	Lower Basin	Skagit Valley	Upper Basin
Eastern Washington	12.48%	2.03%	6.22%
Western Washington	14.29%	6.81%	6.40%
Oregon	2.31%	4.35%	1.25%
California	14.58%	40.72%	11.85%
Idaho	0.00%	0.00%	34.33%
States West of Mississippi	22.01%	13.30%	12.76%
States East of Mississippi	24.26%	23.58%	11.99%
Canada	8.85%	7.04%	2.91%
Mexico	0.14%	1.96%	0.25%
Other International	1.09%	0.20%	12.03%

Potatoes destined for Eastern Washington are assumed to be distributed to Moses Lake, Spokane, Kennewick, Warden, Yakima, and Grandview, proportional with population. Destinations in Western Washington are Seattle, Tacoma, Stanwood, and Auburn according to population. Destinations in Oregon and California are served via I-5, I-205, I-82, and Highway 97 equally. Destinations in Idaho are served via I-90, Highway 2, Highway 12, I-82, and I-84 equally. Other destinations in the United States are served via I-90 and I-82 equally. Destinations in Canada are served via I-5 or Highway 9 equally. Destinations in Mexico are served via I-5, I-205, I-82, or Highway 97 equally.

Truck Trips

To convert short tons to truckloads of potatoes, it was assumed that fresh and dehydrated potatoes weigh out at 22,000 pounds. For frozen potatoes, the research team assumed that a truckload can carry 20,000 pounds because of the refrigeration unit necessary. For potato chips, a truckload can carry only 5,000 pounds as the trucks meet volume constraints before weight constraints. Table 4.6 shows the number of truck trips per day between each origin and each destination for each product type estimated using this procedure.

Table 4.5. Major destinations by route of exported Washington-produced potatoes.

Major Destinations	Lower Basin	Skagit Valley	Upper Basin
Eastern Washington	I-90, I-82, Highway 12, 14	I-5, I-90, Highway 2, 405	I-90, I-82, Highway 17
Western Washington	I-90, I-82, I-5, 240, 395	I-5, 405, 167	I-90, Highway 17
Oregon	I-90, I-82, I-84, Highway 97, 395, 597	I-5	I-90, I-82, I-84
California	I-90, I-82, I-5, Highway 97, 395	I-5	I-90, I-82, I-5, Highway 17, 395
Idaho			I-90, I-82, I-84, SR 17, 395
States West of Mississippi	I-90, I-82, 395	I-90, I-80, I-5, I-84, 405	I-90, I-82, I-5, I-84, SR 17, 395
States East of Mississippi	I-90, I-82, 395	I-90, I-80, I-5, 405	I-90, I-82, I-5, I-84, SR 17, 395
Canada	I-90, I-82, I-5	I-5, I-90	I-5, I-90
Mexico	I-82, I-5, Highway 97	I-5	

Table 4.6. Potato truck trips per day.

	Total Production			Fresh			Frozen			Dehydrated			Chips		
Destinations	Skagit Valley	Upper Basin	Lower Basin	Skagit Valley	Upper Basin	Lower Basin	Skagit Valley	Upper Basin	Lower Basin	Skagit Valley	Upper Basin	Lower Basin	Skagit Valley	Upper Basin	Lower Basin
East Washington	0.17	0.06	0.12	2.90	1.71	3.96	0.00	4.17	9.67	0.00	0.24	0.53	0.00	0.27	0.63
Moses Lake	0.01	0.00	0.01	0.13	0.07	0.17	0.00	0.18	0.42	0.00	0.01	0.02	0.00	0.01	0.03
Spokane	0.08	0.03	0.06	1.40	0.83	1.92	0.00	2.02	4.68	0.00	0.12	0.26	0.00	0.13	0.30
Kennewick	0.03	0.01	0.02	0.50	0.30	0.69	0.00	0.72	1.68	0.00	0.04	0.09	0.00	0.05	0.11
Warden	0.01	0.00	0.01	0.13	0.07	0.17	0.00	0.18	0.42	0.00	0.01	0.02	0.00	0.01	0.03
Yakima	0.02	0.01	0.02	0.37	0.22	0.50	0.00	0.53	1.23	0.00	0.03	0.07	0.00	0.03	0.08
Grandview	0.02	0.01	0.02	0.37	0.22	0.50	0.00	0.53	1.23	0.00	0.03	0.07	0.00	0.03	0.08
West Washington	0.06	0.07	0.14	0.99	1.99	4.53	0.00	6.49	14.76	0.00	0.28	0.61	0.00	0.32	0.72
Seattle	0.02	0.02	0.04	0.28	0.56	1.28	0.00	1.83	4.15	0.00	0.08	0.17	0.00	0.09	0.20
Tacoma	0.01	0.02	0.03	0.23	0.47	1.06	0.00	1.52	3.46	0.00	0.07	0.14	0.00	0.07	0.17
Stanwood	0.01	0.01	0.03	0.20	0.40	0.92	0.00	1.32	2.99	0.00	0.06	0.12	0.00	0.06	0.15
Auburn	0.02	0.02	0.04	0.28	0.56	1.28	0.00	1.83	4.15	0.00	0.08	0.17	0.00	0.09	0.20
Oregon	0.03	0.01	0.02	0.57	0.28	0.73	0.00	0.70	1.79	0.00	0.04	0.10	0.00	0.05	0.12
via I-5	0.01	0.00	0.01	0.14	0.07	0.18	0.00	0.17	0.45	0.00	0.01	0.02	0.00	0.01	0.03
via I-205	0.01	0.00	0.01	0.14	0.07	0.18	0.00	0.17	0.45	0.00	0.01	0.02	0.00	0.01	0.03
via I-82	0.01	0.00	0.01	0.14	0.07	0.18	0.00	0.17	0.45	0.00	0.01	0.02	0.00	0.01	0.03
via Hwy 97	0.01	0.00	0.01	0.14	0.07	0.18	0.00	0.17	0.45	0.00	0.01	0.02	0.00	0.01	0.03
California	0.35	0.12	0.15	5.81	3.42	4.63	0.00	8.34	11.29	0.00	0.48	0.62	0.00	0.54	0.73
via I-5	0.09	0.03	0.04	1.45	0.85	1.16	0.00	2.09	2.82	0.00	0.12	0.16	0.00	0.14	0.18
via I-205	0.09	0.03	0.04	1.45	0.85	1.16	0.00	2.09	2.82	0.00	0.12	0.16	0.00	0.14	0.18
via I-82	0.09	0.03	0.04	1.45	0.85	1.16	0.00	2.09	2.82	0.00	0.12	0.16	0.00	0.14	0.18
via Hwy 97	0.09	0.03	0.04	1.45	0.85	1.16	0.00	2.09	2.82	0.00	0.12	0.16	0.00	0.14	0.18

(continued on next page)

Table 4.6. (Continued).

	Total Production			Fresh			Frozen			Dehydrated			Chips		
Destinations	**Skagit Valley**	**Upper Basin**	**Lower Basin**	**Skagit Valley**	**Upper Basin**	**Lower Basin**	**Skagit Valley**	**Upper Basin**	**Lower Basin**	**Skagit Valley**	**Upper Basin**	**Lower Basin**	**Skagit Valley**	**Upper Basin**	**Lower Basin**
Idaho	0.00	0.36	0.00	0.00	10.26	0.00	0.00	25.03	0.00	0.00	1.45	0.00	0.00	1.63	0.00
via I-90	0.00	0.07	0.00	0.00	2.05	0.00	0.00	5.01	0.00	0.00	0.29	0.00	0.00	0.33	0.00
via Hwy 2	0.00	0.07	0.00	0.00	2.05	0.00	0.00	5.01	0.00	0.00	0.29	0.00	0.00	0.33	0.00
via Hwy 12	0.00	0.07	0.00	0.00	2.05	0.00	0.00	5.01	0.00	0.00	0.29	0.00	0.00	0.33	0.00
via I-82	0.00	0.07	0.00	0.00	2.05	0.00	0.00	5.01	0.00	0.00	0.29	0.00	0.00	0.33	0.00
via I-84	0.00	0.07	0.00	0.00	2.05	0.00	0.00	5.01	0.00	0.00	0.29	0.00	0.00	0.33	0.00
West of Mississippi	0.11	0.08	0.22	1.84	2.28	6.98	0.00	5.56	17.05	0.00	0.32	0.94	0.00	0.36	1.11
via I-90	0.05	0.04	0.11	0.92	1.14	3.49	0.00	2.78	8.52	0.00	0.16	0.47	0.00	0.18	0.55
via I-82	0.05	0.04	0.11	0.92	1.14	3.49	0.00	2.78	8.52	0.00	0.16	0.47	0.00	0.18	0.55
East of Mississippi	0.20	0.13	0.24	3.40	3.70	7.70	0.00	9.04	18.79	0.00	0.52	1.03	0.00	0.59	1.22
via I-90	0.10	0.07	0.12	1.70	1.85	3.85	0.00	4.52	9.39	0.00	0.26	0.52	0.00	0.29	0.61
via I-82	0.10	0.07	0.12	1.70	1.85	3.85	0.00	4.52	9.39	0.00	0.26	0.52	0.00	0.29	0.61
Canada	0.06	0.03	0.09	0.99	0.85	2.81	0.00	2.09	6.85	0.00	0.12	0.38	0.00	0.14	0.45
via I-5	0.03	0.02	0.04	0.50	0.43	1.40	0.00	1.04	3.43	0.00	0.06	0.19	0.00	0.07	0.22
via Sumas (Hwy 9)	0.03	0.02	0.04	0.50	0.43	1.40	0.00	1.04	3.43	0.00	0.06	0.19	0.00	0.07	0.22
Mexico	0.02	0.00	0.00	0.28	0.00	0.04	0.00	0.00	0.11	0.00	0.00	0.01	0.00	0.00	0.01
via I-5	0.00	0.00	0.00	0.07	0.00	0.01	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
via I-205	0.00	0.00	0.00	0.07	0.00	0.01	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
via I-82	0.00	0.00	0.00	0.07	0.00	0.01	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
via Hwy 97	0.00	0.00	0.00	0.07	0.00	0.01	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
Other International	0.00	0.13	0.01	0.00	3.70	0.35	0.00	9.04	0.84	0.00	0.52	0.05	0.00	0.59	0.05
Port of Seattle	0.00	0.00	0.00	0.00	3.70	0.35	0.00	9.04	0.84	0.00	0.52	0.05	0.00	0.59	0.05

Step 3—Estimate Missing Data

The process of estimating missing data was implemented in several of the interim steps of this process. For example, no information was available regarding the location of dehydration processing for potatoes grown in the Skagit Valley. Dehydration could take place at processing facilities in either the Upper or Lower Basins. On the basis of the expert knowledge of the Washington State Potato Commission, it was estimated that 25 percent of these potatoes would be dehydrated in the Upper Basin, and 75 percent would be dehydrated in the Lower Basin. Potatoes grown in the Upper Basin were assumed to be processed in the Upper Basin, and potatoes grown in the Lower Basin were assumed to be processed in the Lower Basin. ConAgra staff also was used as industry experts in other cases in which missing data had to be estimated.

4.5 Next Steps

This chapter provides a detailed description of several aspects of how to develop subnational commodity flow from supplemental sources on local economic activity. It might be useful at this point to identify some local commodity flow data sources. Many of these sources can be identified by conducting an Internet search using publicly available search engines. Additionally, it might be helpful to consider private-sector participants. Consider what organizations they may belong to and how they might be able to help identify subnational commodity flow data sources for future freight planning efforts. Also, consider how these local sources of data might compare to commodity flow data captured from other methods such as establishment and roadside intercept surveys.

Refer back to the Playbook section to identify the next portion of this *Guidebook* that will be most relevant to a particular stage in the data collection process.

CHAPTER 5.0

Developing Subnational Commodity Flow Data Using Disaggregation

5.1 Introduction

This section provides an examination of how to develop subnational commodity flow data using data disaggregation. The disaggregation of freight flow data is the process of taking a preexisting freight flow database and dividing it into further detail to generate a more refined freight flow database. Disaggregation can occur across any dimension of the database (e.g., commodity, geography, or season). For freight flow databases, disaggregation most commonly occurs for geography. Transportation planners often take national-level freight flow data and generate county-level or more refined freight flow data to assist in freight planning efforts. Another common occurrence is travel demand modelers purchasing county-level proprietary freight flow data and disaggregating them to the TAZ level as part of the development of truck trip tables.

The primary pieces of information needed to conduct disaggregation of commodity flow data are an aggregated commodity flow database (e.g., FAF or TRANSEARCH), defined boundaries of the disaggregation region, and a disaggregation variable that is defined at the level of the disaggregated region. The same process would be used for disaggregating both FAF and TRANSEARCH. This *Guidebook* identifies the following five steps for administering a commodity flow data disaggregation technique:

- Step 1—Identify aggregated commodity flow database
- Step 2—Determine geographic boundaries for disaggregated database
- Step 3—Specify time dimension for disaggregated database
- Step 4—Select disaggregation variable
- Step 5—Select and implement disaggregation technique

Many of these steps are interrelated, but the *Guidebook* discussion of each step is ordered as shown in the above bulleted list. The description of each step is structured to focus on the following four key elements described in the Playbook (Chapter 6.0):

1. **Key Considerations—**A brief description of the main issues encountered and tradeoffs that will need to be made for the step.
2. **Implementation Process—**A detailed description of how to implement the step.
3. **Example—**An example of how this step has been implemented in other studies. Note that Steps 1, 2, and 4 are sufficiently straightforward such that no examples are needed. Step 3 shows an example for disaggregating commodity flow data along a time dimension. Examples of the full disaggregation technique are provided in Step 5.
4. **User's Guide Worksheet Punch List—**Simple bulleted instructions that *Guidebook* users can check off to ensure that they have implemented each of the major steps involved in conducting a commodity flow data disaggregation.

Each of these four elements focuses on different aspects of data disaggregation. For transportation agencies that are considering hiring a contractor to disaggregate FAF or TRANSEARCH data, reading the "Key Considerations" section of each step will likely provide enough information for the generation of an RFP on the topic. Transportation agencies that want to understand the details of how to conduct a commodity flow data disaggregation should focus on the "Implementation Process" sections in addition to the "Example" section. The "Example" section also will provide specific reference to efforts that have been conducted in other regions that can be compared to what already has been done in the agency's region and to responses to an agency's RFP. After transportation agencies have a sufficient background in all of the aspects related to conducting a commodity flow disaggregation, the "User's Guide Worksheet Punch List" sections can be used to walk the agency through all of the specific steps that need to be done to disaggregate a database. This section also can be compared to responses to an agency's RFP.

5.2 Step-by-Step Process for Disaggregating Commodity Flow Data

This section provides a comprehensive examination of the steps involved in developing subnational commodity flow data by disaggregating a preexisting commodity flow database. This section provides a detailed description of all of the necessary steps, including addressing relevant implementation issues that a typical regional planning office, local agency, or state department of transportation may experience when considering and implementing a data disaggregation.

Step 1—Identify Aggregated Commodity Flow Database

Key Considerations

The databases that most often serve as a starting point for disaggregation are the FHWA FAF database and the Global Insight TRANSEARCH database. The FAF database is publicly available at no cost. Its primary limitation tends to be that it has 114 predefined geographic regions that are not refined enough to match with the full spectrum of many transportation agencies' freight planning applications. TRANSEARCH is a proprietary commodity flow database that is often purchased at the county level, which provides much more refined geographic data than FAF. However, the process for developing TRANSEARCH is not as replicable as FAF, and TRANSEARCH's estimates do not include statistically validated confidence intervals.

Implementation Process

The primary action in this step is to review the features of the two most commonly used commodity flow databases: the FHWA FAF database and TRANSEARCH. FAF includes estimates of the weight and value of commodity flows by origin, destination, commodity, and mode for a base year of 2007 and a forecast year of 2040. FAF data are a prime candidate for disaggregation because this database provides comprehensive geographic and commodity coverage of commodity flows and is publicly available. The currently available version of FAF covers all flows for 123 regions (major metropolitan areas and balances of states), 17 additional metropolitan areas that serve as major international gateways, and 7 international regions. Disaggregation of FAF data often includes allocating freight flows from these default regions to regions that are compatible with state and metropolitan-level planning regions such as counties, zip codes, or TAZs.

FAF data can be obtained at http://ops.fhwa.dot.gov/freight/freight_analysis/faf/index.htm.

The FAF 2007 base year database is built entirely from public data sources. The starting point for FAF is the BTS CFS. The CFS is an establishment survey of 100,000 businesses in the United States that is then statistically expanded to estimate freight flows across 43 commodities and all freight modes. This survey is conducted every 5 years with the most recently released data being from 2007. The 2002 survey included only 50,000 businesses.

As discussed in Chapter 1.0 of the *Guidebook,* while the CFS is the most comprehensive establishment survey conducted in the United States, there are known sources of inaccuracy when considering the CFS as a comprehensive commodity flow database. These inaccuracies include business categories that are not surveyed, shipment types that are not included, and cells that are estimated.

Business types that are not included in the CFS sample are construction, service establishments, retail, farm-based businesses, logging, printing and publishing, retail, and household and business moves. Not including these establishments leads to an undercounting of commodity flows in the CFS. A study by the Oak Ridge National Laboratory (ORNL) completed in 2000 estimated that the 1997 CFS captured only 75 percent of total U.S. freight shipments measured in tons, 74 percent when measured in ton-miles, and 81 percent when measured in value. The 2002 CFS was estimated to have captured only 54 percent of total U.S. freight shipments when measured in tons, 67 percent in ton-miles, and 63 percent in value. FAF adjusts for this undercounting by developing estimates of these other sources through other means. Table 5.1 shows the process used for estimating the freight flows for each of the sectors missing from the CFS. Note that VIUS refers to the U.S. Census Bureau Vehicle Inventory and Use Survey. This database

Table 5.1. Variables for state and county allocations for CFS out-of-scope shipments.

Out-of-Scope CFS Business Sector	National to State Allocations	Commodity Used for State Allocation	State to County Allocation (Origin)	State to County Allocation (Destination)
Construction	VIUS Industry Sector	All	CBP Sector Employment	CBP Sector Employment
Services	VIUS Industry Sector	All	CBP Sector Employment	CBP Sector Employment
Retail	VIUS Industry Sector	All	CBP Sector Employment	Population
Farm-Based	VIUS Commodity	Animals	Value in Farm Sales (USDA)	CBP Animal Slaughtering and Processing Employment
Farm-Based	VIUS Commodity	Cereal	Value in Farm Sales (USDA)	CBP Grain and Oil Seed Milling Employment
Farm-Based	VIUS Commodity	Other Agriculture	Value in Farm Sales (USDA)	CBP Food Mfg Employment
Logging	VIUS Commodity	Logs and Other Wood	Round Wood Production (NFS)[a]	CBP Wood Products Employment
Printing	CBP Industry Employment	Printed Materials	CBP Industry Employment	Population
Fisheries	CBP Industry Employment	Live Fish	CBP Industry Employment	CBP Seafood Products Employment

[a]NFS = National Forest Service.
Source: "Out-of-Scope" reports developed for FAF by Macrosys and ORNL.

includes information on commercial vehicles registered by state along with truck type and operating characteristics for each state. VIUS was discontinued after the 2002 publication.

The CFS also only requests information on outbound shipments. This is sufficient for capturing domestic flows, since both the origin and destination occur within the CFS survey area. However, there is no survey coverage of import flows, because these shipments originate in foreign countries outside of the sample area covered by the CFS.

Additionally, there are several origin-destination-commodity-mode combinations that are not captured in the CFS, which is typically due to relatively smaller amounts of freight being moved and the randomness of the collected data. FAF compensates for these "zero cells" by estimating these combinations using other techniques. For example, although CFS information is not available for fertilizer shipments from Iowa to Memphis, Tennessee, CFS information is available on total fertilizer shipments from Iowa to all other CFS regions and for all commodities shipments from Iowa to Memphis, Tennessee. Using these broader shipment patterns, the fertilizer shipments from Iowa to Memphis are interpolated, and this estimate is provided within the FAF database.

Each of these estimation procedures has its own set of impacts on disaggregating FAF data. FAF data that are developed based on the expanded CFS sample data are the ideal type of commodity flow data to disaggregate as there are actual raw shipment data to support the estimates. FAF data that are estimated to fill in missing cells in the CFS survey may be problematic if they are used as the starting point of additional disaggregation, because there are no raw data to support the estimates. These estimated cells tend to have small shipment volumes and further disaggregating them can lead to potentially large errors.

FAF data that are the result of supplementing missing sectors in the CFS through the use of publicly available sources also presents issues. FAF estimates these missing sectors through a combination of economic data and allocation to smaller geographies using employment data. Further disaggregating these estimated data intensifies the potential for errors.

Many transportation agencies acquire commodity flow data by purchasing the Global Insight TRANSEARCH database. The methodology for creating the TRANSEARCH database is constantly evolving, but it has typically involved a mix of extrapolated survey data, economic output analysis, supply chain analysis, and geographic allocation. The database is typically purchased using county-level geographies, but also can be purchased at smaller levels such as zip codes or Census blocks. To provide subcounty-level commodity flows, TRANSEARCH tends to rely heavily on disaggregation. Therefore, the issues described related to FAF disaggregation also are applicable to TRANSEARCH subcounty commodity flows.

Example

No example is provided for this step. A full example of commodity flow disaggregation is provided in Step 5. Note that snippets of the FAF database can be found in Section 1.3. The full database can be found at http://ops.fhwa.dot.gov/freight/freight_analysis/faf/index.htm.

Examples of the TRANSEARCH database can be found by contacting Global Insight.

User's Guide Worksheet Punch List

- Review key features of FAF and TRANSEARCH database. Consider tradeoffs inherent in both databases.
- Review any other available commodity flow databases for the region of interest.
- Select aggregated commodity flow database.

Step 2—Determine Geographic Boundaries

Key Considerations

There are two geographies to consider when conducting a commodity flow disaggregation. First, is the desired geographic unit the most conducive for freight planning activities? For a truck-rail diversion analysis, county-level freight flows are likely to be sufficient. However, for travel demand modeling, the desired geographic unit is often the TAZ. Other potential geographic units include zip codes, BEA regions, or entire MPO regions. However, geographies that change over time, such as MPO boundaries, create additional estimation challenges. The second type of geographic consideration is the geography for which disaggregation variables are available. In some instances, the size of the geographic area can be used as a disaggregation variable. Typically, the disaggregation variable is some type of socioeconomic data or vehicle activity data. Regardless of which disaggregation technique is selected, the accuracy of each technique decreases as the geographic scale decreases.

Implementation Process

To determine which geographic area should be used for a data disaggregation, consideration must be given to the types of freight planning efforts that are likely to be supported by the disaggregated data. Table 5.2 provides a list of freight planning activities and commonly used levels of detail needed in disaggregated commodity flow data to support the planning activity. However, consideration should be given to the specific nature of the freight planning activity conducted in each region. In most cases, the geographical context also is constrained by existing governmental or administrative boundaries such as state, metropolitan region, urban area, or local planning region.

Disaggregation requires some relationship to be developed between the aggregate data and the subnational data desired. Establishing this relationship typically involves identification of one or more activity variables, which limits to some extent the geographical disaggregation to those levels for which the activity variable is reported or available. Disaggregation to geographic areas with limited or no information is difficult and therefore constrains the degree to which an accurate and reasonable disaggregation may occur.

Table 5.2. Level of commodity flow disaggregation commonly used for freight planning activities.

Freight Planning Activity	Level of Commodity Flow Disaggregation Commonly Used
Travel Demand Modeling	TAZ
Trading Partner Analysis	State, MPO, or metropolitan region
Truck-Rail Diversion Analysis	County
Long-Haul Corridor Analysis	County
Short-Haul Corridor Analysis	TAZ or other subcity zone system such as zip codes
State-Level Freight Plan	County
MPO or County-Level Freight Plan	TAZ or other subcity zone system such as zip codes

It is generally recommended that a geographic scale be selected that is as large as possible while still sufficiently refined to meet the planning objectives of the transportation planning agency. This recommendation is made because the accuracy of the data disaggregation decreases as the geographic scale decreases. Additionally, further disaggregation can occur at a later date if future planning applications require more geographic detail.

Example

No example is provided for this step. A full example of commodity flow data disaggregation is provided in Step 5.

User's Guide Worksheet Punch List

- Identify current freight planning applications for the region of interest.
- Determine the geographic level of data disaggregation needed for each application.
- Select the largest level of geography needed to meet all current freight planning applications.
- Note that this step will be revisited following Steps 3 and 4 related to the time dimension and disaggregation variables needed for this analysis.

Step 3—Specify Time Dimension

Key Considerations

The time dimension is closely related to the geographic scale desired and the freight planning application being considered. The time dimension also influences the accuracy of data disaggregation. Given the dynamic nature of freight flows over time, significant inaccuracies may result if the aggregate data are reported using different time segments than the desired time segmentation of the disaggregated data. A typical example of a time dimension mismatch is when national freight flow data report annual flows, while monthly, weekly, daily or hourly flows are desired for the disaggregated data. Another aspect of the time dimension is the need to forecast disaggregated commodity flow data for the future. A smaller set of disaggregation variables is available to support forecasting activities. It also is important to account for any seasonal distortions that may be present in the aggregated data set or disaggregation variable when considering the time dimension. For example, a commodity flow database constructed by expanding establishment survey data that were collected in a single season will have seasonal bias built into the database. The CFS database was developed using 1 week of data in each of the four seasons in an attempt to minimize seasonal bias in the data set.

Implementation Process

For the vast majority of freight planning applications, annual disaggregated freight flow data are sufficient. This is consistent with the annual format of most national freight flow data. One notable exception is truck trip tables for travel demand models, which tend to require average daily or average weekday estimates of truck trips and often even estimates of truck trips by time period. Another potential exception is if peak flows are desired. Freight flows in different regions and different commodities peak at different times of the day, week, and year, so care must be taken in developing conversion factors for moving from an annual database to a more temporally refined database. Potential sources for temporal conversion factors include permanent vehicle classification count data and interviews with local shippers and carriers. The example provided below illustrates how a commodity flow database can be disaggregated along the dimension of time using a combination of classification count data and stakeholder outreach.

Example

As a hypothetical example of disaggregating commodity flow data along the time dimension, consider an MPO that wants to know the peak daily and hourly truck volumes from a recently expanded intermodal railyard in its area. As a first step, the MPO contacts the railroad that operates the railyard for some information, and the MPO is provided with the following:

- Approximately 300,000 containers will be moved at full build-out in 2020
- 90 percent of the containers move during the week
- Monthly fluctuations prior to the intermodal yard expansion are as shown in Table 5.3
- Daily and hourly fluctuations are roughly the same throughout the year
- This situation is similar to a very simplified commodity flow database with two regions (the intermodal yard and an external region), one mode (trucking), and one commodity (intermodal goods) and an annual flow between regions of 300,000 units
- Using the information provided by the railroad, the MPO decides to conduct truck counts just outside the intermodal railyard gates for 1 week during the month of April. The hourly flows from these counts are shown in Table 5.4.

The first step is to calculate the peak monthly volume. The 300,000 annual container estimate is multiplied by the peak monthly volume in September of 13 percent (see Table 5.3). This provides an estimate of 39,000 containers during the peak month. This is then divided by 4.3 (52 weeks divided by 12 months) to get the average number of weeks in a month. Therefore, the

Table 5.3. Hypothetical estimate of monthly percentage of activity at intermodal railyard.

Month	Percent of Traffic
January	4%
February	4%
March	6%
April	8%
May	9%
June	10%
July	10%
August	11%
September	13%
October	11%
November	9%
December	5%
Total	**100%**

Table 5.4. Hypothetical two-way truck count data from intermodal railyard.

Period	Monday	Tuesday	Wednesday	Thursday	Friday	Total
8 a.m. to 9 a.m.	50	85	80	60	90	365
9 a.m. to 10 a.m.	65	90	95	65	85	400
10 a.m. to 11 a.m.	75	120	130	95	80	500
11 a.m. to 12 p.m.	100	125	125	110	95	555
12 p.m. to 1 p.m.	50	55	45	65	35	250
1 p.m. to 2 p.m.	125	140	135	135	95	630
2 p.m. to 3 p.m.	125	145	130	130	100	630
3 p.m. to 4 p.m.	65	95	75	100	50	385
4 p.m. to 5 p.m.	65	90	80	95	55	385
5 p.m. to 6 p.m.	90	125	105	140	75	535
6 p.m. to 8 a.m.	0	0	0	0	0	0
Total	**810**	**1,070**	**1,000**	**995**	**760**	**4,635**

peak week container volume is 9,070. Table 5.4 indicates that the peak day of the week is Tuesday with 1,070 (23 percent) of the total 4,635 weekday trucks during the 5 days of data collection in April. Therefore, the peak daily trucks also is on a Tuesday with peak daily flows being 9,070 multiplied by the 90 percent of trucks during the week and then multiplied by 23 percent for the trucks that typically occur on a Tuesday. This amounts to an estimate of 1,877 trucks during the peak day in September in 2020.

The hourly data collection shown in Table 5.4 indicates that peak hourly flows occur between 2:00 p.m. and 3:00 p.m. on Tuesdays with 14 percent of the daily flows occurring during this hour. Therefore, the peak hourly volume estimate is 1,878 multiplied by 14 percent, or 263 trucks.

User's Guide Worksheet Punch List

- Identify current freight planning applications for the region of interest.
- Determine the time dimension requirements for each application.
- Select the smallest unit of time needed to meet all current freight planning applications.
- Note that this step will be revisited following Step 4, which is related to disaggregation variables.

Step 4—Select Disaggregation Variable

Key Considerations

Disaggregation variables are activity variables that are used as a proxy to represent economic activity and freight flows at the desired subnational level. Some of the most common disaggregation variables are employment, output, warehouse space, population, revenue, personal income, and geographic area. When using a disaggregation technique to develop subnational freight forecasts, it also is necessary to have a forecast of the disaggregation variable. For example, if employment is used as the disaggregation variable, then there must be a forecast of employment in the forecast year to develop a forecast of freight flows for the study area. The limited availability of forecasted disaggregation variables is often the primary factor in identifying an appropriate disaggregation variable. For example, the difficulty in developing a long-term forecast for warehousing space at the geographic level consistent with most study areas precludes warehouse square footage from being a useful disaggregation variable.

The selection of the disaggregation variable also significantly impacts the accuracy of the disaggregation, because for each commodity the variable is only partially relevant or there are other influences that are not incorporated into the model and these relationships are likely to change over time. Depending on the availability of activity variables at different geographies, it may be necessary to conduct the disaggregation in multiple steps.

Implementation Process

To select a disaggregation variable, a thorough review of activity variables available at different geographies should be conducted for the region. As mentioned above, some of the most common variables that are used for disaggregation include employment, output, warehouse space, population, revenue, personal income, and geographic area. However, less commonly used variables can be considered depending on the specific commodities being disaggregated. For example, the location and size of landfills can be used to disaggregate commodity flows of waste materials. In this manner, the local data sources that are discussed in Chapter 4.0 can generally be used for disaggregation of national commodity flow data. The most common disaggregation variables are discussed below.

Employment. The number of employees is the variable most commonly used to disaggregate national data, due to its availability at the local, regional, and state level and the availability of forecast employment data in most regions. Total employment is available from Census data, and the number of employees per industrial classification is often available from the state department of revenue at the county or zip code level. Travel demand models typically have estimates of employment at the TAZ level. Traditionally, the relationship between employment and shipments has been assumed to be linear. However, recent research indicates that for several industries, large companies have fewer shipments per employee than smaller industries (Holguín-Veras and Ban 2010). This may be a result of efficiencies of scale that are achieved at larger companies relative to their logistics practices. There also are several reasons why the number of employees may not be a very accurate prediction of freight flows or shipment activity including the following:

- There is sometimes a lack of correlation between employment and output. Some companies within the same industry may rely more on equipment than on personnel for their output, which can generate widely different output-per-employee rates. A recent research paper conducted a statistical analysis of employment relative to tonnage output by commodity using the FAF data set. The correlation was found to vary widely, with some industries having a high correlation (as reflected in a high R^2 value) and others having a low correlation. Table 5.5 shows the correlation by industry.

Table 5.5. Correlation between employment and tonnage output by industry (based on FAF data).

SCTG	Dependent Variable	Explanatory Variable(s)	R^2
20-23	Various[a]	Chemical mfg	11%
10-15	Various[b]	Mining (except oil and gas)	13%
9	Tobacco Products	Beverage and tobacco product mfg	15%
1	Live Animals/Fish	Support activities for agriculture and forestry	17%
16	Crude Petroleum	Oil and gas extraction	21%
38	Precision Instruments	Miscellaneous mfg	34%
24	Plastics/Rubber	Plastics and rubber products mfg	43%
2	Cereal Grains	Food mfg, farm acres	48%
8	Alcoholic Beverages	Beverage and tobacco product mfg	50%
39	Furniture	Furniture and related product mfg	56%
4	Animal Feed	Support activities for agriculture and forestry	60%
31	Nonmetallic Mineral Products	Nonmetallic mineral product mfg	61%
6	Milled Grain products	Food mfg	62%
19	Other Coal and Petroleum Products	Oil and gas extraction, petroleum and coal products mfg	62%
34	Machinery	Fabricated metal product mfg, machinery mfg	63%
40	Misc. Manufactured Products	Miscellaneous mfg	64%
3	Other Agriculture Products	Food mfg, farm acres	65%
33	Articles of Base Metals	Fabricated metal product mfg	65%
25	Logs	Forestry and logging, support activities for agriculture and forestry, wood product mfg	70%
35	Electronic and Electrical	Machinery mfg, computer and electronic product mfg, electrical equip and appliance and component mfg	70%
27	Newsprint/Paper	Forestry and logging, printing and related activities	73%
30	Textiles/Leather	Textile mills, textile product mills	73%
36,37	Various	Transportation equipment manufacturing	74%
7	Other Foodstuff	Food mfg, chemical mfg	75%
26	Wood Products	Wood product mfg	75%
32	Base Metals	Primary metal mfg, machinery mfg	75%
18	Fuel Oils	Petroleum and coal products mfg	77%
28	Paper Articles	Paper mfg, printing and related activities	81%
17	Gasoline	Petroleum and coal products mfg	83%
29	Printed Products	Paper mfg, printing and related activities	85%
5	Meat/Seafood	Food mfg	86%
41	Waste and Scrap	NAICS 115, 221, 321–327, 331–339	86%
43	Mixed Freight	NAICS 321-327, 481, 483–488, 492–493	86%

[a] SCTG 20-23 is Basic Chemicals, Pharmaceutical Products, Fertilizers, and Chemical Products and Preparations n.e.c.

[b] SCTG 10-15 is Monumental or Building Stone, Natural Sands, Gravel and Crushed Stone, and Nonmetallic Minerals n.e.c.

- There also is often a wide variation of companies within a single industry even at refined levels of industry codes. For example, NAICS code 326111 is for Plastic Bag and Pouch Manufacturing, while NAICS code 326122 is for Plastics Pipe and Pipe Fitting Manufacturing and NAICS code 326211 is for Tire Manufacturing. All of these activities are within the same three-digit NAICS code of 326, but are likely to have very different output-per-employee rates.
- There can be a "headquarters" issue, in which employees are recorded in one location while the freight activity associated with their employment occurs in a different location. In the agricultural industry, it is common for employees to be recorded in an urban area where the company's

administrative activities occur while the freight-related aspects of the employee's employment often occurs hundreds of miles away in a rural location.

- At finer geographic levels, there are typically fewer employment categories available than desired. For example, at the TAZ level, employment is frequently at the one-digit SIC level. At the county level, several of the two-digit employment categories that are publicly available through the U.S. Census Bureau are suppressed due to privacy or accuracy concerns. It is often necessary to extrapolate commodities based on limited employment specificity.
- Employment databases are typically industry based, while freight flow databases are typically commodity based. Therefore, companies that produce goods in multiple commodity categories are typically classified as being associated with a single industry. For example, Hewlett-Packard sells a combination of electronic hardware, electronic software, and information technology services. However, if the company is listed as only being in electronics, there is a danger that all of the employees within the company will be misrepresented as being involved in freight-intensive activities. Additionally, the industry codes in employment databases and the commodity codes in commodity flow databases are not always a one-to-one match, thereby generating the need to shift goods between categories that are not ideally compatible.
- There is often a need to consider matching both inputs and outputs to employment categories. For example, a food manufacturing employee can reasonably be associated with food manufacturing output. However, the inputs associated with food manufacturing are a combination of other manufactured foods, raw agricultural products, and other commodities likely in smaller quantities. This matching of inputs to output categories can be handled through the application of input-output factors to industry categories, as described in Chapter 1.0.
- The number of employees needed 20 years ago to produce a car, computer chip, or potato chip is likely not the same as it is today or will be 20 years in the future. A productivity factor is recommended for employment forecasts to account for additional goods produced per employee in many industries.

Revenue. Revenue data are typically available from the state department of revenue. These data are typically available at the state level and by industry or even establishment type. Sometimes revenue data are available by county. Revenue has many of the same drawbacks as employment as a disaggregation variable. The issues can be further exacerbated when using revenue because there are several bulk commodities that generate low-revenue sales, but have high tonnage amounts.

Population. Population also is a widely used disaggregation variable, due to its availability and accessibility at various geographic scales. The U.S. Census Bureau provides very detailed population estimates for cities, towns, urban areas, counties, and zip codes making population easily applicable for subnational commodity flow data disaggregation. Travel demand models typically have employment estimates at the TAZ level. Population can be an accurate reflection of the consumption of select commodities that are typically sold at the retail level, including gasoline, food products, and other products that are transferred to retail stores through warehouses and distribution centers. However, as with other activity variables, commodity flows are generated and affected by more than population. In particular, there are several production activities that are barely related to population, and the consumption of industrial commodities is typically not well correlated to population. Developing different relationships between population and commodity flows by zoning category (residential, commercial, manufacturing, construction, etc.) may help mitigate these inaccuracies.

Personal Income. Income data also are available from the U.S. Census Bureau. This activity variable is available at various subnational geographies such as state, city, county, BEA zone, zip code, and urban/metropolitan area. Income is an indicator of per capita purchasing power and can impact the specific type of freight activity in a region. Personal income as a disaggregation variable has most of the same drawbacks as population, especially in being an indicator of consumption and much less linked to production.

Output (Gross Domestic Product). Gross domestic product, or output, is another activity variable that may be used to disaggregate national data. It is geographically limited to either state or BEA zone, thus is not applicable where more localized disaggregation is sought. At the county level, many of the output data are suppressed for proprietary reasons. However, while between and within industries there may be significant variation in the relationship of output and freight activity, across all industries increases in output should translate into increases in freight activity. Using output as a disaggregation variable generates similar concerns to those generated using revenue or employment—issues with industry and commodity classification, headquarters issues, and issues with matching inputs to outputs.

Warehouse Space (Square Feet). The amount of warehousing space also may be used as an activity variable for disaggregation of national freight data. This information is often available through the state department of revenue and is used as an indicator of freight activity and intensity. This indicator variable works best for commodities with supply chains that feature warehouse usage. To develop commodity-specific disaggregation factors, it may be necessary to contact warehouse operators to determine the types of goods that are stored at the facility.

One of the more challenging considerations of data disaggregation and, subsequently, the calibration of disaggregated data, is the degree to which data disaggregation accurately represents reality. The degree of accuracy or inaccuracy is often difficult to assess given that a thorough understanding of subnational commodity flows at a local and regional level is often limited or unknown and comparisons to flows from disaggregated national data are difficult in this information vacuum. The issue of accuracy is further complicated by the fact that the national data source itself is often an estimation with various geographical and spatial limits. As discussed in Chapter 4.0, accuracy can be improved by validating estimates from a variety of published local data or primary data collection efforts.

Example

No example is provided for this step. A full example of commodity flow data disaggregation is provided in Step 5.

User's Guide Worksheet Punch List

- Review available disaggregation variables in the region of interest.
- Select the appropriate disaggregation variable for each commodity in your desired database.
- Review Steps 1 through 3 to determine whether changes are needed in the aggregated database used, the geographic boundaries of the region of interest, or the time dimension used for the disaggregated database.

Step 5—Select and Implement Data Disaggregation Technique

Key Considerations

There are several disaggregation techniques that are available for consideration. The simplest of the techniques are easy to implement, but do not allow for consideration of all available freight flow and economic information. The more complex methods can require significant resources to implement, but allow for use of a broader range of data sources to shape the disaggregated commodity flow database. From simplest to most complex, the techniques are geographic allocation, regression models, iterative proportional fitting, and input-output methods. When multiple rounds of disaggregation are used to develop a subnational commodity flow database (e.g., FAF region to county, then county to zip code), more than one of these methods may be incorporated. Alternatively, it is possible that different techniques could be used for different commodities within a single round of disaggregation.

Implementation Process

There are four disaggregation techniques that are discussed in this section: (1) geographic allocation, (2) regression methods, (3) iterative proportional fitting, and (4) input-output methods.

Geographic Allocation. This technique involves disaggregating commodity flows to smaller geographies based on features of each of the smaller geographic units. In the simplest sense, one can imagine taking a county-level commodity flow database, dividing up a county within that database into two equally sized halves, and then allocating half of the full county's commodity flows to each of the half counties. This new allocation represents the subcounty commodity flow database. Other, more complex, examples revolve around this basic concept and can include the following:

- Allocating commodity flows to subnational levels based on the physical size of the disaggregated geographies
- Allocating commodity flows based on some socioeconomic data variable within each of the subregions such as total employment or employment in all freight-related sectors
- Allocating commodity flows to the subregional level based on industry-specific or commodity-specific activity data within each of the subregions

For each of these examples, the commodity flows are allocated to the subregions based on unique characteristics of each subregion.

Regression Models. Regression models are used to establish a statistical relationship between two or more variables. Regression analysis includes a dependent variable that is the function of different levels of independent variables. For the purpose of developing subnational commodity flow databases, the independent variables are the commodity flows for a subregion typically defined by commodity type. The dependent variables are unique characteristics that are available at the subregional level. The regression analysis develops mathematical equations that define the commodity flows that are likely to occur for varying values of the dependent variables.

Using the analysis conducted in a report prepared for FHWA, it is possible to apply regression equations to the FAF database and to establish relationships between commodity flows and several different combinations of dependent variables (Cambridge Systematics 2009). In the report, the independent variables were tonnage output by two-digit SCTG commodity code. The dependent variables included industry-specific employment, population, and other industry-specific factors where relevant such as farm acres and oil/gas extraction. The regression was run for each of the 89 domestic zones that are part of FAF. The regression was used to identify the best variable(s) to predict the tonnage flows for each commodity. The best fit variables and level of confidence in the relationship are shown in Table 5.5. As the table demonstrates, for some commodities, a strong statistical relationship between commodity flows and socioeconomic activity variables was identified. For other commodities, there was no variable that was found to have a strong predictive relationship with commodity flows. The relationships established through this process can be used to estimate tonnage flows at the subnational level. The estimates can actually be done for any geographic level for which the independent variables can be identified. In the report prepared for FHWA, commodity flow estimates were developed at the county level using the regression results and county-level activity data (Cambridge Systematics 2009).

Regression methods are essentially similar to the methods discussed in the establishment survey section that are used to extrapolate commodity flows from sampled companies to the universe of companies in a region. See Chapter 2.0 for a description and examples of this process.

Iterative Proportional Fitting. The most common approach to disaggregating national freight data is to apply some form of the iterative proportional fitting process, first developed by William Edwards Deming and Frederick F. Stephan (1940). This technique is ideal for two-dimensional tables where the marginal (column and row) totals are known (or estimated through an activity variable), but the distribution throughout the matrix is unknown.

For a commodity flow database, the columns and rows correspond with origins and destinations. The totals of commodity flows for an origin at a large aggregation level are known, while the commodity flows for subregions would not be known. Iterative proportioning can be used to develop estimates of flows at the subregional level. If the subregional commodity flow database has I rows and J columns, then we can create an I x J table. Assuming independence amongst the origin-destination freight flows at the subregional level and some multinomial distribution of freight flows between subregions, it can be estimated that the flow between a specific origin *i* and a specific destination *j* is $\widehat{m}_{ij}$, where $m_{ij} = a_i b_j$ for all *i* and *j*. The percentage of the total commodity flow in row I is a_i, while the total commodity flow in column J is b_j.

$$\widehat{m}_{ij}^{(2n-1)} = \frac{\widehat{m}_{ij}^{(2n-2)} x_{i+}}{\sum_{k=1}^{J} \widehat{m}_{ik}^{(2n-2)}} \qquad \text{(Eq. 10)}$$

$$\widehat{m}_{ij}^{(2n)} = \frac{\widehat{m}_{ij}^{(2n-1)} x_{+j}}{\sum_{k=1}^{I} \widehat{m}_{kj}^{(2n-1)}} \qquad \text{(Eq. 11)}$$

Notice that the row and column totals are constant, denoted by

$$x_{i+} = \sum_j x_{ij} \qquad \text{(Eq. 12)}$$

$$x_{+j} = \sum_i x_{ij} \qquad \text{(Eq. 13)}$$

This particular form of the iterative proportional fitting model is often used at the national level for completing unknown cell values for which historical values or values from other cells are available, and $\hat{m}_{ij}$ is estimated using either a maximum likelihood estimation technique or utilizing a log-linear model (Lee and Viele 2001).

Unfortunately, at the subnational level, historical table values are not always available, and estimation using one of these approaches is difficult. Therefore, what is usually applied is a simplified case of the model above assuming quasi independence and resulting in a two-step factor estimation process where the initial $\hat{b}_j^{(0)} = 1$ and $n \geq 1$ and the resulting cells are estimated such that:

$$\hat{a}_i^{(n)} = \frac{x_{i+}}{\sum_j \hat{b}_j^{(n-1)}} \qquad \text{(Eq. 14)}$$

$$\hat{b}_j^{(n)} = \frac{x_{+j}}{\sum_i \hat{a}_i^{(n)}} \qquad \text{(Eq. 15)}$$

Input-Output Methods. One approach to developing disaggregated data is through input-output analysis. Input-output analysis is the process of relating the quantity and type of products produced at a given location to the quantity and type of products supplied to the location. The location can be a single facility, a collection of unrelated facilities within a region, or a specific industry within a region. The theory of input-output analysis is that the relationship between inputs and outputs is relatively constant across different geographies and time periods, and therefore inferences can be made about both the inputs and outputs if just one of the factors is known at a single location.

Input-output data are publicly available from the U.S. Department of Commerce Bureau of Economic Analysis for the year 2002 at http://www.bea.gov/industry/io_benchmark.htm. The data available include commodities that are generated by specific industries in what are termed "make tables." It also includes commodities that are consumed by specific industries in what are termed "use tables." Make tables tend to be straightforward in that industries tend to produce only one or two commodities that are directly affiliated with their industry. For example, the forestry and logging industry in 2002 was found to produce $18.9 million of crop products and $32.0 billion of forestry and logging products, while producing none of the other 138 commodities that are tracked in the summary database. An industry as diverse as basic chemical manufacturing produces 22 different commodities, but 94 percent of their production by value is in two commodities: basic chemicals (82 percent) and resins/rubber/artificial fibers (12 percent). Note that the detailed input-output database contains thousands of commodities and may be useful for analysis of very specific goods.

Use tables show the commodities consumed by a specific industry and reflect much higher levels of diversity. Table 5.6 shows the use table for three industries: crop production, forestry and logging, and basic chemical manufacturing. The value in the top row of the crop industry column shows that the U.S. crop industry purchased $893 million of basic chemicals in 2002. In total, the crop industry purchased items from the 16 commodities shown in Table 5.6 and from the 64 commodities that have been combined into the "other commodities" category. Four of these commodities represent over 10 percent of the total commodities purchased by the crop industry. The forestry and logging industry has a more consolidated use table than the crop industry, but it still purchased 7 different commodities listed in the table and another 47 that have been combined into the "other commodities" category. The basic chemicals industry purchased 136 different commodities, including over 100 that have been combined in the "other commodities (products and services)" category.

The term "commodities" is used to refer to all items that are purchased by companies in these industries. For example, included in commodities is everything from petroleum and coal products to truck transportation to real estate. To use an input-output make or use table for a commodity flow input-output analysis, it is necessary to remove all of the items that would not be listed as a commodity in the SCTG format or other similarly structured commodity codes. Then, the dollar value of commodities used and made by each industry can be converted to tonnages using ton-value ratios from a source such as the BTS CFS or local sources, if available. The tonnage values can then be used to understand the quantities and types of input commodities needed to produce a unit of output of each commodity. As mentioned above, these input-output relationships can then be applied to individual facilities, a group of companies, or an entire industry.

More recent input-output data are available from proprietary sources. The current and most widely used input-output package (IMPLAN Professional Software Version 2) includes nearly 500 industry sectors and allows geographic aggregation at the state, county, subcounty, and zip

Table 5.6. Use table for three industries (dollars in millions).

Commodity	Select Industry: Crop Production	Percent of Total	Forestry and Logging	Percent of Total	Basic Chemical Manufacturing	Percent of Total
Basic chemicals	893	1%	–	0%	28,779	34%
Real estate	14,249	19%	115	1%	242	0%
Support activities for agriculture and forestry	10,761	15%	2,811	14%	–	0%
Forestry and logging products	–	0%	12,924	62%	22	0%
Petroleum and coal products	4,476	6%	187	1%	6,233	7%
Wholesale trade	4,491	6%	1,267	6%	5,054	6%
Crop products	8,063	11%	5	0%	886	1%
Agricultural chemicals	7,897	11%	29	0%	581	1%
Monetary authorities, credit intermediation and related activities	6,415	9%	292	1%	344	0%
Electric power generation, transmission, and distribution	2,681	4%	10	0%	3,697	4%
Management of companies and enterprises	–	0%	–	0%	6,224	7%
Natural gas distribution	683	1%	1	0%	3,455	4%
Truck transportation	1,725	2%	471	2%	1,369	2%
Scientific research and development services	–	0%	–	0%	2,850	3%
Plastics and rubber products	738	1%	21	0%	1,501	2%
Rights to nonfinancial intangible assets	118	0%	3	0%	2,076	2%
Other fabricated metal products	67	0%	19	0%	1,543	2%
Maintenance and repair construction	778	1%	32	0%	759	1%
Rail transportation	451	1%	71	0%	866	1%
Other commodities (products and services)	8,863	12%	2,562	12%	17,556	21%
Total intermediate inputs	**73,351**	100%	**20,819**	100%	**84,037**	100%
Compensation of employees	14,569		4,115		15,324	
Taxes on production and imports, less subsidies	(6,707)		1,135		898	
Gross operating surplus	38,308		5,990		5,742	
Total Value Added	**46,171**		**11,240**		**21,964**	
Total Industry Output	**119,522**		**32,060**		**106,001**	

Note: Numbers may not add up due to rounding.

code level. The IMPLAN data consist of (1) a matrix of industry-specific technical coefficients that specifies the quantity of inputs necessary to produce a given unit of output and (2) sector-specific final demand, final payments, industrial output, and employment. This combination of industry-specific activity for a given geographical area allows a more robust estimation of freight activity for that region by industry sector.

Input-output analysis does have limiting issues that are worthy of consideration. The technical coefficients are treated as constants, thus they do not account for the real-life variability in the number of specific inputs per product produced across various firms or regions. In reality, firms are constantly adjusting and substituting inputs as market conditions change, technologies change, labor productivity changes, prices for labor and equipment change, and the structure of the industry changes. This limitation is especially problematic if this type of modeling approach is applied to longer period forecasting.

Example 1—Geographic Allocation for San Joaquin Valley Truck Trip Table Development

In 2004, the San Joaquin Valley Council of Governments (COG) developed a truck component to its travel demand model. Generating a truck trip table at the TAZ level was a critical step in the development of this model. To develop the intercounty portion of the truck trip table, the San Joaquin Valley COG disaggregated county-level California Intermodal Transportation Management System (ITMS) freight flow data—first to the zip code level, then to the TAZ level, using a series of economic relationships and conversion factors. The ITMS database was a county-level freight flow database managed by the California Department of Transportation generated primarily from TRANSEARCH data. The specific steps of this process are described below.

For intercounty truck trips, the first step in this process converted the truck tons in the ITMS database into truck trips using average payloads for each of the commodities in the ITMS data. The average payload data were developed from the 1997 Vehicle Inventory and Use Survey. Application of the payload matrix to the ITMS data created a county-level truck trip table from the truck tonnage data for the state of California.

The ITMS truck trip data were then grouped geographically to create relevant regions for the truck model. Internal regions were based on the eight counties that constitute the San Joaquin Valley study area. Regions external to the Valley were developed to correspond to each of the external cordons that can be used for trucks exiting the study area. Next, the county-level ITMS commodity flow truck trip data were allocated to zip codes. This allocation was performed using Dun & Bradstreet employment data from 2000. These data include the number of employees by zip code for each of the eight counties in the San Joaquin Valley for thousands of different employment categories based on the SIC system at a four-digit level. For the agricultural industry, estimated farm acreage was used rather than employment to allocate the ITMS tons to each zip code. Farm acreage was used because it is more representative of the location where the goods are actually produced than employment data. Employment data in the agricultural industry are often inaccurate due to the seasonality of much of the industry's employment. Employment data also can be geographically inaccurate due to the tendency to report employment at company headquarters in urban areas rather than at the rural locations where the goods are actually produced. The allocation from agriculture was then combined with the allocation from other industries to create two zip-code-to-county tonnage tables. One table contained the tonnage originating in each zip code destined for each county, while the other table contained the tonnage destined for each zip code originating in each county.

The zip-code-level tonnage data were then allocated to the TAZs in the truck model. This allocation was done based on employment data from the statewide model combined with the areas of geographic overlap between the zip codes and the TAZs. This process developed the final TAZ-level truck tonnage table for the 1996 ITMS data. This truck trip table was then projected to the year 2000 based on the freight tonnage growth derived from the FHWA FAF data for the state of California.

Example 2—Using the Regression Method to Disaggregate the FAF Database

For this example, we will consider an MPO whose metropolitan region is not specified as a FHWA FAF region. The MPO wants to develop a county-level flow database for the wood products commodity. We will assume that this region is composed of three counties: Aspen, Birch, and Cedar. The MPO region is bordered by an ocean to the west and mountains to the east, so it is assumed that the vast majority of its wood is going either north or south.

The first step in this process is to review Table 5.5 (provided previously in this chapter) to determine whether there are established variables that have a strong correlation with wood products production. Table 5.5 shows that the correlation between wood products employment and wood products tonnage produced is 75 percent. This indicates that wood products employment is a good predictor of wood products tonnage at the regional level. It is therefore reasonable to assume that it also is a good predictor at the county level.

The next step is to identify county-level employment for the three counties in the MPO region. The MPO first researched the U.S. Census Bureau web site, but found that employment data in all three counties were suppressed due to proprietary concerns. However, the MPO was able to secure county-level employment data from its state department of commerce by industry that showed 2007 employment for the three counties in wood products was 1,000, 2,000, and 3,000 for Aspen, Birch, and Cedar counties, respectively.

Based on this information, the MPO performed a regression with FAF state-level wood products tonnage as the dependent variable and state-level wood products employment as the independent variable. For simplicity, it is assumed that this regression equation is calculated as the following:

Wood Products Tonnage = 10,000 tons * Wood Products Employment

This regression equation can then be applied to the county-level employment data such that a table can be developed (see Table 5.7).

The step shown above is the core of the use of regression in this estimation process. The remaining work needed to develop a commodity flow database is to determine the origin-destination patterns of the wood products that are produced along with the modes used. This can be done using one of several different techniques that are discussed Chapters 2 through 4. These options include the following (generally sorted from the least resource-intensive option to most resource-intensive option):

- Leverage FAF origin-destination patterns for the state where the MPO resides.
- Interview local industry experts and adopt their estimates (see Chapter 4 for details on this method).
- Conduct roadside truck surveys on the main roadways used by trucks to exit the MPO (see Chapter 3 for details on this method).
- Conduct an establishment survey of the wood products manufacturing companies in the MPO (see Chapter 2 for details on this method).

It should be noted that a similar process can be used to estimate wood products entering the MPO region. Regression equations can be developed for attractions as well as productions as

Table 5.7. Estimating wood products tonnage output for hypothetical MPO.

County	Wood Products Tonnage Employment	Estimated Wood Products Tonnage Output
Aspen	1,000	10,000,000 tons
Birch	2,000	20,000,000 tons
Cedar	3,000	30,000,000 tons

Table 5.8. Initial origin-destination matrix.

		Destinations				Marginal Row Totals
		Lewis	Clark	Adams	Lincoln	
Origins	Lewis	1	1	1	1	34
	Clark	1	1	1	1	63
	Adams	1	1	1	1	18
	Lincoln	1	1	1	1	26
Marginal Column Totals		23	41	62	15	141

Note: Cells with a light diagonal background are values to be estimated and are initially set to 1. "Marginal Row Totals" and "Marginal Column Totals" (light shading) are the known origin and destination totals for each county.

demonstrated in a report prepared for FHWA (Cambridge Systematics 2009). Additionally, input-output analysis can be used to determine the full range of products consumed by the wood products industry as a first step in the development of a full supply chain for the wood products industry within the MPO.

Example 3—Using Iterative Proportional Fitting: Limited Prior Information

This section provides an example of how commodity flow data may be disaggregated where limited or no prior information is available using the iterative proportional fitting approach. Table 5.8 shows a four-by-four matrix of hypothetical origins and destinations corresponding to Lewis, Clark, Adams, and Lincoln. These locations could be counties, states, cities, or any other geographical points or regions where aggregate information from a national source is available, but information on the distribution of flows or shipments between origins and destinations is unavailable.

In the initial matrix (see Table 5.8), the lightly shaded areas (the column titled "Marginal Row Totals" and the row titled "Marginal Column Totals") are known and correspond to the previously presented equations:

$$x_{i+} = \sum_j x_{ij} \text{ and } x_{+j} = \sum_i x_{ij} \qquad \text{(Eq. 12, 13)}$$

The cells with a light diagonal background in the center of Table 5.8 represent information that is sought. Initially, when there is no information, all cell values are equal to "1." Then, beginning with the first iteration of factor adjustments, the row values for each cell are divided based on the proportional total for each cell out of the row total. In the row adjustment table (see Table 5.9), this corresponds to 8.50, 15.75, 4.50, and 6.50 for all Lewis, Clark, Adams, and Lincoln destinations, respectively.

Table 5.9. First iteration—row adjustment.

		Destinations				Marginal Row Totals
		Lewis	Clark	Adams	Lincoln	
Origins	Lewis	8.50	8.50	8.50	8.50	34.00
	Clark	15.75	15.75	15.75	15.75	63.00
	Adams	4.50	4.50	4.50	4.50	18.00
	Lincoln	6.50	6.50	6.50	6.50	26.00
Marginal Column Totals		35.25	35.25	35.25	35.25	141.00

Table 5.10. First iteration—column adjustment.

		Destinations				Marginal Row Totals
		Lewis	Clark	Adams	Lincoln	
Origins	Lewis	5.55	9.89	14.95	3.62	34.00
	Clark	10.28	18.32	27.70	6.70	63.00
	Adams	2.94	5.23	7.91	1.91	18.00
	Lincoln	4.24	7.56	11.43	2.77	26.00
Marginal Column Totals		23.00	41.00	62.00	15.00	141.00

Then a proportional column adjustment occurs based on the new cell values derived from the previous row adjustment and relative to each column total. Thus, in order to arrive at the 5.55 value in the first row and column cell of Table 5.10, the row-adjusted value of 8.50 is divided by the column sum (35.25) (see Table 5.9) and then multiplied by the initial column total of 23 (see Table 5.8). The resulting matrix after the column adjustment, now lightly shaded, represents the estimated flows from each origin and to each destination, yet still summing to the column and row totals. This represents one complete iteration. No further iterations are necessary since they will result in identical cell values. The final origin-destination matrix is shown in Table 5.10.

The iterative proportioning process has been automated within several software systems, including recent travel demand modeling software, so this process can be fully captured within the current set of travel demand models available to most MPOs and state DOTs.

User's Guide Worksheet Punch List

- Select commodity flow data disaggregation technique.
- Implement commodity flow data disaggregation technique.

5.3 Next Steps

This chapter describes the data disaggregation process. It illustrates that data disaggregation can be done using basic or very complex techniques. The level of complexity will depend on the precision needed in the answer to the question being asked by the transportation agency along with available data and resources available within the transportation agency. Disaggregation can be considerably powerful because it develops comprehensive data that can cover a wide range of modes, geographies, and commodities within a single analysis. This process can be combined with the supplemental data collection/assembly techniques described in Chapters 2.0, 3.0, and 4.0 to develop subnational commodity flow data that are comprehensive and accurate for industries and geographies of the most importance to local transportation agencies.

Refer back to the Playbook section to identify the next portion of the *Guidebook* that will be most relevant to where your transportation agency is in the data collection process.

CHAPTER 6.0

Playbook

6.1 Introduction to the Playbook

This section of the *Guidebook* is designed to help users develop a structured approach to diagnosing their commodity flow data needs and figuring out what methods are best applied to their freight transportation planning problem. This section is called the "Playbook." Football terminology is used here because this section allows the users to construct a "game plan" from among the many "plays" (or methods) that they have available to them. The game plan will be tailored to users' specific circumstances and the problems they are trying to solve. In many cases, users will want to combine multiple methods to take advantage of the strengths and compensate for the weaknesses of individual methods. The Playbook helps guide users through diagnosing their problem and data needs, identifying their resources and skills, and exploring their relationship with the private-sector freight community (the source of much potentially useful data) in order to help craft the right game plan for their situation. The Playbook provides some real-life examples of problems that planners are dealing with that require commodity flow data and uses these examples to illustrate the thought process involved in developing an approach and selecting the appropriate commodity flow data development methods.

6.2 Getting Started—What Should Be Known About Problems and Resources

It cannot be stressed strongly enough that in order to apply the major techniques for developing commodity flow data properly, the user needs a specific question to be answered. The more specific the question, the more likely the right sequence and techniques will be applied. It is recommended that prior to moving forward the user and any stakeholders deemed important to the outcome of the process spend time explicitly developing the key questions to be answered.

Another important dimension is the designation of the perspective for the user. For example, if the user is involved in issues at the national level, the primary source of commodity flow information, the CFS, provides answers for those questions on the commodities covered in the survey. The sampling strategy used was designed to make estimates at the national level sufficient for planning applications. FAF also is designed for federal questions, with accompanying visualizations. Additional strategies are needed when the questions needing to be answered involve a commodity not included in the survey. Reaching out to trade associations and other major information providers, as described in Chapter 4, could provide sufficient answers.

If the user is asking questions regarding a multistate regional problem, CFS data from the states in the region could be aggregated, or FAF regional geographies could be sufficient to answer the questions. Careful observation of the geographies used for special aggregations of data sets needs to be reconciled to the boundaries of the region of concern. State-level users can directly apply the state totals from CFS and FAF and other state-level data sets. However, as previously noted, if the desired commodity is outside of the scope of these data sets, another approach is needed, which may include locating trade association and other industry-specific sources.

Only a small subset of MPO areas is included in CFS data. All others are part of a "rest of state" total, which then requires use of one of the disaggregating techniques described in Chapter 5.0 to attempt to localize commodity information. This also is the case for any attempts to use CFS data for local uses. Many users of the *Guidebook* may be able to use relatively low-cost techniques to disaggregate CFS or FAF data for an MPO region that is part of the "rest of state" total to get an initial sense of key commodities and trading partners so that data collection can focus on industries and commodities that are of most interest. Since new primary data collection can be very expensive, particularly if the application of the data requires very geographically disaggregate information, using disaggregated CFS or FAF data to target the primary data collection on specific industries and commodities is a good way of combining inexpensive disaggregation techniques with primary data collection.

Many users of freight data also report the need to validate or improve upon national data sets, such as TRANSEARCH. Since TRANSEARCH can be purchased with county-level or even zip-code-level geographic detail, it is an attractive option for many states and MPOs who need this level of detail. But users often ask how accurate the data are at this level of geographic disaggregation. Using information from Chapters 4.0 and 5.0 of the *Guidebook,* users should be able to identify local economic data sets and economic input-output model data that will give them an idea of the relative value of shipments of different types of commodities for their region as a whole. These data can be compared to TRANSEARCH data to identify apparent anomalies. Then, more focused surveys of the industries that produce and consume these commodities can be conducted using the establishment survey techniques described in Chapter 2.0 as a way of improving upon the TRANSEARCH data. These are just a few ways that the different methods described in the *Guidebook* can be combined once users clearly define the question they are trying to answer and the specific data that will be important.

The time available to answer a question affects the amount of attention it will be given and the depth of the investigation. In emergency situations, often the easiest way to obtain data is used, while special studies can be conducted if sufficient time is available to plan and execute such a study.

After determining the perspective to be used, the next understanding necessary is the skill set of the work force tasked with answering the question. This includes previous experience with CFS, FAF, and other freight data sources. It includes relationships with freight community members. Some users have met formally with freight community members for years (e.g., Seattle Freight Roundtable), while other users have never previously included freight community members in their transportation planning tasks. In cases where there is extensive trust, it is more likely that data sources can be shared and analysis can be reviewed for reasonableness both by public-sector and private-sector individuals. More than any other area of transportation planning and analysis, freight issues are best addressed with a dual approach—public- and private-sector data partnerships. At the same time, there are many analyses that can be conducted by the public sector whether the private sector is involved or not. Including the freight community could help with efficiency of effort.

To summarize the key points of this section, to effectively use the *Guidebook,* it is important to know the following:

- What needs to be learned? (explicit description with specific questions).
- Who is asking for this knowledge? (description of the agency/firm, including perspective and geographic constraints).
- What resources are available? (e.g., available data and staff skill set with these resources)

6.3 Considerations in Answering Sample Freight Questions

This section of the Playbook provides several examples of problems that require different types of subnational commodity flow data. Each example provides a description of the problem, why commodity flow data are needed, and what types of commodity flow data are needed. This section also presents an approach to using different methods, often in combination, to address these needs. References to relevant *Guidebook* chapters appear in bold.

After the problems are introduced in this section, the next sections of the Playbook are set up in more of a workbook format, first describing a self-assessment process that can be used to determine the types of methods that will generally work for a particular problem, followed by some sample problems presented in a format that models how users can describe their own problem and prepare a game plan. After the examples are presented in workbook format, a blank Playbook worksheet and instructions are provided so that users can construct a game plan to attack their subnational commodity flow data problem.

6.3.1 Example #1. Analyzing the Market for a Short Sea Shipping Service

The situation in Example #1 is a group of states and MPOs on the West Coast of the United States who are interested in whether it would be possible to develop a short sea shipping service on the West Coast. The group wants to examine whether there is a sufficient market to support a service, which port pairs should be considered in the service, and what the potential impacts will be on truck and rail flows in some critical corridors.

- **Need for and types of subnational commodity data**
  - Identify the specific commodities that would be most likely to be carried by different types of short sea services (containerized, bulk, etc.).
  - Once the commodities to be carried are determined, identify the volume (tonnage and value) of commodities that will be carried between the port pairs that are under consideration as locations for the short sea services. The origins or final destinations of the commodities should be within drayage distance from their ports of departure and arrival (probably 100 to 150 miles), respectively, which could include multiple counties surrounding the ports (county-to-county flows).
  - Examine the modes that would be affected by the new short sea shipping service (i.e., determine the modal shifts). This will require modal details on the existing commodity flows.
- **Possible methods**
  - Use disaggregated FAF or TRANSEARCH data to get an initial idea of what commodities move between the states, or, in this case, a user can probably get metro area flows in total. Identify commodities of interest/or that are most plentiful **(see Chapter 5.0 for methods).**
  - Look to see whether there are local sources of data on particular commodities (like agricultural or mineral products) such as state government sources or trade associations that address the commodities of interest **(see Chapter 4.0 for methods).**
  - Once particular commodities of interest have been determined, do an establishment survey of shippers of these commodities within the counties of interest to get more accurate data on the flow volumes. Since this will be an expensive step, it may be prudent to wait until it has been determined from previous steps that there is a potentially viable market with sufficient volume.
  - To adequately understand the potential for shifting the commodity flows to the short sea mode, determine competitiveness by investigating the rates currently charged by mode. Such rates could be obtained through an establishment survey.
  - Conduct a special study to understand the operating costs of the short sea shipping service.

Some of these actions are sequential and others can be conducted simultaneously. The lead agency will need to structure a comprehensive approach that takes advantage of available strengths and mitigates weaknesses. This customization can rely on the methods in the *Guidebook,* but also will require a smart application of techniques that will work for a particular situation, with all of its ramifications.

6.3.2 Example #2. Developing an Advanced Freight Model

Another example of a real-world need is the development of inputs for, and calibration of, a supply chain freight transportation model in the Chicago metro area (or any other major intermodal hub and trans-shipment location). It is well-known that Chicago is home to a large number of logistics parks, intermodal terminals, and railyards. The region is building a mesoscale model (zones in the region are the scale of townships—smaller than counties but larger than TAZs) of supply chains that move products through the metro region. One of the applications for the model is to look at how various proposals for new logistics hubs and intermodal terminals will affect freight flows through the region, including potential changes in the patterns of rail and truck flows through specific corridors.

One approach to developing a supply chain model starts with commodity flows from a source like FAF that shows all the commodity flows that have an origin-destination in the region (including wholesale trade and third-party logistics transfers) and allocates these to specific industry supply chains (shipper-receiver industry pairs). Individual business decisions are modeled such that receivers can choose suppliers from a pool of available suppliers (controlled to the commodity flow totals between the industries and at the township level) and then logistics paths are chosen to maximize supply chain objectives (typically a combination of logistics cost minimization and some other service variable).

- **Need for and types of subnational commodity flow data**
  - The construction of the supply chains with shipper-receiver pairs requires information about commodity flows with origin-destination detail within the Chicago metro area that gets down to the township level with a high level of commodity detail. It also requires an understanding of the quantity of these commodity flows that move through wholesalers and other intermediate handling locations.
- **Possible methods**
  - FAF flows can be allocated to the county level using County Business Pattern employment data and to townships using Quarterly Census of Employment and Wages data **(see Chapter 5.0 for methods).** While this approach may not be as accurate as conducting a new establishment survey (**see Chapter 2.0 for methods**), it may be sufficient for high-level regional modeling.
  - Some industry sectors, particularly warehousing, are not well correlated with employment and will need additional investigation to allocate flows. This might require a special set of interviews or an establishment survey of particular types of firms (**see Chapter 2.0 for methods and Chapter 5.0 for discussion of the limitations of disaggregation techniques using employment data**).
  - In addition, since flows through intermediate handling facilities are not likely to be accurately represented in this allocation process, conduct specific establishment surveys of logistics parks and intermodal terminals (**see Chapter 2.0 for methods**).
  - After a model is developed and an initial estimate of township-level flows has been developed, identify the township zones with the highest flow values or commodities that represent the most significant share and conduct establishment surveys focused on these geographic areas or industries. Also, drill down on rail flows using the STB Waybill data (**see Chapter 2.0 for methods**).

This example suggests the use of a prioritization process that focuses the most attention on the places experiencing the highest flows or on commodities representing the most significant levels of activity. If there are other local considerations (e.g., development questions for a particular property), additional investigations could be required, including detailed site-selection elements considered by the real estate industry (e.g., direct access to highways for a particular site).

6.3.3 Example #3. Understanding Development Potential for New Distribution Centers and Supporting Infrastructure Needs

While similar in outcome to Example #2, the development of new distribution centers and supporting infrastructure as economic development projects for the Phoenix and Reno areas is actually more complex as it involves two states. Arizona and Nevada know that there are major national and regional distribution centers in California (Northern and Southern), but that space for expansion is getting difficult to find. The state DOTs in partnership with economic development agencies are interested in looking at the potential for developing logistics parks outside of Phoenix and Reno and want to know how much supporting infrastructure investment might be needed.

- **Need for and types of subnational commodity flow data**
  - It will be important to know current and projected flow volumes of through traffic related to product distribution on I-80 through Reno and I-10 through Phoenix.
  - In addition, it will be important to establish the nature of competitive costs between these places and in relation to other places.
  - Depending on the timeframe, a full supply chain analysis of assembly, processing, and distribution of the candidate industries needs to be compiled.
- **Possible methods**
  - CFS is unlikely to provide any information on the quantity of the commodity flows related to distribution, and, while FAF may include these flows, specifically identifying the distribution center flows in FAF is not possible. Conducting a roadside intercept survey along I-80 or I-10 would allow for collection of this type of information (**see Chapter 3.0 for methods**).
  - Once the major commodities and volumes and origin and destination are identified from the roadside survey, FAF commodity flow data assigned to the road network could be calibrated to these flows, and then the FAF forecasts could be used to project the potential market (**see Chapter 5.0 for methods**).
  - To forecast distribution center demand, the growth rate of several surrogate variables can be considered including employment within the warehouse or wholesale industries, projected growth of industries most likely to use distribution centers, or overall growth of end markets targeted by distribution centers.

Again, if specific local questions need to be addressed, communications with the real estate industry should be pursued to clarify issues with site selection.

6.3.4 Examples #4 and #5. Evaluating Modal Diversion Potential

A real-world example of an attempt to understand the potential for a modal diversion from truck to rail concerns I-81 in Virginia. This example is similar to one involving the assessment of the modal diversion potential for a new Cross Harbor Rail Tunnel between New Jersey and Long Island. In the case of I-81, there already is a recognition that congestion and truck volumes are growing rapidly. The Virginia DOT (VDOT) is interested in working with the railroads to determine the effect of a potential diversion from truck to rail if the state works in a public-private partnership to improve north-south rail service in the corridor.

- **Need for and types of subnational commodity flow data**
  - Would like to know current and projected volumes of commodities that are potentially shippable by rail that are now shipped by truck due to lack of good rail service.
  - Need to identify the commodities that are shipped long distances on I-81 by truck now and develop a method for forecasting. Need to know rates and costs and capacity of alternative modes.
- **Possible methods**
  - Use previous research to identify the appropriate commodities and characteristics of these commodities relative to the performance of each mode and to appropriate distances. Conduct a roadside survey to determine current volumes on I-81 (**see Chapter 3.0 for methods**).
  - Identify FAF flows assigned to the road network and then calibrate these to the roadside data. Use FAF growth forecasts to project the flows (**see Chapter 5.0 for methods**).

All of these examples have certain similarities in terms of how the user should address the problem:

- Combinations of methods are very helpful when solving subnational commodity flow problems because they allow the user to work with methods on a continuum from very detailed primary data collection to less intensive techniques that use what is readily at hand.
- The process involves defining the problem by thinking about which industries and commodities are going to be most important, whether the problem covers a large area or is focused along specific corridors, and understanding which modes are important. Armed with this information, any primary data collection can be focused on specific industries, geographic areas, and modal flows to reduce the size of the sample needed and the overall cost of primary data collection.
- The process often involves manipulation/disaggregation of existing data sets as a starting place to help focus primary data collection. Always try to determine whether it is possible to answer part of a question with data at hand.

6.4 Self-Diagnosis of Available Freight Resources

Below is a set of questions that need to be answered to help sort out the various dimensions that would call for different approaches. These questions help define the problem, the users' level of experience, and their resources with respect to the dimensions that will be important as methods are selected. Your responses to these questions can be compared to the questions that you are trying to answer to identify which portions of the recommended analysis can be conducted immediately, which portions require new data or training, and which portions might be suitable candidates to request outside support.

USER IDENTIFICATION STRATEGY

1. Problem identification: Describe the problem as specifically as possible, with a summary of the desired outcome(s) and three specific questions that need to be answered:

____

____

____

____?

____?

____?

2. User classification:
   a. Your agency is:
      i. ☐ Federal ☐ State
         ☐ Regional (multi-MPO) ☐ Multicounty MPO
         ☐ Single County MPO ☐ Local
   b. You need the answer(s) for your problem(s):
      i. ☐ Immediately (emergency) ☐ 3 weeks ☐ 6 months
         ☐ 1 year ☐ 1 to 3 years ☐ 3 to 5 years
   c. Staff resources available:
      i. ☐ None ☐ Part-time 1 person
         ☐ Full-time 1 person ☐ More than 1 full-time person
   d. Previous staff experience with similar problem(s):
      i. ☐ None ☐ Some experience
         ☐ Sufficient experience ☐ Extensive experience
         ☐ All such work would be conducted on contract
   e. Previous experience with freight community:
      i. ☐ None ☐ Web site ☐ Listserv
         ☐ Informal outreach ☐ Formal outreach ☐ Freight task force
         ☐ Permanent freight “roundtable” ☐ Private-sector initiative
   f. Previous staff experience with freight resources (**N**one, **L**imited, **S**ufficient, **E**xtensive):

i. National data:				
a. CFS	☐ N	☐ L	☐ S	☐ E
b. FAF	☐ N	☐ L	☐ S	☐ E
c. CBP	☐ N	☐ L	☐ S	☐ E
ii. Commercial data:				
a. TRANSEARCH	☐ N	☐ L	☐ S	☐ E
b. InfoUSA	☐ N	☐ L	☐ S	☐ E
c. Dun & Bradstreet	☐ N	☐ L	☐ S	☐ E
iii. State data:				
a. State model	☐ N	☐ L	☐ S	☐ E
iv. Relevant MPO data:				
a. MPO model	☐ N	☐ L	☐ S	☐ E
v. Local data:				
a. Geocoded tax parcels	☐ N	☐ L	☐ S	☐ E
vi. Freight studies:	☐ N	☐ L	☐ S	☐ E
vii. Freight plans:	☐ N	☐ L	☐ S	☐ E

Note that if your agency has little experience or staff availability to address key elements of this self-assessment, and these elements are directly related to the questions that you are trying to answer, then consideration should be given to hiring outside assistance.

3. Specific circumstances for this problem(s):
   a. This problem is located in a
      i. ☐ Broad multiregional area
         ☐ Single region
         ☐ Corridor within a multiregion
         ☐ Corridor entirely within the region
   b. Previous staff experience with this specific problem: ☐ N ☐ L ☐ S ☐ E
   c. Commodity Flow Survey (CFS) location designation:
      i. ☐ CFS metro area ☐ Rest of state ☐ Don’t yet know
   d. Does this problem involve interregional or interzonal trade flows without regard to specific infrastructure being used? ☐ Y ☐ N

e. Mode(s) involvement:
   i. ☐ Rail ☐ Truck ☐ Rail/Truck ☐ Water
   ☐ Water/Rail ☐ Water/Rail/Truck ☐ Air ☐ Air/Truck
f. Is there a possibility for "competition/diversion" between two or more of these modes? ☐ Y ☐ N
g. If "Yes," are the reasons for the competition/diversion documented? ☐ Y ☐ N
h. If "Yes," are there specific parameters to measure this competition/diversion?
   ☐ Y ☐ N
i. If "Yes," what is the source of the parameters? ______________________________
j. Number of industries involved:
   i. ☐ Single industry ☐ 2–5 industries
   ☐ 5–10 industries ☐ More than 10 industries
   ☐ Don't yet know________
k. Number of commodities involved:
   i. ☐ Single commodity ☐ 2–5 commodities
   ☐ 5–10 commodities ☐ More than 10 commodities
   ☐ Don't yet know

Note that if the nature of your problem is beyond your jurisdiction, then interagency partnerships should be considered. Additionally, if staff is unfamiliar with the data, industry, or modal requirements of the problem, then consideration should be given to hiring outside assistance.

6.5 Examples of Preparing a Game Plan to Answer Freight Questions

This section of the Playbook provides some generic examples of ways to address specific freight issues. These examples are for illustrative purposes and should not be taken as direct instruction for any one freight problem. Customizing an appropriate approach to any freight issue is recommended in all cases. Freight issues create a need for some specialized information to address particular circumstances in an appropriate manner. As a result, there are no "one-size-fits-all" solutions. Using the User Identification form helps identify elements for consideration, but not all the various classifications come into play, depending on the situation.

The sample problems presented in this section are offered in a worksheet format so that users can see illustrations of how to set up a problem and determine the type of data approach that may work for them. While the *Guidebook* is designed as a comprehensive collection of methods and procedures for collecting and using commodity flow data, the Playbook provides structured steps to help users choose the right approach and evaluate whether any particular strategy is appropriate for the purposes desired. At the end of the section, a worksheet is provided with instructions that will help users in setting up a game plan for the commodity flow data problems that they face.

The steps for developing a game plan are as follows:

- **Determine the circumstances generating the concerns.** Describe the background and context for the concerns. The Playbook samples begin with a short description of the players or partners and the circumstance that is generating the need for subnational commodity flow information.

- **Formulate question(s).** Questions need to be formally written so that everyone knows the level of specificity required and the limitations of this particular process for dealing with a freight issue.
- **Identify stakeholders.** Which agencies or organizations should be involved in the process? Look specifically at public-sector players and private-sector players particularly helpful for freight issues. The "business" of freight is primarily a private-sector function, while the provision of infrastructure and operations decisions often fall to the public sector.
- **Lay out an action plan**, including identifying potential sources of data, reviewing the relevant portions of the methods chapters in the *Guidebook,* and following the appropriate steps.
- **Reflect on outcomes and additional strategies** to confirm that the data collection/analysis strategy will be sufficient to answer the questions asked, identify additional uses of the work effort, and bolster the overall likelihood of success.

6.5.1 Playbook Sample Problem #1—Short Sea Shipping

Determine the circumstances generating the concerns.

A group of states and MPOs on the West Coast of the United States are interested in determining whether it would be possible to develop a short sea shipping service on the West Coast.

Formulate question(s).

- Is there a sufficient market to support a service?
- Which port pairs should be considered in the service?
- What would be the potential impacts on truck and rail flows in some critical corridors?

Identify stakeholders.

- Public Sector: State Departments of Transportation, MPOs, port authorities, port employee unions.
- Private Sector: Shipping lines; terminal operators; regional, local, national, and international shippers.

Lay out an action plan.

One task of the action plan is to identify specific commodities that would most likely be carried by types of short sea services (containerized, bulk, etc.).

To identify sources of relevant national data:

- **Review Chapter 5.0** to determine whether FAF or TRANSEARCH can be used to identify a set of commodities moving between the states/MPO regions considered as origin and destination pairs.
- Review User Identification Strategy form for level of expertise with FAF or TRANSEARCH data.
- Learn more about FAF and TRANSEARCH at the following links:

http://www.ops.fhwa.dot.gov/freight/freight_analysis/faf/
http://www.TRANSEARCH.com/

To identify sources of relevant local data:

- **Review Chapter 4.0** to determine whether there are local sources of data on these commodities such as state government agencies (e.g., agricultural or mineral products) or trade associations.
- Review User Identification Strategy form for level of expertise with local data sources.

The next task is to estimate the volume (both in tons and value) of these commodities that currently move between possible port pairs (with a drayage limitation of 100 to 150 miles) using county-to-county flows. The action items for this task are the following:

- **Review *Guidebook* Chapter 5.0** to determine whether FAF or TRANSEARCH can be used to quantify the tons and values moving between the states/MPO regions considered as origin and destination pairs.
- Extract disaggregated data for the chosen set of commodities moving between origin and destination pairs at the county level, if possible.
- Identify a set of port origin-destination pairs (using the 100 to 150 mile drayage limitation) for multiple counties. Calculate the tons and values moving between this set of port pairs.

The final task is to identify which modes local firms currently are using to move these commodities (e.g., truck or rail). The section in the *Guidebook* that can be used to fill this need is **Chapter 2.0**. Within this chapter the key elements are the following:

- Follow instructions in Steps 1 through 6 to set the stage, making sure to be able to identify firms that move the commodities of interest.
- Conduct the survey using instructions in Step 7.
- Follow instructions for Steps 8, 9, and 10 to assemble the database.
- Query the database for modes used to move the commodities of interest.

Reflect on outcomes and additional strategies.

This analysis can assist in identifying the level of public support that is appropriate to encourage and develop short sea shipping facilities. The analysis should be able to generate estimates of tonnage potentials and industries impacted by the development of a short sea shipping service. Since this analysis may be of value to a number of jurisdictions, consider partnering with other agencies for this effort, if possible. This is particularly important when calculating the cost of conducting local surveys to understand the identified commodities.

6.5.2 Playbook Sample Problem #2—Enhancing MPO Modeling Capacity

Determine the circumstances generating the concerns.

A medium-sized MPO wants to enhance their freight modeling capacity using local freight data.

Formulate question(s).

- Can commodity flow data be gathered at the local level?
- Can the data be used to model commodities moving:
  - Within the MPO region?
  - To and from ports within the MPO region?
  - To and from outside of the MPO region?

Identify stakeholders.

- Public Sector: MPOs and ports.
- Private Sector: Local firms.

Lay out an action plan.

One element of the action plan is to identify a method for gathering local commodity flow data.

To identify sources of relevant local data:

- **Review and use Chapter 2.0.**
- Follow instructions in Steps 1 through 6 to set the stage for the survey.
- Conduct the survey using the instructions in Step 7.

The next element is to estimate the volume (both in tons and value) of these commodities that move within the MPO region. The action items for this element are

- **Using Chapter 2.0.**
- Following instructions for Steps 8, 9, and 10.
- Querying the database for commodities whose origins and destinations are within the MPO region.

The next element is to estimate the volume (both in tons and value) of these commodities that move to or from ports in the MPO region. The action item for this element is to query the database for commodities whose origins and destinations are the port locations within the MPO region.

The final element is to estimate the volume (both in tons and value) of these commodities that move to or from the area outside the MPO region. The action item for this element is to query the database for commodities with origins and destinations that are outside the MPO region.

Reflect on outcomes and additional strategies.

This process should provide greater confidence in the trucking component of a regional travel demand model. It will enable a more accurate depiction of truck generation and movement throughout the region and the impacts of pursuing various project and policy initiatives on a corridor-specific and regional basis. Private-sector cooperation is critical for the success of this effort. Encourage dialogue and enthusiasm where possible as this could increase response rates, making your database more valuable for understanding commodity movement.

6.5.3 Playbook Sample Problem #3—State DOT Examining Trade Flows in Portion of the State That Is Dependent on Goods Movement

Determine the circumstances generating the concerns.

A state DOT wants to better understand freight flows in a portion of the state and on specific corridors.

Formulate question(s).

- What commodities are moving on the major east-west highway in the southern end of the state?
- What commodities are moving on the major north-south highway in the southern end of the state?

Identify stakeholders.

- Public Sector: State DOT, local MPO(s), highway operators, and enforcement agencies such as state highway patrol.
- Private Sector: Truck drivers using major highway networks and shippers and receivers relying on local corridors to transport goods.

Lay out the action plan.

One element of the action plan is to identify a method for gathering state-level flow data on specific highway segments. To identify sources of relevant local data:

- **Review Chapter 3.0.**
- Determine the feasibility of conducting a roadside survey based on the issues identified in the chapter.

The next element is to determine which commodities are moving on the major east-west highway in the southern end of the state. The action items for this element are the following:

- Site selection
  - Using a GIS map, locate all truck weigh stations, rest areas, truck pull-out areas, and points of entry along the east-west highway network in the southern area of the state.
  - Determine the best sites on the map for stopping trucks for an intercept surveying effort.
- Questionnaire design
  - Identify the most important information to collect in the survey.
  - Make any necessary modifications to the sample survey instrument (see Figure 3.3 of the *Guidebook* for an example questionnaire).
  - Determine whether you want to use electronic or paper survey instruments.
- Selection of survey dates and times
  - Consider any seasonal behaviors in the state that might be important to consider in the surveying effort.
- Data collection
  - Follow Steps 4 through 9 in Chapter 3.0 of the *Guidebook* to complete data collection.

The final element is to determine which commodities are moving on the major north-south highway in the southern end of the state. The action items for this element are the following:

- Site selection
  - Using a GIS map, locate all truck weigh stations, rest areas, truck pull-out areas, and points of entry along the north-south highway network in the southern area of the state.
  - Determine the best sites on the map for stopping trucks for the intercept surveying effort.
- Selection of survey dates and times
  - Consider any seasonal behaviors in the state that might be important to consider in the surveying effort.
- Data collection
  - Follow Steps 4 through 9 in the *Guidebook* to complete data collection.
- Consider alternative modes in the region and/or corridor(s)
  - Identify parallel, non-highway transportation infrastructure that also moves goods in the corridor. Rail, marine, and air cargo flows—should they exist—are likely operated by, at most, a handful of operators. Therefore, it is likely that interviews with these non-highway modal operators will generate sufficient information to understand relevant flow activity.

Reflect on outcomes and additional strategies.

The outcome of this effort should include an understanding of how to budget and prioritize the limited resources allocated for truck survey data collection. Additionally, the information can be used to improve several planning activities including scenario development, project development, and diversion studies. The information might also identify key industries or origin-destination patterns for further study. Conducting a roadside surveying effort is labor- and resource-intensive. Consider partnering with other agencies if possible. To encourage cooperation, consider contacting local trucking associations and providing information about the surveying effort to local commercial establishments where trucking industry members might learn about the importance of participating in the program. It might also be important to include a set of pilot deployments prior to rolling out the entire surveying effort. Pilots will help identify any previously unidentified issues and provide an opportunity to "cure" these problems.

6.5.4 Playbook Sample Problem #4—Determination of Impacts of Intermodal Railyard Expansion

Determine the circumstances generating the concerns.

A large MPO is concerned about the impacts of an intermodal railyard expansion plan. Amongst several concerns, the MPO would like to confirm that its locally designated National Highway System (NHS) connectors to the yard have been appropriately identified.

Formulate the question(s).

- What will be the hourly truck volumes related to this planned expansion?
- What will be the peak daily truck volumes related to this planned expansion?

Identify stakeholders.

- Public Sector: MPO, county, city, local neighborhood, State DOT for road maintenance, FHWA for NHS designation.
- Private Sector: Railyard operator and shippers relying on rail service.

Lay out the action plan.

One element of the action plan is to identify a method for calculating truck volumes related to the railyard planning expansion. The action items for this element are the following:

- **Review Chapter 5.0.**
- Prepare an outreach plan to establish a working relation with the railyard operators and set up an appointment or telephone interview to gather the following information:
  - Expected number of containers to be moved when the facility is fully operational.
  - Percentage of these containers moved during the week.
  - Expected monthly fluctuations prior to expansion plans.
  - Stability of daily and hourly fluctuations throughout the year.
- Collect truck count data and calculate parameters:
  - Based on the information obtained from the railyard operator, collect information on truck movements outside the intermodal railyard gates (as closely as can be done safely) in a month with sufficient stability.
  - Assemble the collected data into hourly volumes.
  - Calculate peak monthly volumes using the total number of containers to be moved under full operations and multiplying it by the percent of the highest month to determine an estimate of the number of containers to be moved in peak month.
  - Calculate the average number of containers by week by dividing the number of containers in the peak month by 4.3 (52 weeks/12 months = 4.3). This will provide the impact of the railyard expansion on weekly truck traffic.
  - Review truck counts to identify differences by day of week and establish the peak day of the week.
  - Calculate the peak daily volumes by multiplying the average number of containers per week by the percentage of the highest typical day of the week.

The next element of the action plan is to identify a method for calculating peak daily truck volumes related to the railyard planning expansion. The action items for this element are the following:

- Calculate parameters:
  - Determine peak hourly flow from hourly data (see above).
  - Calculate the peak hourly flows by multiplying the daily peak by the percentage of the peak hour. This will provide the impact of the railyard expansion on peak hourly truck traffic.

The final element of the action plan is to share impacts as understood by the modeling staff with intermodal railyard operators to increase the opportunity for mitigating any adverse outcomes.

Reflect on outcomes and additional strategies.

The implementation of this example would be beneficial across several efforts, including MPO Transportation Investment Plan development, NHS connector updates, and community relations building between the railroad and the local community. Specifically, this effort would determine the need to expand a roadway based on near-term or long-term growth at the facility. It would also identify which industries and companies would most benefit from a railyard expansion. Finally, this effort could identify locations of potential truck-automobile conflict on the local roadway and point the way to needed operational or design improvements on local roads. This effort would require cooperation with railyard operating staff. Establishing a working relation with private-sector freight community members and having an opportunity for regular communications increases the likelihood that critical data elements could be obtained for calculating subnational commodity flows.

6.5.5 Playbook Sample Problem #5—State DOT Working with State Economic Development Agency to Enhance and Promote the Production of a Critical Commodity

Determine the circumstances generating the concerns.

A medium-sized state wants to know the transportation patterns of a particular commodity within the state.

Formulate question(s).

- How many truck trips on a typical day are generated between specific locations (origins and destinations) within the state for this commodity?

Identify stakeholders.

- Public Sector: State DOT.
- Private Sector: Representatives from the facilities involved in marketing, production, storage, distribution, and consumption of the commodity of interest.

Lay out the action plan

One element of the action plan is to identify data sources:

- **Review Chapter 4.0:**
  - Determine the classification code(s) associated with the commodity of interest and employment code(s) for industries involved in the production, storage, distribution, and consumption of the commodity.
- Gather available reports on the industries involved in the production, storage, distribution, and consumption of the commodity of interest.
- Contact and arrange a set of interviews with representatives from facilities involved in the marketing, production, storage, distribution, and consumption of the commodity of interest.
- Develop a schematic of how the various facilities relate to each other (e.g., raw materials, processors, storage, etc.).
- Develop a GIS map that includes the locations of the major facilities and depicts their relative capacities (e.g., small, medium, and large facilities).

The next element of the action plan is to simulate commodity flows:

- Identify the facilities in the facility network that are closest to each other (shortest travel time on the network).
- Match origins and destinations and define "preferred service" areas for identified facilities for production, storage, distribution, and consumption of the commodity of interest.

The final element of the action plan is to estimate the daily truck trips.

Reflect on outcomes and additional strategies.

This analysis will identify local industries that are most reliant on the transportation system. Additionally, it should identify the routes used by the target industries for state economic development agencies. The analysis can also be used as the starting point for identifying key future bottlenecks for targeted industries. Private-sector firms often have high-quality fleet information with large quantities of operations data. If possible, consider opening a dialogue with these firms to determine their willingness to share their operations data, provided these data can be sufficiently aggregated to limit exposure of firm-specific details.

6.5.6 Playbook Sample Problem #6—Development of a Truck Component to a COG Travel Demand Model

Determine the circumstances generating the concerns.

A COG wants to develop a truck component for its travel demand model.

Formulate question(s).

- What data resources and procedures are needed to develop an intercounty truck trip table?

Identify stakeholders.

- Public Sector: COG or large MPO.
- Private Sector: Consulting firm with county-to-county freight flow database.

Lay out the action plan.

One element of the action plan is to identify available data sources:

- **Review Chapter 5.0.**
- Prepare and execute the data assembly plan.
  - Determine the counties to participate in model development.
  - Obtain a commercially generated, disaggregated, county-level freight flow database.
  - Obtain a commercial (e.g., Dun & Bradstreet) or state database with zip-code-level employment locations, including three-digit NAICS codes with number of employees.
  - Download 2002 VIUS microdata (http://www.census.gov/svsd/www/vius/products.html).
  - Obtain estimates of farm acreage by zip code for participating counties.

The next element of the action plan is to calculate the average payload:

- Determine commodities in the county-to-county database.
- Use the VIUS data to determine average payload for each of these commodities.
- Create an average payload matrix.

The next element of the action plan is to allocate county-to-county commodity flows to zip codes:

- Using the number of employees by zip code for each county, allocate tonnage by county.
- Using farm acreage, allocate county-to-county tonnage for agricultural industries.

- Combine these two zip code allocations into two zip-code-to-county tables:
  - Tonnage originating in each zip code destined for each county.
  - Tonnage destined for each zip code originating in each county.

The next element of the action plan is to allocate the zip code tonnage to TAZs:

- Using employment data for TAZs (available from passenger travel models or statewide model staff).
- Combining areas with geographic overlap between zip codes and TAZs.

The final element is to apply the average payload matrix to the tonnage at the TAZ level; this creates a truck trip table for the intercounty model.

Reflect on additional strategies.

Developing the truck component to a regional travel demand model will assist in identifying truck-intensive locations and corridors. This, in turn, will improve the estimate of capacity constraints across the region and also identify locations of truck-auto conflict. It may also indicate locations and corridors that should be targeted for truck-friendly operational and design characteristics. Introducing truck trips into the travel demand modeling environment requires additional training for staff. Consider contacting agencies that have developed truck trip tables and incorporated them into a travel demand model.

6.6 Sample Game Plan Worksheet

Determine the circumstances generating the concerns.

(Try to be as brief as possible, capturing the critical components from the user classification information from Section 6.4.)

Formulate question(s).

- (Be sure to articulate exactly the problem to be solved and the necessary geographical scope of the data collection effort)
- (Break questions down to more easily identify what information is being sought.)

Identify stakeholders.

- Public Sector: (Think strategically about the appropriate stakeholders for the geography identified in the circumstances description and all the appropriate public partners that may need to be included)
- Private Sector: (Try to identify exactly which members of the freight community may be stakeholders for the geography identified in the circumstances description and where exactly in the problem solving strategies their assistance may be needed.)

Lay out the action plan.

Break down the approach to answering each question into as many specific elements as needed. Remember, what is not included most likely will not get completed.

Review the *Guidebook* chapters that seem to be most relevant to the circumstances and question type. For example, if it is necessary to understand the flow of commodities at the local level, research might begin with a review of the steps for conducting a local establishment survey in Chapter 3. Again, users should refer to their user classification components to determine the resources available to tackle such an undertaking. Check out Chapter 5 to see if using a disaggregated approach might be more within the realm of the agency or available consultancies. Consider using a “hybrid” approach in which some guidance from one chapter and some from another are used.

Reflect on additional strategies.

Always consider any additional suggestions or "lessons learned" from other sources.

6.7 Playbook Recap and Conclusions on Developing Subnational Commodity Flow Data

The Playbook section of the *Guidebook* was designed to help users structure a "game plan" for tackling commodity flow data problems by

- Understanding the nature of a problem and the types of data needed.
- Understanding available capabilities and resources.
- Following a systematic approach for using available data and less expensive techniques to help target new primary data collection methods.

The illustrative examples should help users understand how to structure their approach and how to usefully reference the more detailed and technical descriptions provided in Chapters 2.0 through 5.0 of the *Guidebook.* Some users start with the introduction to the *Guidebook* and then skip to the Playbook section. These users may benefit from going back and reading the methods sections, paying more attention to these more technical sections. These users may also want to click on the hyperlinks to specific national databases provided in these sections to learn more about what is available in these databases and how they can be used. Users may also want to review information describing the importance of freight data, freight performance measures, and freight planning in meeting MAP-21 rules and guidance.

There is no simple answer and/or approach that will satisfy all subnational commodity flow data needs. However, armed with the information provided in the *Guidebook,* users should be able to get the most out of the best practices available today in addressing their freight planning needs.

References

Cambridge Systematics, Inc. 2009. *Development of a Computerized Method to Subdivide the FAF2 Regional Commodity OD Data to County-Level OD Data,* prepared for FHWA, unpublished.

Cambridge Systematics, Inc. 2007. *Quick Response Freight Manual II.* FHWA-HOP-08-010. U.S. DOT, Washington, D.C. http://www.ops.fhwa.dot.gov/freight/publications/qrfm2/qrfm.pdf.

Deming, W. E., and F. F. Stephan. 1940. "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals Are Known." *Annals of Mathematical Statistics,* 11(4), 427–444.

FAF. (n.d.) *Report No. S6—Petroleum Products National Totals.* (http://ops.fhwa.dot.gov/freight/freight_analysis/faf/faf2_tech_document.htm) (Accessed June 24, 2013).

Holguín-Veras, J., and J. Ban (Rensselaer Polytechnic Institute). 2010. *Feasibility Study for Freight Data Collection.* University Transportation Research Center—The City College of New York and New York Metropolitan Transportation Council, New York, NY.

ICF International. 2007. *Review of Pipeline Utility Corridor Capacity and Distribution for Petroleum Fuels, Natural Gas and Biofuels in Southwest Washington.* Prepared for the State of Washington Energy Facility Site Evaluation Council. Olympia, WA.

Lee, H., and K. Viele. 2001. "Loglinear Models and Goodness-of-Fit Statistics for Train Waybill Data." *Journal of Transportation Statistics,* Vol. 4, No. 1, Paper 5. http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/journal_of_transportation_and_statistics/volume_04_number_01/paper_05/index.html.

Miller, R. E. and P. D. Blair. 2009. *Input-Output Analysis: Foundations and Extensions, 2nd ed.* Cambridge University Press, New York, NY.

Rhodes, S. S., M. Berndt, P. Bingham, J. Bryan, T. J. Cherrett, P. Plumeau, R. Weisbrod. 2012. *NCFRP Report 14: Guidebook for Understanding Urban Goods Movement.* Transportation Research Board of the National Academies, Washington, D.C.

Thompson, J. S., K. D. Yarbrough, M. D. Anderson, G. H. Harris, and K. Harrison. 2010. "Approach to Collecting Local Freight Information." *Transportation Research Record: The Journal of the Transportation Research Board, No. 2160,* Transportation Research Board of the National Academies, Washington, D.C., 162–168.

WSDOT. September 2008. *Washington Transportation Plan Update Freight Movement.* http://www.wsdot.wa.gov/NR/rdonlyres/67530525-3531-4552-A198-BA4255AADAA7/0/WTPSeptember_2008web.pdf (Accessed June 24, 2013).

WSDOT. February 2008. *Washington State Freight and Goods Transportation System (FGTS) 2007 Update.* http://www.wsdot.wa.gov/NR/rdonlyres/3CF227DF-C152-440D-AD0E-4BADDDB8C470/0/0208FGTS2007Final.pdf (Accessed June 24, 2013).

List of Acronyms

AADT	Average annual daily traffic
AST	Above-ground storage tank
BEA	Bureau of Economic Analysis
BTS	Bureau of Transportation Statistics
CBP	County Business Patterns
CFN	Commercial Fueling Network
CFS	Commodity Flow Survey (BTS)
COG	Council of Governments
CSA	Consolidated Statistical Area
CVISN	Commercial Vehicle Information Systems and Networks
DMS	Domestic
DOT	Department of transportation
ECY	Washington State Department of Ecology
EWITS	Eastern Washington Intermodal Transportation Study
FAF	Freight Analysis Framework
FGTS	Freight and Goods Transportation System
GIS	Geographic information system
GPS	Global positioning system
HS	Harmonized system
INEGI	Instituto Nacional de Estadística, Geografía Informática
ISIC	International Standard Industrial Classification System
ITMS	Intermodal Transportation Management System
LBCS	Land-based classification standards
MOU	Memorandum of understanding
MPO	Metropolitan planning organization
MSA	Metropolitan statistical area
NAICS	North American Industry Classification System
NFS	National Forest Service
NHS	National Highway System
NYMTC	New York Metropolitan Transportation Council
ORNL	Oak Ridge National Laboratory
PIERS	Port Import-Export Reporting Series
REMI	Regional Economic Models, Inc.
RFP	Request for proposals
SCTG	Standard Classification of Transported Goods
SFTA	Strategic Freight Transportation Analysis

SIC	Standard Industrial Classification
STB	Surface Transportation Board
STCC	Standard Transportation Commodity Code
TAZ	Traffic analysis zone
TEU	20-foot equivalent units
TRS	Township-range-section
TSA	Transportation Satellite Account
USDA	U.S. Department of Agriculture
UST	Underground storage tank
VDOT	Virginia DOT
VIUS	Vehicle Inventory and Use Survey
WCSC	Waterborne Commerce Statistical Center
WSDOT	Washington DOT

Abbreviations and acronyms used without definitions in TRB publications:

A4A	Airlines for America
AAAE	American Association of Airport Executives
AASHO	American Association of State Highway Officials
AASHTO	American Association of State Highway and Transportation Officials
ACI–NA	Airports Council International–North America
ACRP	Airport Cooperative Research Program
ADA	Americans with Disabilities Act
APTA	American Public Transportation Association
ASCE	American Society of Civil Engineers
ASME	American Society of Mechanical Engineers
ASTM	American Society for Testing and Materials
ATA	American Trucking Associations
CTAA	Community Transportation Association of America
CTBSSP	Commercial Truck and Bus Safety Synthesis Program
DHS	Department of Homeland Security
DOE	Department of Energy
EPA	Environmental Protection Agency
FAA	Federal Aviation Administration
FHWA	Federal Highway Administration
FMCSA	Federal Motor Carrier Safety Administration
FRA	Federal Railroad Administration
FTA	Federal Transit Administration
HMCRP	Hazardous Materials Cooperative Research Program
IEEE	Institute of Electrical and Electronics Engineers
ISTEA	Intermodal Surface Transportation Efficiency Act of 1991
ITE	Institute of Transportation Engineers
MAP-21	Moving Ahead for Progress in the 21st Century Act (2012)
NASA	National Aeronautics and Space Administration
NASAO	National Association of State Aviation Officials
NCFRP	National Cooperative Freight Research Program
NCHRP	National Cooperative Highway Research Program
NHTSA	National Highway Traffic Safety Administration
NTSB	National Transportation Safety Board
PHMSA	Pipeline and Hazardous Materials Safety Administration
RITA	Research and Innovative Technology Administration
SAE	Society of Automotive Engineers
SAFETEA-LU	Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (2005)
TCRP	Transit Cooperative Research Program
TEA-21	Transportation Equity Act for the 21st Century (1998)
TRB	Transportation Research Board
TSA	Transportation Security Administration
U.S.DOT	United States Department of Transportation