MEASURING COMMODITY FLOWS IN CANADA: OPTIONS AND CONSIDERATIONS

Georgiana Madar and Lawrence McKeown, Statistics Canada

Introduction

Transportation plays an instrumental role across the economy as a cost that must be incurred to complete almost any market transaction. As Winston (2013) asserts, transportation is so intertwined with almost every part of the economy that it is vital for government to continually assess system performance and to consider improvements.[1] As a trade-reliant nation with its population spread over a vast landscape, Canada is particularly dependent on a transportation system that is efficient, reliable, innovative, responsive to change, and resilient to disruptions.[2] It is evident that quality statistical information is required to assess the national transportation system and its ability to move both people and goods.

A Freight Analysis Framework (FAF) is a method for estimating commodity flows and related transportation activities by different modes over a network or system. While conventional, industry-based surveys of transportation provide data to meet national accounting needs, a FAF requires activity-based surveys to estimate commodity origin-destination flows. This study briefly reviews freight analysis methods and data before sketching out a possible Canadian FAF with consideration of geographic precision, commodity detail, modal characteristics and other factors. In the second part, the study assesses two possible approaches to gathering the necessary commodity flow data: A carrier-based and a shipper-based approach.

Review

According to the Transportation Research Board (TRB),[3] a FAF can be used for understanding domestic and international goods trade

patterns, exploring growth patterns in freight transportation, conducting economic analyses, observing traffic volumes and their network effects, and for analyzing impacts of transportation policies. Using commodity origin-destination (O-D) information benchmarked to a reference year, the FAF is then updated periodically with annual data in order to determine network capacity required in future years or under different assumptions. For example, a FAF can be used to determine if the transport network can handle the "surge capacity" of certain commodities (e.g. grain, potash, oil).

However, among the many active datasets maintained by Statistics Canada and Transport Canada, there are only limited data available to enable freight analyses for marine, rail or road. There is a variety of methods used to collect these data, ranging from establishment-based and commodity flow surveys to carrier-based and road side surveys (Table 1). While each method is useful for capturing some data, none is able on its own to satisfy all FAF data requirements.

Table 1: Types of Data Collected by Different Survey Instruments[4]

	Type of Data			
Type of Surveys	Flows at establishment	Goods by vehicle	Vehicle routing	Origin of vehicle trips
Establishment Survey	Y	N	N	S
Commodity Flow Survey	Y	S	N	S
Freight Operator Survey	N	U	U	S
Driver Survey	N	Y	U	Y
Roadside Interview Survey	N	Y	U	Y
Vehicle Observation Survey	S	S	N	N
Parking Survey	N	N	N	N
Trip Diary	N	Y	U	Y
GPS Tracking	N	N	Y	Y
Suppliers Survey	N	U	U	S
Service Providers Survey	N	U	U	S
Traffic Count	N	N	N	N

Y = Data is commonly collected; S = Data is sometimes collected;
U = Data could be collected but it is not commonly carried out; and
N = Data cannot be collected using this type of survey

There are five general categories of data required to construct and maintain a FAF: 1) Shipment O-D information, 2) Commodity characteristics, 3) Mode of transportation, 4) Routing and timing, and 5) Vehicle type and related information. An overview of data collection instruments for urban freight in selected European countries is summarized in Table 2.[5]

Table 2: Urban Freight Data Collection in Select European Countries

	Country										
Type of survey	Belgium	France	Germany	Hungary	Italy	Netherlands	Portugal	Spain	Sweden	Switzerland	United Kingdom
Commodity Flow Survey	Y	N	N	N	N	N	N	N	Y	Y	N
Goods Vehicle Activity Survey	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Shipper Survey	Y	Y	Y	N	Y	N	N	Y	N	Y	N
Receiver Survey	Y	Y	Y	N	Y	Y	N	Y	N	Y	Y
Port Freight Traffic Data in Urban Area	Y	Y	Y	N	N	Y	N	U	U	U	Y
Rail Freight Traffic Data in Urban Area	Y	U	Y	N	N	Y	N	U	U	U	Y
Inland Waterway Freight Traffic Data	N	Y	Y	N	N	Y	N	U	U	U	Y
Airport Freight Traffic Data in Urban Area	Y	Y	Y	N	N	Y	N	U	U	U	Y

Y = Freight data is collected; N = Freight data is not collected;
U = Uncertain

In this table, a Commodity Flow Survey refers to a carrier-based survey since two of the three countries (Belgium and Switzerland) that report conducting such a survey explicitly mention surveying carriers. Moreover, it also seems that this type of survey refers specifically to trucking. Establishment-based surveys are listed in the table as the shipper survey and its reciprocal the receiver survey.

In the United States (U.S.), the FAF is a Bureau of Transportation Statistics (BTS) managed program that provides estimates of total volumes and values of freight moved among and within regions in the United States.[6] Although the FAF relies on data from several sources, baseline data are provided by an establishment-based Commodity Flow Survey (CFS), undertaken every five years by the United States Census Bureau (USCB).

The CFS provides a set of freight flow matrices reported in weight and value of goods transported for a base year (i.e. 2007 and now 2012). The base year flow matrices are updated annually using data sources such as output and employment by industry. CFS data consist of shipment characteristics including modes of transportation from a sample of cross-economy business establishments. There are certain limitations however, such as excluding some economic sectors (e.g. crude oil extraction) as well as shipments of imports.

The genesis for both the American FAF and CFS was a committee tasked with studying transportation data needs in the early 1990s. It recommended the creation of an agency (i.e. the BTS) to compile and integrate system-wide transportation data for the Department of Transportation (DOT).[7] The committee further recommended that the first priority for this new agency should be the development of national passenger and commodity flow surveys.

The Transportation Association of Canada has examined the American freight data framework and concluded that Canada would benefit immensely from acquiring a national level FAF. [8] A more recent study of best practices in freight transportation analysis undertaken for Ontario proposed a sequential modelling system that would allocate commodity flow forecasts by mode and assign them to

networks.[9] The study noted that success of the modelling system would depend on data availability and recommended a strong federal role in gathering the required commodity flow information.

The commodity flow information necessary to develop a FAF can be obtained in a variety of ways. To help understand and determine this information need, it is necessary to identify current data available and limitations.

A Canadian Freight Analysis Framework

To assess data needs, it is useful to work backward by sketching out a plausible Canadian FAF. In general, such a framework would consist of a series of matrices to estimate commodity flows, both in terms of weight and value, by mode among a set of zones or districts within the country (Table 3).

Table 3: Possible Canadian Freight Analysis Framework

O/D	**ER_1**	**ER_2**	**...**		**ER_j**	**...**		**ER_{76}**	**EX**	$\sum$
ER_1										
ER_2										
⋮										
ER_i					F_{xy}^{ij}					
⋮										
ER_{76}										
IM										
$\sum$										$\sum\sum$

where F_{xy}^{ij} = *Flow (tonnes or $) of Commodity x by Mode y from Origin i to Destination j.*

Geography
The framework uses Canada's 76 Economic Regions (ER) as a starting point. The ER, comprised of complete Census Divisions (CDs), is small enough to permit regional analysis yet large enough to release a broad range of statistics. However, since Census Metropolitan Areas (CMA) consist of Census Sub-divisions (or municipalities), ERs may not capture a CMA in its entirety. While the ER is the geographic unit for collection and estimation, some ERs may have to be grouped or collapsed for analysis and dissemination.

Commodity
Key dimensions of freight flows are the weight and value of shipments by type of commodity. Again, the principal data source for freight flows in the American FAF is the CFS. In 2012, the CFS asked respondents to report using the 5-digit Standard Classification of Transported Goods (SCTG) but the FAF disseminates using the 2-digit SCTG. For 2017, there has been some consideration about the commodity classification since the Economic Census may be using a variant of the North American Product Classification System (NAPCS). In any event, it is imperative for a Canadian FAF to employ the same commodity classification.

Modal Characteristics
In the American FAF, shipments are assigned either a single (i.e. truck, private or for-hire; rail; deep sea or inland marine; pipeline) or a multi mode designation based on CFS data. Average distance, essential for estimating tonne-kilometres, is not reported but rather derived; a program – GeoMiler – is used in to calculate or impute the distance travelled by mode from shipper to customer based on data provided. Using carrier-based surveys, it would only be possible to collect modal-based segments of multi-modal trips.

Other Considerations
To assemble O-D commodity flows by mode, cross-economy data is required. A Shippers survey would target establishments that ship products, those from the Mining, Manufacturing, Wholesale, Retail, and Transportation and Warehousing industries. This approach excludes imports but efforts could be supplemented by international

trade data. Carrier-based surveys target establishments that transport products. This approach includes imports but excludes shipments undertaken by non-transportation establishments (i.e. private trucking).

With 5 modal categories (marine, rail, truck, air, multi / other) and 42 2-digit SCTG classifications, estimating flows of both value and weight would require data to populate roughly 420 (76 by 76) matrices (i.e. 42 x 5 x 2). Again, the principal data (i.e. O-D commodity flows by mode) used to calibrate this Canadian FAF can be collected from either the carriers or from the shippers. In the following section, the study begins to investigate how data could be gathered for a Canadian Freight Analysis Framework (CFAF) using these two approaches.

Carrier-Based FAF Data

Statistics Canada has collected OD commodity flows using carrier-based data. For the Rail industry (NAICS 483), a census of federally regulated carriers collects OD flows by tonnage and commodity using 64 SCTG groupings; a monthly survey collects the number of cars, tonnage of revenue freight and commodities based on 7-digit Standard Transportation Commodity Classification (STCC). Complete flow information is available from the Energy Statistics Program on pipelines (NAICS 486). However, there are no surveys for certain transport industries (e.g. NAICS 488 Support Services, 492 Couriers, and 493 Warehousing). [10]

For trucking, there is an annual Trucking Commodity Origin and Destination (TCOD) survey to collect freight movements from a sample of large establishments (> $1.3 million annual revenue) classified to the for-hire trucking industry (NAICS 484). Data collected includes tonnage, distance, commodity (5-digit SCTG), O-D, revenue, and number of shipments. TCOD only includes the trucking segment of a multi-modal movement and excludes shipments provided by ancillary units of non-trucking establishments (i.e. private trucking).

The question is whether data from these carrier-based programs can be used as input for a CFAF. A definite advantage is cost in that these surveys already exist. However, there are some technical challenges given incongruence in concepts, measures and classifications. Also, carrier-based surveys typically only collect weight of shipments and not value. However, work has been undertaken to assign value to commodity shipments with TCOD[11] and with rail data.

To study a carrier-based approach for FAF data, we began by assessing trucking commodity flows at a provincial level. Using 2012 TCOD, O-D commodity flows were extracted by weight for each of the 42 2-digit SCTG classifications among the 10 provinces and 3 territories. In order to assess the coverage, a value of 1 was assigned if an O-D pair had flows present, regardless of the volume, and a value of 0 was assigned to all pairings that did not record any commodity flows.

To summarize the 42 matrices (i.e. one for each SCTG code), the values for each O-D pair were summed over all of the SCTG classes (Table 4). The O-D pairs with the highest sum values are shaded darker indicating that the freight flows were observed for the O-D pair for a larger number of commodities, with 42 indicating all possible commodity flows among the O-D pairings.

As expected, values along the principal diagonal (i.e. intra-provincial flows) are relatively high as are imports to and exports from many provinces. Also as expected, Ontario has the highest coverage by origin (averaging 41 of 42 flows to all other provinces, excluding territories) and as a destination (averaging 39 of 42 flows). Conversely, the fewest commodity flows occur to and from the territories.

Table 4: Number of SCTG Commodities with Flows between Provinces

Origins		*Destinations*													Exports	**Total**
		10	11	12	13	24	35	46	47	48	59	60	61	62		
		NL	PE	NS	NB	QC	ON	MB	SK	AB	BC	YT	NT	NU		
10	NL	39	18	33	32	39	37	19	15	27	19	2	1	2	34	317
11	PE	23	34	32	35	29	34	14	13	18	19	1	0	0	29	281
12	NS	37	34	39	38	37	37	28	23	33	28	12	10	2	37	395
13	NB	37	35	36	39	38	40	27	25	29	25	10	1	1	40	383
24	QC	41	35	39	40	41	41	37	38	39	37	20	21	16	42	487
35	ON	42	36	42	41	42	42	41	39	42	41	24	31	12	42	517
46	MB	30	22	29	26	37	39	40	40	39	40	21	21	6	38	428
47	SK	16	11	21	20	33	37	40	42	41	36	11	7	4	39	358
48	AB	30	25	30	28	39	39	39	41	42	41	26	31	14	41	466
59	BC	30	24	35	30	38	41	37	39	41	42	30	26	11	41	465
60	YT	0	0	1	0	7	6	6	0	10	10	25	1	1	10	77
61	NT	9	1	8	1	11	8	2	1	10	5	2	10	1	10	79
62	NU	1	0	1	1	0	2	0	1	1	1	0	1	3	3	15
Imports		41	39	42	42	42	42	42	41	42	42	19	16	11	0	461
Total		376	314	388	373	433	445	372	358	414	386	203	177	84	406	

Source: Statistics Canada, TCOD Survey, 2012

The same exercise was then repeated at the ER level, creating a series of 76 by 76 commodity flow matrices. Another way to view coverage is by commodity. For each 2-digit SCTG, the proportion of for-hire trucking flows by economic region pairings are summarized in Table 5.

Certain commodity classes, such as SCTG 26, 34 and 40 (wood products, machinery, and miscellaneous manufactured products respectively), perform very well, with flows present among over one half (53%) of the ER-level O-D pairs. Other classes, such as SCTG 15 (coal) and SCTG 16 (crude petroleum), do not, since they are only produced in certain regions of the country and are not generally shipped long distances by truck.

Table 5: Proportions of ER O-D Pairs with Flows by Commodity Category

SCTG category	% with flows		SCTG category	% with flows
1	14%		22	11%
2	13%		23	35%
3	25%		24	47%
4	30%		25	15%
5	22%		26	53%
6	19%		27	23%
7	35%		28	37%
8	11%		29	32%
9	2%		30	38%
10	8%		31	32%
11	6%		32	30%
12	9%		33	53%
13	16%		34	53%
14	3%		35	41%
15	2%		36	45%
16	2%		37	17%
17	8%		38	27%
18	7%		39	35%
19	19%		40	53%
20	26%		41	17%
21	20%		42	49%

Source: Statistics Canada, TCOD Survey, 2012

With less complete coverage at this geography, it implies the need for supplementary data or for aggregation by either geography or commodity.

Shipper-Based FAF Data

Alternatively, estimates of commodity O-D flows by mode can be obtained directly from the shipper. To investigate this approach, we begin with the American CFS which collects data on type, O-D, value, weight, mode, distance and ton-miles of commodities shipped for establishments in selected industries (e.g. manufacturing). The CFS is collected every 5 years as part of the Economic Census. For the most recent cycle in 2012, a targeted advance survey (pre-canvass) of 99,828 establishments was conducted in 2011 to determine first, if certain establishments conduct shipping activity, and second, to obtain an accurate measure of their shipping activity.

For establishments selected into the CFS sample, a questionnaire is mailed for each of its four reporting weeks to obtain the requisite information on shipments originating in the U.S. The Economic Census is collected for the same reference year and results are used to calibrate the CFS sample weights to reflect the current U.S. industrial structure. Based on a review of the American FAF / CFS publications[12], ongoing discussions between Statistics Canada and Transport Canada, and consultations with the BTS and USCB, we provide some preliminary estimates and considerations for developing a CCFS.

Geography

In the U.S., a CFS area (n = 134) is defined as either a Metropolitan Area, based on population and its importance as a transportation gateway, the rest of state (ROS) or entire states. For a CCFS, a sampled establishment would be asked to provide the city, province and postal code for domestic shipments which would be coded by the Standard Geographic Classification (SGC). For both establishment locations and domestic shipments, origins and destinations can be coded into ERs using the SGC. For exports, the respondent would be asked to provide the state for U.S. destination and the country for others, along with the domestic port, airport or border crossing of exit. From this reported information, a suitable typology of destinations would be developed for exports.

Survey and Sample Design

In the U.S., for the 2012 CFS an advance survey (pre-canvass) was sent to target industries across the economy in order to construct a sample frame (> 760,000) of establishments with shipments. The Economic Census is used for the *post hoc* calibration of sample weights by type of establishment. For a CCFS, there would be three stages beginning with a Nature of Business Survey (NBS) for those establishments not part of a regular survey (e.g. NAICS 488) to ensure the Business Register (BR) is updated. Next, once the target sample by NAICS is finalized, a pre-canvass stage would be used to identify shipping activity and the correct location / respondent for reporting shipping activity. The final stage would be the survey collection itself.

In the U.S., the 2012 CFS had a three stage sample design. First, a sample size of roughly 100,000 establishments was based on size, NAICS and geography with three strata: normal (target industries), auxiliary (e.g. warehouses), and hazardous material. Second, each selected establishment was systematically assigned to four reporting weeks – one in each quarter. Third, for each reporting week, the respondent was asked to select a systematic sample of its outbound shipments, based on the total number of such shipments. Even with this large sample, there was a need for the BTS to add "statistical noise" to its O-D flow by mode estimates for specific CFS area pairings.[13]

In Canada with a more disperse geography and a linear trading pattern, even more concentration of commodity flows among the O-D pairings is anticipated. As such, a sample size of perhaps one fifth or 20,000 establishments is a reasonable starting proposition. However, the underlying Canadian economic structure is different from that of the U.S.[14] Only after industry coverage by type (NAICS) is finalized can methodological work commence to determine the sample size necessary to provide certain coverage a pre-determined by geography, commodity and mode.

Collection and Processing

For the American CFS, each establishment sampled was mailed a questionnaire for its four reporting weeks and asked to provide information about each selected shipment including value, weight, SCTG commodity, temperature control, UN/NA code for hazardous material, domestic shipment destination or gateway along with city and country of destination for export shipments, mode of transport, and use of rush delivery services. For 2012, roughly 60% of reported questionnaires were filed electronically. Even so, a staff of 20 worked full-time for one year to assist with respondent queries and for initial data grooming.

After several CFS cycles, the USCB has developed robust tolerance edit and imputation regimes applied during both collection and processing. For example, E&I programs identify incompatible commodity and mode as well as industry and commodity combinations for follow-up or statistical treatment (imputation). In addition, a team of BTS employees spent 18 months using a piece of custom software – *GeoMiler* – to assign distances and routes to the shipment data. Development of a CCFS would certainly learn and benefit from the BTS and USCB experiences to date.

Estimated Resources

In the U.S., cost estimates for the five year CFS cycle range from $25 to $30 million depending on what elements are included (e.g. FAF). These costs are shared by the BTS (80%) and the USCB (20%) as part of prior arrangements pursuant to the Economic Census program. In summary, the American CFS is the "Cadillac" of freight flow estimations, being able to collect all pertinent data dimensions - geographic, commodity and modal. However, it is excludes shipments of imports which are relatively more important in Canada.

At this point, a number of specific research and costing studies are needed for estimating CCFS costs. In order to align the CCFS with the 2017 American CFS, it would be necessary to pilot test an instrument in 2016. This requires a series of research and costing studies during fiscal year 2015-16 including, for example, cognitive testing of the content, questionnaire and sample design and BR

cleaning for selected industries. It is only after the results from these studies are analysed that a true cost estimate is possible.

Summary

In developing a framework for a national freight data program in the U.S, it was asserted that, while data alone cannot guarantee good transportation policy and investment decisions, informed choices are not possible without good data.[15] To inform policy and investment choices aimed at ensuring that the Canadian transportation system continues to have the capacity and resiliency to meet freight transport demand, a Canadian FAF is considered as a useful decision tool.

This study has identified two broad options, along with some of the key considerations, for obtaining the O-D commodity flows by mode estimates that serve as baseline data for such a framework. Finally and in the context of increasing availability of "big data", there is a need to investigate the possible contributions of administrative and other non-survey sources of information for the development of a Canadian FAF.

Endnotes

[1] Winston, C. (2013), On the performance of the U.S. transportation system: Caution ahead, *Journal of Economic Literature*, **51** (3) 773-824.

[2] Canada Transportation Act Review Discussion Paper, accessed September 15, 2014. (http://www.tc.gc.ca/media/documents/ctareview/ctar_discussion_paper_en.pdf).

[3] Schmitt, R. & Tang, T. 2006, "Freight Analysis Framework and the Commodity Flow Survey", Transportation Research Board Transportation Research Circular: Commodity Flow Survey Conference, No. E-C088, 140-143.

[4] Adapted from Allen, J., Browne, M., & Cherrett, T. (2012), Survey Techniques in Urban Freight Transport Studies, *Transport Reviews*. **32** (3), pp. 287-311.

[5] Browne, M. & Allen, J. (2004), Best Urban Freight Solutions II: Best Practice in data collection, modelling approaches and application fields for urban commercial transport models I, Urban freight data collection – synthesis report.

[6] Oak Ridge National Laboratory (2010), The Freight Analysis Framework, Version 3: Overview of the FAF3 National Freight Flow Tables. Report prepared for the Federal Highway Administration of the U.S. Department of Transportation.

[7] Transportation Research Board (1992). Data for Decisions: Requirements for National Transportation Policy Making, *Special Report 234*. Washington: National Research Council.

[8] Kriger, D., McCumber, M., & Mucsi, K. (2009). What Can We Do to Improve Urban Goods Movement Data Collection in Canada? Project on the Framework for the Collection of High Quality Data on Urban Goods Movement, *Annual Conference of the Transportation Association of Canada*. Vancouver.

[9] Parsons Brinkerhoff & MMM Group (2013). Best Practices in Freight Transportation Forecasting Modelling. Report prepared for the Ontario Ministry of Transportation.

[10] Statistics Canada's International Marine Origin Destination Commodity data program (SCTG-based) was terminated in 2012 as part of the federal Deficit Reduction Action Plan (DRAP).

[11] See Anderson, W. and M. Brown (2012). Trucking Across the Border: The Relative Cost of Cross-border and Domestic Trucking, 2004 to 2009. *Economic Analysis Research Paper Series*. Statistics Canada: 11F00027.

[12] See for example, Bureau of Transportation Statistics and United States Census Bureau (2015). 2012 Commodity Flow Survey (Draft). Washington: U.S. Departments of Transportation and Commerce (2015).

[13] In 2010, just 10% of the FAF-based trade corridors moved almost 80% of all goods in the U.S.; Tomer, A. and Kane, J. (2014). Mapping Freight: The Highly Concentrated Nature of Goods Trade in the United States. Washington: The Brookings Institute.

[14] The underlying structure also points to the relatively more important role of import shipments in the Canadian economy.

[15] Transportation Research Board (2003). A Concept for a National Freight Data Program, *Special Report 276*. Washington: National Academies.