1 Abstract
## Table of Contents

1 Abstract .......................................................................................................................... 2
2 Table of Contents ........................................................................................................... 3
3 Introduction .................................................................................................................... 4
4 Freight Transport Modelling .......................................................................................... 4
5 Freight Demand Models and Their Data Needs .............................................................. 4
5.1 Global Freight Transport Models ............................................................................... 5
5.2 Intercity Freight Transport Models ............................................................................ 6
5.2.1 At Aggregate Level ............................................................................................... 6
5.2.2 At Disaggregate Level ........................................................................................ 7
5.3 Urban Freight Transport Models ............................................................................... 8
6 Freight Data Sources Today ........................................................................................... 8
6.1 Australia ..................................................................................................................... 8
6.1.1 Freight Movement Survey .................................................................................... 8
6.1.2 Freightinfo ........................................................................................................... 9
6.1.3 Vehicle Movements and Fleet Characteristics ..................................................... 9
6.2 UK .............................................................................................................................. 10
6.2.1 Survey of Heavy Goods Vehicles ......................................................................... 10
6.2.2 The Transport of Goods by Road in Great Britain .............................................. 10
6.2.3 Waterborne Freight in the UK ............................................................................. 10
6.2.4 Focus on Freight .................................................................................................. 11
6.2.5 Road Goods Vehicles Travelling to Mainland Europe: Quarterly Bulletins .......... 11
6.2.6 The Origin and Destination Survey of UK International Trade ......................... 11
6.3 EC ............................................................................................................................ 11
6.3.1 The COST 310 Freight Transport Logistics ......................................................... 12
6.3.2 COST 321 Urban Goods Transport .................................................................... 12
6.4 USA ........................................................................................................................... 13
6.4.1 Commodity Flow Survey ..................................................................................... 13
6.4.2 Transborder Surface Freight Dataset .................................................................. 13
6.4.3 Motor Freight Transportation and Warehousing Survey .................................... 14
6.4.4 Vehicle Inventory and Use Survey ....................................................................... 14
6.4.5 TRANSEARCH and Freight Locater .................................................................. 15
7 Data Needs for Freight Modelling .................................................................................. 15
7.1 Data for analysing freight transport at global scale .................................................. 16
7.1.1 Economic Activity ............................................................................................... 16
7.1.2 Transport Modes and Infrastructure ................................................................... 16
7.1.3 Transport Services ............................................................................................. 17
7.1.4 International Trade Conditions ........................................................................... 17
7.2 Data for analysing freight transport at intercity scale ............................................... 17
7.2.1 Modal Split ......................................................................................................... 17
7.2.2 Fleet’s Attributes and Composition ..................................................................... 17
7.2.3 Network Characteristics ..................................................................................... 18
7.3 Data for analysing freight transport at urban scale ................................................... 18
7.3.1 Vehicle Fleet ....................................................................................................... 18
7.3.2 Vehicle Flows ..................................................................................................... 18
7.3.3 Commodity Flows ............................................................................................. 19
7.3.4 Major Freight Generators .................................................................................. 19
7.3.5 Major Freight Corridors ..................................................................................... 19
8 Key Issues in Freight Data Collection ............................................................................ 19
9 Information Technology and Freight Data Collection .................................................. 22
10 Final Remarks ............................................................................................................. 24
11 REFERENCES .............................................................................................................. 25
3 INTRODUCTION

Most of the analysis and evaluation of transport projects involve the determination of some kind of quantitative data that cannot be directly measured “on the field” but needs to be estimated with the aid of mathematical models. Such is the case of the estimation of equilibrium flows; understanding by “equilibrium” a system’s state in which the vehicle flows are such that enable compatibility of the level of service perceived by the transport users at all stages in the modelling process. In order to compute equilibrium flows, demand and supply (network) models are needed for all the modes and commodities participating in a given transport system. Both types of models are generally intensive in data use, though requiring different types (and quantity) of data and data processing methods. Demand models usually require a wider variety of data than their supply counterpart, mostly because they are essentially based on econometric techniques. For these reasons, the aim of this article is to review several aspects of transport data from the freight demand models (FTD) perspective.

4 FREIGHT TRANSPORT MODELLING

While passenger transport modelling has attained a certain degree of maturity, the freight counterpart is still in an earlier stage of development. The latter may be due to several factors such as lack of efficient methods and tools to solve large-scale problems, the difficulty of identifying the decision makers involved in the process, and the lack of relevant modelling data (sometimes as a result of the private sector agents opposition to disclosing information with competitive commercial value).

Freight transport is commonly measured and described using (loosely) two terms: commodity flows (CF), and vehicle flows (VF). CF are represented by an OD matrix and the focus is on type and quantity of moved goods vis-à-vis VF, represented by traffic flows in different modes, where the focus is on the vehicle and its operation. From the economics perspective, freight transport demand constitutes a derived demand, meaning that its existence is derived from the need to move goods between different points in space. Consequently, the primary focus of attention should be the CF as they are motivated by goods consumption by the community, representing the actual demand. Nevertheless, VM are the result of logistics decisions made by the carriers and therefore they are useful in identifying the assignment paradigms needed for the supply models. Thus, both types flows are needed for FTD as they will be used for different purposes within the modelling framework.

5 FREIGHT DEMAND MODELS AND THEIR DATA NEEDS

Regan and Garrido (2000) classify freight transport models by spatial considerations into three broad categories: global (international), intercity, and urban. Therefore, the FTD models (and data needs) will be presented accordingly.
5.1 Global Freight Transport Models
At this level, the aim is to model the goods movement between different countries. This type of transport has experienced an explosive growth in the last decade due to the operations of multinational firms which operate (logistically) dispersed all over the world taking advantage of competitive prices for both materials and labour. The goods components are manufactured in different locations and hence need to be transported to a certain location to be assembled and shipped abroad again.

Haralambides and Veenstra (1998) identify three main approaches to model global demand for shipping. The first approach follows the standard theory of international trade (see Cassing, 1978). The second approach relies on an aggregate cost function for a given industrial sector, from which a demand function for shipping is derived (see Friedlander and Spady, 1980; Oum, 1979). The demand function is such that minimises the cost function. This approach allows to work with an analytical expression for the demand function. Nevertheless, this approach has two main drawbacks: the data requirement and the computational complexity of the solving process (a non-linear multidimensional minimization). Indeed, the cost function for each sector requires an enormous amount of data, which are often times private and not disclosed due to its potential strategic impact. For instance, Friedlander and Spady (1980) model uses transport rates and shipment characteristics for 96 three-digit manufacturing industries, for five geographic regions, production, materials and capital costs, shares of the cost attributable to each mode, capital stock, average wage rates, average length of haul by mode, average size of loads by mode and density of transported commodities (among others), all of which add up to an unmanageable amount of data.

The third approach is the use of spatial interaction models to estimate trade flows. So far, the most widely used model in this approach is the gravity model (Wilson, 1974; Hartwick, 1974; Nijkamp, 1975), generally used to estimate bilateral trade flows (see Black, 1972). This approach (unlike the previous two) models vehicle movements directly (instead of modelling the demand for commodities) which makes them attractive for practical use in the medium and short run. However, as they are cross-section models they are not adequate for forecasting purposes. From the data viewpoint, these models are not too demanding, as they reflect mere tendencies in spatial distribution according to an impedance function. However, these models are incapable of capturing behavioural aspects of freight demand and thus are less powerful than the more data-demanding approaches discussed in this article.

Markusen and Venables (1998) put forward a theoretically appealing model to estimate international trade, based on an industrial-organisation approach, which they call the “new trade theory”. The model endogenously generates both national and multinational enterprises and goods flows. The model, though theoretically attractive, has not been applied in practice yet due to the extensive (and expensive) data requirement. Indeed, the model requires estimates of demand
elasticity for each type of good, wages, transport cost factors, as well as utility functions to represent the consumers in each country.

Garrido (2000) describes another type of model within this category where the truck flows through the Texas-Mexico border are modelled. A space-time autorregressive moving average model was calibrated for the system of border crossing along the Texas-Mexico border. The model data needs are series of international vehicle flows at different points in space, which is easily measured and available for public use. However, this model does not capture behavioural aspects of the freight movement.

Input-output analysis has been used for both intercity and global freight transport modelling. The basic input-output model consists of a table that accounts the amount of a good involved in the production of another good, which is reflected by the purchases of an industrial sector from the rest of the industries within a given market. The standard approach (Leontief, 1941) assumes constant technical coefficients (i.e. the share of each good involved in the production of a given good), constant trade coefficients (i.e., ratio between the production of a good in a given location and the total production of that good), and constant modal split. These assumptions are very convenient from the data viewpoint, because they significantly lower the amount of data to be collected and their corresponding mathematical treatment. However, these assumptions rarely hold in practice and hence the prediction capability of this approach is rather scarce. Liew and Liew (1984) relaxed the original Leontief’s assumptions. This model allows the estimation of the amount of a given commodity \(i\) produced in a given region \(s\) and delivered to the industrial sector \(j\) in a region \(r\) by mode \(k\), as a function of transport cost/prices. This is an attractive FTD model, very flexible and may reflect behavioural aspects not captured by the original Leontief’s approach, however, it needs an enormous amount of disaggregate data for each product, sector, location, and transport mode which makes it almost impossible to apply in a practical situation. A more practical approach is the one followed by Inamura and Srisurapanon (1998). They estimated a rectangular input-output model with fixed coefficients but disaggregated not only by products but also by region of origin and region of destination. The latter gives the model more flexibility than the original fixed implementation by Leontief but is less data demanding than the Liew & Liew model.

5.2 Intercity Freight Transport Models
This level of analysis is the most widely addressed in the literature. At this level Winston (1983) classifies the models into aggregate and disaggregate levels.

5.2.1 At Aggregate Level
One of the first aggregate models reported in the literature is the so-called "abstract mode" model (Quandt and Baumol, 1966), which assumes that the FTD for a mode depends on the attributes of that specific mode and the attributes of the available "best mode". This model does not present any challenge from the data viewpoint but it is not very useful either. Another early approach is the
"aggregate logit" modal split model (see Morton, 1969; Boyer, 1977; Levin, 1978). This model is a log-linear regression whose dependent variable is the ratio between the market shares of two modes. The model's structure is very simple and not computationally demanding, making it attractive in practical applications, especially for large-scale problems. Nevertheless, its main drawback is the lack of theoretical underpinning.

Oum (1979) analysed two aggregate modal split models used in practice: the "price-difference" and "price-ratio" models. Oum showed that both specifications have weaknesses from the economic point of view, arising when logit models are estimated with aggregate data.

5.2.2 At Disaggregate Level

Two classes of disaggregate FTD models are reported in the literature: the so-called "behavioural" and "inventory" models. Behavioural models focus on the mode choice decision made by either the consignee or the shipping firm, whereas inventory models analyse the FTD from the viewpoint of an inventory manager.

The behavioural models attempt to explain the FTD as the result of a process of utility maximisation made by the decision-maker (DM). The data needs are the components of the level of service offered by the different modes, such as rates, travel time, flexibility of the service, reliability, insurance costs, etc. In addition, the “choice set” of each DM must be known to the modeller. One of the drawbacks of this approach (from the data viewpoint) is that the DM must be identified before the data are gathered. This is not an easy task, especially for complex enterprises within a complex supply chain, where some decisions are simultaneously made (for instance transport mode and shipment size) and many different actors participate in the decision.

The second type of disaggregate models are the so-called inventory based models. These models attempt to integrate the mode choice and the production decisions made by a firm (see Baumol and Vinod, 1970; Das, 1974; McFadden and Winston, 1981). These type of disaggregate models can take the simultaneity of the decisions into consideration (see for example McFadden et al., 1985).

Abdelwahab and Sargious (1992), developed a discrete-continuous joint decision model for mode choice and shipment size. The model was calibrated using data from the Commodity Transportation Survey, of the USA Bureau of Census. The model considered only two modes: rail and road. However, in practical applications the shipper chooses among different logistic services rather than pure transport modes. The latter may dramatically increase the number of alternatives to be modelled; the dataset needed for estimation would increase not only in size but also in the complexity of the level of service attributes to be measured.
5.3 Urban Freight Transport Models

Unlike the case in the passengers market, the urban context is the less developed in freight transport studies. Indeed, there is only a handful of published works (in scientific journals) addressing the freight movement in the urban scope. Despite the fact that commodity flow seems to be the core of FTD, most of the literature deals primarily with vehicle flows; especially truck flows (see for example He and Crainic, 1998; Gorys and Hausmanis, 1999; Fridström, 1998.) One exception is the model put forward by Harris and Liu (1998) which predicts purchases and sales for different commodity categories within and outside the city limits. Unfortunately, the model outputs are expressed in currency units (British pounds) instead of quantities. This inconvenience might seem easy to circumvent if the average value per ton were known. However, at disaggregate level this transformation (from currency to loads) does not work, because a given monetary value transported from one origin to a destination may come up as the result of a large number of shipment possibilities (e.g. several JIT small shipments or a few TL shipments). Therefore data transformation at this level might be quite risky.

6 Freight Data Sources Today

The aim of this chapter is to present some of the main data sources available for public use. It is not intended to be an exhaustive list of data sources but a description of the main characteristics of “representative” data sources in Australia, USA, UK, and the EC.

6.1 Australia

In Australia, freight transport data sources can be described either as regular or ad-hoc surveys (Luk and Chen, 1997). There are four regular surveys, which are briefly discussed below.

6.1.1 Freight Movement Survey

It is carried out quarterly by the Australian Bureau of Statistics (ABS) since June 1994 (unfortunately this survey has been suspended). This survey collects freight movements by commodity group, mode (including road, rail, sea, and air), weight, and origin-destination. The sample consists of all the non-road carriers and about 5,000 road freight carriers (including all road freight carriers operating more than 20 vehicles).

The methodology followed in this survey is evolving, as it is a relatively new survey. The latter might be a potential problem when comparing data on a dynamic basis. One caveat is that over-counting may easily occur, as the units of measure are individual movements and hence loads transported with more than one operator may be counted more than once. Another potential problem is the exclusion of light commercial vehicles (less than 3.5 tonnes) and short movements (less than 25 km).
6.1.2 Freightinfo

This is a privately owned data base carried out by the firm FDF Management Pty Ltd. It is a commodity flow survey containing detailed information about the movement of commodities between zones called Statistical Divisions. This survey collects inputs and outputs from commodity producers, as well as transport modes, origin-destination, and some pipeline and conveyor movements.

The methodology used in this survey involves the disaggregation of the commodity database produced by the ABS. The disaggregation is adjusted for imports and exports using external data sources. Modal split is also inferred using external data sources (from rail, ports and airline operators). The data sources combination becomes the main limitation of this database, as its statistical error depends strongly on the range and diversity of data sources involved. Another problem is the accuracy of the road freight movement, which is estimated as the remaining freight after accounting for all the movements made by the other transport modes.

6.1.3 Vehicle Movements and Fleet Characteristics

In addition to the surveys that collect data on vehicle or commodity movements, there are other two regular surveys that are especially relevant to the freight transport task. These are the Survey of Motor Vehicle Use (SMVU), carried out by the ABS, and the Motor Vehicle Census (MVC).

The SMVU is collected every three years since 1971 (with some delays due to budgetary shortages). The main objective is to collect detailed information on vehicle usage for all types of vehicles (passengers and freight). Annual values of distance travelled, fuel consumption, and freight moved by commodity group and area are recorded annually (the survey considers only vehicles registered for road use). The distance travelled in a year is collected through an odometer survey, which may introduce significant errors in the sample.

The MVC is carried out by the ABS every three years since 1971 (published as Motor Vehicles in Australia since 1994). This census records the number of vehicles registered by state and territory, by vehicle type, weight, make, and fuel propulsion in one year. This database is a key element for the estimation of the sample size used for the SMVU. One of the caveats of this survey is the fact that vehicles are classified according to the motor register systems used by each state, which do not have a consistent definition of each vehicle type. Also, off-road vehicles are not required to register and hence are not included in the survey.

Luk and Chen (1997) describe 28 different ad-hoc surveys related to the freight transport task, including surveys of trucking operations, hazardous materials, and vehicle costs among others.
6.2 UK
One of the most extensive examples of publicly available freight data sources in Europe is the case of UK, where at least the following databases are available.

6.2.1 Survey of Heavy Goods Vehicles
The movement of goods by road has been recorded in the UK since the 17th century with the aim of maintenance of common highways and the payment of tolls by vehicles using a certain number of horses or oxen. Later, there were several official efforts to collect relevant transport data in an organised manner; e.g. the Road and Rail Traffic Act in 1933, the Transport Act in 1947, surveys of road freight haulage carried out by the Ministry of Transport in 1952, 1958, 1962 and 1967/68. Then, in 1970 the transport function was put in charge of the Department of the Environment, which established regular quarterly surveys to provide transport related information on a regular basis. In 1976 the Department of Transport began to compile a database of road goods movement by heavy vehicles and to publish series of reports summarising its findings. Later on, in 1997 the Department of the Environment, Transport and the Regions (DETR) was created and took over that task.

6.2.2 The Transport of Goods by Road in Great Britain
The activity of heavy goods vehicles (which account for 95 per cent of all freight moved by road) is recorded in the Continuing Survey of Road Goods Transport (CSRGT).
This survey is based on a weekly sample of vehicles, where the vehicle’s driver is asked to inform of all trips undertaken in a week, including the domestic part of any trip starting or ending outside Great Britain.
The CSRGT provides cross sectional data on detailed activity by type and weight of vehicle, and by public and own-account operation, length of haul, and carried commodities.

6.2.3 Waterborne Freight in the UK
This is a bulletin containing statistics on freight flows (measured in tonnes lifted and tonne-kilometres moved) within the UK by water transport. The database is in charge of the DETR. The statistics cover traffic carried by both barges and seagoing vessels along the inland waterway system and around the coast of the UK.

The main data sources are: the survey of port and barge operators, carried out by MDS-Transmodal (a supplier of consultancy and information services to the transport industry), the Port statistics collected by the DETR, and the ship arrivals data supplied by Lloyd’s Maritime Information Services Ltd.
The statistics are compiled by MDS-Transmodal under a contract with the DETR. In addition to data collection and combination from all the sources, these data are used for the estimation of an origin-destination matrix by ports, and also for estimates of coastwise tonne-kilometres.

### 6.2.4 Focus on Freight

This is a series of occasional publications (started in 1998) aimed to report on the freight and logistics industry in Great Britain. This publication is compiled mostly by the staff of Transport Statistics Freight division on the basis of previous publications in Transport Statistics Great Britain. However, it also includes data drawn from various external sources; for example, the data in the 'Maintaining Standards' section from the Vehicle Inspectorate, the Maritime and Coastguard Agency and the Civil Aviation Authority. In addition, unpublished material is offered upon request as customised analyses for clients.

### 6.2.5 Road Goods Vehicles Travelling to Mainland Europe: Quarterly Bulletins

These bulletins report statistics derived from quarterly interviews with the roll-on/roll-off ferry operators, such as number of powered vehicles and unaccompanied trailers carried on each route to mainland Europe, disaggregated by the country of registration. The other data source is monthly information supplied by Eurotunnel.

### 6.2.6 The Origin and Destination Survey of UK International Trade

International origin and destination surveys are periodically carried out by the DETR, covering the movement of both passengers and freight. These surveys have been carried out in 1978, 1986, 1991 and 1996. The main objective is to provide a database of UK international traffic flows, especially for cross-Channel traffic. The 1996 survey had a secondary objective, which was to collect the data needed for calibrating mode choice models as part of the European Commission’s Strategic European Multi-Modal Modelling project. The intention of the DETR was that the methodology employed for the 1996 survey would be similar to the one used for the 1991 survey. However, this was not possible. It was necessary this time to use different sample selection procedures due to the fact that, with the introduction of the EC single market in 1993, detailed information about individual shipments was no longer recorded for trade within the EC. Thus the sampling frame available for carrying out the previous surveys was no longer suitable. The methodological differences may adversely affect the quality of the data collected and the range and validity of analyses that can be performed.

### 6.3 EC

One interesting experience on a regional joint effort in the EC is the European Co-operation in the Field of Scientific and Technical Research (COST) project. COST is a framework for scientific and technical co-operation, co-ordinating national research on a European level aimed to facilitate the development of a strong position in the field of scientific and technical research for peaceful purposes in Europe. Nowadays there are nearly 200 COST Actions involving about 40,000
scientists from 32 European member countries and from nearly 50 participating institutions from additional 14 countries.

There are two COST actions of particular interest for the freight transport data and analysis community: the COST 310 Freight Transport Logistics, and the COST 321 Urban Goods Transport.

6.3.1 The COST 310 Freight Transport Logistics

It was a three years research involving these 13 countries: Finland, France, Germany, Ireland, Italy, Norway, Portugal, Spain, Sweden, Switzerland, The Netherlands, the UK and Yugoslavia. The research was aimed to present a forward-looking analysis of freight transport logistics in Europe and to make recommendations for the development of an integrated European freight transport network.

As a result this research gives the main characteristics of freight transport, the current status of the transport system and logistics, transport demand, environmental factors. Finally some measures to be taken are proposed relating to the logistics chain and resources involved.

The final report makes various recommendations but unfortunately, data issues are only mentioned implicitly within the, so-called “flanking measures”, which include standardisation of equipment, packaging and information transfer techniques.

6.3.2 COST 321 Urban Goods Transport

It was a four years research involving the following 12 countries: Denmark, France, Finland, Germany, Greece, Italy, The Netherlands, Slovenia, Spain, Sweden, Switzerland, and UK. The main objective was studying the design and operation of innovative measures to improve the environmental performance of freight transport in urban areas. Analysing how the air pollution, noise and energy consumption are reduced by optimising the use of trucks in city traffic, through the application of modern logistical tactics and appropriate administrative measures.

Administrative measures and logistical methods employed in the operation of truck fleets were examined to see their effectiveness in reducing environmental impact. The results of this Action helped to widen (internationally) the knowledge of the effects and acceptability of different pollution control policies in urban settings, and prepare the introduction of appropriate measures in Europe as a whole. This Action also helped to increase public awareness of the problems caused by urban goods movement and the need for international co-operation in this field. Again, unfortunately the data collection topic is not explicitly treated.

It seems clear that a serious problem arises when different countries collect the same freight data in a strongly connected region (as Europe). Indeed, the imports recorded by a given country A, coming from a country B, may not necessarily match the exports recorded by the country B as entering the country A. This type of problem happens surprisingly often when comparing statistics.
collected by different countries. This topic however, has not been tackled within the COST activities.

6.4 USA

6.4.1 Commodity Flow Survey
The Commodity Flow Survey (CFS) provides data on the flow of goods and materials by mode of transport. The CFS is sponsored by the U.S. Department of Transportation (DOT), Bureau of Transportation Statistics and U.S. Census Bureau, Department of Commerce, and performed by the U.S. Census Bureau, Department of Commerce and Oak Ridge National Laboratory.

The CFS follows a series of publicly available datasets from 1963 through 1997. Samples of domestic establishments engaged in mining, manufacturing, wholesale, auxiliary establishments (warehouses) of multi-establishment companies, and some selected activities in retail and service were used to collect the data through the completion of a questionnaire. The current version of the CFS contains a geographic coverage of data at national level, stratified by State and Metropolitan Area. The 1993 and 1997 CFS have an expanded coverage of intermodalism compared to previous versions. The survey reports all modes used for a shipment (for-hire truck, private truck, rail, inland water, deep-sea water, pipeline, air, parcel delivery or U.S. Postal Service, other mode or unknown). With previous available data (a Mode-Distance Table developed by Oak Ridge National Laboratory) route distance for each mode for each shipment was imputed. The ton-mileage by mode of transport was computed using the travelled distance.

6.4.2 Transborder Surface Freight Dataset
The Transborder Surface Freight Dataset provides North American merchandise trade data by commodity type, by surface transport mode (including pipeline) with geographic detail for U.S. exports to and imports from Canada and Mexico, updated on a monthly basis. Its objective is to provide transportation information on North American trade flows. The source is the official U.S. international merchandise trade dataset.

Currently, these data are being used to monitor transborder freight flows since the beginning of the North American Free Trade Agreement (NAFTA) in 1994. Other uses of this database are: trade corridor studies, transportation infrastructure planning, logistics strategy analyses amongst other purposes.

The dataset is compiled from the Census Foreign Trade Statistics Program. Import and export data are collected from administrative records required by the Departments of Commerce and Treasury of the US.
Most of the imports data from Canada and Mexico, are collected electronically via an Automated Broker Interface, and the Customs entry documents collected by the Customs Service and transmitted to the Census Bureau. Information on U.S. exports of goods from the U.S. to all countries (except Canada) is compiled from copies of Shipper's Export Declarations (SEDs) and data collected from shippers, forwarders or carriers. On the export side about half of the data are collected electronically, through a U.S./Canada Data Exchange agreement and the Automated Export Reporting Program.

The official U.S. import and export statistics provide information on shipments of merchandise between foreign countries and the U.S. Customs Territory, U.S. Foreign Trade Zones, and the U.S. Virgin Islands, without regard to whether or not a commercial transaction is involved. The statistics record the physical movement of merchandise between the United States and foreign countries.

### 6.4.3 Motor Freight Transportation and Warehousing Survey

The Motor Freight Transportation and Warehousing Survey is carried out annually by the US Census Bureau. It provides detailed estimates of operating revenues and expenses for the for-hire trucking and public warehousing industries, as well as inventories of revenue-generating freight equipment for the trucking industry at the national level in the US. The survey excludes private motor-freight carriers operating as auxiliary establishments to non-transportation companies and independent owner-operators with no paid employees.

The survey covers all employer firms with one or more establishments that are primarily engaged in providing commercial motor freight transportation or public warehousing services. The results of this survey are published in a report where statistics are summarised by kind-of-business classification based on the Standard Industrial Classification (SIC) Manual issued by the Office of Management and Budget.

### 6.4.4 Vehicle Inventory and Use Survey

The aim of this survey (collected by the U.S. Census Bureau every five years since 1963) is to measure the physical and operational characteristics of the truck’s population in the US. It covers private and commercial trucks registered (or licensed) in the US, excluding vehicles owned by Federal, state, or local governments; ambulances; buses; motor homes; farm tractors; unpowered trailer units; and trucks reported to have been sold, junked, or wrecked prior to July 1 of the year proceeding the survey. The dataset on physical characteristics include date of purchase, weight, number of axles, overall length, type of engine, and body type. The operational characteristics data include type of use, lease characteristics, operator classification, base of operation, gas mileage, annual and lifetime miles driven, weeks operated, commodities hauled by type, and hazardous materials carried.
Several private and public agencies use these data on a regular basis. Examples of public usage are: analysis of cost allocation, safety issues, proposed investments in new roads and technology, user fees, estimation of per mile vehicle emission, vehicle performance and fuel economy, fuel conservation practices of the trucking industry, among others. Examples of private sector’s usage are the following. Tire manufacturers use the data to estimate the duration of products and to determine the use and applications of their products. Heavy machinery manufacturers use the survey data to track the importance of various parts distribution and service networks. Truck manufacturers use the data to determine the impact of equipment on fuel efficiency.

In addition to the public-domain data sources described above, there are two data bases privately owned and commercialised:

6.4.5 TRANSEARCH and Freight Locater

The TRANSEARCH database contains origin-destination freight movements in the US, covering major modes of transport. It is compiled and produced on an annual basis since 1980 by the firm Reebie Associates. Records are kept for freight traffic shipments across geographic markets and commodities for seven modes of transport, including truckload, less than truckload (LTL), private truck, rail, intermodal, rail carload, waterborne, and air. The database contains the freight activity of U.S. domestic, Canada/U.S. and Mexico/U.S.

The same firm commercialises the database called Freight Locater which contains detailed information on type of freight being transported and who is shipping it at different levels of spatial aggregation. It also contains information on annual tons and sales, and employees by individual establishments. The establishments are divided by industry, commodity, and vehicle type requirements.

While it is obvious that privately owned (and managed) databases are returning profits, doubt is cast on the “profitability” of public databases. Indeed, the role that the state should play in collecting and making information available to the citizens is rather controversial (Rhind, 1992). No clear answer is found on what is considered acceptable on issues like budget, cost-effectiveness, or continuity and reliability regarding public data sources.

7 Data Needs for Freight Modelling

Data needs are generally dependent on the specific modelling techniques to be used. Nevertheless, some common data requirements can be identified in most cases. After all, high-quality data are required to effectively model any stage in the process of freight transport. Chapter 4 showed that most of the data sources available nowadays have significant limitations in scope as they focus on a specific mode, commodity, or a given spatial aggregation level. Further more, several analyses in the freight market require “external” data sources not directly related to the freight transport task. For instance, hazmat analysis usually require information on route characteristics (road design
aspects), environmental attributes and parameters, accident records, expected consequences, etc. Therefore, the combination of data sources, including demographic and socio-economic variables, and economic activity at industry sectors levels, arise as the natural approach to follow to overcome the data limitations. The data sources to be combined need common records to link and reference one and other, especially when geo-referencing is needed. The referencing process could be very expensive to achieve if the data sets were collected independently and without knowledge of the scope of their further use. On the other hand, planning ahead the attributes to be collected, the methodology to be used, and the spatial and temporal resolution, may dramatically decrease the cost of further data combination and database construction.

The combination of various types of freight surveys (Casavant, et. al., 1995; Matherly, 1996) may yield a variety of new problems, both at the theoretical and practical levels. Excluding the external data sources, the most common types of surveys to be combined are decision makers (DM) surveys and drivers’ surveys. The DM’s surveys involve direct interviews with shippers, carriers or consignees gathering valuable information about the freight transport decision-making processes at operational level. At the tactical level (for instance, to obtain O-D matrices), it is more efficient to use drivers’ surveys. These surveys may be conducted in the form of mail-back questionnaires or roadside intercept surveys at toll stations, rest areas, border crossings or others sites where the willingness to participate might be relatively high. Handheld computer technology has proven to be especially successful in collecting freight data in weigh stations (see Anderson, 1997). The response rates tend to be lower for the mail-back questionnaires, but they are easier to carry out. Response rates are high for intercept surveys, but they are more labour and cost-intensive (Pendyala and Shankar, 2000).

7.1 Data for analysing freight transport at global scale

7.1.1 Economic Activity

Most of the models identified in Chapter 3 have exogenous variables that measure the level of economic activity of the countries involved in the process of international transport. Thus for example, models based on the theory of international trade make use of national income, rewards on labour and capital factors, technical coefficients of production, employment, etc.; models based on aggregate cost functions need at least production, materials and capital costs.

7.1.2 Transport Modes and Infrastructure

International freight movement is highly sensitive to the available physical infrastructure. This is especially true when two countries establish trading but different standards are found in the level of infrastructure. For instance, considerable congestion can be observed in some bridges at the border crossings between the United States and Mexico due in part, to different standards in technology and organisational procedures at both sides. A similar situation occurs when technology becomes
an obstacle to the freight transport process. An example of the latter is the different rail gauges between Spain and France, which may significantly delay the rail freight transport.

Therefore, accurate data related to the five typical freight transport modes: water, air, road, rail, pipelines, and the conditions of the corresponding infrastructure are fundamental for the modelling task.

7.1.3 Transport Services
In addition to the physical infrastructure, the commercial freight services already operating between two countries may play a crucial role in promoting trade through discount rates and services, especially as a result of policies to reduce trips with empty vehicles. In fact, when two regions establish trading in the presence of competitive arbitrage, prices in the importer and exporter regions differ by only the cost of transport, and hence changes in the freight charges may have a dramatic impact on the volumes being traded. Therefore, data on the transport rates structure and economies of scale and scope in the market under analysis are needed for modelling at the global level.

7.1.4 International Trade Conditions
Freight volumes moving from one country to another depend also on the bilateral commerce regulations, administrative procedures and efficiency of the customs and other public agencies involved in the acceptance of imported (exported) goods. Thus, the set of legal and administrative processes related to the trading conditions is essential to set the constraints in any modelling framework. These pieces of information form one of the many external databases to be combined with freight data itself in order to model the actual global transport phenomenon.

7.2 Data for analysing freight transport at intercity scale

7.2.1 Modal Split
At the aggregate level models typically regress the proportion of market shares (between pure modes) against some aggregate attributes such as prices, travel time and cost, etc. Therefore, at the very least, the modeller should have accurate data on modal split and some level of service attributes. However, the modeller must proceed with caution when designing the data collection method and strategy because it has been proved that severe arbitrary constraints are imposed on the model elasticities when discrete choice models, whose dependent variable is the ratio of market shares, are estimated with aggregate data (Oum, 1979).

7.2.2 Fleet’s Attributes and Composition
Intercity freight flows are often times analysed as vehicles flows. This is especially true for models at disaggregate level, where mode choice has been the prevalent dependent variable. In fact, most of the behavioural models attempt to forecast the DM’s mode choice and hence data on the vehicle
characteristics are needed. Furthermore, the type of technology and organisational characteristics of the fleet define the sample segmentation and expansion methods. Indeed, operational characteristics such as LTL or TL have significant differences, which preclude modelling their services together. Another example is the competition between rail and road transport when different hauling distances are considered, in this case not only the fleet composition would be important but also the trips length distribution.

7.2.3 Network Characteristics

Intercity models are especially sensitive to the network resolution and level of service. Routing options as well as the costs at each arc have a tremendous impact on the quality of model results. The network costs structure is especially relevant when the freight flows are found as a result of an equilibrium process -usually under Wardrop’s second principle (see for example Friesz et al, 1983). In fact, the cost structure (e.g. symmetry, mathematical representation, continuity, etc.) defines the existence and uniqueness of the equilibrium solution. This solution is found through a relatively complex iterative process where the quality of the solution is strongly dependent on the cost functions and the quality of the data on which these cost functions are calibrated.

7.3 Data for analysing freight transport at urban scale

Ogden (1992) divides the information needs for urban freight data collection into five areas:

7.3.1 Vehicle Fleet

Information is needed on the number, type and size of vehicles as well as their ownership pattern (e.g. for-hire operations, owner-drivers, and private operations). All this information should be obtained at dissaggregate level, keeping in mind the possibility georeferencing and linkage with sociodemographic data.

7.3.2 Vehicle Flows

It is perhaps the most relevant information for infrastructure management. Vehicle flows are also instrumental for identification of freight activities in different locations and temporal variations of activities within a city. These type of data are nowadays widely available with a high degree of accuracy, due to the ample use of electronic counting devices. Vehicle counts are a main source of data to synthesize an origin-destination truck matrix through traffic assignment models.

Vehicle flows should be accompanied with data on origin-destination land use, time of the day and hopefully type of commodity and weight (weight in motion techniques should be used in conjunction with traffic count devices whenever is possible).
7.3.3 Commodity Flows
Unlike vehicle flows, commodity flows are the direct manifestation of consumption and hence are a crucial piece of information, as far as modelling is concerned. Unfortunately, they cannot be observed from the roadside and some form of interview must be used to obtain this information. Typically, for each load being moved, origin-destination land use or industrial activity and commodity classification data should be collected, as well as load size (weight and/or volume), type of packing and handling, ownership and responsibility for transport (shipper, forwarder, carrier, etc.) , and method of despatch (depot, radio, phone, EDI, Internet etc.).

7.3.4 Major Freight Generators
This information is essential for modelling the economic impact of freight movement. According to Ogden (1992) there are two approaches for the analysis of freight generators. First, a metropolitan-wide analysis, including both freight-related data (e.g., volume and type of freight generated and its associated vehicle trips) and economic data (e.g. income, population density) to establish relationships between explanatory variables. The second approach is to focus upon specific freight generators, such as a truck terminal or airport. Useful information about such facilities includes origins and destination of truck movements, and accessibility (e.g. delays, routes, etc.). Specific data may include weight by commodity by origin/destination, type of freight (general, containerised, bulk, etc.) and mode of transport to/from the facility.

7.3.5 Major Freight Corridors
For planning purposes it is needed to identify the major corridors for the movement of goods within a metropolitan area. Freight corridors may include more than one mode (e.g. roads and rail tracks) therefore the estimated level of service offered in a corridor may exhibit significant variation and therefore it should incorporate all the modal possibilities. Identification of the corridors can be done upon the analysis of truck and commodity flows described above. Important characteristics of the corridors are commodity types being moved, vehicle types or services using them.

8 Key Issues in Freight Data Collection
Hutchinson (1985) points out three distinctive aspects of freight transport that make difficult the application of data collection methods commonly used for passenger’s transport. First, the identification of the actor who actually makes commodity shipping and service decisions is not an easy task (in the passengers case the identification is quite clear). Second, most of the pertinent variables used to describe goods movements are very different from those used to describe passenger flows (e.g. weight/volume of a shipment, distance shipped, annual tonnage shipped, value of commodity, potential risk, ownership, etc.). In addition, often times it is difficult to find the right person to be interviewed regarding all the pertinent variables. Finally, shipper’s knowledge and perceptions about the characteristics and attributes of modes may be significantly different from objective measures of level of service properties. In the case of freight transport, the knowledge of the level of service offered by a mode not being used by the shipper (or the corresponding DM) is
rather limited compared to the passenger counterpart. For instance, in the passengers case, commuters have a relatively accurate knowledge of travel time, cost, comfort, etc. for all the available modes, whereas in the freight sector, the DM would need to actually use the available alternatives to find out the level of service attributes, which may result in large unwanted costs.

Hutchinson (1985) also emphasises the concept of total distribution logistics of a firm, when freight transport is analysed and measured. Accordingly, shipment decisions and service characteristics should be analysed beyond the transport function, i.e. imbedded in the context of the total transport supply environment including interactions between routes, inventory, vertical integration and other strategies of the companies participating in the corresponding supply chain. The latter is equivalent to the consideration of “activities” instead of simply “trips” in the passengers case. That author also provides general recommendations, some of which are of interest to this paper:

- In many studies, the main need is for surveys aimed to understand the nature of shippers, carriers and facilitators behaviour rather than the collection of formal quantitative data.
- Surveys of shippers, carriers and facilitators usually involve personal interviews with a knowledgeable set of respondents, therefore the interviewers should be specialists in freight transport since they will probably deal with open questionnaires and interactions with the interviewees.
- Freight surveys should recognise that quantitative data may be difficult to obtain due to their commercial value. On the other hand, qualitative data elicited from expert panels might be an initial way of understanding freight transport processes.
- A commodity classification scheme is a major necessity before the data collection begins. However, this scheme must be transport-related (standard commercial classification schemes might not be appropriate for further transport analysis, especially for modelling purposes).
- A key statistical issue is the identification of the population of interest. In fact, survey results need to be expanded to the proper population, which in the case of freight movement might not be as clear as their counterpart in the passengers’ case. For this purpose, a clear understanding of the supply chain components and extension is needed.

With the differences between the freight transport market and its passenger counterpart, the standard four-step transport model become inappropriate as a framework for modelling and consequently it should not be used as a framework for data collection either. In fact, the four-step standard model does not reflect the actual phenomenon taking place in the freight movement as the following discussion shows:

**Trip Generation**
The stage of production and attraction of personal trips from and to the spatial units of interest does not necessarily apply to the freight movement. Indeed, unlike a passenger trip, a shipment movement is the “result” of several previous trips with origins and destinations different from the
one observed for a given shipment. The amount of previous trips depends on several factors such as type of commodity, stage of production (raw material, semi-finished, finished, etc.), or value. For example, when a computer set is sent from a seaport to a computer store in a shopping mall, the standard four-step model would record the seaport as the origin and the shopping mall as the destination. However, neither the seaport is the relevant origin of the shipment nor the shopping mall is the final destination where the computer set will be used by the buyer (the one who truly “originates” the production and transport of the computer set).

What this examples shows, is that each time a trip is generated in one origin or attracted to a destination, it would be necessary to account for several other trips made between the origin and destination of all the primary and intermediate goods involved in its production. Also all the intermediate logistical movements (e.g. cross-docking) should be recorded.

**Trip Distribution**

The standard approach followed in the four-step model is to distribute trips between spatial units, according to the attractiveness of both origin and destination and an “impedance” between their spatial locations, following either a simple or double flow conservation rule (either the productions or attractions must be replicated exactly by an aggregation of the distribution model). This scheme does not necessarily hold in the freight transport case. Depending on the time frame to be considered, spatial units could “accumulate trips”, i.e. several shipments could arrive to a given destination where a production plant is located and then they might stay there for several weeks (or even months or years) depending on external factors such as production technology, inventory policy and demand from another link within the corresponding supply chain. For instance, in the wine industry, some of the primary and intermediate inputs may stay in the plant for several years before a wine shipment goes out. Another example is the heavy freight traffic between plants and warehouses located in different countries, where the impedance between origin and destination is large and hence standard distribution models would assign only a small fraction of trips. The latter is due to the fact that the decision making process is based on total cost criteria, including but not restricted to transport cost within the supply chain. Therefore, the trip distribution, in the case of freight transport, does not fit in the same structure as the personal trips.

**Modal Split**

The core of modal split modelling in personal trips is the discrete choice analysis based on random utility theory (McFadden, 1981; Ben-Akiva and Lerman, 1985, Chapter 3), which requires that the DM chooses only one alternative from an available choice set. The latter is totally applicable to the personal trips due to the nature of the unit to be travelling (a unique individual), whereas in the freight transport case, the DM may choose more than one of the available alternatives, i.e. a shipper may send a given load by more than one mode at the same time; this situation happens when the shipper is making large shipments of high valued products or shipments of products whose lose can have a significant impact on the company’s profit (Daughety, 1979). In addition, competition between modes is, sometimes, more common than the case of passengers mode choice. Indeed,
intermodalism plays an important role in every freight transport decision, which has led to a restructuring of the carriers companies, regarding horizontal and vertical integration, into mega-carriers (Browne, 1992), therefore the concept of competition between the modes available in each DM’s choice set is blur.

**Traffic Assignment**

In the standard four-steps framework, vehicles are typically assigned to network paths according to the Wardrop’s principles (either user equilibrium or system equilibrium, see Sheffi, 1985 Section 1.3) which assume that the DM chooses mode and route and is completely known to the modeller, and the decision is made on the basis of transport costs alone. These assumptions do not hold in the case of freight transport because in many cases different DM’s choose the mode and route for a given shipment (e.g. the shipper may choose the mode and the carrier the route), besides there are a more hierarchical stage of decisions involving several DM’s with different goals. For example, a remote buy may involve transport service selection (by the consignee), carriers set selection (by the shipper), route selection (strategically chosen by the carrier), network links selection (operationally decided by the delivery truck driver). For these reasons, often the DM is (are) not known to the modeller in the case of freight transport. Finally, when planning deliveries the DM’s choice is not based purely on transport costs but on more general criteria such as total production-inventory-distribution costs (transport costs rarely exceed 30% of the total cost).

The above discussion shows that the four-steps modelling framework is not appropriate for freight transport and hence the data collection methods developed for transport planning and modelling must be revised and adapted to be able to capture the phenomenon of interest.

9 **INFORMATION TECHNOLOGY AND FREIGHT DATA COLLECTION**

Most of the tactical and operational surface freight transport models rely heavily upon data on traffic flows, speed, and type of trucks. These variables had been traditionally measured through equipment controlled by human beings (e.g. manual counts, speed guns, instrumented vehicles, field classification surveys, etc.). Later, the advent of automatic and semi-automatic equipment brought about new possibilities for accurate and continuous data collection (e.g. magnetic loops). Recently, new technology for measuring and transmitting traffic data has been developed and is currently being used for studying passenger vehicles flows (wireless communication from vehicle detectors to central computers). However, the freight transport field does not seem to be exploiting the benefits of these systems, even though some of them are in use for operational purposes (other than data collection). Examples of the latter are the use of transponder-equipped road and rail vehicles entering the USA territory from Canada and Mexico for customs preclearance, and wireless checking of containers in some ports (see www.apl.com). Both examples show applications where the main objective is not data collection itself but they can be easily used as a relatively inexpensive source of real time data.
The replace of standard magnetic loops detectors for camera-based automatic vehicle identification (AVI) systems began as a (seemingly expensive) way to increase the efficiency of the urban road network in locations where the weather conditions preclude the use of the buried standard sensors (RCOC, 2000). However, nowadays AVI is seen as a helpful tool for reducing the cost of vehicular flow counting and identification. The use of AVI in the freight transport field could go way beyond traffic counting and identification. Indeed, environmental data can be gathered through remote sensing of trucks emissions, using an existing light-vehicle remote sensing system with minor modifications (see Boulter, 1999); the system would be able to simultaneously retrieve trucks speed, acceleration, make and model as well as exhaust gas pollutants. Another example of the great potential of AVI in this field is the estimation of dynamic origin-destination matrices, which are essential for real-time network models involving trucks assignment to urban roads (see for example Asakura et al, 2000).

The freight transport is a process where the space and time dimensions are especially relevant and hence they ought to be included in modelling (Garrido and Mahmassani, 1998, 2000). However, dealing with spatial data had been traditionally cumbersome due to practical and theoretical (econometric) problems. The advent of geographical information systems (GIS) has solved most of those problems and has become a standard tool for many transport studies allowing an efficient analysis of spatially linked data.

As mentioned in Chapter 4, different data sources are needed for freight modelling and analysis. These data sources can be easily combined with within a GIS. Nevertheless, freight data sets are not commonly presented in a GIS framework.

In general, new technological advances bring a major change in the way information is perceived, needed, distributed and used. For instance, intelligent transport systems (ITS) deal with information that is more disaggregate than any of the usual temporal resolutions seen in transport planning studies (e.g. time dependent O-D matrices). Spatial resolution may as well be dramatically different from the standard zoning system used for transport planning purposes (e.g. raster based GIS). These changes will yield an abundance of data (some times with duplication of the same data by several users) and bring about new problems regarding standardisation of data exchange, data structures, technological compatibility, and privacy-confidentiality issues, among many others. In this new paradigm, automatic data collection is the rule and discrimination criteria to discern what is to be kept (as valuable data) and what to be discarded, may draw the line between what is economically feasible for data collection agencies and users.

Data distribution under these new trends plays an important role. Indeed, one of the main operational problems in the data acquisition field is the difficulty of identifying whether a particular data set already exists and if so, how to get those data (Pienaar and Brakel, 1999). Therefore, the distribution channels for data will have to tackle the problems of data transmission and marketing (public awareness) together. The latter includes not only the communication of the data
characteristics and attributes but also those of the databases and navigation software that handle the data --stressing the capabilities of association with other data sets outside the scope of freight transport. Distribution channels will be inevitably linked to aspects such as communication protocols, copyrights, charging, access control and privacy, among others.

The combination of technologies such as remote sensing devices, wireless communication systems (e.g. GPS, cellular transmission) and specialised software (e.g. GIS) can offer an invaluable help to push the state of practice in freight data collection beyond the standard field survey-traffic count matching. However, one gap that technology is not able to overcome is the development of relationships between operators who have the data (and not necessarily measure them) and analysts, planners and researchers who need those data. The task of successfully developing those relationships depends on the ability (of the actors involved) to meet challenges of privacy, disclosure issues, proprietary data, security, timeliness and handling of massive volumes of data. In the future, the main challenge will be too much data rather than inadequate or scarce data (Lockwood, 1997).

10 Final Remarks

In this article several aspects of freight data collection and use were discussed with emphasis on the main differences between the passenger and freight cases regarding the use of data for modelling.

A wide variety of freight transport models were described, along with their data needs. Three spatial resolution categories were analysed in terms of their most relevant data needs: global, intercity, and urban scale.

Various data sources, around the world, were also described in order to provide a snapshot of what is typically available (or lacking) in some developed countries. Knowing about these data sources and their pros and cons might be especially helpful for developing countries where freight data collection is still in an early stage of development.

A set of key issues in the freight transport system and the corresponding data collection were identified and discussed in light of the classical four-step transport model widely used for personal trips analysis. In addition, some key issues in new technologies applied to data collection in the freight transport field were discussed.

It can be concluded that the relevant issues on freight data will shift from availability (the chronic syndrome of lack of data) to accessibility and management (especially under the demands from ITS), as well as defining who are the pertinent stakeholders and what their real data needs are.

Massive volumes of data are being collected everyday around the globe, hence the co-ordination between different actors collecting similar data has become a new challenge, as the same data
measured by different entities do not always match. Methodological challenges are also rising nowadays, regarding cost-effectiveness measures (directly related to the public awareness on the importance of quality in freight data), proper charging, and budgetary decisions at public administration level.

The subject of legal liability induced by selling or supplying data is being debated nowadays (Pienaar and Brakel, 1999). In fact, the liability of data suppliers may become fuzzy as more added value is incorporated to a database that has been modified by several different users (private or public). Along this line, the magnitude and type of errors in the data, and their propagation characteristics will become an issue of primary interest for potential data buyers/users. Copyright aspects need also to be addressed more systematically than the personal-trips case, because of the high commercial value of freight data.

ACKNOWLEDGMENTS

The author wants to express his gratitude to the Chilean National Fund for the Development of Science and Technology for the financial support provided through the project FONDECYT 1000585-2000.

11 REFERENCES


